DS Journal of Digital Science and Technology
Volume 5 Issue 1, 1-13, Jan - Mar 2026
ISSN: 2583-5416/ https://doi.org/10.59232/DST-V5I1P101

Original Article

Integrating Knowledge Graph with Retrieval-Augmented Generation for Vietnamese Legal Question Answering

Vuong T Pham[1*], Minh Phan[1]

[1]*Saigon University, Ho Chi Minh City, Vietnam.*

*vuong.pham@sgu.edu.vn

Received: 20 November 2025; Revised: 26 December 2025; Accepted: 18 January 2026; Published: 30 January 2026

Abstract - Large Language Models (LLMs) are increasingly used for question answering, but legal applications require grounded answers with article-level citations and minimal unsupported statements. This paper presents a framework that integrates an ontology-driven Knowledge Graph (KG) with Retrieval-Augmented Generation (RAG) for Vietnamese legal question answering. The method builds a document-aware KG that preserves the hierarchical structure of statutes (Document → Chapter → Section → Article → Clause → Point) and encodes semantic and cross-reference relations. Graph-aware retrieval is combined with a hybrid ranking function that balances semantic similarity and graph-derived confidence, and the retrieved legal provisions are provided to an LLM using a constrained prompting template to generate answers with explicit citations. Experiments on three Vietnamese law domains (Electronic Transaction Law, Land Law, and Labor Law) report 85.7% answer accuracy, outperforming BM25 keyword search (62.3%) and a pure LLM baseline (71.5%), while improving citation quality. The framework is intended to increase traceability by grounding generation in retrieved statutory provisions.

Keywords - Knowledge Graph, Large Language Models, Legal Question Answering, Ontology, Retrieval-Augmented Generation, Vietnamese NLP.

1. Introduction

Knowledge bases are widely used to support knowledge-intensive applications that require explicit domain facts and relations, especially in settings that demand verifiable reasoning [1, 2]. In the legal domain, question answering systems must map user queries to precise statutory provisions and provide traceable citations. Although LLMs provide fluent natural-language responses, deploying them for legal question answering remains difficult due to hallucination, limited transparency, and the risk of producing statements that are not supported by the underlying legal text [3]. These risks motivate approaches that explicitly ground generation in curated knowledge sources and structured representations [4, 5].

Retrieval-Augmented Generation (RAG) mitigates part of this problem by retrieving relevant passages from external corpora and conditioning generation on them [6, 7]. However, conventional RAG pipelines typically retrieve independent text chunks and do not explicitly model (i) the hierarchical structure of statutes, (ii) cross-references between provisions, and (iii) the requirement that citations point to the exact legal unit (article/clause/point). As a result, retrieved context can be incomplete or structurally inconsistent, which reduces citation fidelity and answer reliability.

This study addresses these gaps by integrating an ontology-driven KG with RAG for Vietnamese legal question answering. The KG preserves statutory structure and encodes semantic and reference relations, enabling retrieval at the appropriate legal granularity. A hybrid scoring mechanism combines embedding-based similarity with graph-based confidence to prioritize provisions that are both semantically relevant and structurally supported.

The contributions are: (1) An ontology-based KG schema for Vietnamese legal texts that captures hierarchy and cross-references; (2) A graph-aware retrieval and hybrid ranking strategy for grounding RAG in statutory structure; (3) An evaluation across three Vietnamese legal domains with answer-accuracy and citation-quality metrics; and (4) a discussion of deployment considerations and limitations, including planned ablation and user studies. The rest of the paper is organized as follows: Section 2 reviews related work; Section 3 describes the proposed method; Section 4 reports results and discussion; Section 5 concludes.

2. Related Work

Retrieval-augmented generation was introduced by Lewis et al. [6] and has become a standard approach for grounding LLM outputs in external corpora. Dense Passage Retrieval (DPR) [7] and related embedding-based retrievers enable scalable semantic retrieval, but retrieval alone does not guarantee citation fidelity in structured domains.

Graph-Based RAG (GraphRAG) extends RAG by exploiting graph-structured knowledge for retrieval, organization, and generation [12, 13]. Recent surveys summarize GraphRAG design patterns, including graph indexing, graph-guided retrieval, and graph-enhanced generation [12]. Knowledge graphs also support reasoning and explainability through explicit entity-relation structure [9, 10]. Prompting strategies that use structured knowledge (e.g., KG-aware prompting or constrained generation) have been studied to reduce unsupported statements and improve factual grounding [5, 11].

In the legal domain, specialized language models and benchmarks have been proposed to evaluate and improve legal NLP, such as LEGAL-BERT [15] and LegalBench/LawBench [14, 16]. For Vietnamese legal resources, ontology and linked-data efforts have been reported to support downstream legal applications [17], and recent work has explored Vietnamese legal knowledge graph construction and structured representations for legal QA [20, 21]. The approach in this paper differs from prior work by focusing on document-aware graph construction for Vietnamese statutes and by combining graph-derived confidence with semantic retrieval in a unified ranking function to improve traceable citation grounding.

3. Design a Legal Question-Answering System based on Integrating Ontology, Knowledge Graph, and RAG

3.1. Ontology Model for Vietnamese Legal Texts

The ontology model is grounded in the Rela-model formalism [2], which provides a mathematical framework for representing domain knowledge. Formally, a legal knowledge base is defined as:

$$K = (C, R, Rules) + (Concept, Rela)$$

Where C represents the set of legal concepts, R represents relations, Rules represents inference rules, and (Concept, Rela) provides schema definitions.

Concept nodes are encoded with a compact record that is easy to inspect and index. In this study, each concept stores (i) an identifier (Name), (ii) a short gloss (Meaning), (iii) Domain Attributes (Attrs), (iv) Keywords/Phrases (Keyphrases), and (v) optional links to near-synonyms or related concepts (Similar).

c = (Name, Meaning, Attrs, Keyphrases, Similar)

For example, Name provides a stable key (e.g., "Giấy phép lái xe"/Driver's License), while meaning gives a brief definition, and Attrs holds domain-specific fields used during matching. Keyphrases support lexical recall, and Similar connects variants that appear in practice. Relations are stored in the same spirit: a relation record keeps its label and meaning, plus references to the subject/object concepts and relation-specific properties needed for querying and scoring.

r = (Name, Meaning, ConckeyS, ConckeyO, Prop, Keywords)

Capturing semantic relationships between concepts.

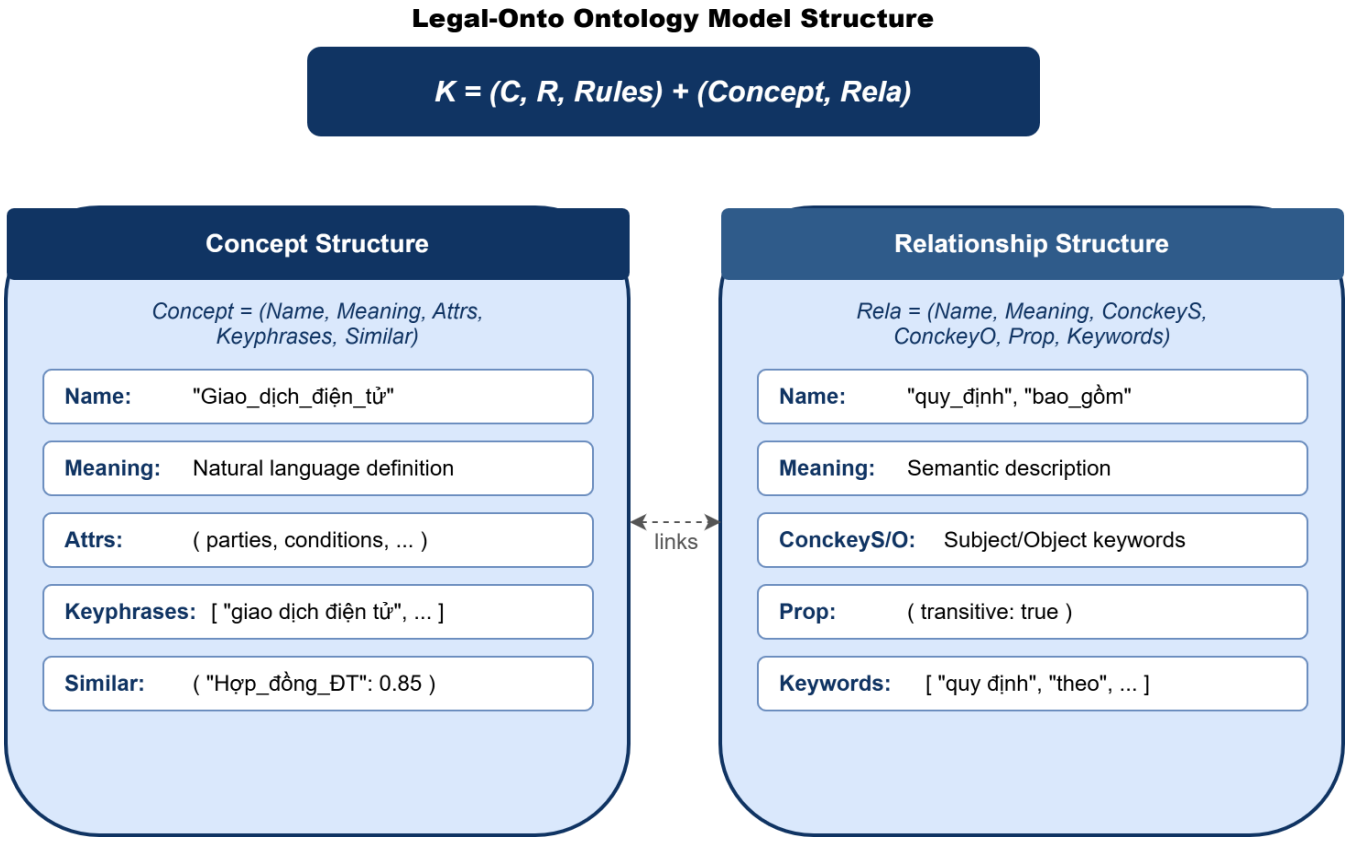

Fig. 1 Ontology model structure showing concept and relational schemas with their attributes

3.2. Vietnamese Legal Document Structure

Vietnamese statutory texts are organized as a nested hierarchy mandated by Circular 25/2011/TT-BTP. In practice, legal content is addressed from a document level down to fine-grained provisions (chapter/section/article/clause/point). The knowledge graph mirrors this nesting via containment edges (e.g., HAS_CHAPTER, HAS_SECTION, HAS_ARTICLE, HAS_CLAUSE, HAS_POINT), which makes it possible to retrieve evidence at the exact granularity required for legal citations. Figure 2 provides a concrete instance from the Electronic Transaction Law (2023).

Chapters contain sections, sections contain articles, and articles are further decomposed into clauses and points. For instance, Article 10 ("Legal Validity") includes multiple clauses; Clause 1 specifies conditions under which an electronic transaction is considered valid.

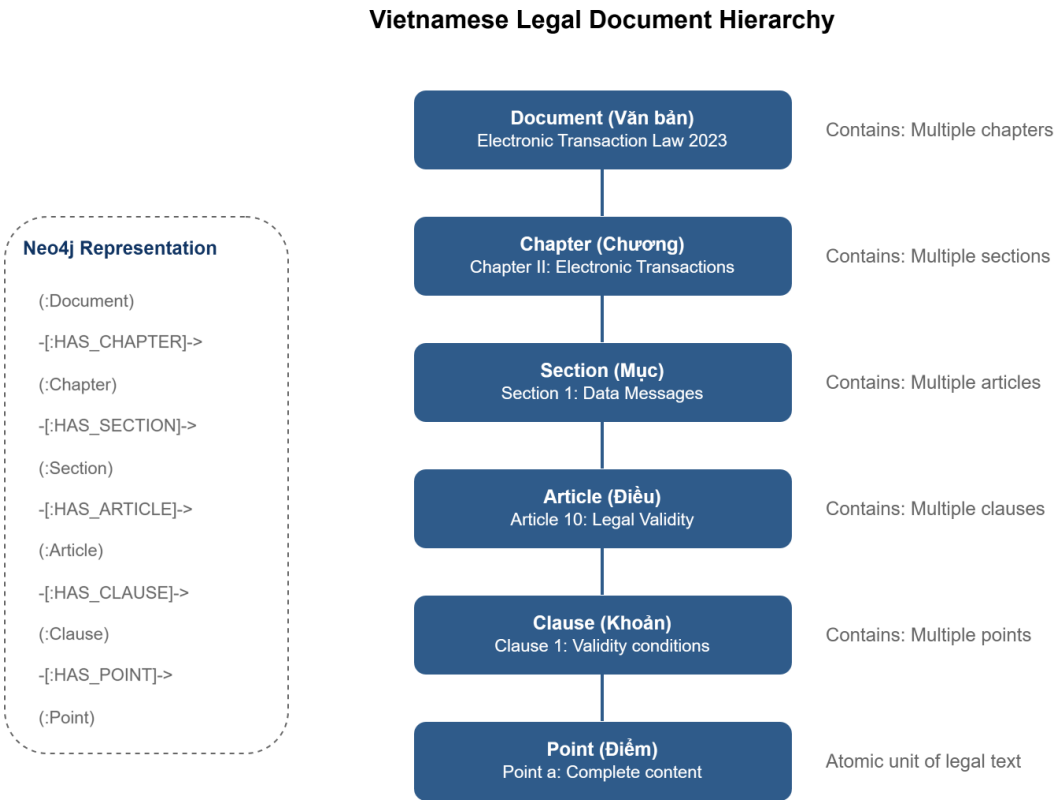

Fig. 2 Concrete instance from the electronic transaction law

3.3. Types of Concepts and Relations

This section specifies the concept/relation taxonomy used in the graph. The goal is pragmatic: the taxonomy should be expressive enough to represent statutory structure and semantics, while remaining simple to query and maintain. Two concept families are used. **Structural concepts** capture the legal outline (Document, Chapter, Section, Article, Clause, Point). **Domain concepts** capture the entities and terms that appear in provisions and questions. Examples include “Chữ ký điện tử” (Electronic Signature) and “Chứng thư số” (Digital Certificate) in the Electronic Transaction domain; “Hợp đồng lao động” (Labor Contract) and “Tiền lương” (Salary) in Labor Law; and “Quyền sử dụng đất” (Land-Use Rights) and “Giấy chứng nhận” (Certificate) in Land Law.

Relations are grouped into (i) **structural relations** that encode nesting (e.g., HAS_CHAPTER, HAS_ARTICLE, HAS_CLAUSE, BELONGS_TO), (ii) **semantic relations** that express normative meaning (e.g., DEFINES, REGULATES, REQUIRES, PROHIBITS), and (iii) **reference relations** that connect provisions across the corpus (e.g., REFERS_TO, AMENDS, REPLACES). The resulting graph is stored in Neo4j, where node/edge properties are chosen to support efficient Cypher queries.

3.4. Integration of RAG with Knowledge Graph

This section describes how Retrieval-Augmented Generation is coupled with the legal knowledge graph. Rather than treating the corpus as flat chunks, the pipeline uses (a) Structured Query Interpretation, (b) Graph-Aware Evidence Retrieval, and (c) Citation-Grounded Generation. Given a natural-language query q, the first step identifies the legal entities and relations mentioned in the question and represents them as a small set of semantic triples T = {(subject, predicate, object)}. Next, T is used to locate a focused neighborhood in the knowledge graph. The output is a subgraph $G' \subseteq K$ that contains candidate provisions (and, when needed, their referenced neighbors) that are likely to answer the query.

Finally, the model generates an answer conditioned on G′ and q, and the answer is required to cite the relevant legal units (article/clause/point). Operationally, the system proceeds in three stages: (1) PhoBERT-based NER and VnCoreNLP syntactic features are used to extract entities/relations and form T; (2) candidates are ranked with a hybrid score that combines SBERT similarity with a graph-derived confidence signal; and (3) the top-ranked provisions are passed to the LLM under a constrained instruction, followed by a lightweight citation-consistency check before returning the response.

3.5. Corpus Construction and Annotation Protocol

Legal texts were sourced from the Vietnamese government's official legal database (vbpl.vn) and restricted to statutes and subordinate instruments relevant to the three evaluation domains. Each document was cleaned (encoding normalization and boilerplate removal) and split along the hierarchy in Circular 25/2011/TT-BTP so that each Article/Clause/Point becomes an addressable unit. These units were imported into Neo4j as nodes connected by containment edges. References inside provisions—such as amendments or replacements—were detected with rule-based patterns and linked as cross-reference edges (REFERS_TO/AMENDS/REPLACES) whenever the target provision could be resolved unambiguously.

For evaluation, a set of natural-language legal questions was compiled to reflect common user intents (definitions, obligations, prohibitions, conditions, and procedural requirements). Gold answers and gold citations were annotated by reviewers with legal expertise using the statutory text as the sole source of truth. Disagreements were resolved through adjudication to produce a single reference answer and citation set per question. This protocol supports both answer-accuracy scoring and citation precision/recall as defined in Section 4.4.

3.6. Implementation Details and Key Hyperparameters

Neo4j is used as the back-end store, and Cypher queries retrieve candidate provisions together with their local neighborhoods. VnCoreNLP [19] handles Vietnamese tokenization/normalization, while PhoBERT [18] is applied for legal entity recognition. Initial retrieval is embedding-based: SBERT encodes the query and each provision, and candidates are ranked by cosine similarity. A separate graph-derived confidence term is computed from the retrieved subgraph (e.g., shorter/high-weight paths increase confidence), and this term is combined with semantic similarity in the hybrid scorer.

The hybrid score in Equation (1) uses $\alpha = 0.4$ and $\beta = 0.6$. The regeneration trigger uses a quality threshold $\theta = 0.75$ with a maximum of three retrieval–generation iterations to avoid uncontrolled loops. Top-k context selection is applied before prompting the LLM to keep prompts within context-length constraints. All components are modular so that alternative Vietnamese embedding models or LLM back-ends can be substituted without changing the graph schema.

4. Results and Discussion

4.1. System Architecture

The proposed system implements a modular architecture comprising five primary components, as shown in Figure 3. The Query Processing module handles text normalization and entity recognition using PhoBERT [18] (91.3% F1-score for legal entity extraction). The Knowledge Graph Store is implemented in Neo4j, containing 1,398 concept nodes and 5,257 relation edges. The Graph-based Retrieval module combines SBERT semantic similarity with graph traversal for context retrieval. The Context Augmentation component ranks and selects the top-k relevant provisions. Finally, the LLM Generation module produces a citation-grounded answer and applies lightweight citation verification against the retrieved context.

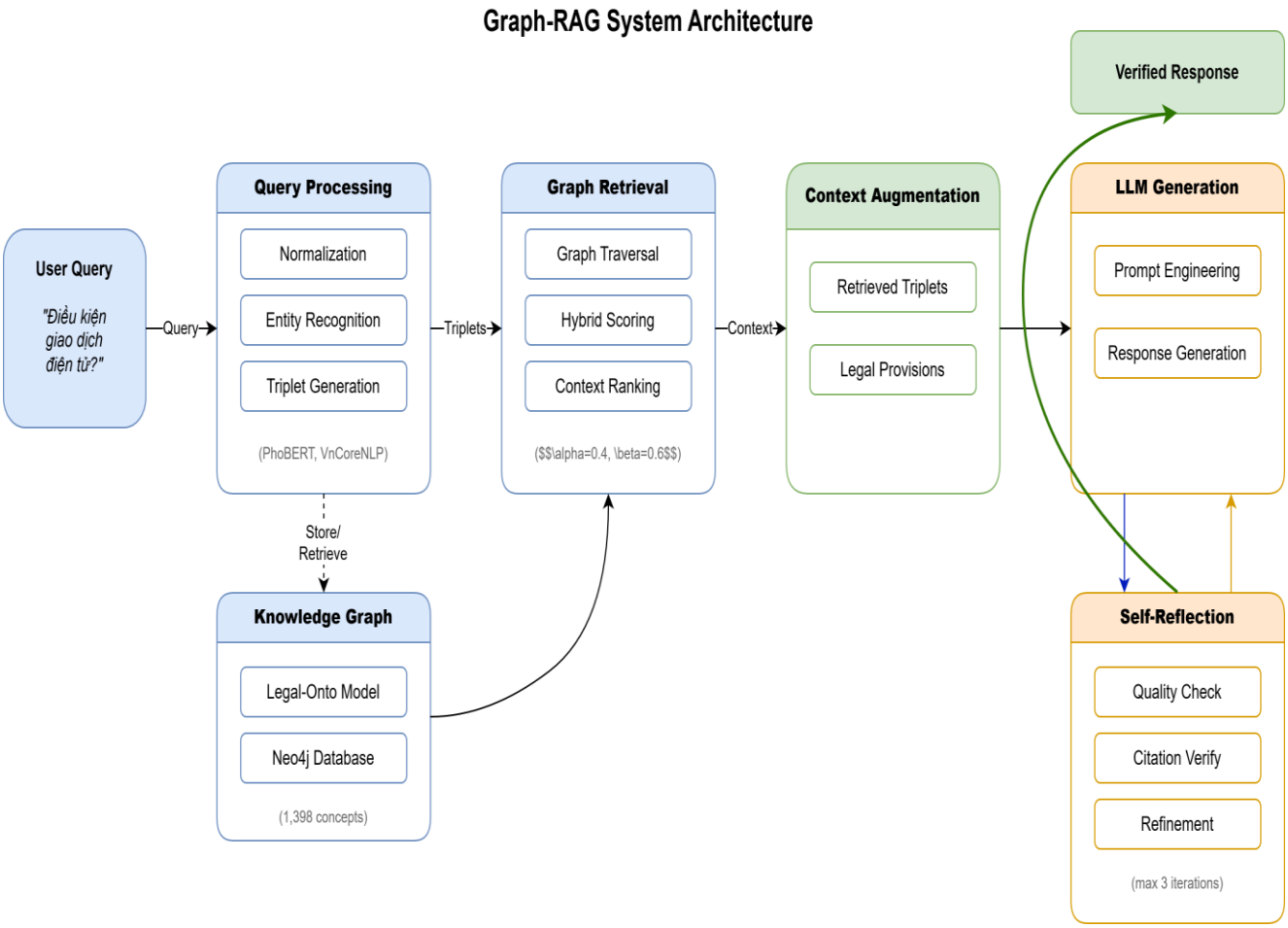

Fig. 3 The query flow of the proposed system architecture

4.2. Query Processing and Hybrid Scoring

Queries are processed in a small pipeline before retrieval. The text is normalized (including spelling/diacritics restoration via VnCoreNLP [19]), entities are detected with PhoBERT NER (BIO tagging), and a lightweight triplet representation is derived to align the query with the graph. Retrieval then uses a hybrid score:

$$score\_total = \alpha \times score\_similarity + \beta \times score\_confidence \quad (1)$$

Here, **score_similarity** comes from cosine similarity between SBERT embeddings of the query and candidate provisions. **score_confidence** summarizes structural support from the graph (path length and relation weights mapped through σ(x) = 1/(1+e^(x))). In the reported experiments, α = 0.4 and β = 0.6 place more weight on graph-derived confidence, reflecting the need to keep answers anchored to the statute hierarchy in legal QA.

Figure 4 Walks Through A Concrete Query ("Thời gian thử việc tối đa là bao lâu?"). After Detecting The Main Concept ("thời gian thử việc"), The Retriever Fetches The Relevant Provision(s) in the Labor Law (e.g., Article 25 on probation) and Supplies Them As Evidence For Generation. Before Finalizing The Response, the System Verifies That Any Cited Article/Clause Actually Appears In The Retrieved Evidence, Reducing Citation Drift.

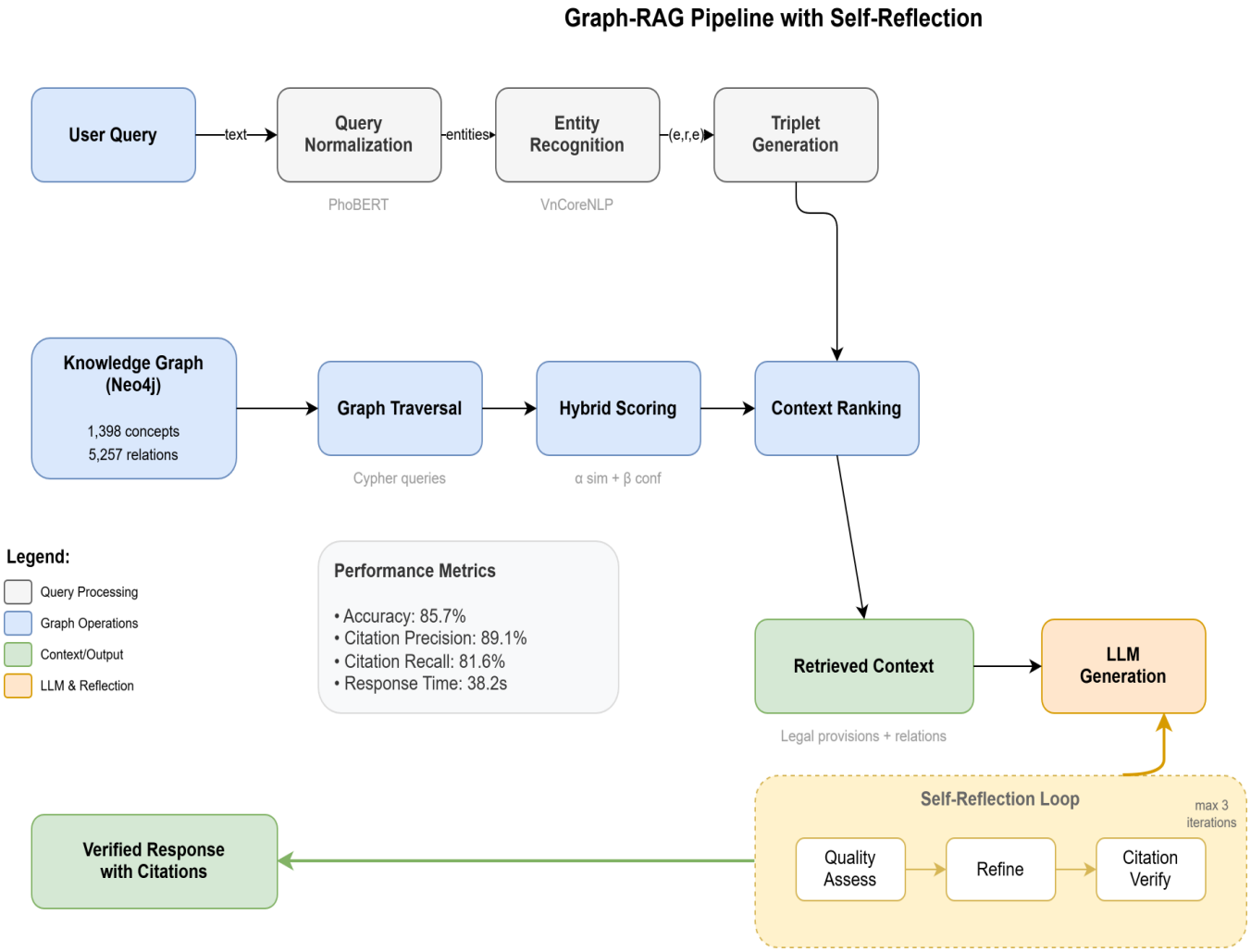

Fig. 4 End-to-end pipeline: query processing, graph-aware retrieval, answer generation, and citation verification

4.3. LLM Integration and Citation-Grounded Generation

Coupling the LLM with the graph-backed retriever is intended to address a common failure mode in legal QA: fluent answers that are not traceable to the statute. In the proposed design, the LLM receives only the evidence selected by the retrieval module and is instructed to cite the legal units that support each statement.

Instead of relying on open-ended prompting, the instruction explicitly constrains the model to (i) use only the provided provisions and (ii) produce citations at the level of article/clause/point. This constraint is the main mechanism for improving citation fidelity in the reported experiments.

Generation is followed by a simple consistency routine. Each cited unit is checked against the retrieved context and the hierarchy represented in the graph. If a citation cannot be validated (missing unit or inconsistent granularity), additional neighborhood evidence is retrieved, and the answer is regenerated, up to three iterations. The trigger uses a quality threshold $\theta = 0.75$ on the hybrid score to decide whether expansion is warranted.

In other words, the system rejects outputs whose citations cannot be grounded in G' and retries with an expanded context window, rather than returning an unverified citation.

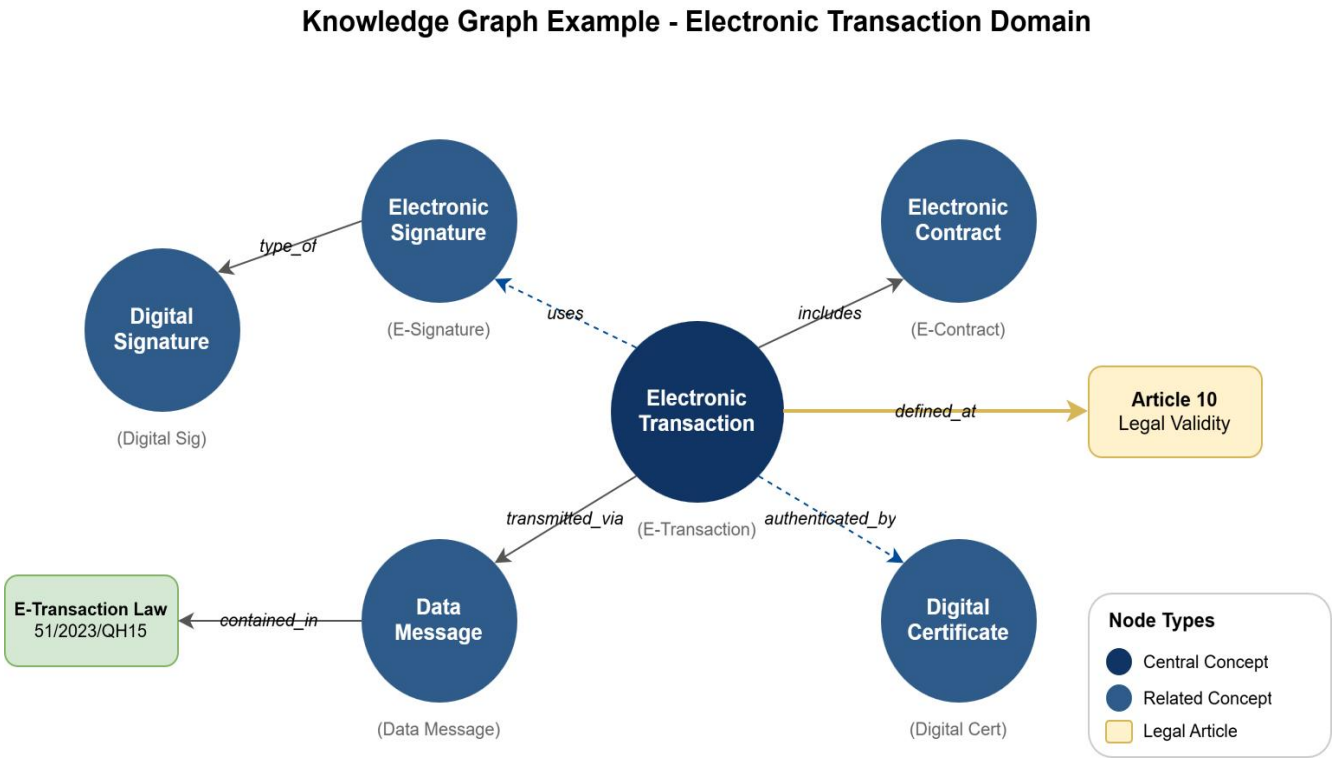

Fig. 5 Knowledge graph example: electronic transaction law showing concepts and relationships

Figure 5 visualizes a small fragment from the electronic transaction law (2023). The Node "Electronic Transaction" is Linked to Related Concepts such as "Electronic Signature" and "Digital Signature," and to the provision where key definitions are stated. When a question asks for a distinction (e.g., Digital vs. Electronic Signature), the retriever follows these links to pull the defining articles and presents them as evidence for the answer, which improves traceability compared to keyword-only retrieval.

4.4. Experimental Evaluation

Evaluation covers three Vietnamese legal domains with different drafting styles and vocabulary: the Electronic Transaction Law, Land Law, and Labor Law. The corpus comprises 56 official legal documents (laws, decrees, and circulars) collected from public sources for these domains.

Table 1. Dataset statistics across three vietnamese legal domains

Domain	Documents	Concepts	Relations
Electronic Transaction Law	15	342	1,247
Land Law	23	567	2,134
Labor Law	18	489	1,876
Total	**56**	**1,398**	**5,257**

The proposed method is compared against three baselines: (i) Keyword Search using BM25 for term-based retrieval; (ii) Pure LLM (GPT-4) without retrieval augmentation; and (iii) Dense RAG using SBERT embeddings over a document vector store without knowledge graph structure.

Evaluation metrics are defined as follows:

Answer Accuracy is computed as the ratio of correctly answered queries to total queries, where correctness is determined by legal experts who manually verify if the answer addresses the query based on actual legal provisions:

Accuracy = (Number of correct answers) / (Total queries)

Citation Precision measures the percentage of cited legal articles that are actually relevant to the answer:

Precision = (Relevant citations) / (Total citations)

Citation Recall measures the percentage of relevant legal articles that are correctly cited:

Recall = (Relevant citations) / (Total relevant articles)

Table 2. Main experimental results comparing the proposed method with baselines

System	Accuracy	Cit.Prec.	Cit.Rec.	Time(s)	Satisf.
Keyword Search	62.3%	71.2%	45.8%	2.1	2.8
Pure LLM	71.5%	N/A	N/A	8.3	3.4
Dense RAG	78.9%	82.4%	68.3%	24.6	3.9
Proposed Method	**85.7%**	**89.1%**	**81.6%**	38.2	**4.3**

Figures 6–8 summarize the quantitative results in Table 2. Editable figure sources (SVG and EMF) are provided as supplementary files. The observed accuracy and citation gains are consistent with three design choices: (i) splitting evidence in a way that respects the Article–Clause hierarchy, (ii) explicitly following statutory cross-references during retrieval, and (iii) ranking candidates using both semantic similarity and graph-supported confidence. The same design increases latency (graph expansion + verification), but it materially improves citation fidelity—often the deciding factor for legal usability.

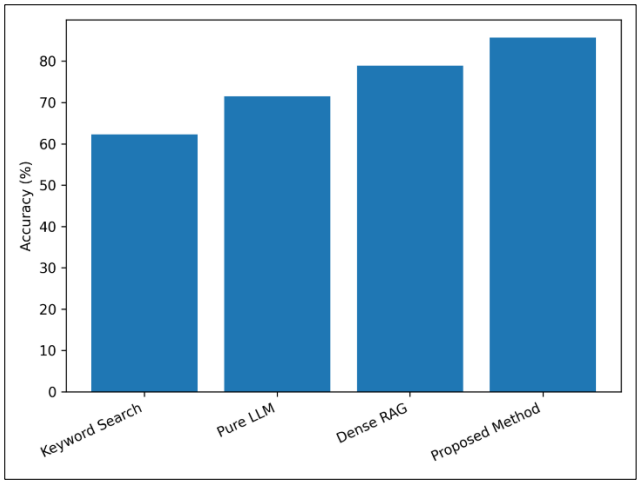

Fig. 6 Answer accuracy across baselines and the proposed method

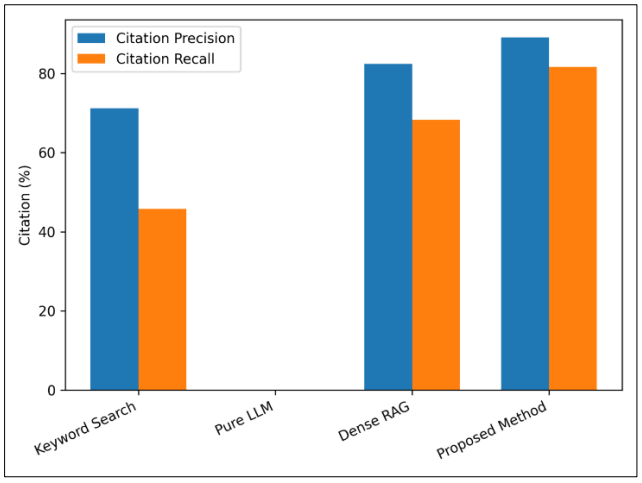

Fig. 7 Citation precision and recall across systems (N/A indicates not applicable)

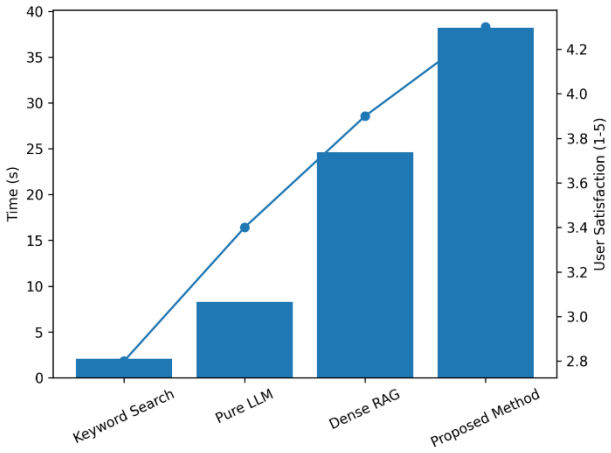

Fig. 8 Latency (seconds) and user satisfaction (1–5) across systems

Table 2 reports 85.7% answer accuracy for the proposed method, compared with 62.3% for keyword search and 78.9% for dense RAG. A plausible explanation is that the graph representation keeps retrieval aligned to article/clause granularity and surfaces cross-referenced provisions that a pure vector index may miss. The generation stage is also bounded by an explicit citation requirement and a post-check against retrieved evidence, which helps reduce unsupported statements. Citation precision reaches 89.1% and recall 81.6%, demonstrating the value of explicit relationship modeling. The higher response time (38.2s vs 24.6s for Dense RAG) is justified by the significant quality improvement, particularly for complex queries requiring multi-hop reasoning across multiple legal provisions.

4.5. Error Analysis

Manual inspection of incorrect or partially correct cases suggests several recurring patterns. Some questions are underspecified (missing actor/situation), leading to evidence from a nearby but non-applicable provision. Cross-reference chains can also be missed when the referenced document falls outside the current corpus snapshot. In addition, Vietnamese lexical variation (synonyms, abbreviations, domain phrasing) can depress embedding similarity—especially for short queries. Finally, citation mistakes tend to occur when multiple adjacent clauses look relevant and top-k truncation drops the truly supporting clause. These findings motivate stronger query clarification, cross-document expansion for references, and citation-aware context selection.

4.6. Scalability and Deployment Considerations

In the current prototype, the KG is hosted in Neo4j, and hybrid retrieval runs over 56 documents. To scale, the main engineering levers are straightforward: pre-compute and index provision embeddings; add targeted graph indexes for high-traffic relations (hierarchy and references); cache small neighborhoods for frequent concepts; and support incremental updates so amendments/new instruments do not require full re-ingestion. Because embedding retrieval, graph expansion, and generation are loosely coupled, the components can be deployed as separate services to scale and monitor independently.

4.7. Qualitative Assessment

Beyond automatic metrics, practical legal QA also benefits from qualitative assessment of usefulness and trust. In this revision, a manual inspection of representative outputs (including incorrect and partially correct cases discussed in Section 4.5) was conducted along four criteria: (i) Legal Correctness, (ii) Completeness With Respect To The Asked Intent, (iii) Citation Adequacy (whether the cited provisions support the answer), and (iv) Clarity for Non-Expert Users. This qualitative lens helps interpret why KG-aware retrieval improves citation grounding relative to Dense RAG, and surfaces failure modes that guide future evaluation design.

4.8. Limitations and Future Work

Several limitations should be noted. First, component-level ablation studies isolating the contributions of the ontology schema, graph traversal, hybrid scoring, and citation verification are deferred to an extended evaluation due to review-cycle constraints. Nevertheless, the comparison against Dense RAG provides an informative indirect ablation of the graph component, since Dense RAG can be interpreted as the KG-free variant that relies on semantic retrieval without explicit statutory relations.

Second, detailed statistical significance testing is reserved for future work involving larger query sets and repeated runs; we will report bootstrap confidence intervals for accuracy and citation metrics. Third, the evaluation currently focuses on three legal domains and a limited corpus snapshot; broader coverage will require additional domains, more extensive cross-law reference handling, and automated update mechanisms for amended provisions. Finally, the LLM component can still be sensitive to prompt formatting and context window limits; future work will explore more robust constrained decoding and structured intermediate representations for generation.

5. Conclusion

This paper presents a framework that integrates an ontology-driven knowledge graph with retrieval-augmented generation for Vietnamese legal question answering. The approach preserves statutory hierarchy and cross-references in a document-aware graph, and combines graph-based retrieval with LLM generation using a hybrid ranking mechanism that balances semantic similarity and graph-derived confidence. Experimental results on three evaluated legal domains indicate an answer accuracy of 85.7% and a citation precision of 89.1%, exceeding the baselines reported in Table 2.

These results are consistent with the hypothesis that combining semantic retrieval with document-aware graph structure improves citation grounding for Vietnamese statutes. Future work will extend the approach with larger-scale evaluation across additional legal domains, improved handling of multi-document reasoning (e.g., amendments and cross-law references), and more comprehensive human-centered evaluation with legal professionals.

Data Availability

The Vietnamese legal documents used in this study are publicly available from the Vietnamese government's legal document database (vbpl.vn). The knowledge graph schema, Neo4j database exports, and evaluation query sets can be made available upon reasonable request to the corresponding author for academic research purposes.

Funding Statement

This research is funded by Saigon University under grant number CSB.2025.87.

Authors' Contributions

Conceptualization, V.T.P. and P.N.M.; Methodology, V.T.P.; Software, P.N.M.; Validation, V.T.P. and P.N.M.; Formal Analysis, V.T.P.; Investigation, P.N.M.; Resources, V.T.P.; Data Curation, P.N.M.; Writing – Original Draft Preparation, V.T.P.; Writing – Review & Editing, V.T.P. and P.N.M.; Visualization, P.N.M.; Supervision, V.T.P.; Project Administration, V.T.P.; Funding Acquisition, V.T.P.

Acknowledgments

The authors thank the undergraduate students at Saigon University who contributed to the implementation of domain-specific knowledge graphs: Cao Bao Quynh and Truong Cong Phuc (Land Law domain), Nguyen Ngoc Son (Electronic Transaction Law domain), Nguyen Thanh Duy and Tien Minh Vy (Labor Law domain).

References

[1] Abdul Qadir Khan et al., "Knowledge-Based Anomaly Detection: Survey, Challenges, and Future Directions," *Engineering Applications of Artificial Intelligence*, vol. 136, pp. 1-30, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Cornelius T. Leondes, Knowledge based Systems Techniques and Applications, Volume 1, Academic Press. [Online]. Available: https://booksite.elsevier.com/9780124438750/pdf/vol-1-fm.pdf

[3] Zhenyu Li et al., "FlexKBQA: A Flexible LLM-Powered Framework for Few-Shot Knowledge Base Question Answering," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 17, pp. 18608-18616, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Zhaorun Chen et al., "Agentpoison: Red-Teaming LLM Agents via Poisoning Memory or Knowledge Bases," *Advances Neural Information Processing Systems (NeurIPS)*, pp. 130185-130213, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Jinheon Baek, Alham Fikri Aji, and Amir Saffari, "Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering," *Proceedings of the 1st Workshop on Natural Language Reasoning and Structured Explanations (NLRSE)*, Toronto, Canada, pp. 78-106, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Patrick Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 9459-9474, 2020. [Google Scholar] [Publisher Link]

[7] Vladimir Karpukhin et al., "Dense Passage Retrieval for Open-Domain Question Answering," *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6769-6781, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Carl Yang et al., "Knowledge Graph and LLM Co-Learning via Structure-Oriented Retrieval-Augmented Generation," *IEEE Data Engineering Bulletin*, vol. 48, no. 4, pp. 1-38, 2025. [Google Scholar] [Publisher Link]

[9] Marvin Hofer et al., "Construction of Knowledge Graphs: Current State and Challenges," *Information*, vol. 15, no. 8, pp. 1-61, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Aidan Hogan et al., "Knowledge Graphs," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1-37, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Seungone Kim et al., "The CoT Collection: Improving Zero-Shot and Few-shot Learning of Language Models via Chain-of-Thought Fine-Tuning," *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Singapore, pp. 12685-12708, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Boci Peng et al., "Graph Retrieval-Augmented Generation: A Survey," *arXiv Preprint*, pp. 1-41, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Haoyu Han et al., "Retrieval-Augmented Generation with Graphs (GraphRAG)," *arXiv Preprint*, pp. 1-88, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Neel Guha et al., "LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models," *arXiv Preprint*, pp. 1-143, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Ilias Chalkidis et al., "LEGAL-BERT: The Muppets Straight Out of Law School," *Findings of the Association for Computational Linguistics: EMNLP*, pp. 2898-2904, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Zhiwei Fei et al., "LawBench: Benchmarking Legal Knowledge of Large Language Models," *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Miami, Florida, USA, pp. 7933-7692, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Hung Q. Ngo et al., "Ontology Knowledge Map Approach towards Building Linked Data for Vietnamese Legal Applications," *Vietnam Journal of Computer Science*, vol. 11, no. 2, pp. 323-342, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Dat Quoc Nguyen, and Anh Tuan Nguyen, "PhoBERT: Pre-Trained Language Models for Vietnamese," *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 1037-1042, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] Thanh Vu et al., "VnCoreNLP: A Vietnamese Natural Language Processing Toolkit," *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, New Orleans, Louisiana, pp. 56-60, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[20] Thi-Hai-Yen Vuong et al., "Constructing a Knowledge Graph for Vietnamese Legal Cases with Heterogeneous Graphs," *2023 15th International Conference on Knowledge and Systems Engineering (KSE)*, Hanoi, Vietnam, pp. 1-6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Ankita Gupta, and Frank Schilder, "Improving Legal Question Answering through Structured Knowledge Representation," *Proceedings of the First Argument Mining and Empirical Legal Research Workshop (AMELR 2025)*, Chicago, United States, pp. 39-45, 2025. [Google Scholar] [Publisher Link]