International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 04 | Apr 2026 www.irjet.net p-ISSN: 2395-0072

Agentic AI in the Legal Domain: Enhancing Ethical Decision-Making through Specialized AI Agent

Mayuri Patil, Prachiti Parab, Vansha Pandita

Dept. of Data Science UMIT, SNDT University Mumbai 400076, India

---

***Abstract*—*Artificial Intelligence (AI) is increasingly transforming the legal domain by enabling automated analysis, decision support, and intelligent assistance for complex legal workflows. Despite recent advancements in Large Language Models (LLMs), many existing AI-based legal tools lack structured validation mechanisms, explainability, and seamless integration within scalable web-based systems, limiting their practical adoption. This paper presents the design and implementation of RemoteCLO, an AI-powered legal advisory platform developed to deliver structured, context-aware, and accessible legal guidance through modular full-stack architecture. The system integrates a FastAPI backend with a Next.js frontend to facilitate efficient user interaction and reliable API communication. Legal queries are analyzed using Google Gemini LLM orchestrated via LangChain, enabling controlled reasoning and response generation. To enhance contextual relevance and ensure up-to-date information, Tavily Web Search is incorporated for real-time retrieval of legal references and supporting data. A Pydantic-based validation layer enforces structured input–output processing, improving reliability and consistency across system components. Experimental evaluation demonstrates stable system performance, coherent response generation, and effective frontend–backend interoperability. The RemoteCLO framework highlights the potential of combining modern web technologies with AI reasoning models to develop scalable, explainable, and user-centric legal advisory systems suitable for real-world deployment.*

***Keywords*—Legal AI, AI Legal Advisory System, Large Language Models (LLMs), Legal Query Answering, Web-Based Legal Assistance.**

I. INTRODUCTION

The rapid expansion of digital services and online platforms has significantly increased the demand for accessible legal information across individuals, startups, and small organizations. As regulatory environments evolve and legal frameworks become increasingly complex, users are required to interpret statutes, compliance requirements, contractual obligations, and jurisdiction-specific regulations without always having direct access to professional legal assistance. Despite technological advancements in other domains, access to reliable legal guidance remains costly, time-consuming, and often inaccessible to non-expert users [1],[14]

Traditional legal services are primarily designed for enterprises and established organizations, where professional consultation costs are sustainable within operational budgets.

For individuals and small-scale entities, however, obtaining timely legal clarification can be prohibitively expensive. Existing alternatives such as static legal information websites or template-based services provide limited contextual understanding and fail to adapt to user-specific scenarios. Meanwhile, generic AI chatbots powered by large language models frequently generate responses lacking structured validation, jurisdictional awareness, or explainable reasoning, raising concerns regarding reliability and ethical usage in legal contexts [2].

Recent advances in Artificial Intelligence, particularly Large Language Models (LLMs), have demonstrated promising capabilities in legal reasoning, document understanding, and automated question answering. Research has explored specialized legal language models, intelligent conversational agents, and AI-driven legal assistants aimed at improving access to justice and enhancing legal research efficiency [13],[9],[3]. Additionally, emerging studies highlight the importance of modular architectures and collaborative AI frameworks to improve reasoning consistency and scalability in legal applications [5],[8]. However, several challenges remain insufficiently addressed, including the integration of AI reasoning within full-stack web systems, structured validation of legal outputs, and the incorporation of real-time legal information retrieval to maintain response relevance.

To address these challenges, this paper introduces RemoteCLO, an AI-powered legal advisory platform designed to provide structured, context-aware legal assistance through a modular web-based architecture. RemoteCLO integrates a Next.js frontend with a FastAPI backend to enable seamless interaction between users

and AI-driven reasoning modules. Legal queries are processed using the Google Gemini Large Language Model orchestrated through LangChain, while Tavily Web Search enables real-time retrieval of relevant legal information and supporting references to enhance contextual accuracy. A Pydantic-based validation layer ensures consistent data handling and reliable system behavior across components.

The proposed system aims to bridge the gap between advanced AI reasoning capabilities and practical legal accessibility by delivering explainable, scalable, and user-friendly legal assistance through modern web technologies. By combining structured backend validation with real-time information retrieval and interactive frontend design, RemoteCLO demonstrates a practical approach toward deploying AI- assisted legal advisory systems in real-world environments. The remainder of this paper is organized as follows. Section II presents a review of existing AI-driven legal systems that contextualize RemoteCLO within current research. Section III describes the system architecture and methodology.

Section IV details implementation and evaluation results obtained during the development stage. Finally, Section V concludes the study and outlines directions for future work.

II. LITERATURE REVIEW

The application of artificial intelligence to legal practice has a rich and rapidly evolving research history. The following review synthesises fifteen relevant research works, organised thematically, to establish the intellectual context for RemoteCLO.

A. LLM Hallucination and Legal Knowledge Accuracy

A significant challenge in applying Large Language Models (LLMs) to legal applications is hallucination, where models generate responses that appear legally valid but lack factual accuracy or verifiable grounding. Because legal decision-making requires precision and accountability, such inconsistencies reduce trust in AI-generated legal guidance and limit real-world adoption [5],[13].

To improve reliability, researchers have explored structured methods that enhance legal knowledge grounding. Frameworks such as ChatLaw incorporate knowledge graphs and role- based collaborative reasoning to reduce incorrect outputs and improve response consistency [5]. Similarly, domain-adapted models like Lawyer LLaMA demonstrate that training on specialized legal corpora improves contextual understanding and reduces fabricated legal references [9].

Another challenge involves processing lengthy legal documents that exceed the context limits of conventional language models. Architecture such as Lawformer enables efficient handling of long legal texts, improving comprehension of statutes and case materials [16]. While these approaches strengthen model capabilities, they largely focus on model improvements rather than deployable system integration, motivating solutions like RemoteCLO that combine structured validation with real-time legal information retrieval.

B. Real-Time Information Retrieval in Legal Question Answering

Reliable legal question answering requires access to up-to- date and verifiable information, as static model knowledge may become outdated or incomplete over time. Traditional Large Language Models rely primarily on parametric knowledge learned during training, which limits their ability to reference recent legal developments or jurisdiction- specific updates. Prior research highlights the importance of integrating external knowledge sources to improve factual grounding and reduce inaccuracies in AI-generated legal responses [13], [4].

Several studies demonstrate that augmenting language models with retrieval mechanisms significantly improves response reliability. Research on AI legal assistants shows that providing models with relevant statutory context or supporting documents enhances reasoning accuracy and reduces misleading outputs [13]. Similarly, hybrid legal information retrieval approaches combining semantic embeddings with traditional search techniques improve precision in identifying relevant precedents and legal materials, emphasizing the role of dynamic information access in legal AI systems [4].

Inspired by these findings, the RemoteCLO system incorporates Tavily Web Search to enable real-time retrieval of relevant legal references and contextual information during query processing. Instead of relying solely on pre- trained model knowledge, the system supplements user queries with current web-sourced legal data before response generation. This approach improves contextual relevance while maintaining lightweight and deployable architecture, demonstrating a practical

alternative to complex retrieval pipelines for real-world legal advisory applications.

C. Multi-Agent Legal Reasoning

Recent research has explored multi-agent architecture as a method for handling complex legal reasoning by distributing tasks across specialized AI components. Studies indicate that single-agent legal systems often struggle with transparency, reasoning consistency, and adaptability across diverse legal contexts. Yang et al. [8] argue that collaborative multi-agent frameworks improve explainability and task coordination by enabling structured interaction between reasoning modules, thereby enhancing trustworthiness in legal AI applications.

Further advancements highlight the role of human–AI collaboration within multi-agent legal workflows. Meng et al. [11] proposes a multilingual legal terminology mapping framework in which multiple AI agents perform repetitive reasoning tasks while human experts validate outputs for semantic accuracy. This hybrid approach addresses challenges associated with linguistic diversity and domain-specific terminology, demonstrating how coordinated agent systems can improve reliability in specialized legal environments.

Complementary research by Gray et al. [7] introduces human-in-the-loop methodologies where AI systems assist in identifying and refining legal factors from judicial opinions. Their findings show that collaborative AI workflows can uncover patterns and insights beyond manually curated legal analyses. While such multi-agent approaches enhance reasoning depth and scalability, they often introduce architectural complexity. In contrast, RemoteCLO adopts a streamlined single-agent orchestration approach using LangChain, prioritizing deployability and real-time usability while remaining extensible for future architectural enhancements.

D. Access-to-Justice Chatbots and Low-Resource Legal AI

A growing area of legal AI research focuses on improving access to legal information for underserved users through conversational systems. Legal chatbots have been proposed as scalable tools for providing preliminary guidance while reducing reliance on expensive legal consultation. Pardhi et al, highlights the importance of continuous testing, regulatory compliance, and privacy safeguards to ensure trustworthy deployment and prevent misinformation.

Studies addressing low-resource environments demonstrate that effective legal assistance can be achieved even with limited datasets. Queudot et al. [14] develop an access-to- justice chatbot using publicly available legal information, showing reliable performance despite sparse training data. Similarly, Mowbray et al. [12] propose sustainable rule-based legal decision-support systems integrated with live legal corpora to provide accessible legal assistance without heavy infrastructure requirements.

Amato et al. [3] further show that semantic-search-based conversational agents can support legal dispute resolution, though dataset diversity remains a key challenge for reliability. Building on these insights, RemoteCLO adopts a web-based advisory approach combining LLM reasoning with real-time information retrieval to provide accessible and context-aware legal assistance.

E. AI Governance, Bias, and Predictive Analytics in Law

Beyond technical performance, researchers have raised important governance concerns regarding the use of AI in legal decision-making. Jain et al. [2] highlight challenges related to algorithmic bias, lack of transparency, and data privacy risks, arguing that explainability and accountability must be embedded into system design rather than introduced after deployment. These considerations influence the ethical design principles adopted in RemoteCLO.

Research on predictive legal analytics further reveals limitations associated with unstructured legal data and restricted training resources. Kumar et al. [1] demonstrate that litigation outcome prediction models often struggle with generalization due to inconsistent data representation, emphasizing the need for structured legal datasets and standardized evaluation practices.

Complementary work by Gray et al. [7] proposes semi-automated methods for identifying legal factors from judicial texts, enabling more structured analysis for predictive modeling. Together, these studies underline the importance of transparency, structured data handling, and responsible AI deployment, which motivate the development approach followed in RemoteCLO.

F. Research Gaps Addressed by RemoteCLO

Synthesizing the research works reviewed in the previous section reveals several consolidated gaps in

current AI-driven legal assistance systems that motivate the development of RemoteCLO.

First, many existing legal AI systems rely primarily on pre-trained model knowledge or static datasets, limiting their ability to provide up-to-date legal information. While retrieval-based approaches improve grounding and accuracy [13], most studies evaluate models using fixed legal corpora rather than dynamically incorporating current legal resources. Consequently, real-time integration of web-sourced legal information within deployable advisory platforms remains insufficiently explored. RemoteCLO addresses this gap by incorporating Tavily Web Search to retrieve relevant legal references during query processing, improving contextual relevance.

Second, existing legal AI solutions often focus either on model-level improvements or conversational interfaces without emphasizing full-stack system integration. Prior work has explored legal question-answering agents and conversational assistants independently [3], [14], yet seamless coordination between frontend interaction, backend validation, and AI reasoning pipelines is rarely demonstrated in practical implementations. RemoteCLO bridges this gap through a modular architecture integrating a Next.js frontend with a FastAPI backend and structured API-based communication.

Third, many legal AI applications target either professional legal practitioners or public assistance scenarios, leaving broader accessibility challenges insufficiently addressed [12]. Systems designed for expert environments may lack usability for non-specialist users, while lightweight public-facing tools often sacrifice reasoning transparency. RemoteCLO aims to balance accessibility and reliability by providing structured, explainable legal responses within an intuitive web-based interface.

Fourth, governance concerns such as explainability, bias mitigation, and accountability remain underrepresented at the architectural level of legal AI systems. Research emphasizes the importance of embedding ethical safeguards directly into system design rather than treating them as post-processing considerations [2]. RemoteCLO incorporates structured validation through Pydantic models and controlled reasoning orchestration via LangChain to promote consistent and transparent outputs.

Collectively, these gaps highlight the need for scalable, explainable, and deployable AI-driven legal advisory systems. RemoteCLO is designed to address these challenges by combining real-time information retrieval, modular system architecture, and structured validation mechanisms within a practical web-based framework.

III. Methodology

A. Research Design

This research follows a design science methodology, in which the primary artefact—the RemoteCLO AI Legal Advisory System—is designed, implemented, and iteratively refined through practical development and evaluation. The methodology integrates system architecture design, backend–frontend implementation, and qualitative assessment of system functionality. The study focuses on developing a working prototype that demonstrates how large language models can support ethical and explainable legal assistance through structured workflows and real-time information retrieval. As the project is currently at an advanced prototype stage, large-scale quantitative benchmarking is reserved for future work. The present methodology emphasizes architectural decisions, implementation strategies, and functional validation to establish a strong foundation for subsequent performance evaluation and system enhancement.

B. System Architecture Overview

The proposed RemoteCLO system follows a modular client–server architecture that integrates a web-based frontend with an AI-driven backend for legal query processing. The frontend, developed using Next.js and React, enables users to submit legal questions and view structured responses through an interactive interface. The backend, implemented using FastAPI, manages API routing, request handling, and communication with AI services. An AI integration layer orchestrated by LangChain coordinates interactions with the Google Gemini large language model[6] and Tavily Web Search[15] to obtain relevant and up-to-date legal information. Input and output data are validated using Pydantic models to ensure structured and reliable processing. This layered architecture promotes scalability, maintainability, and efficient communication between system components while supporting explainable legal response generation.

C. Data Sources and Retrieval Strategy

RemoteCLO does not rely solely on the static knowledge embedded within pre-trained language models; instead, it enhances response generation

through real-time information retrieval. When a user submits a legal query, the system utilizes Tavily Web Search to dynamically retrieve relevant information from reliable online legal resources, public regulatory portals, and verified legal information sources. The retrieved contextual data is incorporated into the reasoning workflow managed by LangChain[10], where it is combined with the user query before being processed by the Google Gemini large language model. This integration enables the system to generate responses grounded in current and contextually relevant information rather than relying exclusively on previously learned model parameters. By incorporating real-time external knowledge during inference, RemoteCLO improves response accuracy, contextual relevance, and practical usability while maintaining a lightweight, scalable, and deployable architecture suitable for web-based legal advisory applications.

D. LLM Backend

The system interfaces with the Google Gemini Large Lan- guage Model through an API to perform natural language understanding and legal response generation. User queries, together with contextual information retrieved via Tavily Web Search, are incorporated into a structured prompt managed by LangChain. The prompt defines the system role as a legal advisory assistant and guides the model to generate clear, context-aware responses tailored to the user's legal query.

The prompting strategy encourages step-by-step reasoning and structured output generation to improve clarity and reliability. By integrating retrieved contextual information with the user query, the system reduces dependence on static model knowl- edge and improves response relevance. The model is further guided to avoid unsupported claims and produce explanations aligned with the provided context, helping mitigate hallucina- tion risks commonly associated with general-purpose language models [5], [9].

Fig. 1. Backend Interface

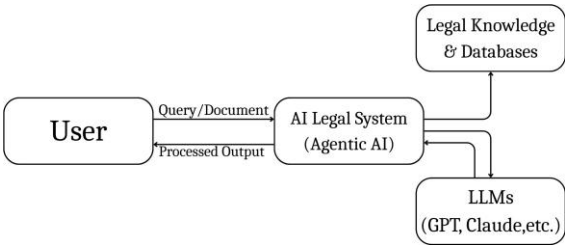

Fig. 2. DFD Level 0: Agentic AI Legal System

E. Prototype Evaluation Approach

The developed prototype was evaluated through structured functional testing and qualitative walkthrough sessions to assess system performance and usability. Test users interacted with the RemoteCLO interface by submitting representative legal queries covering general legal guidance, compliance- related questions, and regulatory information requests. The evaluation focused on response relevance, clarity of explanations, and the system's ability to provide structured and understandable legal insights through the web interface.

System performance was further examined by validating back- end API responses using FastAPI Swagger documentation and frontend interaction testing. Evaluation criteria included response consistency, contextual accuracy, and overall user comprehension of generated outputs. These qualitative obser- vations provide an initial assessment of system effectiveness, while comprehensive quantitative benchmarking is reserved for future work. Results and discussion are presented in Section VI.

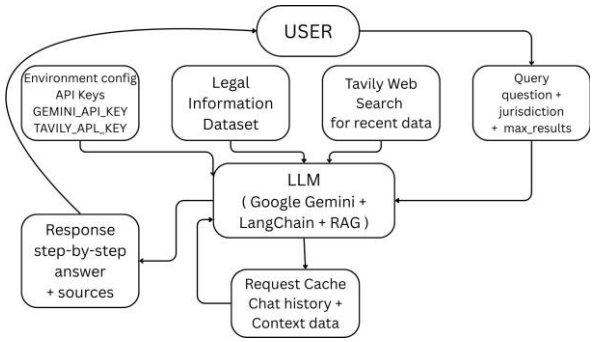

Fig. 3. DFD Level 1: Agentic AI Legal System

F. Proposed Framework

The proposed AI Legal Advisory System follows a modular client–server architecture designed to enable

efficient interaction between users, backend services, and AI reasoning components. The framework is organized into interconnected functional layers that collectively process user queries and generate structured legal guidance.

G. Retrieval and Reasoning Layer

The Retrieval and Reasoning Layer forms the core intelligence component of the proposed legal advisory system. When a user submits a legal query through the frontend interface, the request is transmitted to the FastAPI backend, where it undergoes structured validation using Pydantic models to ensure correctness and completeness of input data. The validated query is then forwarded to the AI orchestration pipeline managed by LangChain.

Instead of relying on static legal datasets, the system retrieves relevant and recent legal information using Tavily Web Search. Tavily dynamically gathers contextual legal references, guidelines, and publicly available legal resources related to the user's query. This retrieved information is combined with the original user input to provide contextual grounding before reasoning is performed by the Large Language Model.

Google Gemini LLM processes the enriched query using structured prompts designed to generate clear, step-by-step legal explanations. The model focuses on producing logically organized and understandable responses rather than deterministic legal judgments. The generated output is then formatted into structured JSON responses and returned to the frontend, ensuring transparency, readability, and consistency in legal guidance delivery.

H. Query Processing and Legal Reasoning Layer

The core functionality of the proposed RemoteCLO system lies in its query processing and legal reasoning layer, which enables users to obtain structured legal guidance through an AI-assisted workflow. When a user submits a legal query through the web interface, the input is transmitted to the FastAPI backend as a structured JSON request containing the legal question and optional contextual parameters.

The backend first validates the request using Pydantic models to ensure data consistency and correctness. The validated query is then forwarded to the AI orchestration pipeline managed through LangChain, which coordinates interaction with external services and the Large Language Model. To enhance contextual relevance, the system retrieves up-to-date legal information using Tavily Web Search, allowing the model to access recent legal references, guidelines, and publicly available sources.

The enriched query is processed by Google Gemini LLM, which performs contextual understanding and generates a clear, step-by-step legal explanation. The response is formatted into a structured JSON output and returned to the frontend, where it is displayed as readable legal guidance. This layered approach ensures reliable processing, improved response clarity, and transparent interaction between user input, external knowledge retrieval, and AI reasoning.

I. User Interaction and Interface Layer (Frontend Layer)

The User Interaction Layer serves as the primary access point through which users communicate with the AI Legal Advisory System. Unlike voice-based systems, the proposed platform operates through a web-based interface developed using Next.js and React, enabling users to submit legal queries and receive structured AI-generated responses through a browser environment.
Users enter legal questions along with optional contextual parameters such as jurisdiction and response preferences. The frontend manages user input validation, interface rendering, and asynchronous communication with backend APIs using HTTP requests. React components and custom hooks handle application state, ensuring smooth interaction and dynamic display of results without requiring page reloads.
Once a query is submitted, the interface presents AI-generated outputs in a structured and readable format, including step- by-step explanations and referenced information. The design focuses on usability, clarity, and accessibility, allowing users to easily understand legal guidance while maintaining transparency in system responses. This layer bridges backend intelligence with intuitive user experience, transforming complex AI processing into a practical legal advisory web application.

J. Orchestration and Safety Architecture (Rewritten for Your System)

The system employs a centralized orchestration workflow managed by the FastAPI backend, which

coordinates communication between the frontend interface, AI integration layer, and external services. When a user submits a legal query through the RemoteCLO web interface, the backend manages session handling, validates request parameters using Pydantic models, and routes the query to the appropriate AI processing pipeline. This orchestration ensures structured data flow, reliable API communication, and consistent response generation across system components.

To promote responsible AI usage, the system incorporates basic safety and validation mechanisms during processing. Input validation ensures that incomplete or malformed queries are filtered before reaching the AI model, while structured prompt formatting guides the Large Language Model to generate informational and context-aware responses rather than definitive legal advice. Additionally, responses are formatted with explanatory context and supporting references retrieved through Tavily Web Search, improving transparency and reducing misleading outputs. These safeguards support ethical deployment principles highlighted in prior legal AI research concerning explainability, accountability, and responsible use of AI in legal assistance systems [4][10].

K. Framework Architecture Diagram

Figure 4 illustrates the high-level data flow of the proposed RemoteCLO framework. A user submits a legal query through the web-based frontend interface developed using Next.js. The query is transmitted as a structured JSON request to the FastAPI backend, which acts as the central orchestration component of the system.

Upon receiving the request, the backend validates the input using Pydantic models and forwards the query to the AI processing pipeline managed by LangChain. The system retrieves relevant contextual information through Tavily Web Search, which gathers up-to-date legal references and supporting in- formation from reliable online sources. The enriched query is then processed by the Google Gemini Large Language Model to generate a context-aware legal response.

The generated output is formatted into a structured response by the backend and returned to the frontend interface, where it is displayed to the user as clear and readable legal guidance. This architecture enables efficient interaction between user input, real-time information retrieval, AI reasoning, and response presentation within a modular and scalable web-based framework.

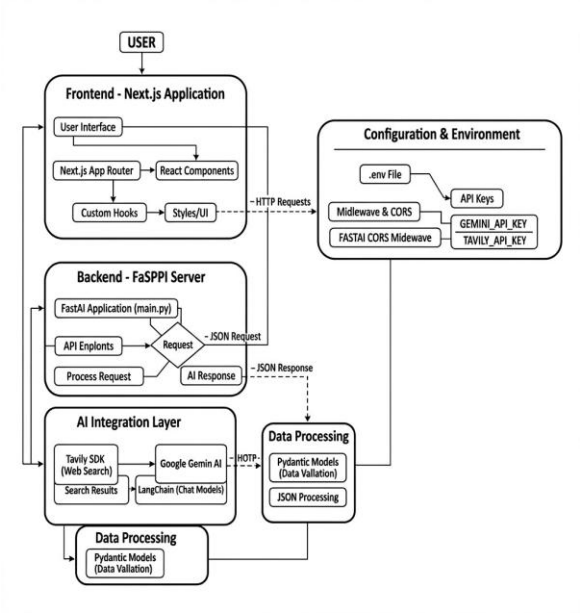

Fig. 4. Architecture Workflow Diagram

IV. Results and Discussions

A. MVP Prototype Overview

The RemoteCLO system is currently implemented as a functional minimum viable product (MVP) focusing on AI- driven legal query processing and ethical response generation. The developed prototype integrates a FastAPI-based backend with Large Language Model reasoning powered by Google Gemini and orchestrated through LangChain. Real-time con- textual information retrieval is supported using Tavily Web Search, enabling the system to generate context-aware legal responses grounded in up-to-date information sources.

The backend implementation has been successfully tested through API validation and structured query execution, demon- strating reliable processing of user inputs and consistent response generation. The front-end interface, developed using Next.js and React, is currently under active development to provide an intuitive web-based interaction layer.

B. Query Answering Performance

During qualitative evaluation, users submitted legal queries related to general legal guidance, compliance requirements, and regulatory understanding. The system generated structured and understandable responses

using Google Gemini LLM supported by contextual information retrieved through Tavily Web Search. In most cases, responses were relevant, clearly explained, and accessible to non-expert users.

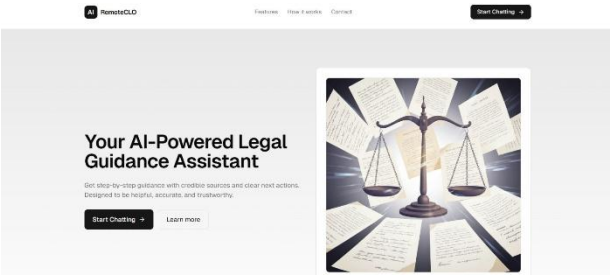

Fig. 5. Frontend Interface

Consistent with findings in prior legal AI studies [13], re- sponse quality improved when relevant contextual information was successfully retrieved. For highly specific or ambigu- ous queries, the system provided generalized informational guidance and indicated the need for professional consultation, supporting responsible AI usage practices . Some limitations were observed for queries involving unclear or mixed legal contexts, highlighting areas for future improvement consistent with challenges identified in legal AI research [8].

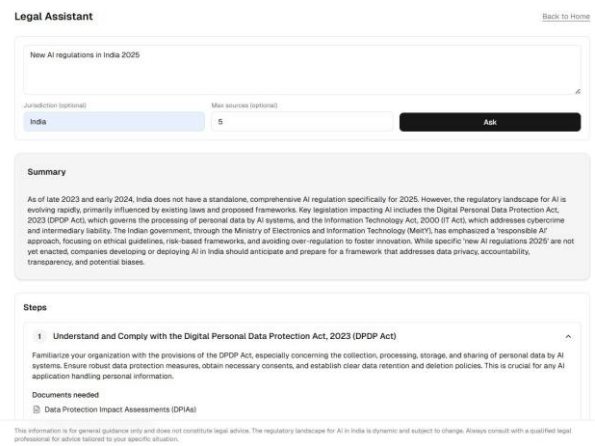

Fig. 6. ChatBot Interface

C. Accuracy Evaluation of the System

To evaluate the effectiveness of the proposed RemoteCLO legal advisory system, a structured evaluation was conducted using a curated dataset of representative legal queries. The evaluation dataset consisted of 30 legal questions covering multiple legal domains including corporate law, intellectual property law, contract law, compliance regulations, and general legal information relevant to startups and individuals.

Each query was submitted through the RemoteCLO web interface, which processes user inputs through Next.js frontend and FastAPI backend and generates responses using the Google Gemini Large Language Model augmented with contextual information retrieved through Tavily Web Search.

The Tavily integration allows the system to access relevant and up-to-date legal information from trusted web sources, thereby improving the reliability of generated responses.

The generated outputs were manually evaluated by comparing them with reliable legal references such as government portals, legal documentation, and verified legal information repositories. Each response was classified into three categories:

Correct – the response accurately explained the legal concept and aligned with authoritative sources

Partially Correct – the response contained generally valid information but lacked contextual completeness

Incorrect – the response contained misleading or irrelevant legal information

Since large language model responses may contain partially correct explanations that still provide useful guidance, a weighted evaluation metric was used to measure overall system performance. In this evaluation approach, correct responses were assigned a score of 1.0, partially correct responses were assigned 0.5, and incorrect responses were assigned 0.

Out of the 30 evaluated queries, 25 responses were categorized as correct, 4 responses were partially correct, and 1 response was incorrect. Using the weighted evaluation method, the RemoteCLO system achieved an overall accuracy of approximately 90 percentage.

The results demonstrate that integrating a large language model with real-time legal information retrieval significantly improves the reliability and contextual relevance of AI- generated legal explanations. The system performs particularly well for general legal information queries, while some limitations were observed for highly jurisdiction-specific questions or ambiguous legal phrasing.

TABLE I ACCURACY EVALUATION OF THE REMOTECLO SYSTEM

Metric	Value
Total Queries Evaluated	30
Correct Responses	25
Partially Correct Responses	4
Incorrect Responses	1
Weighted Accuracy	90%

The weighted accuracy was calculated using the following formula:

$$Accuracy = \frac{C + 0.5P}{N} \quad (1)$$

Substituting the observed evaluation values:

$$Accuracy = \frac{25 + 0.5(4)}{30} \quad (2)$$

$$Accuracy = \frac{27}{30} = 0.90\ (90\%) \quad (3)$$

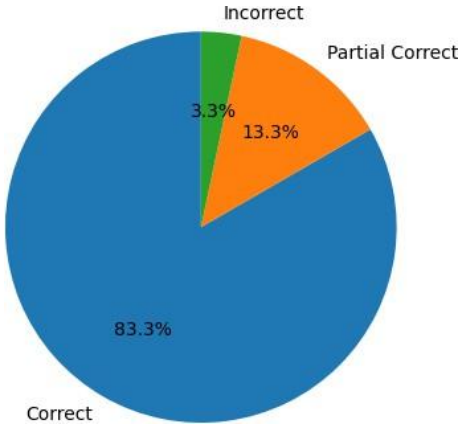

Fig. 7. Distribution of correct, partially correct, and incorrect responses generated by the RemoteCLO system during evaluation.

The distribution shown in the pie chart highlights that most responses generated by RemoteCLO were accurate and con- textually relevant. Only a small proportion of responses were partially correct or incorrect, indicating that the integration of large language models with real-time information retrieval im- proves the reliability of AI-assisted legal information systems.

V. LIMITATIONS AND FUTURE WORK

Several limitations of the current MVP are acknowledged. First, system evaluation has primarily been qualitative, and quantitative benchmarking using standardized legal QA datasets has not yet been conducted. Future work will involve constructing an annotated evaluation dataset to enable objective measurement of response accuracy and reliability. Second, the system currently relies on real-time web retrieval through Tavily, which, while effective, does not yet incorporate advanced retrieval mechanisms such as Retrieval-Augmented Generation (RAG). Integrating RAG-based retrieval and vector indexing is planned to further improve contextual grounding and response consistency.
Additionally, the present implementation focuses on text-based interaction and legal query answering. Future development will extend the platform with document analysis and generation capabilities to assist users in reviewing legal content and drafting standard documents. A multilingual voice interface is also envisioned to improve accessibility and user interaction. Finally, large-scale user feedback and real-world deployment will be essential to identify edge cases, enhance robustness, and evolve the system into a fully functional AI-powered legal advisory platform.

VI. CONCLUSIONS

This paper presented RemoteCLO, an AI-powered legal advisory system designed to assist users in understanding legal information through context-aware and ethically aligned responses. The proposed platform integrates a Next.js- based frontend with a FastAPI backend, combining Large Language Model reasoning through Google Gemini with LangChain orchestration and real-time contextual retrieval using Tavily Web Search. The system demonstrates how modern AI technologies can support legal information access by generating structured, understandable, and explainable responses within a web-based environment.
The developed prototype, currently at the MVP stage, successfully implements the core workflow of legal query processing, contextual information retrieval, and AI-assisted reasoning. Qualitative evaluation indicates that the system provides relevant and accessible legal explanations while maintaining transparency and responsible AI usage practices. The modular architecture enables scalable deployment and supports seamless interaction between frontend interfaces and backend intelligence components.

Although promising, the current system has limitations, including the absence of quantitative benchmarking and restricted feature coverage. Future work will focus on integrating Retrieval-Augmented Generation (RAG) techniques for stronger contextual grounding, expanding capabilities toward document analysis and generation, and introducing multilingual voice interaction to improve accessibility. These enhancements aim to evolve RemoteCLO into a fully functional AI-powered legal advisory platform. Overall, this work demonstrates the potential of responsibly designed AI systems to improve accessibility to legal information. By combining real-time knowledge retrieval with explainable AI reasoning, RemoteCLO contributes toward bridging the gap between complex legal knowledge and everyday users, supporting more informed decision-making in an increasingly digital legal ecosystem.

REFERENCES

[1] Benjamin Alarie, Anthony Niblett, and Albert H Yoon. How artificial intelligence will affect the practice of law. *University of Toronto Law Journal*, 68(supplement 1):106–124, 2018.

[2] Mohammad Ali. Exploring the role of ai in modern legal practice: Opportunities, challenges, and ethical implications. *J. Electrical Systems*, 20(6s):3040–50, 2024.

[3] Flora Amato, Mattia Fonisto, Marco Giacalone, and Carlo Sansone. An intelligent conversational agent for the legal domain. *Information*, 14(6):307, 2023.

[4] Jhanvi Arora, Tanay Patankar, Alay Shah, and Shubham Joshi. Artificial intelligence as legal research assistant. In *Fire (working notes)*, pages 60–65, 2020.

[5] Jiaxi Cui, Munan Ning, Zongjian Li, Bohua Chen, Yang Yan, Hao Li, Bin Ling, Yonghong Tian, and Li Yuan. Chatlaw: A multi-agent collaborative legal assistant with knowledge graph enhanced mixture-of- experts large language model. *arXiv preprint arXiv:2306.16092*, 2023.

[6] Google AI. Google ai studio documentation. https://aistudio.google. com/, 2024. Accessed: Mar. 2026.

[7] Morgan Gray, Jaromir Savelka, Wesley Oliver, and Kevin Ashley. Using llms to discover legal factors. *arXiv preprint arXiv:2410.07504*, 2024.

[8] Changqin Huang, Yihua Zhong, Xizhe Wang, Zhongmei Han, and Tongquan Wei. From single-agent to multi-agent: motivational learning activities design and empirical study supported by llm-based agents. *Journal of East China Normal University (Educational Sciences)*, 43(5):44, 2025.

[9] Q Huang, M Tao, C Zhang, Z An, C Jiang, Z Chen, Z Wu, and Y Feng. Lawyer llama: Enhancing llms with legal knowledge. arXiv preprint arXiv:2305.15062, 2023.

[10] LangChain. Langchain documentation. https://python.langchain.com/ docs/, 2024. Accessed: Mar. 2026.

[11] Lingyi Meng, Maolin Liu, Hao Wang, Yilan Cheng, Qi Yang, and Idlkaid Mohanmmed. Building from scratch: a multi-agent framework with human-in-the-loop for multilingual legal terminology mapping. Artificial Intelligence and Law, pages 1–40, 2025.

[12] Andrew Mowbray, Philip Chung, and Graham Greenleaf. Utilising ai in the legal assistance sector—testing a role for legal information institutes. Computer Law & Security Review, 38:105407, 2020.

[13] John J Nay, David Karamardian, Sarah B Lawsky, Wenting Tao, Meghana Bhat, Raghav Jain, Aaron Travis Lee, Jonathan H Choi, and Jungo Kasai. Large language models as tax attorneys: a case study in legal capabilities emergence. Philosophical Transactions of the Royal Society A, 382(2270):20230159, 2024.

[14] Marc Queudot, E´ ric Charton, and Marie-Jean Meurs. Improving access to justice with legal chatbots. Stats, 3(3):356–375, 2020.

[15] Tavily AI. Tavily web search api documentation. https://docs.tavily.com/, 2024. Accessed: Mar. 2026.

[16] Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu, and Maosong Sun. Lawformer: A pre-trained language model for chinese legal long documents. AI Open, 2:79–84, 2021.