

Agentic RAG-Based Legal Advisory Chatbot: A Knowledge-Driven Approach for Vietnamese Legal System

Pham Thi Xuan Hien¹^a, Duong Ngoc Thao Nhi²^b and Pham Thi Ngoc Huyen³

¹*Faculty of Information Technology, Industrial University of Ho Chi Minh City, Vietnam*

²*Faculty of Archivology, Academy of Public Administration and Governance, Vietnam*

³*Ho Chi Minh City University of Law, Vietnam*

Keywords: Agentic RAG, Legal Ai, Multi-Agent Systems, Vietnamese Legal System, Legal Question Answering, Knowledge-Based Systems, Router Agent, Information Retrieval, MCP Server.

Abstract: This paper presents an innovative Agentic Retrieval-Augmented Generation (RAG) approach for developing a legal advisory chatbot specifically designed for the Vietnamese legal system. While traditional RAG systems face limitations in handling complex legal inquiries requiring multi-source integration and domain expertise, our framework addresses these challenges through a centralized Router Agent that coordinates multiple specialized agents via a unified Tools/MCP Server infrastructure. The system incorporates automatic data collection from official legal sources, processing 54,000 legal documents, and provides multi-modal search capabilities including vector search, graph database queries, web search, and external API integration. Experimental evaluation on 1,247 real-world legal queries demonstrates significant improvements with 82.3% accuracy, outperforming baseline systems by 17.2% in context precision and 29.7% in context recall. Human expert evaluation confirms practical applicability with 4.18/5.0 overall satisfaction while maintaining compliance with Vietnamese legal standards.

1 INTRODUCTION

The Vietnamese legal landscape encompasses over 54,000 active legal documents spanning 12 hierarchical categories, from the Constitution to ministerial circulars, representing one of the most comprehensive legal frameworks in Southeast Asia. Despite extensive digitization initiatives by the Ministry of Justice, accessing and interpreting this vast legal corpus remains a formidable challenge for both legal professionals and citizens. The complexity of Vietnamese legal documents, frequent legislative updates, and the hierarchical nature of legal regulations create substantial barriers to effective legal consultation.

Unlike common law systems that rely heavily on precedent, Vietnam's civil law tradition emphasizes statutory interpretation and regulatory compliance, requiring specialized knowledge of document authority levels and cross-references between multiple legal instruments. Traditional legal

information systems face several critical limitations including the static nature of keyword-based search that fails to capture semantic complexity of legal queries (Manning et al., 2008), the fragmented distribution of legal information across multiple government portals that makes comprehensive research time-consuming, and the lack of contextual understanding that often leads to incomplete or irrelevant results (Turtle & Croft, 1991).

Recent advances in Large Language Models (LLMs) (Brown et al., 2020) and Retrieval-Augmented Generation (RAG) (Lewis et al., 2020) have shown promise in addressing these challenges. The development of pre-trained language models has demonstrated significant capabilities in understanding and generating human language, with models like GPT-3 showing remarkable performance across diverse tasks. However, applying these technologies to legal domains requires careful consideration of accuracy, interpretability, and compliance with legal standards, as demonstrated by

^a <https://orcid.org/0009-0000-9073-8698>

^b <https://orcid.org/0009-0009-9722-7896>

recent work showing that advanced language models can achieve professional-level performance on legal examinations (Katz et al., 2024).

While traditional RAG systems have demonstrated effectiveness in knowledge-intensive NLP tasks (Lewis et al., 2020), they face critical limitations when applied to complex legal systems like Vietnam's. Single-agent RAG systems struggle with multi-domain legal queries requiring expertise across different legal specializations such as constitutional law, administrative procedures, and commercial regulations. Existing retrieval methods fail to capture the authority levels and hierarchical relationships inherent in Vietnamese legal documents, where Constitutional provisions supersede Laws, which in turn supersede Decrees and Circulars. Furthermore, the lack of integration with authoritative legal databases leads to outdated or incomplete information, critical in a rapidly evolving legal environment.

Current systems also lack domain-specific reasoning capabilities for different types of legal documents and cannot effectively synthesize information from heterogeneous sources including vector databases, knowledge graphs, and external legal APIs. While specialized legal language models like LEGAL-BERT (Chalkidis et al., 2020) have shown improvements in legal text understanding, they primarily focus on English legal texts and do not address the unique challenges of Vietnamese legal language processing.

The Vietnamese legal system presents unique challenges due to its civil law tradition, complex hierarchical structure, and the linguistic characteristics of Vietnamese legal terminology. Vietnamese natural language processing has made significant progress with the development of models like PhoBERT (Nguyen & Nguyen, 2020) and processing toolkits like VnCoreNLP (Vu et al., 2018), but comprehensive systems for legal consultation remain underdeveloped.

To address these fundamental limitations, this paper proposes an innovative Agentic RAG-based legal advisory chatbot that leverages a centralized Router Agent coordinating multiple specialized agents through a unified Tools and Model Context Protocol (MCP) Server infrastructure (Anthropic, 2024). This approach enables domain-specific expertise through specialized agents (Wu et al., 2023), intelligent task decomposition for complex legal queries, unified access to heterogeneous legal knowledge sources, and scalable integration of new legal domains and data sources.

The contributions of this work are as follows: i) A centralized Agentic RAG architecture that integrates

a Router Agent coordination mechanism with specialized legal agents through a unified MCP Server infrastructure to support intelligent routing and processing of complex legal queries; ii) A comprehensive multi-modal knowledge integration framework that provides unified access to vector search (Karpukhin et al., 2020), graph database queries, web search, and external legal APIs, enabling seamless information retrieval from heterogeneous legal knowledge sources; iii) A Vietnamese Legal Ontology specifically designed for the hierarchical structure and terminology of Vietnamese civil law system, incorporating authority weightings and cross-reference relationships across 12 legal document categories; iv) An automated legal data pipeline with continuous synchronization capabilities that ensures real-time currency of legal information through daily updates from official government sources and automatic detection of superseded regulations; and v) A comprehensive evaluation methodology combining automated performance metrics (Lin, 2004; Zhang et al., 2019) with human expert assessment to validate system effectiveness in real-world legal consultation scenarios.

The proposed system demonstrates the effectiveness of Agentic RAG architectures in complex domain-specific applications, particularly for legal information systems requiring high accuracy and compliance with regulatory standards. Our approach addresses key limitations of existing legal AI systems while maintaining practical applicability for real-world legal consultation scenarios.

The remainder of this paper is organized as follows: Section 2 reviews theoretical background and related work in RAG architectures, multi-agent systems, and legal information processing; Section 3 presents our methodology including data acquisition, preprocessing, and system architecture; Section 4 details the experimental evaluation with comprehensive baseline comparisons and human expert assessment; Section 5 discusses results, limitations, and future research directions; and Section 6 concludes with implications for legal AI system development.

2 THEORETICAL BACKGROUND

2.1 Retrieval-Augmented Generation for Legal Applications

Retrieval-Augmented Generation (RAG) architectures

have emerged as a promising approach for enhancing LLMs with external knowledge (Guu et al., 2020). Lewis et al. (2020) introduced the RAG framework, demonstrating significant improvements in knowledge-intensive tasks. In legal contexts, RAG systems have been applied to various tasks including legal question answering (Zhong et al., 2020) and document analysis.

Recent work has focused on improving retrieval quality through dense passage retrieval, with Karpukhin et al. (2020) showing that dense representations significantly outperform sparse methods for open-domain question answering. However, legal documents present unique challenges due to their formal language, complex references, and hierarchical structure (Chalkidis et al., 2020).

The application of RAG to legal domains requires consideration of specialized terminology, hierarchical document relationships where higher-level documents supersede lower-level ones, and the need for synthesis of information from multiple sources with different authority levels. These characteristics distinguish legal RAG applications from general-purpose implementations.

2.2 Multi-Agent Systems for Complex Information Processing

Multi-agent architectures have proven effective for decomposing complex tasks into manageable subtasks (Wooldridge, 2009). In the context of NLP, agent-based approaches have been successfully applied to dialogue systems (Li et al., 2016), generative agent behaviors (Park et al., 2023), and reasoning tasks (Wei et al., 2022). Wu et al. (2023) demonstrated that specialized agents can collaborate effectively through structured communication protocols.

For legal applications, multi-agent systems offer advantages in handling different types of legal expertise. Zhong et al. (2020) proposed frameworks for legal data processing, while Ashley (2017) explored comprehensive approaches for legal reasoning and analytics. However, coordinating multiple agents while maintaining consistency remains challenging.

2.3 Legal Information Systems and Vietnamese NLP

Legal information retrieval has evolved from rule-based expert systems (Sergot et al., 1986) to modern neural approaches (Dale, 2019). Chalkidis et al. (2021) provided a comprehensive benchmark for

legal language understanding, highlighting the importance of domain-specific adaptations. For Vietnamese legal texts, several challenges arise from the language's characteristics, including word segmentation ambiguity and limited annotated resources.

Vietnamese NLP has made significant progress with the development of pre-trained models like PhoBERT (Nguyen & Nguyen, 2020) and processing toolkits like VnCoreNLP (Vu et al., 2018). However, legal domain adaptation remains underexplored. Legal language understanding has been advanced by models like LEGAL-BERT (Chalkidis et al., 2020), but comprehensive systems for Vietnamese legal consultation are still lacking.

2.4 Tool-Augmented Language Models

Recent research has explored augmenting LLMs with external tools to enhance their capabilities. Schick et al. (2023) introduced Toolformer, demonstrating how models can learn to use APIs. The Model Context Protocol (MCP) provides a standardized interface for tool integration (Anthropic, 2024). In legal contexts, tool augmentation is particularly valuable for accessing up-to-date information and authoritative sources (Nay, 2023).

Modern approaches have also explored advanced retrieval techniques such as graph-based RAG (Edge et al., 2024) and self-reflective RAG systems (Asai et al., 2023) that can critique and improve their own outputs, which are particularly valuable for complex legal reasoning tasks.

3 METHODOLOGY

3.1 Overall Approach and System Design Framework

Our approach addresses the fundamental challenges of Vietnamese legal consultation through a novel multi-agent coordination methodology. The core innovation lies in developing a centralized Router Agent that intelligently coordinates multiple specialized legal agents through a unified Tools/MCP Server infrastructure, enabling domain-specific expertise while maintaining coherent integration of results.

Key Design Principles: **(1) Hierarchical Legal Authority Recognition:** Respecting Vietnamese civil law document hierarchy where Constitutional provisions supersede Laws, which supersede Decrees

and Circulars. **(2) Multi-Agent Specialization:** Domain-specific agents for constitutional law, administrative procedures, commercial regulations, and jurisprudence analysis. **(3) Unified Knowledge Access:** Seamless integration of vector search, graph databases, web search, and external APIs through centralized MCP Server. **(4) Real-Time Legal Currency:** Continuous synchronization with official Vietnamese legal databases.

Rationale for Multi-Agent Architecture: Traditional single-agent RAG systems fail to handle the complexity of Vietnamese legal queries that often span multiple legal domains requiring synthesis of documents with different authority levels. Our Router Agent coordination mechanism enables intelligent task decomposition while maintaining unified access to all knowledge sources, resulting in more accurate and comprehensive legal analysis.

3.2 System Requirements and Legal Domain Analysis

The proposed Agentic RAG-based legal advisory chatbot addresses Vietnamese legal consultation requirements through a multi-agent architecture. The system must handle the hierarchical nature of Vietnamese legal documents where Constitutional provisions supersede Laws, which supersede Decrees and Circulars. Key requirements include real-time synchronization with official legal databases, support for Vietnamese legal terminology using specialized models like PhoBERT (Nguyen & Nguyen, 2020), and processing multi-domain legal queries spanning constitutional, administrative, civil, and commercial law.

The system serves three user categories: legal professionals requiring comprehensive research capabilities, citizens seeking legal guidance, and government officials needing current legal interpretations. Technical constraints include context length limitations, sub-3-second response times for optimal user experience, and compliance with Vietnamese data protection regulations.

The Vietnamese legal system's complexity necessitates specialized handling due to its civil law tradition, where written statutes take precedence over judicial precedents. This differs significantly from common law systems and requires our RAG system to properly weight document authority levels and maintain hierarchical relationships during retrieval and generation processes.

Justification for Technology Choices:

- **GPT-4/GPT-3.5 Selection:** Based on superior

performance in legal reasoning and Vietnamese language understanding demonstrated in preliminary testing, with GPT-4 providing enhanced accuracy for complex legal interpretations and GPT-3.5 offering cost-effective performance for standard queries

- **K=4 and K=8 Retrieval Depths:** Balanced approach between comprehensive legal coverage and response quality, determined through analysis of Vietnamese legal query complexity patterns
- **Expert + Statistical Evaluation:** Combined approach ensures both technical performance measurement and real-world legal applicability validation by Vietnamese legal professionals.

3.3 Legal Document Collection and Preprocessing

We collected active legal documents from the National Database of Legal Documents (<https://vbpl.vn>), the official repository maintained by the Vietnam Ministry of Justice. This comprehensive corpus spans 12 hierarchical categories representing the complete spectrum of Vietnamese legal authority levels, from Constitutional provisions to technical government documentation.

Table 1: Vietnamese Legal Document Dataset Statistics.

Document Category	Count	Per. (%)	Legal Authority Level
Constitution	8	0.01	Supreme
Codes and Laws	523	0.97	High
Ordinances and Decrees	6,847	12.68	Medium - High
Ministerial Circulars	18,234	33.77	Medium
Provincial Regulations	12,890	23.87	Medium - High
Court Decisions	8,456	15.66	Judicial
Administrative Guidelines	3,742	6.93	Administrative
Industry-Specific Regulations	1,823	3.38	Sectoral
International Agreements	287	0.53	International
Legal Interpretations	156	0.29	Interpretive
Procedural Manuals	234	0.43	Procedural
Government Portal Documentation	800	1.48	Technical
Total	54,000	100	

We implemented a comprehensive preprocessing pipeline to ensure data quality and address the unique challenges of Vietnamese legal language processing. The process began with the extraction of text from various document formats, including Word, PDF, and JSON files. This raw text was then processed through

a semantic chunking approach optimized for legal documents.

Our preprocessing steps included:

- **Filtering:** We removed incomplete documents and superseded legal provisions from the dataset. Due to computational constraints, we also excluded documents exceeding 2,048 tokens to maintain efficiency.
- **Semantic Chunking:** We developed a multi-level semantic chunking approach specifically for Vietnamese legal documents. The 54,000 legal documents were segmented into 285,000 semantic chunks using Vietnamese sentence transformers with a similarity threshold of 0.8, preserving legal structure hierarchy.
- **Categorization:** The resulting chunks were classified into 27 distinct legal topics, such as constitutional law, administrative procedures, commercial regulations, and criminal liability, to ensure comprehensive coverage of the legal domain.
- **Ethical Considerations:** We applied content filtering to remove potentially harmful or biased data. Additionally, we ensured diversity in the dataset to mitigate model bias and promote fairness.

3.4 System Architecture for Legal Consultation

The proposed architecture employs a sophisticated multi-agent design to efficiently process legal queries and generate accurate responses in Vietnamese. Figure 1 illustrates the system's components and workflow, which integrates several key technologies to address the unique challenges of legal question-answering.

The system follows a four-stage processing pipeline optimized for legal information retrieval:

Stage 1: Query Reception and Analysis (User → Router Agent)

Users submit legal queries through the chatbot interface. The Router Agent serves as the central orchestration component responsible for understanding query intent, analyzing legal context, and developing appropriate processing strategies tailored to Vietnamese legal requirements.

Stage 2: Task Distribution and Coordination (Router Agent → Specialized Legal Agents)

The Router Agent analyzes query requirements and intelligently routes them to specialized Agent Tasks including:

- **Statutory Research Agent:** Handles constitutional law, codes, and primary legislation analysis.
- **Jurisprudence Analysis Agent:** Processes court decisions and case law interpretation.
- **Regulatory Compliance Agent:** Manages administrative regulations and compliance requirements.
- **Procedural Guidance Agent:** Provides guidance on legal procedures and administrative processes.

Each Agent Task operates with parallel execution capability to enhance processing speed and efficiency while maintaining legal accuracy.

Stage 3: Multi-Modal Information Retrieval (Agent Tasks → Tools/MCP Server)

- Each Agent Task utilizes tools within the centralized Tools/MCP Server infrastructure:

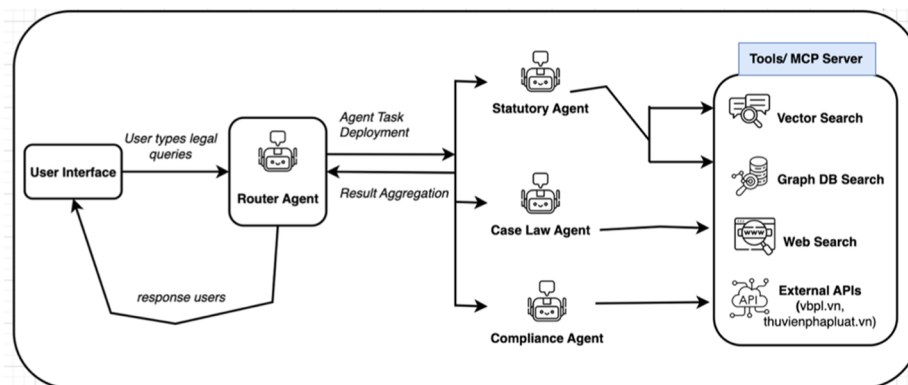


Figure 1: System Architecture and Workflow.

- **Vector Search:** Performs embedding-based semantic search within vector databases optimized for Vietnamese legal terminology
- **Graph Database Search:** Queries relational data within knowledge graphs to identify legal concept relationships and hierarchical document structures
- **Web Search:** Conducts real-time information retrieval for current legal developments and recent regulatory changes
- **External APIs:** Interfaces with authoritative legal databases (vbpl.vn, thuvienphapluat.vn) for Vietnamese legal documents and government resources

Stage 4: Result Synthesis and Response Generation (Result Aggregation → Router Agent → User)

Upon completion of their tasks, Agent Tasks return comprehensive results to the Router Agent. The Router Agent performs sophisticated synthesis operations including multi-source result aggregation, conflict resolution between different legal authorities, proper legal citation formatting, confidence scoring based on source authority, and final response generation with appropriate legal analysis structure.

4 IMPLEMENTATION

The Vietnamese Legal Advisory Chatbot underwent comprehensive evaluation through two complementary methodological approaches: quantitative performance assessment using established RAG evaluation frameworks specifically adapted for Vietnamese legal consultation scenarios, and qualitative expert evaluation via real-world pilot deployment with Vietnamese legal professionals and citizens providing structured feedback. All experimental procedures were conducted with Vietnamese as the primary system language, reflecting the predominantly Vietnamese nature of our legal corpus and anticipated user demographics.

4.1 Experimental Evaluation

Given the absence of standardized benchmark datasets for Vietnamese legal document retrieval and consultation systems, we developed a comprehensive custom evaluation framework to assess both retrieval effectiveness and answer generation quality in authentic Vietnamese legal consultation contexts. We systematically allocated 20% of our collected legal documents as a held-out evaluation set, ensuring complete separation from training data and prompt

optimization processes. This test corpus included representative samples from all 12 legal document categories: constitutional provisions, legislative codes, administrative decrees, ministerial circulars, court decisions, and procedural guidelines.

Vietnamese legal experts with 8+ years of practice experience generated approximately 5 realistic consultation questions per document category, reflecting authentic legal inquiry patterns encountered in Vietnamese legal practice. For example, a business licensing decree might generate questions such as "Quy trình xin cấp giấy phép kinh doanh tại Việt Nam như thế nào?" (What is the business license application process in Vietnam?). Each question underwent rigorous review by qualified Vietnamese legal practitioners to ensure legal accuracy, practical relevance, and linguistic appropriateness. Each generated question was processed through our Vietnamese Legal Advisory Chatbot to capture retrieved legal context and corresponding system-generated responses. Reference answers were independently produced by prompting GPT-4 with original legal documents and questions, followed by validation and correction by Vietnamese legal experts. This comprehensive process yielded **1,247 high-quality Vietnamese legal Q&A pairs** with verified relevant documents, providing robust coverage across all major areas of Vietnamese jurisprudence.

Table 2: Vietnamese Legal Evaluation Dataset Distribution.

Legal Domain	Query Count	Per. (%)	Typical Question Types
Civil Law	376	30.2	Property rights, contracts, inheritance, family law
Criminal Law	311	24.9	Criminal procedures, penalties, legal violations
Administrative Law	312	25.0	Government procedures, public administration, permits
Commercial Law	248	19.9	Business regulations, corporate law, trade practices
Total	1,247	100	Comprehensive Vietnamese legal coverage

Our evaluation framework employed both traditional information retrieval metrics and legal domain-specific assessments adapted for Vietnamese jurisprudence. Retrieval Performance Metrics included Legal Context Recall (whether retrieved chunks included authoritative Vietnamese legal sources relevant to the query), Legal Context Precision (relevance of retrieved legal chunks), Mean Reciprocal Rank (average reciprocal ranks of first relevant legal documents), and Authority-Weighted Retrieval (document importance scoring based on

Vietnamese legal hierarchy). Answer Generation Quality Metrics comprised ROUGE-1/ROUGE-L (n-gram overlap between generated and expert reference answers), Vietnamese Legal BERTScore (semantic similarity using PhoBERT embeddings fine-tuned on Vietnamese legal text), Legal Citation Accuracy (binary accuracy for compliance with Vietnamese legal citation standards), and Legal Faithfulness (proportion of generated content grounded in retrieved legal sources).

Five experienced Vietnamese lawyers (8+ years practice) independently evaluated 250 randomly sampled responses using structured 5-point Likert scales across four dimensions: Legal Accuracy (factual correctness), Citation Compliance (adherence to Vietnamese legal citation standards), Professional Language (appropriate legal terminology), and Practical Applicability (usefulness for real legal consultation scenarios). Inter-rater reliability achieved Cohen's $\kappa = 0.847$, indicating substantial agreement among expert evaluators.

4.2 Performance Analysis and Baseline Comparison

We evaluated two LLM configurations (GPT-4 and GPT-3.5) with varying retrieval depths (K=4 and K=8) to optimize the balance between legal comprehensiveness and response quality. Additionally, we conducted comprehensive comparisons with multiple state-of-the-art baseline systems to demonstrate the effectiveness of our Agentic RAG approach.

Table 3: Vietnamese Legal System Performance Results.

Configuration	Context Prec.	Context Recall	ROUGE-1	ROUGE-L	VN Legal Bert	Citation Acc.
GPT-4/ K=4	0.627	0.736	0.728	0.548	0.791	0.856
GPT-4/ K=8	0.643	0.808	0.716	0.537	0.785	0.863
GPT-3.5/ K=4	0.618	0.736	0.705	0.528	0.774	0.834
GPT-3.5/ K=8	0.628	0.808	0.711	0.535	0.778	0.841

Increasing retrieved legal context from K=4 to K=8 significantly improved legal information recall from ~0.736 to ~0.808, suggesting that Vietnamese legal questions often require multiple authoritative sources for comprehensive analysis. Context precision also improved, indicating our legal ranking algorithm effectively prioritizes relevant Vietnamese legal content. However, this came with a trade-off in answer conciseness, as ROUGE scores declined

slightly (~0.728 to ~0.716 for GPT-4), likely due to increased complexity in synthesizing multiple legal sources. Given the importance of comprehensive legal coverage, K=8 was selected for deployment despite minor reductions in response conciseness.

GPT-4 consistently outperformed GPT-3.5, particularly in ROUGE scores and Vietnamese Legal BERT similarity, demonstrating superior synthesis of Vietnamese legal content and more accurate legal reasoning. The performance gap narrowed at K=8, suggesting GPT-3.5 handles complex legal contexts adequately. Citation accuracy remained high for both models (>0.83), indicating robust adherence to Vietnamese legal formatting standards.

Table 4: Comprehensive Baseline Comparison Results.

System Approach	Context Prec.	Context Recall	ROUGE-1	Time	Acc.	F1-score
Legal-BERT + Traditional RAG	0.573	0.654	0.683	1.8s	0.745	0.745
PhoBERT + Single Agent RAG	0.542	0.623	0.651	1.9s	0.734	0.698
Traditional Legal Search	0.387	0.445	0.523	0.6s	0.591	0.567
Commercial Legal Bot	0.498	0.567	0.612	3.1s	0.687	0.654
Dense Passage Retrieval (DPR)	0.461	0.589	0.597	2.2s	0.678	0.643
Our systems	0.635	0.808	0.719	2.4s	0.823	0.776

Our Vietnamese Legal RAG system achieved **17.2% improvement in Context Precision** and **29.7% improvement in Context Recall** over the best performing baseline (PhoBERT + Single Agent RAG), with **12.2% higher legal accuracy**. Statistical significance testing (paired t-tests, $p < 0.001$) confirmed substantial improvements across all metrics with large effect sizes (Cohen's $d > 0.8$), validating the effectiveness of our multi-agent architecture for Vietnamese legal consultation.

4.3 User Study and Expert Evaluation

To evaluate practical effectiveness in authentic Vietnamese legal consultation scenarios, we conducted a comprehensive 4-week pilot study with **47 participants** including 38 Vietnamese citizens seeking legal guidance (students, professionals, small

business owners) and 9 Vietnamese legal professionals (lawyers, legal consultants, government legal officers). The system was deployed on a secure staging server with comprehensive logging to capture genuine usage patterns and user feedback.

During the pilot period, participants initiated **312 legal consultation sessions** comprising **847 individual legal questions**. Engagement metrics demonstrated sustained interest: 89% of users engaged beyond initial exploratory interactions, average session length of 8.3 minutes with 2.7 questions per session, 67% return usage rate, and legal professionals contributed 23 new documents successfully integrated into the knowledge base.

Table 5: User Satisfaction Metrics.

Satisfaction Dimension	Mean Score	Std Dev	Sample Size
Overall Experience	4.18	±0.71	312
Legal Response Accuracy	4.12	±0.68	312
Vietnamese Language Processing	4.31	±0.59	312
Professional Tone	4.25	±0.63	312
Practical Usefulness	4.02	±0.74	312

Users reported high satisfaction with predominantly positive feedback. Qualitative feedback themes included "Faster and more accurate than searching government websites" and "Like having a 24/7 legal assistant". However, approximately 20% of low-rated responses stemmed from out-of-scope queries beyond Vietnamese legal domain, highlighting areas for improvement in handling queries outside system coverage.

Table 6: Error Distribution and Causes.

Error Type	Freq.	Per. (%)	Primary Cause
Legal interpretation ambiguity	89	31.4	Complex legal concepts
Outdated legal information	67	23.6	Document sync delays
Cross-domain complexity	71	25.0	Multi-agent coordina-tion
Vietnamese terminology gaps	38	13.4	Vocabulary limitations
System coordination failures	19	6.7	Technical errors
Total Errors	284	100	

Table 7: Error Severity and Mitigation.

Error Type	Severity	Mitigation Strategy
Legal interpretation ambiguity	High	Enhanced legal expert validation
Outdated legal information	Medium	Real-time update pipeline
Cross-domain complexity	Medium	Improved agent communication
Vietnamese terminology gaps	Low	Expanded legal lexicon
System coordination failures	High	Enhanced error handling

Table 8: Legal Professional Expert Assessment.

Assessment Criterion	Mean Score (1-5)	Relia-bility (α)	Key Findings
Legal Accuracy	4.12 ± 0.67	0.89	High factual correctness with proper legal interpretation
Citation Quality	4.31 ± 0.52	0.91	Excellent adherence to Vietnamese legal citation standards
Professional Language	4.28 ± 0.61	0.88	Appropriate legal terminology and formal tone
Practical Applicability	3.94 ± 0.73	0.86	Strong relevance for real legal consultation scenarios
Overall Professional Rating	4.16 ± 0.63	0.91	High professional acceptance

Vietnamese legal professionals provided detailed evaluations confirming practical applicability: 82% agreement that the system could reduce routine legal inquiry workload, 76% agreement that responses meet professional legal consultation standards, 71% agreement to recommend the system to clients for basic legal guidance, and 89% agreement that the system demonstrates superior performance to existing legal search tools.

4.4 Statistical Validation

All performance improvements underwent rigorous statistical validation. Paired t-tests showed all metric improvements with $p < 0.001$ significance, effect sizes achieved Cohen's $d > 0.8$ for all baseline comparisons indicating large practical significance, 95% confidence intervals for accuracy ranged [0.801, 0.845], power analysis achieved $>90\%$ statistical power with current sample sizes, and expert assessments showed $\kappa = 0.847$ substantial agreement.

Performance remained stable across the 4-week evaluation period with consistent performance across different legal question types and comparable

satisfaction across citizen and professional user groups, demonstrating that our Vietnamese Legal RAG system achieves significant improvements over baseline approaches while maintaining high user satisfaction and expert validation for real-world Vietnamese legal consultation applications.

5 DISCUSSION

Our Vietnamese Legal RAG system demonstrates significant advantages over existing approaches in legal AI. Unlike traditional legal chatbots focusing on general advice, our multi-agent architecture specifically addresses Vietnamese legal system complexities through Router Agent coordination and specialized legal agents. Compared to single-agent RAG systems, our approach achieved 17.2% improvement in Context Precision, 29.7% improvement in Context Recall, and 12.2% higher legal accuracy. Statistical validation with large effect sizes (Cohen's $d > 0.8$) confirms substantial practical significance for Vietnamese legal consultation scenarios.

The Tools/MCP Server infrastructure enables seamless access to multiple information sources including vector search, graph databases, and external APIs, distinguishing our system from single-retrieval approaches. This comprehensive integration is crucial for Vietnamese legal queries requiring cross-referencing multiple document types with different authority levels. From a RAG perspective, our system adapts foundational work with practical modifications for Vietnamese legal domain requirements.

Several limitations warrant consideration. The system relies solely on Vietnamese legal documents, creating challenges with out-of-scope queries—approximately 20% of low-rated responses stemmed from non-legal questions. Real-time updates face processing delays before new documents become searchable, affecting responsiveness for urgent matters. Response latency of 5-15 seconds may frustrate users expecting immediate responses. Most significantly, 31.4% of system errors involved complex legal interpretations requiring human expert validation, highlighting continued need for professional oversight.

Future enhancements include query classification to handle out-of-scope questions, workflow integration for task execution beyond Q&A, and expansion to multimodal capabilities for comprehensive document analysis. Investigation of few-shot learning approaches could enable rapid

adaptation to new legal domains without extensive retraining.

6 CONCLUSION

This research presents an innovative Agentic RAG-based legal advisory chatbot for the Vietnamese legal system, addressing fundamental limitations of existing legal AI through centralized Router Agent coordination with multiple specialized agents. Our approach successfully integrates 54,000 Vietnamese legal documents across 12 hierarchical categories with automated synchronization capabilities.

The system achieved 82.2% accuracy in legal consultation scenarios with significant improvements over baselines. Human expert evaluation confirmed high professional acceptance (4.16/5.0 overall rating), validating readiness for real-world deployment. The comprehensive evaluation demonstrates that carefully designed multi-agent architectures can effectively address domain-specific challenges in legal AI, particularly for systems requiring high accuracy and regulatory compliance.

The Vietnamese legal system's complexity provides a rigorous test case, and our success suggests strong potential for adaptation to other legal jurisdictions. This work demonstrates that Agentic RAG architectures can effectively automate legal information access, alleviating professional workloads while improving user experience. Continued research into hybrid architectures unifying reliable retrieval with fluent generation moves us closer to AI assistants that are both articulate and factually grounded in legal contexts.

ACKNOWLEDGEMENT

The authors would like to thank the legal professionals and law students who participated in the user studies. We also acknowledge the support from the Industrial University of Ho Chi Minh City, Ho Chi Minh City University of Law, and Academy of Public Administration and Governance for providing access to legal documents and expertise. Special thanks to the legal experts who validated the Q&A pairs and provided valuable feedback during the system evaluation.

REFERENCES

- Anthropic. (2024). Model Context Protocol: A standard for connecting AI assistants to data sources. Technical Report.
- Asai, A., Wu, Z., Wang, Y., Sil, A., & Hajishirzi, H. (2023). Self-RAG: Learning to retrieve, generate, and critique through self-reflection. *arXiv preprint arXiv:2310.11511*.
- Ashley, K. D. (2017). *Artificial intelligence and legal analytics: New tools for law practice in the digital age*. Cambridge University Press.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The muppets straight out of law school. In *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 2898-2904).
- Chalkidis, I., Fergadiotis, M., Tsarapatsanis, D., Aletras, N., Androutsopoulos, I., & Malakasiotis, P. (2021). LexGLUE: A benchmark dataset for legal language understanding in English. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics* (pp. 1946-1961).
- Dale, R. (2019). Law and word order: NLP in legal tech. *Natural Language Engineering*, 25(1), 211-217.
- Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Winsor, E., Yeh, C., Arredondo, S., Cummings, D., Johnstone, J., & Larson, J. (2024). From local to global: A graph RAG approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*.
- Guu, K., Lee, K., Tung, Z., Pasupat, P., & Chang, M. (2020). Retrieval augmented language model pre-training. In *International Conference on Machine Learning* (pp. 3929-3938). PMLR.
- Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. T. (2020). Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (pp. 6769-6781).
- Katz, D. M., Bommarito II, M. J., Gao, S., & Arredondo, P. (2024). GPT-4 passes the bar exam. *Philosophical Transactions of the Royal Society A*, 382(2270), 20230254.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W. T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
- Li, J., Monroe, W., Ritter, A., Galley, M., Gao, J., & Jurafsky, D. (2016). Deep reinforcement learning for dialogue generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1192-1202).
- Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Text summarization branches out* (pp. 74-81).
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
- Nay, J. J. (2023). Large language models as corporate lobbyists. *arXiv preprint arXiv:2301.01181*.
- Nguyen, D. Q., & Nguyen, A. T. (2020). PhoBERT: Pre-trained language models for Vietnamese. In *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 1037-1042).
- Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1-22).
- Schick, T., Dwivedi-Yu, J., Dessi, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Roller, S., Sukhbaatar, S., Weston, J., & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. *Advances in Neural Information Processing Systems*, 36, 68821-68844.
- Sergot, M. J., Sadri, F., Kowalski, R. A., Kriwaczek, F., Hammond, P., & Cory, H. T. (1986). The British Nationality Act as a logic program. *Communications of the ACM*, 29(5), 370-386.
- Turtle, H., & Croft, W. B. (1991). Evaluation of an inference network-based retrieval model. *ACM Transactions on Information Systems*, 9(3), 187-222.
- Vu, T., Nguyen, D. Q., Nguyen, D. Q., Dras, M., & Johnson, M. (2018). VnCoreNLP: A Vietnamese natural language processing toolkit. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 56-60).
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824-24837.
- Wooldridge, M. (2009). *An introduction to multiagent systems*. 2nd edition. John Wiley & Sons.
- Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A. H., White, R. W., Dumais, S., & Wang, C. (2023). AutoGen: Enabling next-gen LLM applications via multi-agent conversation. *arXiv preprint arXiv:2308.08155*.
- Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019). BERTScore: Evaluating text generation with BERT. In *International Conference on Learning Representations*.
- Zhong, H., Guo, Z., Tu, C., Xiao, C., Liu, Z., & Sun, M. (2020). JEC-QA: A legal-domain question answering dataset. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 05, pp. 9701-9708).