A Multi-Agent Expert System for Global Data Regulation Compliance

Hantao Lin [1, 2,†]

1 Northwestern University

2 https://www.linkedin.com/in/hantao-lin/

† hantaolin520@gmail.com

Abstract

As global data protection regulations continue to evolve, ensuring compliance with frameworks such as the General Data Protection Regulation (GDPR), California Consumer Privacy Act (CCPA), and Brazil's Lei Geral de Proteção de Dados (LGPD) poses significant challenges for businesses and legal professionals. Existing AI-driven compliance tools—often reliant on single large language models (LLMs)—struggle with contextual reasoning, multi-jurisdictional comparisons, and adaptability to regulatory changes.

To overcome these challenges, we propose a multi-agent expert system that integrates structured legal retrieval, semantic search, and generative reasoning. The system is composed of four specialized components:

1. Manager Agent - Decomposes complex queries, selects retrieval strategies, and assigns subtasks.
2. Information Retrieval System – Employs Neo4j for structured retrieval and Pinecone for semantic search to fetch jurisdiction-specific legal provisions.
3. Aggregator Agent (TeammateAgent3) – Synthesizes retrieved legal texts, detects information gaps, and issues follow-up queries when necessary.
4. Conversational Memory System – Maintains a long-term memory of past user interactions to provide context-aware legal assistance.

Our system uses Multilingual-E5 Large for embedding generation and DeepSeek-R1:8B for legal synthesis and reasoning. Evaluation demonstrates that the system performs strongly on procedural and comparative legal queries, outperforming single-LLM baselines due to its retrieval-grounded approach. However, it currently underperforms on explanation-based queries, where standalone DeepSeek-R1:8B models yield more comprehensive responses.

By combining graph-based retrieval, semantic similarity search, multi-agent collaboration, and dynamic refinement, this system offers a scalable and intelligent legal compliance solution. It represents a significant step toward context-aware, self-improving legal AI systems capable of navigating complex global data protection laws.

Table of Contents

Abstract.. i

Introduction and Problem Statement.. 1

Literature Review.. 4

Data.. 6

Methods.. 9

Results.. 15

Discussion.. 19

Conclusions... 21

Directions for Future Work.. 22

Code Availability... 23

Bibliography.. 24

Introduction and Problem Statement

1.1 Background & Motivation

In an increasingly digitized world, data privacy laws govern how organizations collect, store, and process personal information. Regulatory frameworks such as the European General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), and Brazil's Lei Geral de Proteção de Dados (LGPD) impose stringent requirements on companies to ensure consumer data protection. However, cross-border compliance remains an ongoing challenge, given that legal frameworks vary significantly in scope, enforcement mechanisms, and specific obligations.

Legal professionals and compliance officers have turned to AI-driven solutions to automate regulatory analysis, risk assessment, and policy enforcement. Traditional AI systems have demonstrated considerable capabilities in processing legal texts and retrieving documents; however, their single-model architectures lack adaptability in dynamic legal landscapes where regulations are frequently updated. Moreover, these systems struggle with complex, multi-step reasoning tasks, such as comparing jurisdictional differences and ensuring legal fact-checking—both of which are critical for accurate legal analysis.

Figure 1. *Limitations of Traditional AI Chatbots in Navigating Global Data Privacy Regulations*

1.2 Research Problem & Challenges

Existing LLM-based legal AI systems face several limitations:

- Limited Multi-Hop Reasoning

  Traditional single-model AI systems struggle to decompose complex legal queries and retrieve relevant jurisdictional laws for comparison. For example, questions like "How does GDPR compare with CCPA in terms of data subject rights?" often yield simplified answers that fail to properly integrate and contrast legal provisions across jurisdictions.

- Knowledge Retrieval Bottlenecks

  Static models rely on pre-trained knowledge, making them susceptible to obsolescence as regulations change. Traditional legal AI systems lack real-time retrieval capabilities, often providing responses that fail to reflect the latest legal amendments, rulings, and interpretations.

- Lack of Context Awareness in Multi-Turn Legal Queries

  Traditional legal AI systems treat each query independently, failing to retain long-term conversation memory and track evolving legal discussions. Many legal inquiries involve sequential questions that require AI to recall previous interactions and refine its responses accordingly. Without memory, users must repeatedly restate details, limiting efficiency in AI-driven legal analysis.

- Difficulty in Handling Ambiguous or Implicit Queries

  Legal AI systems struggle with vague or underspecified queries where users do not explicitly mention jurisdictions, legal sections, or key terms. Traditional retrieval-based approaches often fail to infer missing details, resulting in incomplete or irrelevant responses. AI-driven compliance systems must accurately interpret ambiguous legal questions, infer jurisdictional context, and retrieve legally relevant provisions dynamically.

1.3 Proposed Solution: A Multi-Agent Expert System

To address the limitations of traditional legal AI models, we propose a multi-agent expert system that enhances global data regulation compliance through dynamic knowledge retrieval, multi-hop reasoning, and long-term context retention. Unlike single-model AI systems, which struggle with contextual understanding, jurisdictional comparisons, and evolving legal frameworks, our system leverages specialized agents to improve retrieval accuracy, response synthesis, and adaptability.

Drawing inspiration from recent advances in multi-agent collaboration (Tran et al., 2025; Li et al., 2025), our system integrates four specialized components that work in tandem to generate accurate, jurisdiction-aware, and dynamically updated legal assessments.

- Manager Agent

The Manager Agent serves as the orchestration hub of the system, responsible for query interpretation, retrieval strategy selection, and task delegation. Upon receiving a query, the agent:

  - Parses complex legal questions, identifying key entities such as countries, legal topics, and compliance requirements.
  - Determines the most effective retrieval strategy, prioritizing structured retrieval through Neo4j when explicit country-section references are detected, and semantic similarity search via Pinecone when queries lack structured specificity.
  - Routes subqueries to the appropriate Teammate Agents, ensuring that legal provisions are retrieved, analyzed, and synthesized efficiently.

- Teammate Agents

  The Teammate Agents are specialized components dedicated to retrieving and synthesizing legal information from structured and unstructured sources. Each agent plays a distinct role in ensuring completeness, accuracy, and contextual consistency:

  - Teammate Agent 1: Performs semantic searches in Pinecone to locate relevant legal texts.
  - Teammate Agent 2: Uses Neo4j for structured retrieval, retrieving pre-indexed legal texts based on jurisdiction and section.
  - Teammate Agent 3 (Aggregator Agent): Synthesizes retrieved data, detects missing information, and issues follow-up queries before generating final responses.

- Conversational Memory System

  To enable multi-turn interactions, this system maintains a conversation history buffer, storing past user queries and responses. By leveraging embedding-based retrieval, the system can reference prior discussions and provide continuity in ongoing legal inquiries, ensuring responses remain contextually relevant across multiple interactions.

- Gap Detection and Query Refinement

  Rather than relying solely on AI-generated responses, the system performs gap detection to identify missing legal references in retrieved documents. When information is incomplete, it autonomously generates follow-up queries to refine retrieval results before synthesizing a final response. This iterative approach ensures that multi-jurisdictional comparisons and procedural legal analyses are comprehensive and well-supported by authoritative legal sources.

Literature Review

Legal AI systems must navigate complex regulatory landscapes, requiring robust methods for information retrieval, reasoning, and compliance verification. Recent advancements in multi-agent systems, legal NLP, knowledge graphs, and AI evaluation techniques offer promising directions for enhancing legal decision-making. This section reviews relevant research across these domains, highlighting how multi-agent frameworks improve task distribution, how specialized NLP models enhance legal text processing, how retrieval and knowledge graph methods contribute to structured legal information access, and how AI self-evaluation mechanisms ensure system reliability.

2.1 Multi-Agent Systems in Legal Decision-Making

Recent work on multi-agent collaboration highlights the benefits of distributing complex tasks among specialized agents, particularly in domains requiring nuanced reasoning and decision-making. Tran et al. (2025) discuss how coordinating multiple agents—each responsible for a distinct subtask—can significantly enhance overall system performance compared to single-model architectures. In the context of legal decision-making, this finding supports the design of a Manager Agent that decomposes complex regulatory queries (e.g., determining relevant jurisdictions or identifying specific data laws) and routes subtasks to specialized Teammate Agents. Moreover, Li et al. (2025) provide an example of an agent that dynamically retrieves external knowledge while maintaining coherence in multi-step reasoning tasks, further reinforcing the value of an agent-based approach for legal AI.

2.2 Information Retrieval & Knowledge Graphs for Legal AI

Effective retrieval of legal documents and the ability to incorporate dynamic regulatory updates are essential for addressing the challenges of modern legal compliance. Lewis et al. (2020) highlight the advantages of Retrieval-Augmented Generation (RAG) in improving contextual awareness and reducing factual inaccuracies. Building on this, we implemented Pinecone for semantic search using cosine similarity, enhancing the system's ability to retrieve relevant legal texts. Additionally, we extended this approach by integrating Neo4j for structured retrieval, enabling precise and dynamic querying of legal documents based on jurisdictional and categorical attributes. This combined retrieval strategy ensures that responses are both contextually relevant and legally up-to-date, significantly improving the system's accuracy and reliability in legal compliance analysis.

2.3 AI Self-Improvement and Evaluation in Legal AI Systems

Effective legal AI systems require continuous self-improvement and validation to ensure accuracy, reliability, and compliance. Levi and Kadar (2025) introduce IntellAgent, a multi-agent framework that emphasizes the necessity of a critic agent for internal validation. Their approach highlights the importance of evaluating AI-generated responses against foundational knowledge rather than relying solely on external sources. This ensures that responses remain factually accurate, contextually consistent, and free from hallucinations before being delivered to users. Drawing inspiration from this, our system incorporates gap detection and refinement mechanisms, which function as a form of internal validation by identifying missing legal information before synthesizing a response.

Wei et al. (2022) further demonstrate that chain-of-thought prompting enhances LLMs' reasoning capabilities by guiding them through structured, intermediate steps, leading to more logically sound outputs. Inspired by this, our system employs reasoning-driven refinement, ensuring that retrieved legal texts sufficiently address user queries before final synthesis. Unlike traditional AI models that generate responses in a single pass, our approach dynamically retrieves legal provisions, identifies missing details, and autonomously issues follow-up queries to refine its understanding.

By integrating multi-agent evaluation principles with structured retrieval and reasoning-driven refinement, our expert system minimizes reliance on model-generated assumptions. Instead, it prioritizes retrieval-based grounding, ensuring that responses are rooted in authoritative legal texts. Additionally, by maintaining a conversation history buffer, our system learns from past interactions to improve query refinement, retrieval accuracy, and jurisdictional consistency over time. This iterative approach strengthens legal AI's ability to provide transparent, adaptive, and regulation-compliant insights, drawing inspiration from Levi and Kadar's work on AI evaluation and Wei et al.'s findings on structured reasoning.

Data

3.1 Data

The data for this study was sourced from the publicly available DLA Piper Data Protection Laws of the World repository, accessible at https://www.dlapiperdataprotection.com/ (DLA Piper, 2023). The dataset was obtained as a PDF document containing detailed legal frameworks and data protection regulations from multiple countries. Each country-specific section in the document follows a standardized structure, organizing legal information under key headers: Law, Definitions, National Data Protection Authority, Registration, Data Protection Officers, Collection & Processing, Transfer, Security, Breach Notification, Enforcement, Electronic Marketing, Online Privacy, and Key Contacts.

3.2 Data Preprocessing and Embedding

To enable efficient retrieval and similarity search, the processed legal data was structured and embedded for storage in Pinecone, a high-performance vector database optimized for semantic search. Each document was first summarized using Multilingual-E5 Large in Ollama, ensuring that key legal information was retained while reducing unnecessary verbosity. This summarization step enhances retrieval efficiency by creating a condensed representation of legal provisions while maintaining jurisdictional context.

Once summarized, each document was converted into dense vector embeddings using Multilingual-E5 Large, which was also responsible for generating embeddings optimized for legal text retrieval. Embeddings were processed in batches to manage memory efficiently and normalized using L2 normalization to ensure consistent vector magnitudes for cosine similarity searches. These embeddings were then stored in Pinecone, enabling rapid and scalable retrieval based on semantic similarity.

To further enhance performance, embeddings were upserted into Pinecone in batches, with each document linked to metadata, including the country name and summary. This structured approach ensures efficient jurisdiction-aware retrieval, allowing the expert system to accurately retrieve relevant legal provisions even when query phrasing differs from the original text.

Figure 2 provides an example of the structured embeddings stored in Pinecone, illustrating country-specific legal provisions, associated metadata, and the embedding vector representation.

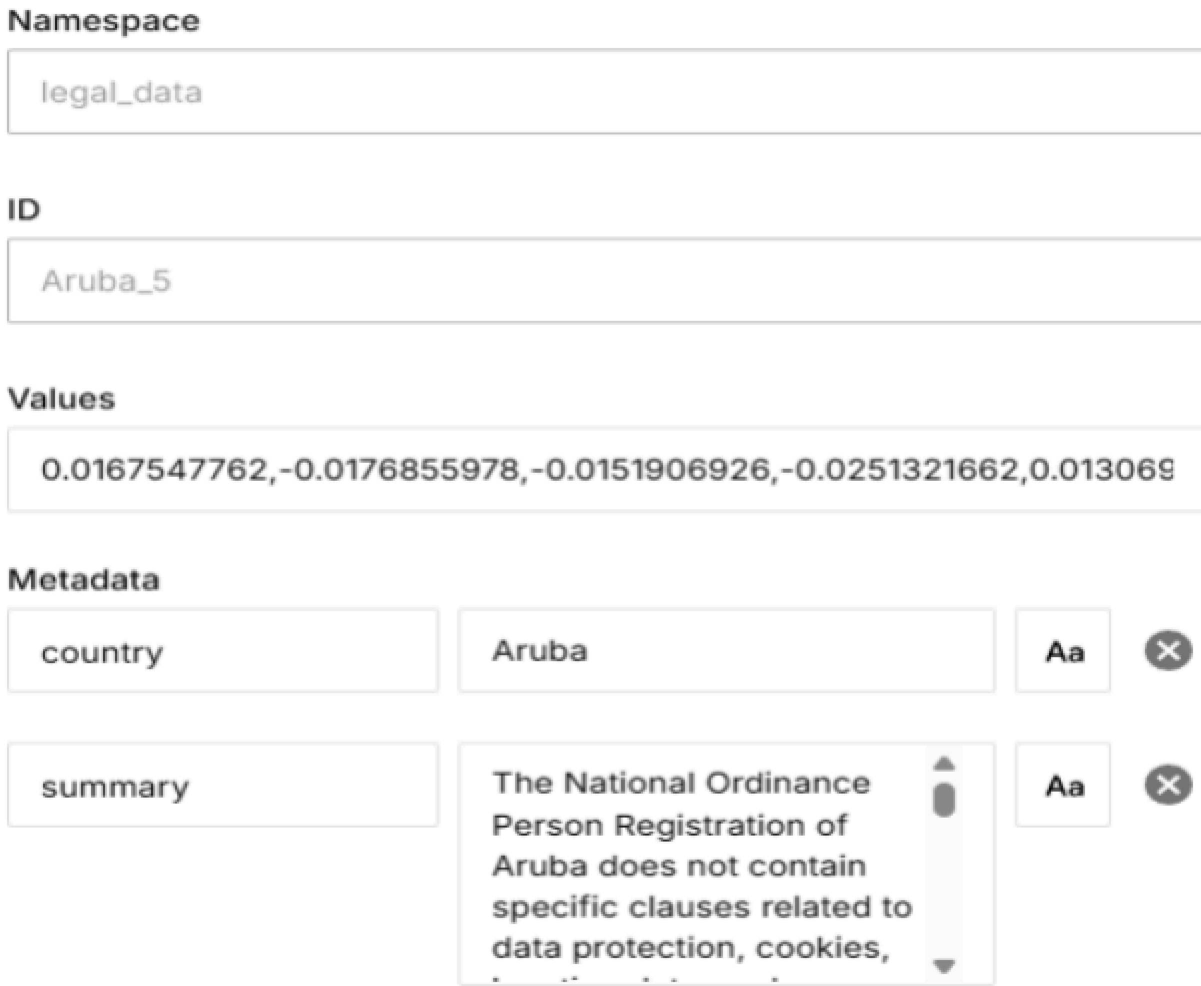

Figure 2. *Example of Aruba Index Record in Pinecone*

3.3 Graph Database Structure

In parallel, the preprocessed legal data was structured and stored in Neo4j, enabling efficient querying and structured relationship mapping between legal provisions across jurisdictions. Each legal provision is represented as a node with attributes such as a unique ID, the associated country (e.g., Czech Republic), the relevant legal section (e.g., Law), and the full legal text of the provision. This graph-based approach allows for flexible legal information retrieval while preserving jurisdictional structure.

The graph schema organizes data into three key node types:

- Country nodes, representing jurisdictions.
- Section nodes, representing distinct legal categories such as Law, Security, and Breach Notification.

- Text nodes, storing specific legal provisions.

These nodes are linked through two primary relationships:

- (:Country)-[:HAS_SECTION]->(:Section) → Links each country to its corresponding legal sections.
- (:Section)-[:HAS_PROVISION]->(:Text) → Connects each legal section to its respective legal provisions.

Unlike many graph models where sections are shared across jurisdictions, in this structure, each section belongs to only one country, and each legal provision is uniquely linked to a single section. This design ensures a structured representation of country-specific legal frameworks while allowing efficient retrieval of compliance requirements.

Figure 3 demonstrates this structure by illustrating how the Czech Republic connects to various legal sections (e.g., Law, Security, Breach Notification), while each section exclusively links to its corresponding legal provisions. For readability, multiple sections have been hidden from the figure, but the underlying schema supports additional legal sections per country. This approach highlights both country-specific regulations and structural legal interdependencies, facilitating jurisdictional comparisons and compliance assessments.

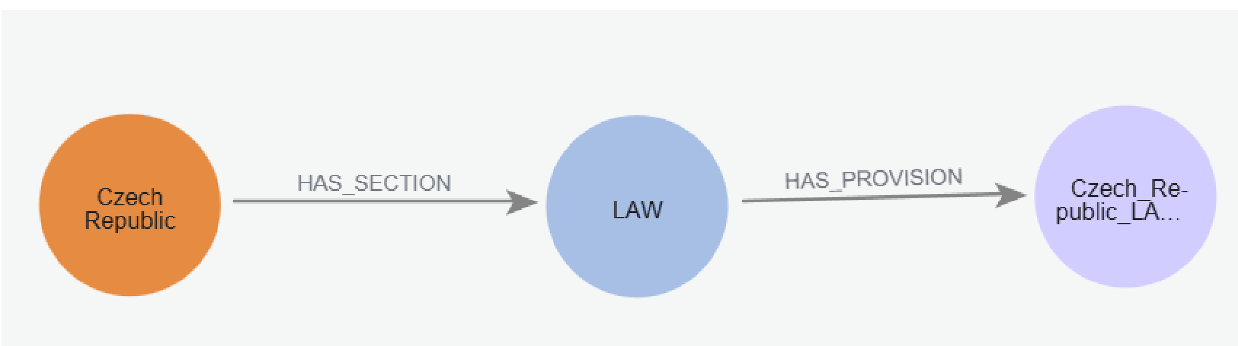

Figure 3. *Graph Representation of Czech Republic's Data Protection Framework in Neo4j*

Methods

This section describes the architecture, multi-agent workflow, query processing strategy, knowledge representation, and evaluation framework of the expert system. The system is designed to dynamically retrieve, synthesize, and refine legal information across multiple jurisdictions while maintaining high accuracy and adaptability. Figure 4 provides an overview of the system architecture.

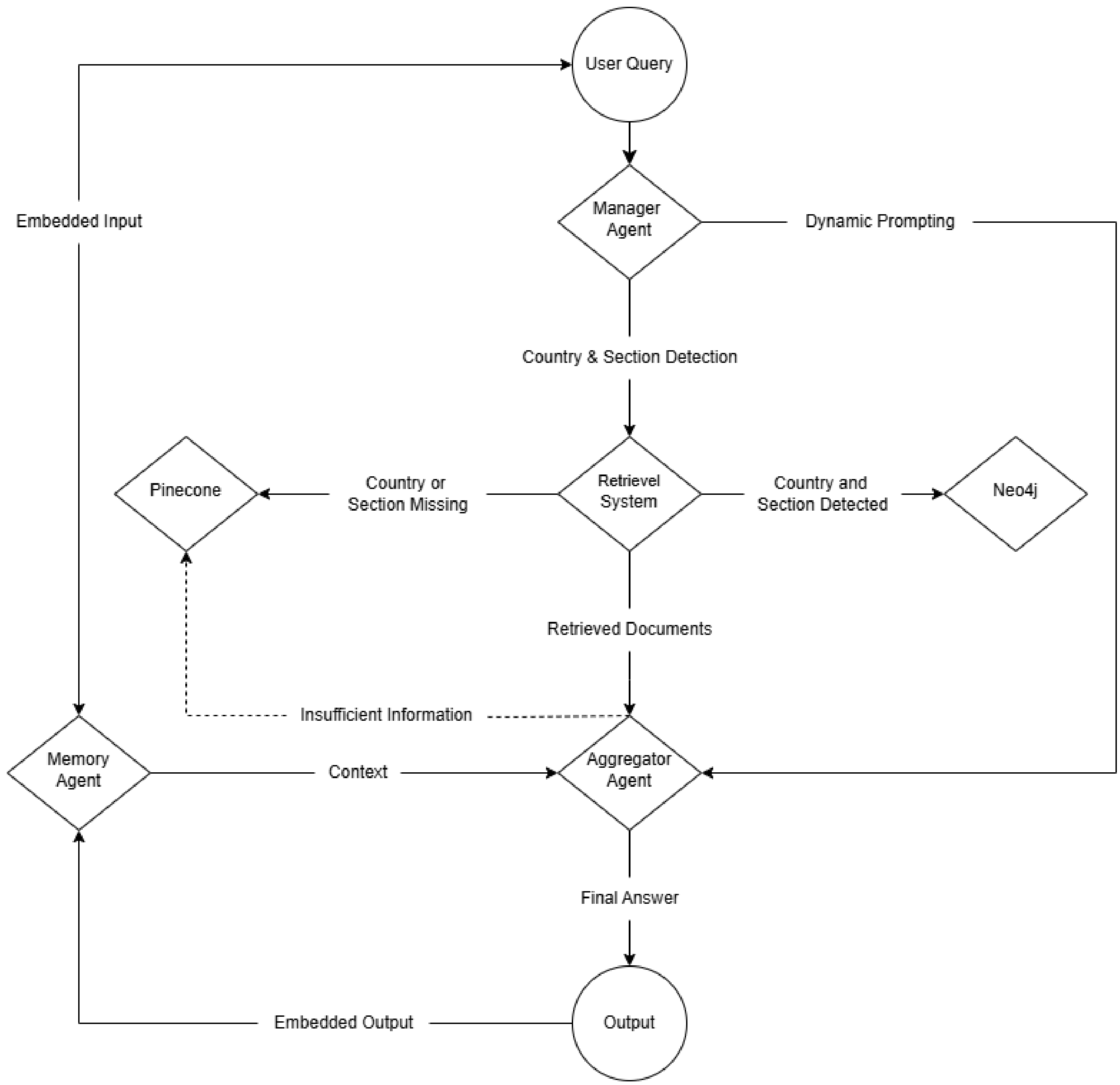

Figure 4. *System Architecture*

4.1 System Architecture

The expert system is structured as a multi-agent framework that integrates structured retrieval, semantic search, and iterative refinement to process complex legal queries. The core component is the Manager Agent, which interprets user queries, determines retrieval strategies, and orchestrates information synthesis. Supporting the retrieval process are three Teammate Agents, each assigned to a specialized retrieval task.

- Teammate Agent 1 conducts semantic search using Pinecone to retrieve relevant legal provisions based on contextual meaning rather than keyword matching.
- **Teammate Agent 2** performs **structured retrieval** from Neo4j, retrieving pre-indexed legal texts based on jurisdiction and section.
- **Teammate Agent 3** synthesizes retrieved data, refines responses, and ensures coherence before presenting a final answer.

To maintain contextual continuity across multi-turn interactions, the system incorporates a Memory System, which tracks past queries and retrievals. This enables the chatbot to recall prior discussions, improving consistency in ongoing legal inquiries. Unlike static legal AI models that generate responses in a single pass, this architecture dynamically refines answers, issuing follow-up queries where necessary to ensure completeness and accuracy.

4.2 Multi-Agent Workflow

The system workflow begins when a user submits a query. The Manager Agent first processes the query by detecting country and legal section references. If both are explicitly mentioned, structured retrieval from Neo4j is prioritized; otherwise, the system defaults to semantic search in Pinecone to identify relevant legal provisions.

Once the initial retrieval is complete, the system evaluates the retrieved documents for completeness. If missing information is detected, follow-up queries are issued to fill the gaps. The retrieved legal texts are then aggregated by Teammate Agent 3, which synthesizes a structured and legally coherent response. The Memory System stores the query-response pairs, allowing the chatbot to maintain context and improve consistency in future interactions.

Unlike conventional AI-driven compliance tools, which rely on static knowledge, this multi-agent framework dynamically retrieves and refines legal provisions in real time. Its modular structure enables adaptability to changes in regulatory frameworks by allowing updates to stored legal texts without retraining the entire model.

4.3 Query Processing and Retrieval Strategy

When a query is received, the system follows a structured retrieval pipeline:

1. *Country and Section Detection*: The system first determines whether a jurisdiction and legal section are explicitly mentioned using string-matching techniques. If both are detected, structured retrieval is prioritized.
2. *Retrieval Prioritization*: If both a country and section are identified, the system fetches legal provisions directly from Neo4j. If either is missing, the system defaults to Pinecone for semantic similarity-based retrieval.
3. *Query Categorization*: The query is classified into one of three categories—Comparison, Procedural, or Explanation—which influences how the response is structured but does not affect the retrieval process.
    a. Comparison Queries focus on analyzing jurisdictional differences.
    b. Procedural Queries outline compliance steps.
    c. Explanation Queries provide interpretations of legal concepts.
4. *Gap Detection and Refinement*: Once retrieved, the legal texts are assessed for completeness. If gaps exist, the system issues additional retrieval requests to ensure a comprehensive response before synthesis.

This retrieval pipeline ensures that structured lookups are prioritized when precise legal provisions are available, while semantic search enables flexibility in handling ambiguous queries.

4.4 Knowledge Representation and Retrieval

The system employs a dual retrieval approach, integrating structured knowledge graph retrieval (Neo4j) and semantic similarity search (Pinecone) to optimize legal information access.

Neo4j is used for structured retrieval, where legal provisions are stored as a knowledge graph with three primary node types: Country nodes representing jurisdictions, Section nodes categorizing legal provisions, and Text nodes containing the full legal text. These nodes are connected through two primary relationships: (:Country)-[:HAS_SECTION]->(:Section), linking each jurisdiction to its respective legal categories, and (:Section)-[:HAS_PROVISION]->(:Text), connecting each section to its corresponding legal texts. Unlike many legal knowledge graphs where sections are shared across multiple jurisdictions, each section in this system is uniquely assigned to a single country, ensuring precise retrieval.

Pinecone is used for semantic retrieval, allowing the system to retrieve relevant legal texts even when exact country-section matches are unavailable. The system first summarizes legal provisions using Multilingual-E5 Large to enhance retrieval efficiency, then embeds them as dense vector representations optimized for cosine similarity search. When users submit queries without explicit country or section references, Pinecone retrieves the most contextually relevant provisions.

This hybrid retrieval strategy ensures that responses are both precise and adaptable to variations in user queries. By leveraging Neo4j for structured lookups and Pinecone for flexible semantic search, the system maintains a high level of accuracy in legal information retrieval while ensuring adaptability to diverse legal inquiries.

4.5 System Evaluation

To assess the expert system's effectiveness, we designed a benchmark consisting of seven legal queries, categorized into Comparison, Procedural, and Explanation queries. These categories test different aspects of retrieval accuracy, synthesis quality, and the system's ability to handle complex legal reasoning. An edge case query is included to evaluate how well the system detects missing or ambiguous legal information.

Evaluation Questions and Categories

- Comparison Queries: Test the system's ability to retrieve and compare legal provisions across multiple jurisdictions.
    - How do data protection requirements differ between companies operating in the European Union and those in the United States?
    - Compare the definitions of personal data in the United Kingdom and Germany.
    - List the enforcement measures for non-compliance with data protection laws for a multinational corporation operating in the US, UK, and Canada.
- Procedural Queries: Evaluate whether the system provides clear, step-by-step compliance guidance based on jurisdiction-specific legal texts.
    - What steps must a company follow to comply with breach notification regulations in India?
    - What are the key compliance requirements for electronic marketing in Australia, and are there any missing provisions?
- Explanation Queries: Assess the chatbot's ability to interpret legal concepts and deliver well-structured, jurisdiction-aware explanations.

    - Can you explain the role of Data Protection Officers under the new GDPR guidelines?
    - What does the law require for online privacy?
- Edge Case Query: tests how well the system identifies missing or ambiguous legal information and whether it issues appropriate follow-up queries.
    - What are the key compliance requirements for electronic marketing in Australia, and are there any missing provisions?

Evaluation Methodology

Each query was processed by three different systems:

- The Expert System (DeepSeek-R1:8B-based) – Uses structured retrieval from Neo4j and semantic search from Pinecone, combined with a multi-agent synthesis process.
- A chatbot using DeepSeek-R1:8B directly – Uses the same foundational model but without structured retrieval or multi-agent coordination.
- A chatbot using Multilingual-E5 Large – Uses embeddings for retrieval without a dedicated synthesis mechanism.

The generated answers were evaluated by ChatGPT and DeepSeek, focusing on factual accuracy, completeness, and coherence. The key evaluation criteria include:

- Legal Accuracy: Does the response align with authoritative legal texts?
- Jurisdictional Specificity: Does the system correctly reference country-specific legal provisions?
- **Completeness:** Does the response fully address the question, or are critical details missing?
- **Comparative Depth:** For comparison queries, does the answer clearly articulate jurisdictional differences?
- **Procedural Clarity:** For compliance-related questions, does the response outline actionable steps?
- **Handling of Edge Cases:** Does the system recognize missing legal provisions and attempt to refine the response?

By systematically analyzing the retrieval efficiency, response synthesis, and query refinement, this evaluation framework highlights the strengths of structured retrieval (Neo4j) in jurisdiction-specific queries and the advantages of semantic search (Pinecone) in handling underspecified legal inquiries. The results inform future improvements in retrieval optimization, legal reasoning, and self-improving AI compliance tools.

Results

To evaluate the performance of the multi-agent expert system, we designed a benchmark consisting of seven legal compliance queries categorized into Comparison, Procedural, and Explanation types. Each query was processed by three systems: (1) our Expert System (DeepSeek-R1:8B-based with Neo4j and Pinecone retrieval), (2) a baseline DeepSeek-R1:8B chatbot, and (3) a Multilingual-E5 Large retrieval-based system. Responses were independently assessed by ChatGPT and DeepSeek evaluators using a 1–10 scale, measuring accuracy, completeness, and coherence.

5.1 Quantitative Evaluation

Figure 5 presents the average rating for each system across the seven queries. The Expert System achieved the highest scores on questions 1, 2, and 7, which focused on multi-jurisdictional comparisons and procedural guidance. However, performance dipped on question 3—an explanation-based query—where the baseline DeepSeek-R1:8B model outperformed due to its superior generative reasoning capabilities. The Multilingual-E5 Large system maintained consistent mid-range performance, excelling in semantic retrieval tasks but falling short in response synthesis depth.

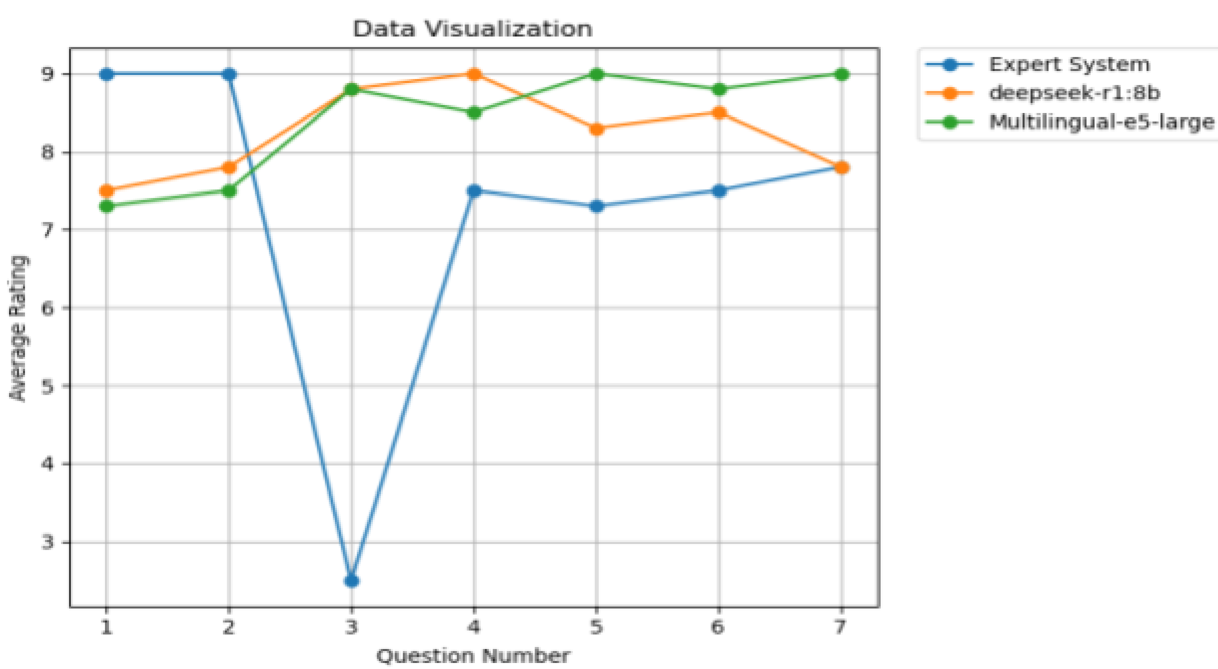

Figure 5. *Average Ratings per Evaluation Question Across Systems*

5.2 System Performance Breakdown

- Expert System: Demonstrated strong performance in structured retrieval and gap refinement, particularly for complex multi-jurisdictional comparisons (e.g., questions 1 and 7). The lower score on question 3 reflects its current limitations in purely explanatory tasks that require nuanced, free-form reasoning without direct retrieval support.
- DeepSeek-R1:8B Baseline: Outperformed in explanation queries due to its advanced generative capabilities but lacked jurisdiction-specific precision and regulatory context without retrieval-grounding.
- Multilingual-E5 Large: Showed stable performance, especially on queries with vague or underspecified inputs, leveraging its semantic matching capabilities. However, it struggled with response synthesis and lacked depth in legal interpretation.

5.3 Summary of Findings

The evaluation underscores the strengths of a multi-agent, retrieval-grounded system for legal compliance tasks requiring precision, procedural clarity, and cross-jurisdictional comparison. The performance gap in explanation tasks highlights an opportunity for future integration of advanced generative language modeling or hybrid approaches for conceptual legal analysis.

Discussion

This section analyzes the implications of the evaluation results, exploring how the multi-agent expert system performs across different query types and examining its operational strengths and limitations in real-world legal compliance scenarios.

6.1 Performance Across Query Types

Evaluation findings reveal that the expert system excels in procedural and comparative queries, where accurate retrieval and jurisdictional specificity are critical. The combined use of Neo4j for structured querying and Pinecone for semantic search enabled precise retrieval of jurisdiction-specific legal provisions, which were effectively synthesized into coherent, regulatory-compliant responses.

In procedural queries, the system reliably outlined jurisdiction-specific compliance steps by detecting country and section references and prioritizing Neo4j retrieval. The gap detection mechanism added robustness, autonomously issuing follow-up queries when initial retrievals were incomplete. This ensured comprehensive responses, enhancing legal reliability.

Conversely, the system underperformed in explanation queries, where free-form reasoning and conceptual depth take precedence over strict retrieval. In such cases, the DeepSeek-R1:8B baseline produced more coherent and contextually rich responses, indicating a limitation in the expert system's ability to handle interpretative legal queries without structured data inputs.

6.2 Strengths of the Retrieval-Based Multi-Agent Framework

The system's hybrid retrieval architecture—blending graph-based querying (Neo4j) with semantic similarity search (Pinecone)—proved effective in handling diverse queries. Structured retrieval ensured accuracy and jurisdictional alignment, while semantic search provided flexibility in interpreting underspecified or ambiguous inputs. This dual capability allowed the system to dynamically adapt to user intent, outperforming static AI legal tools.

The Memory System further enhanced interaction quality by maintaining contextual continuity across multi-turn conversations. Embedding-based retrieval of prior queries enabled context-aware legal assistance, ensuring consistency and reducing redundancy.

6.3 Limitations and Areas for Improvement

While the gap detection and refinement process improved completeness, it occasionally led to over-retrieval, introducing tangential or redundant data that diluted response clarity. Future work should refine follow-up query thresholds and implement relevance scoring to prioritize the most pertinent legal texts during synthesis.

The reliance on retrieval-based synthesis limits system performance in open-ended legal reasoning tasks. Integrating fine-tuned generative capabilities—such as training DeepSeek-R1:8B on explanatory legal content—may address this gap.

Moreover, the current system's retrieval prioritization is solely based on country and section detection, with no influence from query categorization. Enhancing the retrieval strategy to incorporate query type (e.g., defaulting to semantic search for explanation queries) could further optimize performance.

6.4 Implications for Legal AI Systems

The findings suggest that multi-agent, retrieval-grounded systems hold significant promise in legal compliance domains, especially for delivering jurisdiction-aware, factually grounded, and procedurally accurate insights. As global data regulations evolve, scalable and adaptable AI tools will become critical. The expert system's ability to dynamically retrieve, refine, and synthesize legal knowledge in real time positions it as a forward-looking solution for regulatory compliance across complex legal landscapes.

Conclusions

This research introduces a multi-agent expert system that integrates semantic search, structured graph retrieval, and iterative refinement to address the challenges of AI-driven legal compliance. By combining the strengths of Neo4j and Pinecone, along with advanced language models like Multilingual-E5 Large and DeepSeek-R1:8B, the system dynamically retrieves, evaluates, and synthesizes jurisdiction-specific legal information.

Evaluation results highlight the system's strengths in handling procedural and comparative legal queries, particularly in retrieving precise legal provisions and constructing accurate, jurisdiction-aware responses. The modular design of the system enables real-time adaptability to regulatory updates and supports long-term conversational context retention, marking a substantial improvement over static, single-model AI systems.

However, the system exhibits limitations in generating explanation-based responses, often providing less comprehensive outputs compared to standalone LLMs. Additionally, the reliance on follow-up queries introduces latency, highlighting the need for optimization in query refinement mechanisms.

Overall, this work demonstrates the efficacy of multi-agent collaboration in legal AI systems and sets the foundation for future advancements in accurate, adaptive, and transparent legal compliance tools.

Directions for Future Work

Building upon the current framework, several directions are identified for future research:

1. **Improve Explanation-Based Queries** – Enhance the reasoning capabilities of the Aggregator Agent to better handle open-ended legal interpretations and explanation-driven tasks.
2. **Explore Model Hallucination in Multi-Model Systems** – Investigate how hallucinations emerge when multiple models contribute to a single response and develop techniques for mitigation.
3. **Making the Expert System More Agentic** – Enable greater autonomy and coordination among agents, allowing dynamic task delegation and adaptive decision-making.
4. **Develop a More Robust Evaluation Framework** – Establish standardized metrics and diverse benchmarks to comprehensively evaluate legal AI performance, including edge cases and ambiguity handling.
5. **Implement Online Search** – Integrate real-time legal database access to enhance the retrieval of the most up-to-date legal information and amendments.
6. **Long-Term Memory Management** – Develop scalable and efficient mechanisms for managing and retrieving long-term conversational data, ensuring context-aware responses in extended interactions.

These future directions aim to advance the scalability, robustness, and intelligence of legal expert systems, contributing to the broader field of legal AI and compliance automation.

Code Availability

The code supporting this study is publicly available at https://github.com/Hantao-Lin/Expert-System---World-Data-Regulation-Chatbot.git. This repository includes scripts for data preprocessing, knowledge graph construction, semantic search, multi-agent query processing, and system evaluation. As no license has been specified, the code is available for non-commercial academic use only. For other use cases, please contact the corresponding author.

Bibliography

DLA Piper. 2023. "Data Protection Laws of the World." Accessed November 6, 2023. https://www.dlapiperdataprotection.com/.

Levi, Elad, and Ilan Kadar. "IntellAgent: A Multi-Agent Framework for Evaluating Conversational AI Systems." arXiv preprint arXiv:2501.11067 (2025).

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." Advances in Neural Information Processing Systems 33 (2020): 9459-9474.

Tran, Khanh-Tung, Dung Dao, Minh-Duong Nguyen, Quoc-Viet Pham, Barry O'Sullivan, and Hoang D. Nguyen. "Multi-Agent Collaboration Mechanisms: A Survey of LLMs." arXiv preprint arXiv:2501.06322 (2025).

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." arXiv preprint arXiv:2201.11903 (2023). https://arxiv.org/abs/2201.11903.