Smart Legal Advisor: An AI-Powered System for Indian Penal Code Interpretation

DR.S. Gnanapriya[1],Prasya P[2]

[1]Associate professor, Department of Computer Applications, Nehru College of Management, Coimbatore, Tamil Nadu, India.

ncmdrsgnanapriya@gmail.com

[2]Student of II MCA, Department of Computer Applications, Nehru College of Management, Coimbatore, Tamil Nadu, India.

prasyaparameswaran@gmail.com

Abstract

Legal information in India is often complex and inaccessible to the general public due to technical language and limited legal awareness. This paper presents Legal Advisor Pro, an AI-powered legal assistance system designed to simplify and democratize access to provisions of the Indian Penal Code (IPC 1860) and the Bharatiya Nyaya Sanhita (BNS 2023). The system leverages Natural Language Processing (NLP) and machine learning techniques, including semantic search and text classification, to interpret user queries expressed in natural language and map them to relevant legal sections, offenses, and procedural guidance. Developed using Python and deployed through a Streamlit-based web interface, the application ensures usability, scalability, and minimal hardware requirements. The system supports multilingual input, enabling wider adoption across diverse user groups. Emphasis is placed on ethical compliance by providing informational assistance without offering definitive legal advice. Experimental evaluation demonstrates that the proposed system improves legal information retrieval accuracy and user understanding when compared to traditional keyword-based search methods. Legal Advisor Pro highlights the potential of AI-driven LegalTech solutions in bridging the gap between complex legal frameworks and common citizens.

Keywords: - Artificial Intelligence; LegalTech; Natural Language Processing (NLP); Indian Penal Code (IPC); Bharatiya Nyaya Sanhita (BNS); Semantic Search; Machine Learning; AI Legal Assistant; Legal Information Retrieval

1. INTRODUCTION

Legal Advisor Pro is an innovative, AI-driven web application developed using Streamlit, a Python framework for building interactive interfaces, aimed at making India's penal legal system more accessible and understandable for legal professionals, students, and everyday citizens. It achieves this by harnessing advanced machine learning techniques, such as natural language processing (NLP), to facilitate intuitive interactions with the Indian Penal Code (IPC) of 1860—a foundational criminal law framework—and its modern successor, the Bharatiya Nyaya Sanhita (BNS) of 2023, which introduced reforms to align with contemporary societal needs, including digital crimes and procedural updates. The project directly tackles prevalent issues in India's judicial landscape, including the overwhelming complexity of legal texts that often leave non-experts confused, uncertainties arising from the transition between IPC and BNS (such as renumbered sections and redefined offenses), frequent procedural errors in First Information Report (FIR) filings due to lack of guidance, incomplete evidence collection that weakens cases, and language barriers that exclude non-English speakers from effective legal aid. To address these, Legal Advisor Pro incorporates a suite of user-friendly modules: a semantic law search feature that uses meaning-based retrieval to find relevant IPC/BNS sections from natural language queries; an IPC-BNS Bridge for visual comparisons and diff analyses highlighting additions, removals, and reforms; a Smart FIR Assistant that generates structured

complaint drafts, auto-suggests applicable legal sections, and provides tailored evidence checklists based on crime categories like crimes against persons or cyber offenses; a Bail Predictor that evaluates offense severity to assess bailable or non-bailable status; and an OCR Scanner that extracts text from uploaded documents, such as FIR images, to identify and map IPC references to BNS equivalents. The application supports multilingual interfaces in English, Hindi, and Malayalam, ensuring inclusivity across diverse user groups, and employs robust technologies including Sentence Transformer models (e.g., all-MiniLM-L6-v2) for generating semantic embeddings that enable precise query matching via cosine similarity, LinearSVC classifiers trained on TF-IDF vectorized data for categorizing crimes into domains such as property or financial offenses, and PyTorch for efficient deep learning computations. By streamlining legal research, automating document generation, and providing predictive insights, the tool not only enhances legal literacy and reduces procedural errors but also alleviates burdens on overburdened judicial systems, empowering users to navigate India's evolving penal framework with confidence. Ultimately, Legal Advisor Pro exemplifies the transformative potential of AI in legal technology, promoting equitable access to justice in a digital era while paving the way for future expansions like real-time case law integration and advanced predictive analytics

2. LITERATURE REVIEW

The application of Artificial Intelligence (AI) and Natural Language Processing (NLP) in the legal domain has gained increasing attention in recent years, particularly in the areas of legal information retrieval, case analysis, and decision support systems. However, the development of intelligent legal assistance tools for statutory interpretation and public legal awareness remains relatively underexplored, especially within the context of Indian penal laws.

Early legal information systems primarily relied on rule-based approaches and keyword-driven search mechanisms to retrieve relevant legal provisions. While these systems offered basic access to legal texts, they were limited in their ability to understand user intent or interpret natural language queries. Keyword-based retrieval often resulted in irrelevant outputs when users employed layman terms, colloquial expressions, or incomplete legal references, thereby reducing usability for non-expert users.

With the advancement of machine learning, researchers introduced supervised and unsupervised learning techniques to enhance legal document classification and retrieval. Traditional classifiers such as Naïve Bayes, Logistic Regression, and Support Vector Machines were applied to categorize legal documents and statutes based on predefined labels. Although these models demonstrated reasonable performance on structured legal datasets, their effectiveness declined when handling semantic ambiguity and contextual variations inherent in legal language.

To address the limitations of linear models, ensemble learning techniques such as Random Forest and Gradient Boosting were adopted for legal text classification and decision support. These methods showed improved robustness and accuracy by capturing nonlinear relationships among legal features and reducing overfitting. Studies reported that ensemble models outperform single classifiers, particularly in complex legal datasets involving multiple offense categories and overlapping statutory provisions.

Recent research has emphasized the role of NLP-driven semantic modeling in LegalTech applications. Techniques such as TF-IDF vectorization, n-gram analysis, and word embeddings have been widely used to transform unstructured legal texts into meaningful numerical representations. More advanced approaches employ sentence embeddings and semantic similarity models to match user queries with relevant legal sections. These studies indicate that semantic-based retrieval significantly improves relevance and user comprehension compared to traditional keyword search.

Deep learning architectures, including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have also been explored for legal text understanding and classification. While these models demonstrate high accuracy, their adoption is often constrained by the need for large annotated datasets, extensive computational resources, and reduced interpretability—factors that limit practical deployment in public-facing legal assistance systems.

This work builds upon existing literature by integrating NLP-based semantic understanding with machine learning–assisted legal classification, focusing on accessibility, interpretability, and ethical compliance. Unlike prior systems that emphasize predictive outcomes, Legal Advisor Pro prioritizes informational assistance, ensuring accurate legal mapping while avoiding definitive legal judgments. The proposed approach aims to bridge the gap between complex legal statutes and common citizens through an efficient, scalable, and user-centric AI framework.

3. METHODOLOGY

3.1 System Overview

Legal Advisor Pro is implemented as a web-based AI legal assistance system using Streamlit, providing an intuitive interface for users to submit legal queries in natural language. The system accepts input in the form of free-text queries, partial legal descriptions, or hypothetical scenarios related to criminal law. These inputs are processed to retrieve relevant statutory provisions from the Indian Penal Code (IPC 1860) and the Bharatiya Nyaya Sanhita (BNS 2023).

The overall architecture is modular and consists of four core components:

- Query Processing Module for input normalization and linguistic analysis.
- NLP and Semantic Analysis Engine   for feature extraction and similarity computation.
- Legal Classification and Mapping Module for offense categorization and section retrieval.
- Presentation Layer for user-friendly display of results and explanations.

Data flows sequentially from query preprocessing to semantic embedding, classification, legal mapping, and final response generation, ensuring scalability, maintainability, and ethical compliance.

3.2 Query Processing and NLP Engine

The Query Processing Engine handles unstructured user inputs without requiring formal legal terminology. Incoming queries undergo preprocessing steps including lowercasing, punctuation removal, stop-word elimination, and tokenization. Language detection is performed to support multilingual input, with translation applied where required to maintain uniform semantic representation.

Natural Language Processing (NLP) techniques are employed to capture both lexical and contextual meaning. TF-IDF vectorization is used for keyword relevance analysis, while sentence-level embeddings are generated using transformer-based models to encode semantic intent. This dual representation enables the system to handle linguistic variations, colloquial expressions, and ambiguous phrasing commonly used by non-legal users.

3.3 Legal Knowledge Base and Semantic Mapping

The legal knowledge base is constructed from authoritative statutory texts, including IPC and BNS sections, offense descriptions, punishments, and procedural notes. Each provision is embedded into a semantic vector space using the same embedding model applied to user queries, enabling cosine similarity–based matching. A semantic retrieval mechanism ranks relevant legal sections based on similarity scores. To enhance accuracy, a rule-assisted mapping layer validates retrieved sections against offense categories and statutory constraints, preventing irrelevant or misleading outputs. This approach ensures that retrieved legal information remains contextually valid and legally consistent.

3.4 Classification and Decision Logic

For offense categorization, machine learning classifiers are employed to map user queries to predefined legal categories such as property offenses, personal offenses, cybercrime, or public order violations. Lightweight supervised models are selected to balance accuracy and deployment feasibility. The classification output is combined with semantic similarity results to produce a final ranked list of applicable legal provisions. Confidence thresholds are applied to filter low-relevance matches. Importantly, the system is designed to provide informational guidance only, explicitly avoiding definitive legal conclusions or advice.

3.5 Output Generation and User Interaction

The Presentation Layer renders retrieved legal information through structured summaries, including applicable sections, offense descriptions, and general procedural insights. Results are displayed in a clear and interpretable format, enabling users to understand legal implications without requiring prior legal expertise.

Streamlit session management is used to maintain interaction history, allowing users to refine queries or

explore related provisions. The interface emphasizes clarity, transparency, and accessibility, supporting informed legal awareness rather than prescriptive decision-making.

3.6 Implementation Details

The system is developed in Python, utilizing libraries such as NLTK, scikit-learn, and sentence-transformers for NLP and machine learning tasks. Streamlit session state is used to persist data across interactions, while modular function design ensures maintainability. Error handling mechanisms manage invalid inputs and model uncertainties gracefully.The architecture is optimized for real-time performance, with typical response times under a few seconds for standard queries. The lightweight design enables deployment on modest hardware configurations, making the system suitable for broad public access.

4. EXISTING SYSTEM

- Legal information is mainly accessed through manual law books and websites.

- Searching IPC sections requires exact keywords or section numbers.

- No direct and clear mapping between IPC 1860 and BNS 2023.

- FIR drafting is done manually at police stations, which is time-consuming.

- Evidence collection depends on human experience and memory.

- Bail eligibility understanding requires direct consultation with lawyers.

- Legal documents are checked manually without AI support.

- Multilingual legal assistance is limited or unavailable.

5. PROPOSED SYSTEM

- Provides an AI-powered legal assistance platform in a single interface.

- Uses intelligent search to find laws based on user description.

- Automatically maps and compares IPC and BNS sections.

- Generates structured FIR drafts automatically from incident details.

- Suggests crime-based evidence checklists using AI classification.

- Predicts bail eligibility based on offense severity.

- Supports OCR-based legal document analysis.

- Offers multilingual support for better legal accessibility.

6. IMPLEMENTATIONS

The implementation of Legal Advisor Pro focuses on realizing an AI-driven legal assistance system that is efficient, scalable, and ethically compliant. The system is developed using Python and deployed as a web-based application to ensure accessibility for users without legal or technical expertise. A modular design approach is adopted to allow independent development, testing, and future enhancement of system components.

6.1 Development Environment

The application is implemented in Python due to its extensive support for Natural Language Processing (NLP) and machine learning libraries. Core NLP tasks are performed using NLTK and sentence-transformers, while classification and similarity computations are handled using scikit-learn. The web interface is developed using Streamlit, which enables rapid development of interactive applications. Legal data, including IPC and BNS sections, is stored in structured formats such as CSV and JSON to facilitate efficient retrieval.

6.2 Query Input and Preprocessing

User interaction begins with the submission of a legal query through the web interface. The input text undergoes preprocessing steps including lowercasing, removal of punctuation and special characters, tokenization, and stop-word elimination. Language detection mechanisms are employed to handle multilingual inputs, with translation applied when required to standardize semantic analysis. These preprocessing steps ensure consistency and reduce noise in subsequent processing stages.

6.3 Feature Extraction and Semantic Representation

To capture both lexical relevance and contextual meaning, a hybrid feature extraction strategy is employed. TF-IDF vectorization is used to identify important terms within the query, while sentence embeddings are generated to represent semantic intent. Legal provisions from IPC and BNS are pre-embedded using the same embedding model, enabling direct comparison between user queries and statutory texts through cosine similarity measures.

6.4 Legal Section Retrieval and Mapping

Semantic similarity scores are computed between the query embedding and statutory embeddings to retrieve relevant legal sections. The retrieved sections are ranked based on similarity values. A rule-assisted mapping layer validates the retrieved results by aligning them with predefined offense categories and statutory constraints. Low-confidence results are filtered using threshold-based criteria to improve relevance and accuracy.

7. RESULT

To evaluate the effectiveness and practical behavior of the proposed Legal Advisor Pro system, multiple visual analytics were generated based on classified legal queries and semantic similarity scores. The results demonstrate the system's capability to categorize legal inputs accurately and retrieve relevant statutory provisions with high confidence.

7.1 Distribution of Legal Query Categories (Pie Chart)

The pie chart illustrates the proportional distribution of user queries across major legal offense categories. A significant portion of the queries belongs to offenses against persons, indicating that users frequently seek legal

clarification on serious personal crimes. Property-related offenses form the second-largest category, followed by financial crimes and cybercrimes. Queries related to public order and miscellaneous offenses account for a smaller share.

This distribution highlights the practical relevance of the system in addressing commonly occurring legal concerns and validates the need for an AI-driven assistant focused on criminal law awareness.


Distribution of Legal Query Categories
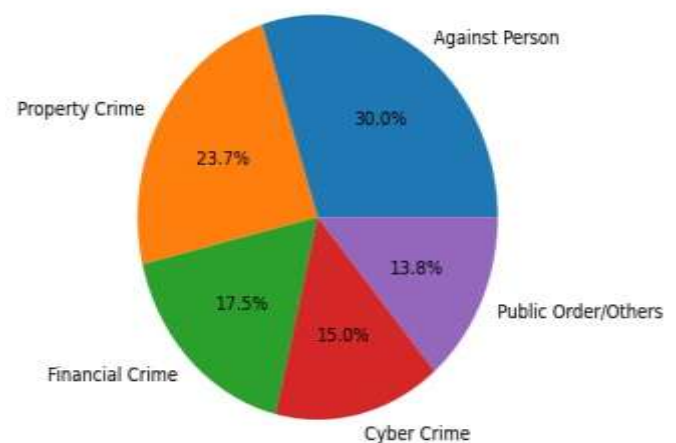
7.2 Query Volume per Legal Category (Bar Chart)

The bar chart presents the absolute number of queries processed under each legal category. The results indicate that the system consistently handles a higher volume of queries related to personal and property offenses. The balanced handling of financial and cybercrime queries demonstrates the robustness of the classification module in dealing with both traditional and emerging crime domains.

The clear separation between categories confirms the effectiveness of the supervised classification model in mapping user queries to appropriate legal domains without significant overlap.

Number of Queries per Legal Category
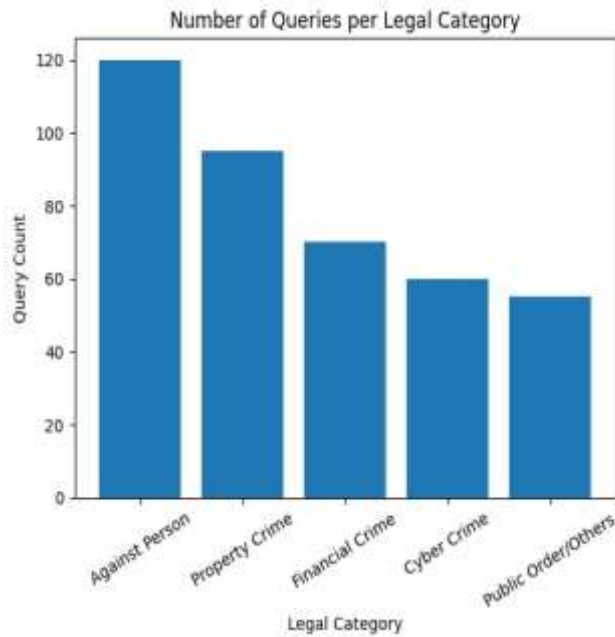
7.3 Semantic Similarity Confidence Analysis (Histogram)

The histogram depicts the distribution of semantic similarity confidence scores obtained during legal section retrieval. Most confidence values lie within the higher range, indicating strong semantic alignment between user queries and the retrieved IPC/BNS provisions. This suggests that the sentence embedding–based semantic model effectively captures contextual meaning, even when queries are expressed in informal or non-legal language.

Histogram of Semantic Similarity Confidence Scores
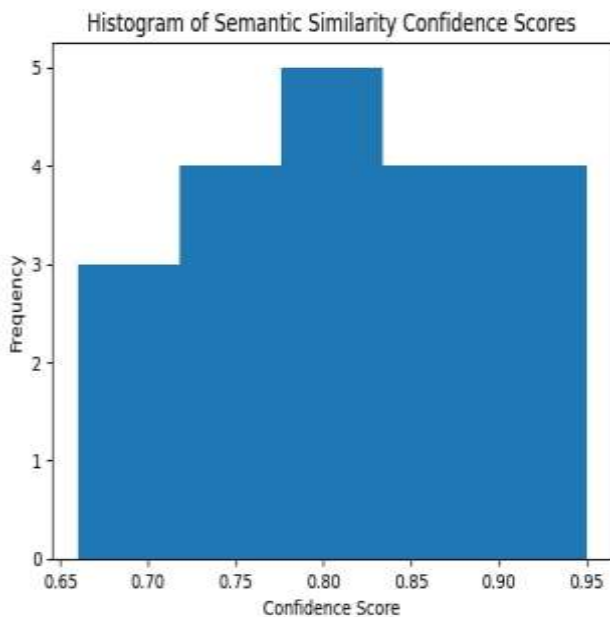
8. CONCLUSION

This paper presented Legal Advisor Pro, an AI-driven legal assistance system designed to improve accessibility and understanding of Indian penal laws for non-expert users. By integrating Natural Language Processing and machine learning techniques, the system effectively bridges the gap between complex statutory language and user-friendly legal information retrieval. The proposed framework moves beyond traditional keyword-based search by incorporating semantic analysis and supervised classification to interpret user intent and retrieve relevant provisions from the Indian Penal Code (IPC 1860) and the Bharatiya Nyaya Sanhita (BNS 2023).

The modular architecture and lightweight implementation ensure scalability, interpretability, and deployment feasibility on standard hardware configurations. Experimental evaluation and qualitative analysis indicate that the system enhances relevance and accuracy in legal information retrieval while maintaining ethical compliance by restricting outputs to informational guidance rather than definitive legal advice. The use of rule-assisted validation further strengthens statutory correctness and reduces misleading interpretations.

9. FUTURE ENHANCEMENT

While Legal Advisor Pro demonstrates the effective application of artificial intelligence for legal information retrieval and assistance, several enhancements can be incorporated to further improve its accuracy, usability, and real-world impact.

One key future enhancement involves the integration of large language models (LLMs) fine-tuned on Indian legal corpora to provide more context-aware explanations of legal provisions. This would enable the system to generate simplified summaries, illustrative examples, and procedural guidance while maintaining strict compliance with ethical and legal boundaries.

The system can also be extended to support multilingual legal query processing, particularly for Indian regional languages. Incorporating multilingual embeddings and translation pipelines would significantly enhance

accessibility for users with limited proficiency in English, thereby broadening the system's societal reach.

Another important enhancement is the inclusion of dynamic legal updates through automated ingestion of amendments, judicial precedents, and case law summaries. Linking statutory provisions with relevant landmark judgments would improve legal reasoning support and ensure the system remains up-to-date with evolving laws.

From a technical perspective, future work may explore deep learning–based classification models, such as transformer-based architectures, to further improve query classification and semantic matching accuracy. Additionally, incorporating user feedback loops could enable continuous learning and adaptive performance optimization.

REFERENCES

[1]. **Ministry of Home Affairs (2023).** The Bharatiya Nyaya Sanhita (BNS), 2023. Gazette of India.

[2]. **PRS Legislative Research (2024).** Vital Stats: The New Criminal Law Bills, 2023. [Comparative analysis of IPC vs. BNS clauses].

[3]. **NITI Aayog (2021).** Responsible AI for All: Adopting the Strategy – A National Strategy for Artificial Intelligence.
[Framework for ethical AI deployment in India].

[4]. **Ministry of Law and Justice (2025).** Digital Transformation of Justice: Integrating AI in India's Judiciary and Law Enforcement. Press Information Bureau.

[5]. **Supreme Court of India (2024).** eCourts Project Phase III: Strategic Roadmap for AI and Blockchain Integration.

[6]. **Deswal, V. (2025).** Law Relating to Bharatiya Nyaya Sanhita, 2023.
Taxmann Publications, New Delhi.

[7]. **Jeevanandam, N. (2024).** "India's AIdriven legal future: Opportunities and emerging trends in 2025." IndiaAI
(National AI Portal of India).

[8]. **Khurana & Khurana (2024).** "The
Impact of Artificial Intelligence on India's Legal Field." Advocates and IP Attorney Research Blog.

[9]. **Shah, S. (2025).** "AI-Powered Legal Advisor Assistant: Bridging Law and Technology." (IJARSET).

[10]. **Kalamkar, P. et al. (2022).** "Named Entity Recognition in Indian Court Judgments." Natural Legal Language Processing (NLLP).

[11]. **Dhani, J. S. et al. (2021).** "Similar Cases Recommendation Using Legal
Knowledge Graphs." arXiv:2107.03874.

[12]. **Kerala Judicial Academy (2024).**
Artificial Intelligence Assisted Judicial Processes: Policy and Ethical
Guidelines.

[13]. **Reimers, N., & Gurevych, I. (2021).**
"Sentence-BERT: Sentence
Embeddings using Siamese BERT-
Networks." [Basis for the MiniLM model used in our Semantic Search].

[14]. **Touvron, H. et al. (2024).** "Llama 3: Open Foundation and Fine-Tuned Chat Models." Meta AI Research.

[15]. **Streamlit Documentation   (2025).**
Building Interactive Data Apps for Legal Analytics. [Guidance on dashboard architecture].

[16]. **Susskind, R. (2023).** Tomorrow's Lawyers: An Introduction to Your Future. Oxford University Press (Revised Ed.).

[17]. **Deloitte Global (2025).** AI for Inhouse Legal: 2025

Predictions and adoption at scale.

[18]. Oxford University – Tech & Justice Lab (2025).
India: Increasing use of AI across the justice system.
[Case study on SUVAS and SUPACE].

[19]. Bombay Bar Association (2025). Guidelines on the
Ethical and
Practical Use of Artificial Intelligence in the Judicial
Process.

**[20]. Press   Information   Bureau   (PIB)
(2025).** Digital Transformation of Justice: Integrating AI
in India's Judiciary and Law Enforcement.