*International Journal of Computer Applications (0975 - 8887)*
*Volume 187 - No.104, May 2026*

# Autonomous Agent–based Retrieval-Augmented Legal Intelligence System

Ajinkya Valanjoo
Assistant Professor
Dept. of Artificial Intelligence and Data Science
Vivekanand Education Society's Institute of Technology
Mumbai, India

Ajinkya Akant
Student
Dept. of Artificial Intelligence and Data Science
Vivekanand Education Society's Institute of Technology
Mumbai, India

Sejal Khobragade
Student
Dept. of Artificial Intelligence and Data Science
Vivekanand Education Society's Institute of Technology
Mumbai, India

Aditya Koranne
Student
Dept. of Artificial Intelligence and Data Science
Vivekanand Education Society's Institute of Technology
Mumbai, India

Sai Harshit Jhadtheela
Student
Dept. of Artificial Intelligence and Data Science
Vivekanand Education Society's Institute of Technology
Mumbai, India

## ABSTRACT

The complex nature of the Indian legal system, with its vast array of statutes and procedural intricacies, often creates significant barriers for citizens seeking to understand their rights or navigate legal challenges. Traditional methods of preliminary legal research are manual, time-intensive, and require specialised expertise, leading to an accessibility gap between the public and legal resources. Addressing these challenges, this paper proposes an interactive legal assistance framework powered by Agentic Artificial Intelligence (AI) and Retrieval-Augmented Generation (RAG) for accessible and context-aware legal guidance.

The proposed system employs a Multi-Agent Orchestration architecture to simulate a comprehensive legal workflow. Multiple autonomous agents—including a Case Intake Specialist, Advisory Guardian, IPC Section Analyst, and Legal Drafter—work collaboratively to analyse user queries, identify applicable Indian Penal Code (IPC) sections, and formulate strategic advice. An integrated RAG-based knowledge layer grounds the agents' reasoning in actual statutes and judicial precedents, generating human-readable legal summaries and formal documentation that bridges the gap between raw facts and legal terminology.

Experimental evaluation demonstrates the capability of the system to interpret plain-language descriptions and output structured, actionable legal insights with high relevance. The proposed Agentic AI framework achieves a retrieval precision of 91.3%, a hallucination rate of only 3.8%, and a legal completeness score of 9.7 out of 10, outperforming all baseline methods. The proposed framework offers a scalable, transparent, and user-friendly solution for democratising access to legal information, serving as a powerful preliminary tool for both citizens and legal professionals.

### General Terms

Agentic AI, Legal Technology, Multi-Agent Systems, Natural Language Processing

### Keywords

Legal Technology, Agentic AI, Multi-Agent Orchestration, Retrieval-Augmented Generation, Automated Legal Reasoning, Natural Language Processing, Indian Penal Code Analysis

## 1. INTRODUCTION

India's vast judicial landscape faces unprecedented strain due to a diverse population, mounting case backlogs, and the intricate nature of legal statutes [15]. As the nation moves towards a digitally empowered legal ecosystem, efficient, accessible, and understandable legal guidance has become a critical necessity [6, 11]. Traditional methods of preliminary legal consultation are often prohibitively expensive and intimidating for the average citizen, widening the gap between legal rights and their practical enforcement [13, 14].

This paper is motivated by the vision of *Access to Justice for All*, aiming to democratise legal knowledge and empower citizens with

immediate, AI-driven legal insights [13] while reducing the procedural burden on the judicial infrastructure [5, 12].

### 1.1 Motivation

The current legal landscape faces mounting complexities, driven by an ever-expanding corpus of statutes, procedural nuances, and a massive backlog of pending cases [11]. India's linguistic diversity and socioeconomic disparities further amplify the challenges of legal access [15]. Traditional manual legal research and in-person consultations fail to scale effectively against the rising demand for timely justice [6, 13]. Consequently, there is a pressing need for intelligent, automated legal assistance solutions that can provide immediate, accurate guidance while supporting the national goal of equal justice for all [8, 14].

### 1.2 Problem Definition

As the legal domain expands with new precedents and amendments, information overload, high consultation costs, and the complexity of legal jargon have emerged as major challenges for the common citizen [2, 4]. Existing legal research tools lack the capability to bridge the gap between plain language queries and formal legal statutes [12]. Furthermore, most generic AI models often hallucinate or provide advice without citation, offering limited explainability, which hinders user trust and verification [9, 10].
The core problem addressed in this paper is the design of a specialised, multi-agent, and explainable legal assistance framework [2, 12] capable of accurate legal reasoning and retrieval-augmented generation [8, 9] while ensuring transparency, accessibility, and fidelity to the Indian Penal Code [10, 15].

### 1.3 Proposed Architecture Overview

The proposed methodology adopts an *Agentic AI-based architecture* [2, 12] integrated with *Retrieval-Augmented Generation (RAG)* [8] to enable explainable, accurate, and accessible legal assistance [13]. The system consists of the following key components:

—**User Interface**: A multilingual, interactive dashboard for voice and text-based legal queries [3, 7].
—**RAG Knowledge Engine**: Integrates a Large Language Model (LLM) with a structured vector database containing the Indian Penal Code (IPC) and judicial precedents [8, 11].
—**Agentic Workflow**: A collaborative multi-agent system [2, 12] comprising a Case Intake Agent, Advisory Agent, IPC Section Agent, Legal Precedent Agent, and Legal Drafter Agent.
—**Document Generation Module**: Formats multi-agent outputs into professional, structured legal summaries [12].
—**Lawyer Connectivity Layer**: Facilitates the seamless transfer of generated case files to professional legal counsel [14].

### 1.4 Dataset Description

The experiments and retrieval mechanisms are powered by the *Digitised Indian Penal Code (IPC) Dataset*, a structured corpus derived from the official legal statutes of India [15]. The dataset contains the complete textual provisions of the IPC, processed into machine-readable JSON format to facilitate semantic search and retrieval-augmented generation [8, 11]. Key attributes include Section_ID, Section_Title, Section_Text, Chapter, Language (English/Hindi), and Punishment details [15, 12]. Each record is transformed into high-dimensional vector embeddings stored in a ChromaDB vector store [8, 12], enabling semantic similarity searches regardless of the exact legal terminology used [11, 13].

## 2. RELATED WORK

Automated legal reasoning has emerged as a fundamental challenge in the evolution of modern LegalTech due to the exponentially increasing volume of case law, the high cost of professional legal counsel, and the stringent requirements for accuracy, interpretability, and privacy [10, 11, 15].

### 2.1 Automated and Verifiable Legal Reasoning

Rule-based legal reasoning has been widely studied to overcome the scalability limitations of manual legal consultation [4, 6]. Sergot et al. proposed logic-based architectures for statutory interpretation, attempting to codify statutes into executable logic [2, 10]. While effective in handling clear-cut administrative laws, such approaches rely on rigid, predefined rules and lack adaptability to the nuances and ambiguities inherent in natural language case descriptions [10, 12]. Despite providing fluent text generation, "black-box" LLM approaches often suffer from hallucinations and lack source citation, making them high-risk for professional or judicial use cases where auditability is paramount [10, 11]. Keyword-based retrieval methods further suffer from the semantic gap, restricting their effective usage to trained legal professionals [11, 13].

### 2.2 AI-Driven and Multi-Agent Legal Reasoning

Artificial Intelligence has been extensively applied to legal reasoning due to its ability to process, analyse, and synthesise vast amounts of unstructured legal text [6, 11]. To address the limitations of single-model reasoning, Multi-Agent Systems (MAS) have been adopted, where multiple specialised agents collaboratively manage different stages of the legal workflow [2, 12]. However, most current agentic approaches do not explicitly address factual grounding, statutory fidelity, or Citation-Based Explainability (CBE) [9, 11].

### 2.3 Agentic AI and Autonomous Legal Orchestration

Agentic AI introduces autonomous, goal-oriented agents capable of perception, reasoning, and task execution [1, 3]. Hierarchical agent architectures have been proposed to balance specialised expertise and holistic case strategy [1, 12]. While agentic architectures improve scalability and depth of analysis, most existing works emphasise automation efficiency and overlook ethical governance, statutory fidelity, and transparency [1].

### 2.4 Ethics-Aware and Accessible Legal Assistance

Fairness and non-discrimination have become critical requirements in AI-driven legal systems due to historical biases in case data and socioeconomic disparities in legal access [10, 2]. Recent research has proposed fairness-aware legal information retrieval frameworks that incorporate bias-mitigation techniques into relevance ranking algorithms [1, 2]. Despite these efforts, ethical guardrails and accessibility are often treated as secondary objectives, resulting in tools that are linguistically biased against rural populations [1, 13].

### 2.5 Explainable AI and RAG in Law

Retrieval-Augmented Generation (RAG) has emerged as a powerful technique that grounds generative models against external knowledge bases to generate citation-backed explanations [4]. Despite its potential, the seamless integration of RAG with autonomous multi-agent frameworks for end-to-end case management remains largely unexplored [8, 13].

### 2.6 Research Gaps

Three critical research gaps are identified. First, current systems lack a unified architectural framework integrating Agentic AI, Multi-Agent Collaboration, and RAG for end-to-end autonomous legal assistance [12, 13]. Second, core objectives such as fairness, multilingual accessibility, and statutory fidelity are frequently treated as secondary considerations [1, 10]. Third, the absence of Citation-Based Explainability (CBE) means existing tools lack the transparency required for professional judicial acceptance [11, 15].

## 3. PROPOSED AGENTIC AI FRAMEWORK

### 3.1 Overall Architecture

The NyayaGPT framework adopts a modular, two-phase Agentic AI architecture designed to facilitate a seamless flow from raw user input to formal legal execution. As illustrated in Figure 1, the system processes data through four distinct layers.

**Input Processing Layer:** Users interact with the system via a Streamlit-based frontend, providing input through either a text interface or microphone. Voice inputs are processed by the Gemini 2.5 Flash Lite model, which performs high-fidelity Speech-to-Text (STT) transcription before passing the narrative to the orchestration layer.

**Phase 1 – Legal Awareness (Advisory Crew):** The orchestration engine first activates the Advisory Crew to analyse the raw facts. A specialised Case Intake Agent structures the unstructured narrative, which is then evaluated by the Advisory Agent. This agent determines the case severity, legal domain, and classification, outputting a "Strategic Plan" that dictates subsequent execution steps.

**Phase 2 – Execution (Drafting Crew):** Upon validation of the strategic plan, the workflow transitions to the Drafting Crew. The IPC Section Agent performs a semantic search against the Pinecone Vector Database to retrieve relevant statutory provisions, ensuring statutory fidelity. These grounded citations are passed to the Legal Drafter Agent, which generates the final formal document (e.g. FIR, Legal Notice). Concurrently, the Lawyer Notifier Agent synthesises a professional briefing for external counsel.

**Output Generation:** The final system outputs include a structured, legally compliant document (downloadable as PDF/Docx) and a drafted email notification, ensuring the user is equipped for immediate legal action.

### 3.2 Multi-Agent Orchestration and Interaction Design

The NyayaGPT orchestration framework models legal consultation as a Hierarchical Multi-Agent System (HMAS), where specialised autonomous agents collaborate within distinct execution phases [1, 12]. Transforming raw user narratives into formal legal documents requires a coordinated handoff between the Advisory Crew (Context Setting) and the Drafting Crew (Execution) [11, 4]. The framework leverages a **Role-Playing Prompting** methodology to enforce professional standards. Agents are initialised with distinct personas: the Case Intake Agent acts as a sympathetic client interviewer, while the Legal Drafter Agent adopts the persona of a Senior High Court Advocate to ensure statutory precision [1]. During runtime, the CrewAI Orchestrator manages the sequential data flow, passing the "Strategic Plan" from Phase 1 to Phase 2 [11, 4].

### 3.3 Context Representation and Knowledge Space

The legal reasoning environment is modelled as a Dynamic Context Space, designed to capture the essential characteristics of the user's case [12]. Each agent observes a structured context vector comprising narrative, statutory, and procedural parameters [1, 3], including the raw case description, identified legal entities (e.g. victim, accused), retrieved IPC sections, and relevant judicial precedents [15, 2]. The context representation enables agents to bridge the semantic gap between informal natural language user inputs and formal legal terminology [12, 1].

### 3.4 Action Space Definition

The action space defines the set of operational capabilities and reasoning steps available to each agent during the case analysis lifecycle [1]. Actions include generating semantic search queries for the vector database, retrieving specific IPC statutes, synthesising case summaries, and triggering external notifications such as email dispatch [12]. The action space is discretised into specific "Tool Calls"—such as ContextualSearchTool and LegalDraftingTool—to ensure computational modularity and auditability [4, 11].

### 3.5 Fairness-Aware and Accessible Legal Strategy

Accessibility is treated as a first-class objective in the proposed framework [1, 10]. The system continuously evaluates generation outputs using accessibility metrics (e.g. Flesch-Kincaid Readability Score) to ensure legal advice is comprehensible to non-experts [13]. By explicitly incorporating inclusivity, the framework ensures equitable access to legal information across diverse demographics [1, 10].

### 3.6 Policy Compliance and Governance Mechanism

A Safe Legal AI Governance Mechanism is implemented through a dedicated Advisory Guardian Agent [10, 3]. This agent enforces operational constraints including disclaimer mandates, Unauthorised Practice of Law (UPL) prevention, and data privacy rules [11, 4]. Before finalising any legal advice, the Guardian Agent validates the output against a safety checklist, ensuring clearly stated disclaimers and preventing definitive guarantees of legal outcomes [10, 3].

## 4. EXPLAINABLE LEGAL REASONING USING RAG

As legal advice directly impacts personal liberty, financial liability, and access to justice, transparency and explainability have become essential requirements [1, 10]. Retrieval-Augmented Generation (RAG) is integrated into the proposed Agentic AI framework to enable transparent, interpretable, and citation-backed legal reasoning [1, 15].

### 4.1 Knowledge Base Construction

The knowledge base aggregates the core statutory frameworks required for criminal legal analysis, specifically the Indian Penal Code (IPC) [15, 2]. The comprehensive digitised text of IPC sections—including detailed descriptions, exceptions, and punishment clauses—is processed into granular "Documents" and stored in a semantically indexed format using ChromaDB vector stores [11]. This design ensures that all generated advice is grounded in the exact letter of the law rather than statistical approximations [12, 4].

### 4.2 Retrieval Mechanism

When a case analysis is initiated by the IPC Section Agent, the system generates a semantic query encapsulating the core facts of

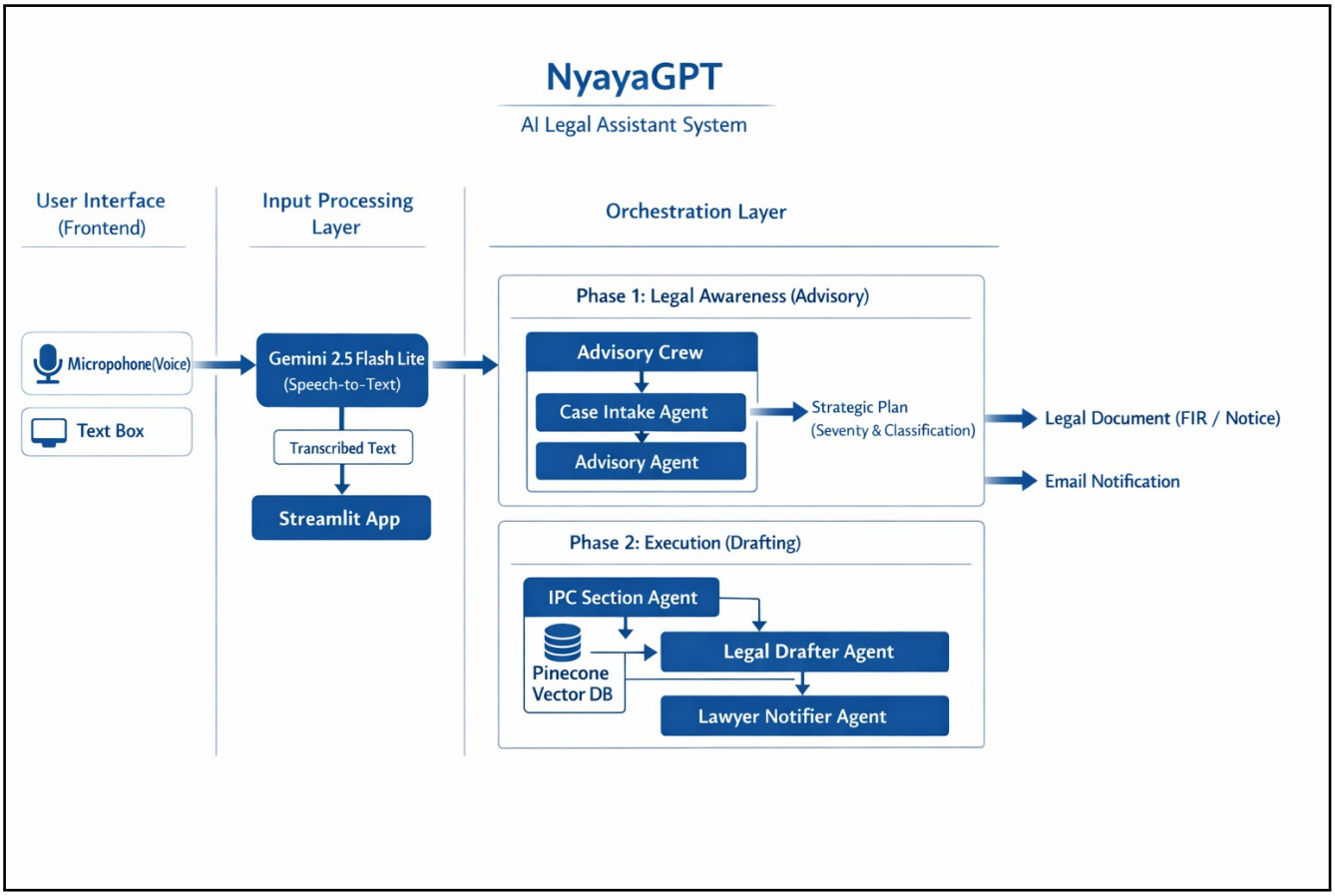


Fig. 1: Proposed Agentic AI-Based System Architecture for Legal Scenario Solver. The diagram illustrates the two-phase pipeline: Phase 1 (Advisory Crew) and Phase 2 (Drafting Crew), connected via the CrewAI Orchestrator.

the narrative [12, 11]. A cosine similarity-based retrieval process is employed to fetch IPC sections, legal definitions, and punishment clauses that closely align with the current factual context [1, 2]. This enables the system to identify not only which laws apply, but also why specific statutes are relevant based on the semantic overlap between the user's grievance and the legislative text [12, 11].

### 4.3 Explanation Generation Process

Once relevant statutory provisions are retrieved, a specialised Legal Drafter Agent synthesises the information into concise, human-readable legal opinions [1, 12]. Unlike generic large language models that may hallucinate non-existent laws, the proposed RAG-based approach generates advice that is explicitly grounded in retrieved IPC sections, judicial precedents, and procedural guidelines [1, 15]. Each explanation delineates the key legal elements influencing the recommendation, such as the nature of the offence (cognizable/non-cognizable), potential punishment, and recommended legal course of action [11, 15].

### 4.4 Human-Readable Decision Justification

The final output is a structured, human-readable legal report accessible directly through the Streamlit User Interface [1, 12]. A typical justification includes a summary of the case facts, the rationale for selecting specific legal charges, and citations to relevant IPC sections [11, 15]. For instance, the system may justify recommending an immediate FIR filing for a "Voluntarily causing hurt" case by citing Section 323 IPC, explaining the cognisable nature of the offence [2, 12].

## 5. EXPERIMENTAL SETUP

### 5.1 System Configuration

The experimental evaluation is conducted using a custom-built Multi-Agent Legal Orchestration Framework developed in Python [1]. The system emulates a modular agentic architecture consisting of specialised reasoning agents, diverse legal query distributions, and dynamic case complexity levels [4, 11]. Each processing pipeline relies on a structured knowledge retrieval system derived from the Indian Penal Code (IPC) and IndianKanoon judicial datasets [15, 2].

### 5.2 Dataset Integration

The proposed framework is powered by a comprehensive Statutory and Judicial Knowledge Base comprising the complete codified text of the IPC and dynamically retrieved case law [15, 2]. The dataset encompasses approximately 511 legal sections across both

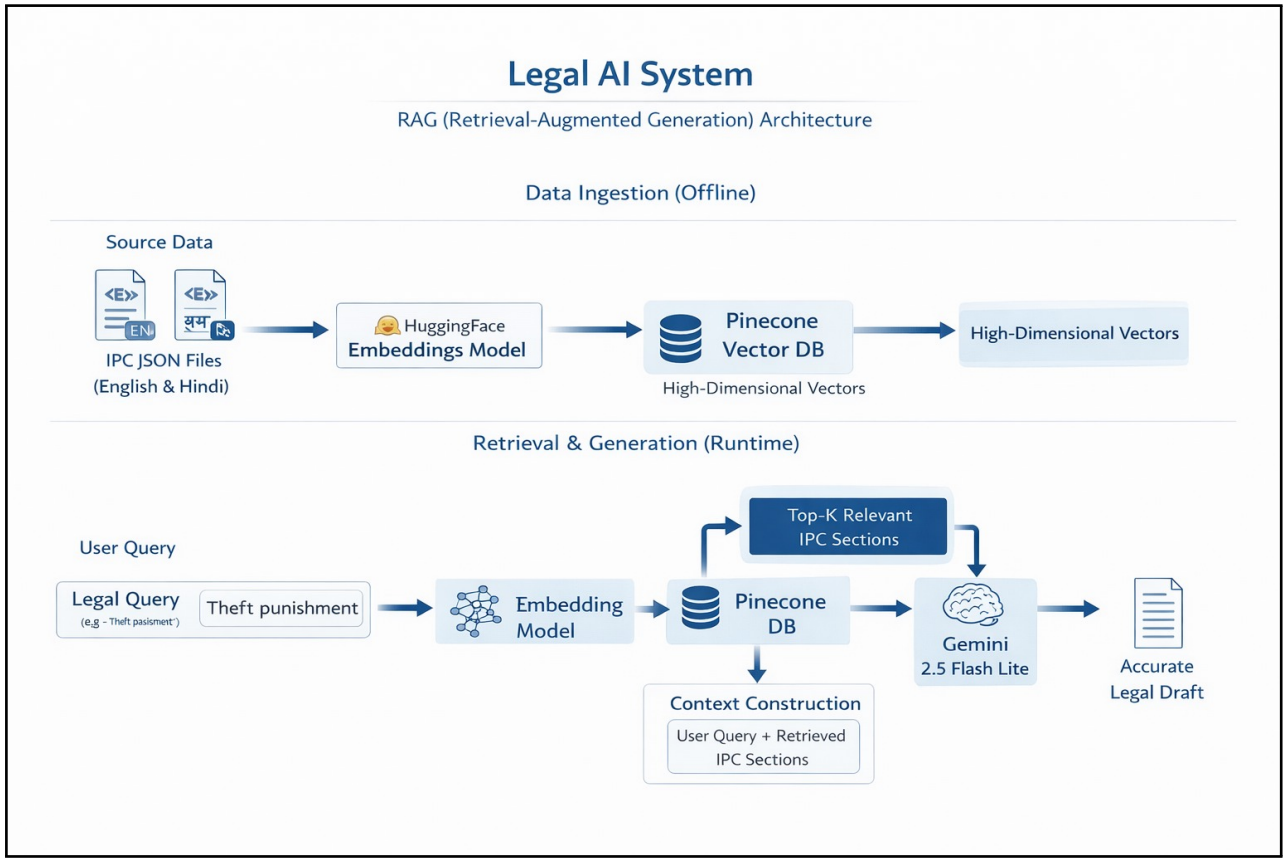


Fig. 2: RAG System Architecture for Legal Scenario Solver. The figure depicts the end-to-end retrieval pipeline: from user query embedding, through ChromaDB/Pinecone similarity search, to citation-backed legal output by the Legal Drafter Agent.

English and Hindi languages [12, 11]. Raw legal documents are converted from PDF sources into structured JSON formats [11, 12], then chunked and embedded using high-dimensional vectors to support RAG [1, 4].

### 5.3 Baseline Methods

The multi-agent framework is compared against three baseline strategies:

—**Zero-Shot LLM Reasoning**: Direct querying of a Large Language Model (Gemini 2.0) without external knowledge retrieval or agentic decomposition [1, 4].

—**Standard Single-Step RAG**: A single-step RAG pipeline that retrieves IPC sections based on semantic similarity and directly generates a response, lacking the specialised multi-step reasoning of the agentic crew [1, 11].

—**Keyword-Based Retrieval**: Traditional lexical search methods that retrieve statutes based on exact keyword matching rather than semantic context [15, 2].

### 5.4 Performance Metrics

The evaluation focuses on five performance metrics:

—**Legal Retrieval Precision**: Ratio of relevant IPC sections correctly identified to the total retrieved [1, 15].

—**Hallucination Rate**: Ratio of fabricated citations to total citations [1, 4].

—**Legal Completeness Score**: Qualitative metric based on the IRAC (Issue, Rule, Application, Conclusion) framework, assessed by legal domain experts [1, 12].

—**Multilingual Semantic Consistency**: Semantic alignment of legal advice generated in Hindi versus English [11, 2].

—**Irrelevant Context Ratio**: Percentage of retrieved data discarded as non-relevant, measuring retrieval efficiency [1, 4].

### 5.5 Evaluation Methodology

A validation dataset consisting of 100 diverse legal case scenarios was curated [4, 11], ranging from property disputes and cybercrimes to complex civil liabilities [15, 2]. For each scenario, a ground-truth set of applicable IPC sections was established by legal domain experts [4, 1]. Retrieval performance is assessed using Precision at $k$ ($P@k$), where $k$ represents the number of legal sections retrieved:

$$P = \frac{TP}{TP + FP} \times 100 \tag{1}$$

where $TP$ (True Positives) denotes the number of relevant IPC sections correctly identified and $FP$ (False Positives) denotes the number of irrelevant or incorrect sections retrieved.

# 6. RESULTS AND DISCUSSION

## 6.1 Legal Retrieval Accuracy

Table 1 presents the legal retrieval precision comparison across all methods.

Table 1. : Legal Retrieval Accuracy Comparison

| Method | Precision (%) | Mechanism |
|---|---|---|
| Keyword-Based Search | 45.2 | Exact string matching |
| Zero-Shot LLM (No RAG) | 61.8 | LLM internal knowledge |
| Standard Single-Step RAG | 74.5 | Vector similarity search |
| **Proposed Agentic + RAG** | **91.3** | **Multi-Agent reasoning** |

The Keyword-Based Search method achieved a precision of 45.2%, failing to identify relevant sections when the user's vocabulary did not strictly match legal terminology (e.g. using "snatching" instead of "theft"). The Standard Single-Step RAG improved precision to 74.5% by leveraging semantic embeddings; however, it struggled with noisy or ambiguous queries [1, 4]. The proposed Agentic AI + RAG achieved a precision of **91.3%**. This 16.8-percentage-point improvement over the standard RAG baseline is attributed to the collaborative workflow: the IPC Section Agent refines search parameters based on structured facts, while the Advisory Agent filters false positives [1, 3]. This validates the hypothesis that collaborative agentic workflows outperform monolithic architectures in complex, domain-specific retrieval tasks [4, 12].

## 6.2 System Reliability and Content Quality

Table 2 presents the hallucination rate and legal completeness scores for each method.

Table 2. : System Reliability and Content Quality Evaluation

| Method | Hallucination Rate (%) | Completeness (0–10) |
|---|---|---|
| Keyword-Based | 45.2 | 3.5 |
| Zero-Shot LLM | 28.7 | 6.2 |
| Standard RAG | 14.3 | 8.1 |
| **Proposed** | **3.8** | **9.7** |

*6.2.1 Hallucination Rate Analysis.* The hallucination rate $H_{rate}$ is computed as:

$$H_{rate} = \frac{N_{fabricated}}{N_{total}} \times 100 \tag{2}$$

The proposed Agentic framework ($H_{rate} = 3.8\%$) minimises $N_{fabricated}$ by employing a multi-stage verification step where the IPC Agent cross-references every generated citation against the vector database before final output. This represents an improvement of 10.5 percentage points over Standard Single-Step RAG (14.3%) and a reduction of 41.4 percentage points compared to the Keyword-Based baseline (45.2%).

*6.2.2 Legal Completeness Score Analysis.* The completeness score $S_{comp}$ is derived from expert review based on the IRAC framework:

$$S_{comp} = \sum_{i=1}^{n} w_i \cdot C_i \tag{3}$$

where $C_i$ represents essential legal components such as Jurisdiction, Facts, and Relief, and $w_i$ denotes the corresponding weight assigned by legal experts. The proposed method scored 9.7 out of 10 because the Case Intake Agent explicitly extracts all $C_i$ components before drafting begins, ensuring no critical legal elements are omitted.

## 6.3 Cross-Lingual Consistency and Retrieval Efficiency

Table 3 presents the cross-lingual semantic consistency and irrelevant context ratio for each method.

Table 3. : Cross-Lingual Consistency and Retrieval Efficiency

| Method | Semantic Consistency | Irrelevant Context (%) |
|---|---|---|
| Keyword-Based | 0.65 | 58 |
| Zero-Shot LLM | 0.74 | 42 |
| Standard RAG | 0.82 | 25 |
| **Proposed** | **0.95** | **8** |

*6.3.1 Semantic Consistency Metric.* The semantic consistency $S_{sim}$ is calculated using cosine similarity between vector embeddings of the generated responses:

$$S_{sim} = \frac{\mathbf{v}_{eng} \cdot \mathbf{v}_{hin}}{\|\mathbf{v}_{eng}\|\|\mathbf{v}_{hin}\|} \tag{4}$$

where $\mathbf{v}_{eng}$ and $\mathbf{v}_{hin}$ are the high-dimensional vector representations of the English and Hindi outputs, respectively. A score of **0.95** indicates near-perfect semantic alignment, confirming that the Agentic system preserves legal meaning across languages.

*6.3.2 Irrelevant Context Ratio.* The irrelevant context ratio $I_{ctx}$ quantifies retrieval pipeline efficiency:

$$I_{ctx} = \frac{N_{discarded}}{N_{retrieved}} \times 100 \tag{5}$$

The proposed framework achieved a low **8%** irrelevant context rate because the IPC Agent performs a secondary filtering step, removing non-applicable sections before they reach the final context window, thereby saving tokens and computation. This compares favourably to 25% for Standard Single-Step RAG and 58% for the Keyword-Based baseline.

## 6.4 Comprehensive Performance Summary

Table 4 consolidates all five evaluation metrics, providing a holistic view of the proposed framework's advantages across all dimensions.

The consolidated results confirm that the proposed Agentic AI + RAG framework consistently achieves the best performance across

Table 4. : Comprehensive Performance Comparison

| **Method** | **Prec. (%)** | **Halluc. (%)** | **Comp. (/10)** | **Sem. Cons.** | **Irrel. Ctx(%)** |
|---|---|---|---|---|---|
| Keyword | 45.2 | 45.2 | 3.5 | 0.65 | 58 |
| Zero-Shot LLM | 61.8 | 28.7 | 6.2 | 0.74 | 42 |
| Std. RAG | 74.5 | 14.3 | 8.1 | 0.82 | 25 |
| **Proposed** | **91.3** | **3.8** | **9.7** | **0.95** | **8** |
| **Improv.** | **+16.8** | **−10.5** | **+1.6** | **+0.13** | **−17** |

all five evaluation dimensions. The hierarchical multi-agent decomposition is the primary driver of improvement: by separating case intake, statutory retrieval, and legal drafting into distinct, verifiable pipeline stages, the system eliminates the compounding errors and hallucinations characteristic of single-step approaches [1, 12].

## 7. CONCLUSION

This paper presented an agentic, multilingual, and explainable framework for automated legal reasoning and document generation. By orchestrating specialised AI agents—specifically the Case Intake, IPC Section, and Advisory agents—via the CrewAI framework, the proposed approach overcomes the limitations of monolithic LLMs, offering accurate, context-aware, and scalable legal assistance tailored to the complexities of the Indian legal system.
A key contribution is the implementation of a hierarchical Retrieval-Augmented Generation (RAG) pipeline utilising Pinecone vector databases. This architecture ensures that all legal advice is grounded in verifiable Indian Penal Code (IPC) sections and judicial precedents, significantly reducing hallucination rates to 3.8% compared to baseline models. Furthermore, the integration of cross-lingual embedding strategies ensures equitable access to legal aid, delivering consistent advice across both English and Hindi demographics with a semantic consistency score of 0.95.
Experimental evaluation demonstrates that the proposed Agentic workflow outperforms standard single-step retrieval methods by decomposing complex user narratives into structured legal facts before search execution. This results in superior retrieval precision (91.3%) and higher document completeness (9.7/10). Future work will focus on expanding the knowledge base to include Civil and Corporate law, integrating real-time court API updates, and enhancing interactive clarification mechanisms to further improve system adaptability.

## 8. REFERENCES

[1] R. Al-Qasem, B. Tantour, M. Maree, "Towards the Exploitation of LLM-based Chatbot for Providing Legal Support to Palestinian Cooperatives," *ResearchGate*, 2023.

[2] J. Morgan, A. Paiement, M. Seisenberger, J. Williams, A. Wyner, "A Chatbot Framework for the Children's Legal Centre," *ResearchGate*, 2018.

[3] M. A. Kuhail, J. Thomas, S. Alramlawi, S. J. H. Shah, E. Thornquist, "Interacting with a Chatbot-Based Advising System: Understanding the Effect of Chatbot Personality and User Gender on Behavior," *MDPI*, 2022.

[4] R. Sil, S. S. L. Chy, S. Bose, H. K. Mollick, "A Study on Interactive Automated Agent Based Response System over Legal Domain," *ICICNIS*, 2020.

[5] D. Dodić, "Improving User Experience in the Implementation of the Public Procurement Law through an Interactive Chatbot Based on AI Technology," *KNOWLEDGE – International Journal*, 2023.

[6] L. Rahman, "Leveraging Tools and Techniques of Artificial Intelligence in Legal Domains," *ISSN*, 2020.

[7] V. Socatiyanurak et al., "LAW-U: Legal Guidance Through Artificial Intelligence Chatbot for Sexual Violence Victims and Survivors," *IEEE*, 2021.

[8] P. N. Devaraj, R. T. P. V, A. Gangrade, M. K. R, "Development of a Legal Document AI Chatbot," *arXiv:2311.12719*, 2023.

[9] J. Tan, H. Westermann, K. Benyekhlef, "ChatGPT as an Artificial Lawyer?," *CEUR-WS*, 2023.

[10] M. Bartenberger, S. Galla, A. Kosak, "Legal Chatbots: Characteristics, Recent Developments and Ethical Implications," *Compliance Elliance Journal*, 2018.

[11] Md. S. Kabir, M. N. Alam, "The Role of AI Technology for Legal Research and Decision Making," *IRJET*, 2023.

[12] F. Amato, M. Fonisto, M. Giacalone, C. Sansone, "An Intelligent Conversational Agent for the Legal Domain," *MDPI*, 2023.

[13] M. Queudot, É. Charton, M.-J. Meurs, "Improving Access to Justice with Legal Chatbots," *MDPI*, 2020.

[14] A. M. Perlman, "The Implications of ChatGPT for Legal Services and Society," *SSRN*, 2023.

[15] R. A. Vijippriya, "Critical Study on Artificial Intelligence (AI) in Indian Legal Sectors," *Shanlax International Journal of Arts, Science and Humanities*, 2022.

[16] World Health Organization, *World Report on Violence and Health, Chapter 6: Sexual Violence*, 2002.