

# **AI VOICE ASSISTANT FOR LEGAL INFORMATION USING GENAI**

## **PHASE I REPORT**

*Submitted by*

**MURSHID AHMED S**

**210701171**

**NIKHIL P**

**210701179**

*in partial fulfilment for the award of the degree of*

**BACHELOR OF ENGINEERING IN  
COMPUTER SCIENCE AND ENGINEERING**



**RAJALAKSHMI  
ENGINEERING COLLEGE**



**RAJALAKSHMI ENGINEERING COLLEGE  
DEPARTMENT OF COMPUTER ENGINEERING  
ANNA UNIVERSITY, CHENNAI  
NOVEMBER 2024**

# **ANNA UNIVERSITY, CHENNAI**

## **BONAFIDE CERTIFICATE**

Certified that this project titled “**AI VOICE ASSISTANT FOR LEGAL INFORMATION USING GENAI**” is the bonafide work of “**MURSHID AHMED S (210701171), NIKHIL P (210701179)**” who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported here in does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

### **SIGNATURE**

Dr. P. Kumar, M.E., Ph.D.,

### **HEAD OF DEPARTMENT**

Professor

Department of Computer Science and  
Engineering

Rajalakshmi Engineering College  
Thandalam, Chennai – 602 105

### **SIGNATURE**

Ms. S. Nagomiya, M.E.,

### **SUPERVISOR**

Assistant Professor

Department of Computer Science and  
Engineering

Rajalakshmi Engineering College  
Thandalam, Chennai – 602 105

Submitted to Project viva-voice Examination held on \_\_\_\_\_.

**Internal Examiner**

**External Examiner**

## TABLE OF CONTENTS

<b>CHAPTER NO</b>	<b>TITLE</b>	<b>PAGE NO</b>
	ACKNOWLEDGEMENT	v
	ABSTRACT	vi
	LIST OF FIGURES	vii
	LIST OF ABBREVIATIONS	viii
1.	INTRODUCTION	1
	1.1 GENERAL	1
	1.2 OBJECTIVE	2
	1.3 EXISTING SYSTEM	3
	1.4 PROPOSED SYSTEM	5
2.	LITERATURE SURVEY	10
3.	SYSTEM DESIGN	16
	3.1 GENERAL	16
	3.1.1 SYSTEM FLOW DIAGRAM	16
	3.1.2 SEQUENCE DIAGRAM	17
	3.1.3 CLASS DIAGRAM	18
	3.1.4 USECASE DIAGRAM	18
	3.1.5 ARCHITECTURE DIAGRAM	19
	3.1.6 ACTIVITY DIAGRAM	20
	3.1.7 COMPONENT DIAGRAM	21
	3.1.8 COLLABORATION DIAGRAM	21
4.	PROJECT DESCRIPTION	22
	4.1 METHODOLOGIES	22
	4.1.1 RESULT DISCUSSION	27

5.	CONCLUSION AND WORK SCHEDULE	33
	FOR PHASE II	34
	REFERENCES	36
	APPENDIX	38

## ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S.MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.**, and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN, Ph.D.**, for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution. Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.**, our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P.KUMAR, M.E., Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Ms. S.Nagomiya, M.E.,** Department of Computer Science and Engineering. Rajalakshmi Engineering College for his valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinator, **Dr. Kumaragurubaran T ,Ph.D** Department of Computer Science and Engineering for his useful tips during our review to build our project.

MURSHID AHMED S 210701171

NIKHIL P 210701179

## **ABSTRACT**

This task affords an AI-driven voice assistant that provides legal facts using Generative AI and natural Language Processing (NLP) technology, specially focused on Indian laws. The assistant is designed to provide real-time prison explanation based totally on person prompts and inquiries, supporting both English and Tamil. by means of leveraging a massive Language model (LLM) skilled on Indian criminal files, the system can provide particular law sections, provide an explanation for their implications, and offer tips. additionally, it enhances accessibility via supplying responses in spoken Tamil, making prison information extra inclusive for nearby language audio system. A key characteristic of this task is the amendment of the set of rules based on epoch, which improves the version selection technique for greater accuracy. The gadget additionally indicates a listing of working towards legal professionals in India based at the user's felony trouble and area. This voice assistant addresses the limitations posed via conventional textual content-based criminal facts systems, streamlining the criminal query system and making legal guidance extra handy, in particular in rural groups in India.

**LIST OF FIGURES**

<b>FIGURE NO</b>	<b>TITLE</b>	<b>PAGE NO</b>
1	PROPOSE SYSTEM	9
2	SYSTEM FLOW	16
3	SEQUENCE DIAGRAM	17
4	USECASE DIAGRAM	18
5	CLASS DIAGRAM	18
6	ARCHETECTURE DIAGRAM	19
7	ACTIVITY DIAGRAM	20
8	COMPONENT DIAGRAM	21
9	COLLABORATION DIAGRAM	21

## LIST OF ABBREVIATIONS

<b>SNO</b>	<b>ABBREVIATION</b>	<b>EXPANSION</b>
1	AI	Artificial Intelligence
2	API	Application Programming Interface
3	AR6	Sixth Assessment Report (IPCC)
4	CNN	Convolutional Neural Network
5	GPT	Generative Pre-trained Transformer
6	IMCC	International Marine Conservation
		Congress
7	LSTM	Long-Short Term Memory
8	LLM	Large Language Model
9	NLP	Natural Language Processing
10	NASA	National Aeronautics and Space
		Administration
11	NER	Named Entity Recognition
12	OPT	Open Pre-trained Transformer
13	RAG	Retrieval-Augmented Generation
14	TTS	Text-to-Speech
15	STT	Speech-to-Text
16	DERA	Dialog-Enabled Resolving Agents

## CHAPTER 1

### 1. INTRODUCTION

#### 1.1 GENERAL

Navigating legal structures may be a frightening mission for individuals lacking formal legal training, specially in inter national locations with complex and multifaceted criminal frameworks like India. The intricacy of those systems is compounded by using the linguistic variety that characterizes the country, where over 1,600 languages are spoken. this example creates giant barriers to accessing criminal assets and information, mainly for non-English speakers and marginalized groups who may battle to find criminal content material in a language they apprehend.

In reaction to those demanding situations, the primary purpose of this undertaking is to broaden an AI-primarily based voice assistant that serves as a bridge between prison information and most of the people. The assistant is designed to offer customers with legal records in a conversational manner, making it more relatable and easier to realise. This method not most effective complements consumer engagement however additionally promotes felony literacy by using demystifying prison jargon and processes.

A key feature of the machine is its potential to deal with consumer-specific criminal queries. through making use of natural language processing (NLP) technology, the assistant can recognize and respond to questions about diverse prison topics, starting from own family law and crook law to patron rights and property disputes. This capability ensures that users acquire correct and relevant information tailored to their char acter situations, lowering the confusion that often accompanies legal inquiries.

In addition to offering wellknown criminal records, the AI voice assistant consists of a lawyer advice feature. through assessing the consumer's legal difficulty and geographic area, the machine can propose working towards legal

professionals who specialize inside the relevant vicinity of law. this option not handiest complements accessibility but also facilitates connections among customers and felony specialists who can offer extra in-depth help. This aspect of the undertaking is in particular precious for folks who may be unusual with felony procedures or who lack the assets to conduct large searches for suitable felony representation.

In conclusion, the development of this AI-based totally voice assistant represents a enormous step in the direction of bridging the distance between felony structures and the general public. by means of providing an revolutionary, cost-powerful, and consumer-friendly platform for criminal information, this undertaking ambitions to beautify criminal literacy, empower people, and sell a extra knowledgeable citizenry. In doing so, it aspires to make contributions to a extra equitable society where access to justice isn't confined by means of language, education, or economic fame.

## **1.2 OBJECTIVE**

The objective of this project is to develop an intelligent and interactive chatbot capable of providing reliable legal information, insights, and recommendations tailored to user needs. The chatbot aims to:

### **1. Integrate Legal Knowledge Frameworks:**

Leverage domain-specific legal datasets and natural language processing models to address queries related to laws, regulations, and case precedents effectively. The chatbot will ensure accurate and relevant responses by integrating authoritative legal resources and multilingual support.

### **2. Enable Real-Time Legal Support:**

Facilitate access to current legal statutes and historical case information through APIs and comprehensive legal databases, ensuring users can make informed decisions on legal matters.

### **3. Enhance User Engagement:**

Provide a conversational interface that supports text and voice interactions in both English and Tamil, allowing users to query legal information dynamically and receive personalized responses suited to their language preferences.

### **4. Support Advanced Query Analysis:**

Enable analysis of legal queries with contextual understanding and structured visual outputs, such as flowcharts and timelines, to help users comprehend complex legal scenarios.

### **5. Promote Legal Awareness:**

Educate users about legal rights, responsibilities, and procedures through visual tools such as decision trees and graphical explanations, encouraging informed and responsible legal actions.

## **1.3 EXISTING SYSTEM**

The LegalBot system addresses critical challenges in disseminating accurate legal information in a conversational format. While large language models (LLMs) like GPT-4 demonstrate impressive generative capabilities, they often struggle to provide legally precise and up-to-date information, particularly in specialized domains like law. Recognizing this gap, Legal Bot is designed to integrate validated legal resources, such as the Indian Penal Code, landmark judgments, and regional laws, ensuring that its responses align with authoritative legal sources.

LegalBot combines the natural language understanding (NLU) and generative capabilities of GPT-4 with a sophisticated retrieval mechanism that ensures its responses are grounded in legal documents and case laws. The system indexes and structures legal texts, allowing for efficient retrieval of relevant sections during conversations. With retrieval-augmented generation (RAG), the chatbot retrieves pertinent legal data and incorporates it into its generative process, ensuring that

responses are accurate, contextual, and citation-backed. This dual-layered architecture enhances the credibility of the chatbot and empowers users—including lawyers, paralegals, and laypersons—to rely on its outputs as reliable representations of legal knowledge.

To uphold the accuracy and ethical responsibility of its outputs, LegalBot includes an expert validation mechanism. Legal professionals periodically evaluate the system's performance and provide feedback to refine its retrieval algorithms and prompts. This iterative feedback loop ensures alignment with the latest legal developments and nuanced interpretations, minimizing the risk of misinformation. Furthermore, the chatbot transparently cites its sources, enabling users to cross-verify information independently and fostering trust in the system.

## **Challenges in Current AI Systems**

### **Lack of Domain Expertise:**

Generic LLMs often lack the ability to understand and interpret intricate legal jargon, statutes, and procedural nuances. This gap can lead to oversimplified or even inaccurate responses that are unsuitable for legal decision-making.

### **Dynamic Nature of Legal Frameworks:**

Laws and regulations frequently undergo updates and amendments, making it challenging for static or generalized AI models to stay relevant without regular retraining.

### **Misinformation and Hallucination:**

A critical drawback of current AI systems is their propensity to generate "hallucinated" responses—fabricated or incorrect answers. In the legal domain, where accuracy is paramount, such errors could lead to serious consequences.

### **Language and Accessibility Barriers:**

Most existing systems lack the ability to cater to multilingual audiences effectively, excluding a significant portion of the population, especially those who prefer interacting in regional languages like Tamil.

## **1.4 PROPOSED SYSTEM**

The proposed system, LegalBot, is designed to revolutionize access to legal information by leveraging advanced AI technologies to deliver accurate, efficient, and user-friendly interactions. Recognizing the limitations of existing systems, LegalBot focuses on overcoming challenges such as the dynamic nature of legal frameworks, multilingual accessibility, and the need for precise, actionable legal insights. By integrating cutting-edge AI architectures, retrieval-augmented generation, and multimodal interaction capabilities, the proposed system aims to empower users from diverse backgrounds to navigate legal complexities confidently and effectively.

### **Key Features of the Proposed System**

#### **Domain-Specific Intelligence:**

LegalBot is specifically trained and designed to handle legal queries with a high degree of accuracy. By integrating legal databases and documents, such as statutes, case law, and procedural guidelines, the chatbot provides contextually relevant and reliable responses tailored to users' needs. Its domain-specific training ensures that the system is well-equipped to handle nuanced legal questions and scenarios.

#### **Dynamic Data Retrieval:**

To address the constantly evolving nature of legal frameworks, LegalBot employs retrieval-augmented generation (RAG). This approach allows the system to fetch the most relevant and updated legal information in real-time from indexed legal resources. Unlike static AI models, LegalBot dynamically integrates retrieved data into its responses, ensuring users receive accurate and current legal guidance.

**Multilingual and Multimodal Interaction:**

The proposed system is built with inclusivity in mind, offering support for multiple languages, including English and Tamil, in both text and voice formats. This feature makes LegalBot accessible to a broader audience, especially in regions where regional languages are predominant. Voice interaction capabilities enhance user experience by catering to individuals with limited literacy or those who prefer conversational engagement.

**Simplified Legal Guidance:**

LegalBot prioritizes user comprehension by translating complex legal jargon into clear and accessible language. It provides step-by-step guidance for procedures, such as filing legal documents or understanding rights under specific laws, using plain language, visual aids, and decision trees. This simplification ensures that even non-expert users can understand and act upon the advice provided.

**Expert Validation and Ethical AI Principles:**

The proposed system incorporates periodic validation by legal experts to ensure that its responses are legally accurate, ethically sound, and aligned with current practices. This feedback loop not only refines the chatbot's performance but also upholds its responsibility to provide impartial and responsible guidance.

**Scalable and Modular Design:**

LegalBot's architecture is modular, allowing for seamless integration with additional legal databases, predictive tools, or multilingual capabilities as the system evolves. This scalability ensures that the chatbot can adapt to the increasing complexity and diversity of legal queries over time.

## **Advantages of the Proposed System**

### **Improved Accessibility:**

By supporting multilingual and voice interactions, LegalBot ensures that legal information is accessible to individuals from diverse linguistic and socio-economic backgrounds, reducing barriers to justice.

### **Real-Time Updates:**

The dynamic retrieval mechanism ensures that users always have access to the latest legal information, mitigating the risks of outdated or inaccurate advice.

### **Transparency and Trust:**

LegalBot emphasizes citation-backed responses, providing users with references to the legal sources it relies upon. This transparency fosters trust and enables users to verify the information independently.

### **User Empowerment:**

By simplifying legal concepts and offering step-by-step guidance, LegalBot empowers users to handle legal issues confidently, whether they are drafting contracts, resolving disputes, or understanding their rights.

### **Cost and Time Efficiency:**

For legal practitioners and businesses, LegalBot streamlines research and compliance processes, reducing the need for extensive manual work or consultations. For the general public, it offers a free or low-cost alternative to seeking basic legal advice.

## **Real-World Applications of the Proposed System**

### **Legal Assistance for Individuals:**

LegalBot assists individuals in understanding their rights, filing complaints, and navigating legal processes. It serves as a first point of contact for legal inquiries, reducing dependence on legal professionals for basic guidance.

**Support for Legal Practitioners:**

The chatbot provides lawyers and paralegals with quick access to statutes, case law, and procedural information, helping them save time and improve efficiency in case preparation.

**Corporate Legal Compliance:**

LegalBot aids businesses in understanding regulatory requirements, contract clauses, and labor laws, ensuring compliance with minimal effort.

**Educational Tool:**

LegalBot can be used in educational settings to teach students about legal principles, procedures, and rights, fostering greater legal literacy among the public.



*Figure 1*

**Figure 1** illustrates the LegalBot chatbot's user interface, designed for simplicity and ease of use. Built with React, the UI supports both text and voice inputs, providing real-time responses. It features multilingual capabilities, allowing interactions in English and Tamil, ensuring accessibility for diverse users. The clean layout and responsive design enhance usability, serving as the bridge between users and the chatbot's AI-powered legal assistance system.

## CHAPTER 2

### 2. LITERATURE SURVEY

1. Get right of entry to to Justice for Marginalized groups Aditya Prakash et al. (2022) tested obstacles to legal access con fronted by means of marginalized communities, highlighting socioeconomic, cultural, and structural factors that perpetuate disparities. Their findings emphasize the want for generation pushed solutions, in particular the ones leveraging AI, to lessen these barriers and promote equitable get right of entry to to criminal assets.
2. AI packages in prison Contexts Kuchina (2024) analyzed regu latory processes in AI-driven criminal structures, underscoring the need for harmonized frameworks throughout jurisdictions. This work gives insights into the demanding situations of enforcing AI in prison settings, specially whilst addressing cross-border criminal frameworks and compliance necessities.
3. NLP in self sustaining systems and decision-Making although normally focused on autonomous cars, studies with the aid of Bernhard Jaeger et al. (2023), Dian Chen et al. (2019), Hao Shao et al. (2022), and Hao Shi et al. (2023) sheds mild on the superior programs of NLP and huge language fashions (LLMs) in interpreting human instructions, making complex choices, and improving multimodal facts integration. these studies provide precious parallels for AI applications in prison contexts, especially in terms of improving human-pc interplay and actual-time decision-making.
4. Transformer-primarily based models for prison assistance Migliorini (2024) added transformer-based totally language fashions tailor-made for prison chatbots, which help in im parting automated legal steering. notwithstanding their performance, those models face limitations in addressing complex felony inquiries that fall outdoor their schooling records. This studies underscores the want for extra adaptable AI fashions in legal assistance tools.

- 5. Knowledge-Driven Approaches in Autonomous Driving** Harith Farhad et al. (2022) introduced a knowledge-driven approach to autonomous driving that leverages LLMs and situational data for enhanced decision-making. While this research focuses on driving, the approach of incorporating real-world knowledge can be adapted to legal AI, potentially improving AI's responsiveness to complex legal scenarios.
- 6. Legal Information Chatbots for Marginalized Communities** Renz et al. (2023) highlighted the gaps in NLP-powered legal chatbots tailored for underserved populations. Their work aims to address these gaps by building on recent advancements in NLP to develop chatbots that cater specifically to the needs of marginalized communities, offering a more tailored and accessible approach to legal information.
- 7. Deep Learning for Case Prediction** Alam et al. (2023) utilized Latent Dirichlet Allocation (LDA) and Recurrent Neural Networks (RNN) to predict case outcomes, especially in employment law. Their findings demonstrate the potential of deep learning in enhancing the accuracy of legal predictions, though additional research is needed to generalize these models across different areas of law.
- 8. NLP and AI for Legal Information and Services** Hesham M. Eraqi et al. (2022) explored the role of AI and NLP in enhancing access to justice by providing legal information through innovative solutions. Their research emphasizes technology's potential in making legal knowledge accessible to non-experts, aligning closely with the objectives of our project.
- 9. Human-Like Interaction with Legal AI Systems** Chitta et al. (2023) explored human-like interactions with autonomous vehicles through LLMs. By enabling natural language interactions, their approach suggests pathways for making legal AI systems more user-friendly, allowing users to communicate with the system in an intuitive, conversational manner.

**10.** Synthesizing Insights for felony AI Xu et al. (2022) emphasized synthesizing insights from numerous studies to design effective AI answers in prison contexts. Their findings advocate for a design technique that integrates expertise from present literature, making sure that AI solutions are complete and meet the desires of marginalized groups. These surveys are truly studied give my knowledge and also linked with my projects.

**11.** The dataset *Climate Fever: A Dataset for Verification of Real-World Climate Crisis Claims* by Digglemann et al. provides a benchmark for automated fact-checking systems. The authors curate a collection of climate-related claims, annotated with evidence and counter-evidence from credible sources. This resource facilitates the development of AI models capable of verifying the authenticity of climate crisis statements. The study underscores the importance of combating misinformation in the climate discourse, demonstrating how AI can enhance the reliability of public communication and decision-making.

**12.** *The Choice of Textual Knowledge Base in Automated Claim Checking* by Stammbach et al. investigates the impact of knowledge base selection on the performance of automated fact-checking systems. The authors analyze several textual datasets, highlighting trade-offs between coverage, specificity, and quality. Their findings suggest that aligning the knowledge base with the domain of claims significantly improves accuracy. This research provides practical guidelines for developing robust fact-checking AI and underscores the importance of curating reliable data sources for informed AI systems.

**13.** *Improving Language Models by Retrieving from Trillions of Tokens* by Borgeaud et al. explores the integration of retrieval mechanisms into LLMs to enhance their generative capabilities. The study demonstrates how accessing external datasets during inference allows the model to generate more accurate and context-aware responses. This retrieval-augmented approach reduces the reliance on extensive pre-training, offering a scalable and efficient alternative for building

intelligent systems. The paper highlights the transformative potential of retrieval-based techniques in advancing the capabilities of LLMs.

**14.** The paper *Generative Artificial Intelligence: Opportunities and Challenges of Large Language Models* delves into the dual aspects of LLM development. It discusses the revolutionary applications of generative AI in areas like content creation, education, and healthcare while addressing ethical dilemmas such as bias, misuse, and environmental impact. The authors propose a framework for responsible innovation, emphasizing transparency, fairness, and inclusivity. This research provides a balanced perspective on leveraging the potential of LLMs while mitigating associated risks.

**15.** The study on *Machine Learning-Based Evidence and Attribution Mapping of 100,000 Climate Impact Studies* presents an innovative approach to analyzing vast amounts of climate-related research. The authors employ machine learning techniques to categorize and attribute climate impacts across various domains, creating a comprehensive evidence map. This resource aids researchers and policymakers in understanding the scope and severity of climate change effects, enhancing evidence-based decision-making. The paper highlights the scalability of AI in synthesizing and interpreting large-scale datasets.

**16.** Kumar et al.'s paper *A Novel Approach for Text Generation Using RNN for Language Modeling* introduces an advanced recurrent neural network (RNN) architecture tailored for text generation. The authors emphasize the model's ability to capture sequential dependencies and produce coherent text outputs. Through extensive experiments, they demonstrate the efficacy of their approach in various generative tasks, showcasing its potential in applications like automated storytelling and chatbot development. The research contributes to the ongoing evolution of generative AI methodologies.

**17.** In their follow-up work, Kumar et al. present *AI-Driven Text Generation: A Novel GPT-Based Approach for Automated Content Creation*, which explores the advancements of GPT models in generating high-quality text. The study focuses

on optimizing the GPT architecture for specific domains, achieving improved fluency and contextual relevance. The authors highlight applications in content creation, marketing, and personalized communication, demonstrating the versatility of GPT-based systems. This research further underscores the transformative impact of generative AI in diverse fields supporting complex requests by disassembling them into sub-tasks assigned to specialized models.

**18. VORADASOCATIYANURAK 1,2, (Student Member, IEEE), NITTAYAPA KLANGPORNKUN 1,3, ADIREK MUNTHULI 1,3, PHONGPHAN PHIENPHANICH 1,3, LALIN KOVUDHIKULRUNGSRI 4, NANTAWAT SAKSAKULKUNAKORN4, PHONKANOKCHAIRAUANGSRI4, and CHARTURONGTANTIBUNDHIT**'s The integration of artificial intelligence (AI) in legal support systems has gained significant traction, with a focus on addressing societal issues such as sexual violence. Previous works in this domain, such as the LAW-U chatbot, have demonstrated the potential of AI to provide tailored legal guidance to survivors of sexual violence. LAW-U utilizes Natural Language Processing (NLP) techniques trained on Supreme Court decisions to deliver accurate, context-specific legal advice. Chatbots like "Sophia" from Kona Club and "Hello Cass" in Australia also highlight advancements in supporting victims of domestic and sexual violence, offering discreet and accessible support. However, limitations persist, such as language restrictions, gender bias, and the lack of localized legal guidance.

The LAW-U chatbot addresses these gaps by employing a gender-neutral design and focusing on legal precision, though it remains limited to the Thai language. Building on such frameworks, this study explores the development of a bilingual AI chatbot tailored for legal consultations in Tamil and English, expanding accessibility and inclusivity. The bilingual capability aims to reduce barriers in understanding legal rights and procedures, empowering users to navigate complex legal landscapes with confidence.

**19.** The application of artificial intelligence (AI) in legal domains has advanced through the integration of deep learning models and natural language processing techniques to predict legal outcomes. Alam et al. proposed an AI-based automated chatbot for predicting the outcomes of employment-related legal cases using deep learning methods such as Latent Dirichlet Allocation (LDA) combined with Recurrent Neural Networks (RNN). This system leverages semantic analysis to extract patterns from historical legal documents, demonstrating high accuracy in outcome prediction.

The study highlights the importance of feature selection, semantic clustering, and data preprocessing to ensure precise predictions. By employing variations in neural networks, such as Gated Recurrent Units (GRU) and using SoftSign activation functions, the proposed models achieved superior performance, with cross-validation showing high precision and recall metrics. This novel approach bridges gaps in legal analytics, providing a scalable solution for preliminary legal assessments.

**20.** The high cost of legal services and the lack of representation for many litigants underscore the need for accessible legal information. Research initiatives like the LegalIA project have aimed to address this by developing AI-powered chatbots. These systems, including chatbots for immigration and corporate legal issues, utilize advanced NLP techniques to bridge the gap in access to justice.

The immigration chatbot, designed to address questions about Canadian immigration processes, is based on a knowledge retrieval framework. It uses publicly available FAQs from government sources, employing StarSpace and other IR-based algorithms to classify user queries and match them with appropriate answers. Similarly, the corporate legal chatbot serves employees of the National Bank of Canada by answering legal queries derived from FAQs curated by legal experts.

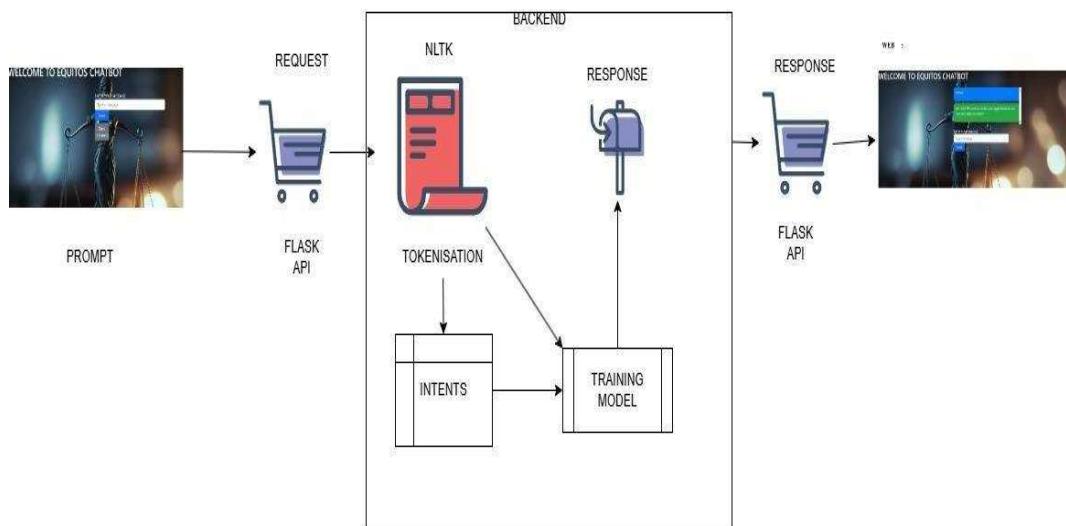
Both systems demonstrate the effectiveness of modern AI techniques, including intent recognition, to address specific legal needs. However, challenges such as data scarcity, the complexity of legal texts, and the need for fine-tuning large pre-trained language models like BERT persist. These limitations highlight opportunities for further refinement in chatbots to make legal assistance both comprehensive and inclusive.

## CHAPTER 3

### 3. SYSTEM DESIGN

#### 3.1 GENERAL

##### 3.1.1 SYSTEM FLOW DIAGRAM



*Figure 2*

### 3.1.2 SEQUENCE DIAGRAM

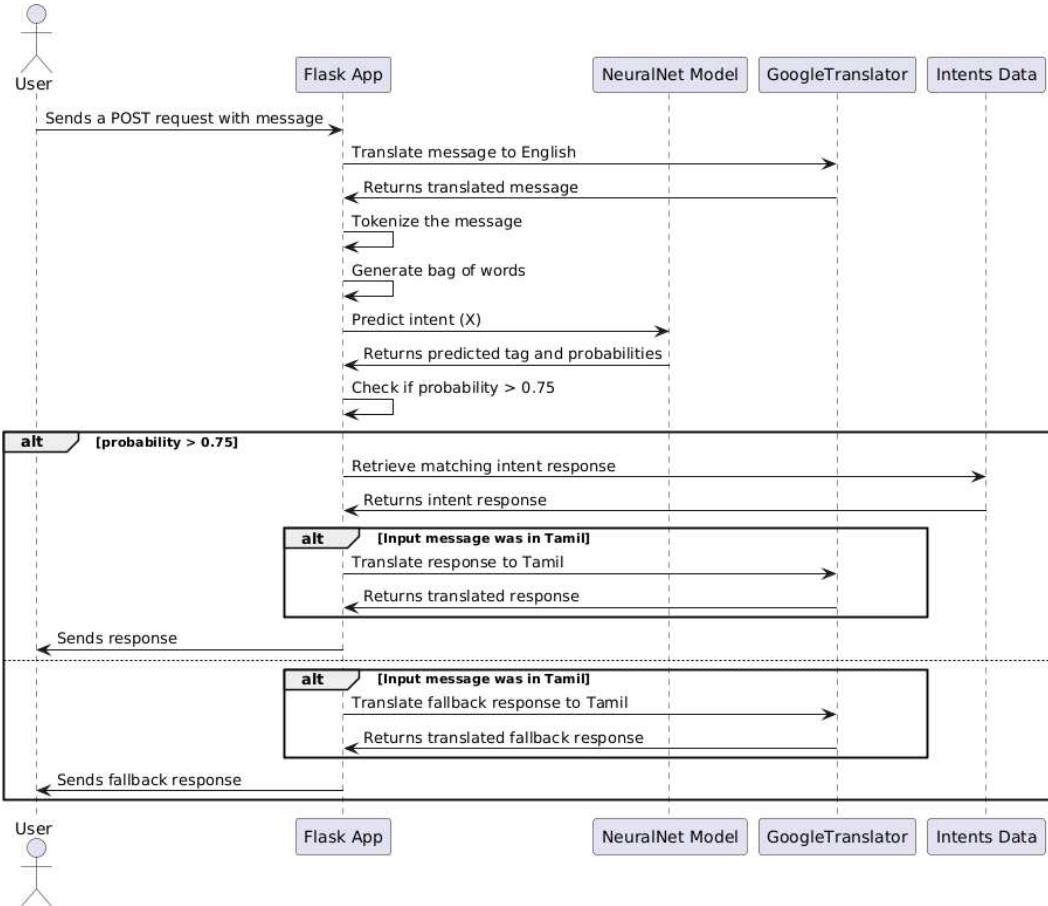


Figure 3

### 3.1.3 CLASS DIAGRAM

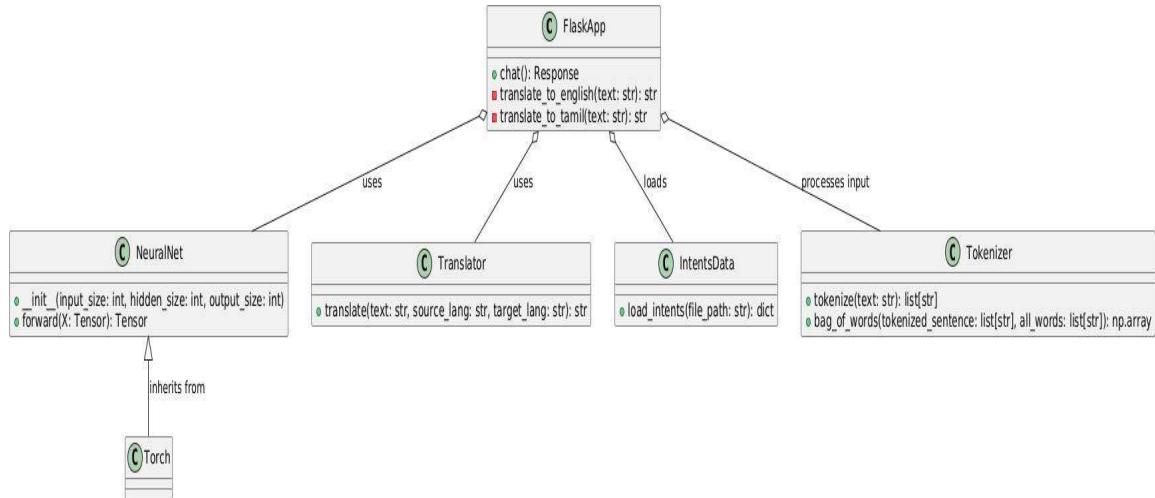


Figure 4

### 3.1.4 USE CASE DIAGRAM

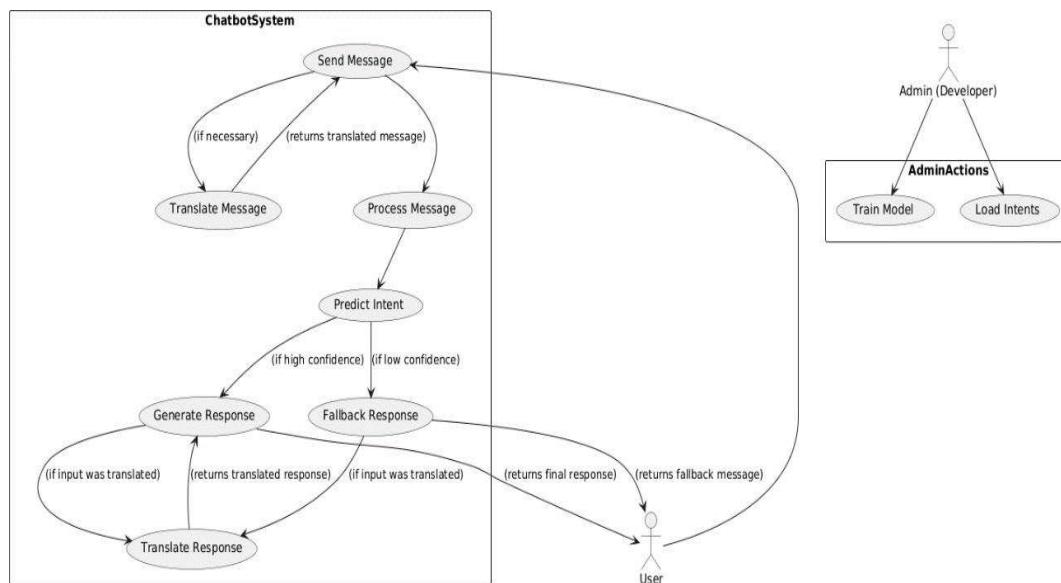


Figure 5

### 3.1.5 ARCHIETECTURE DIAGRAM

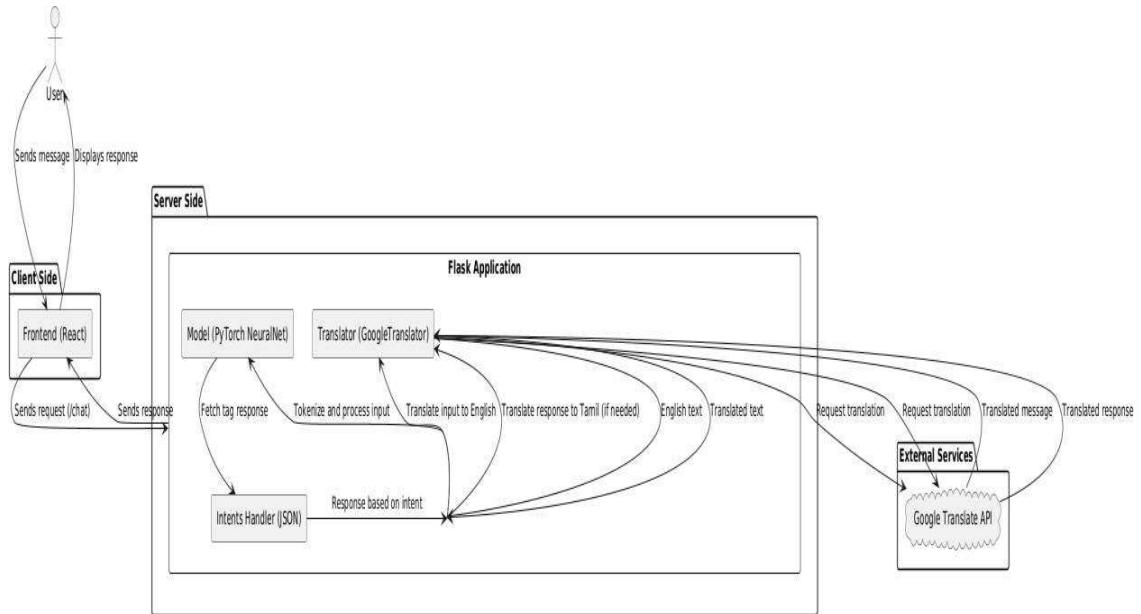


Figure 6

### 3.1.5 ACTIVITY DIAGRAM

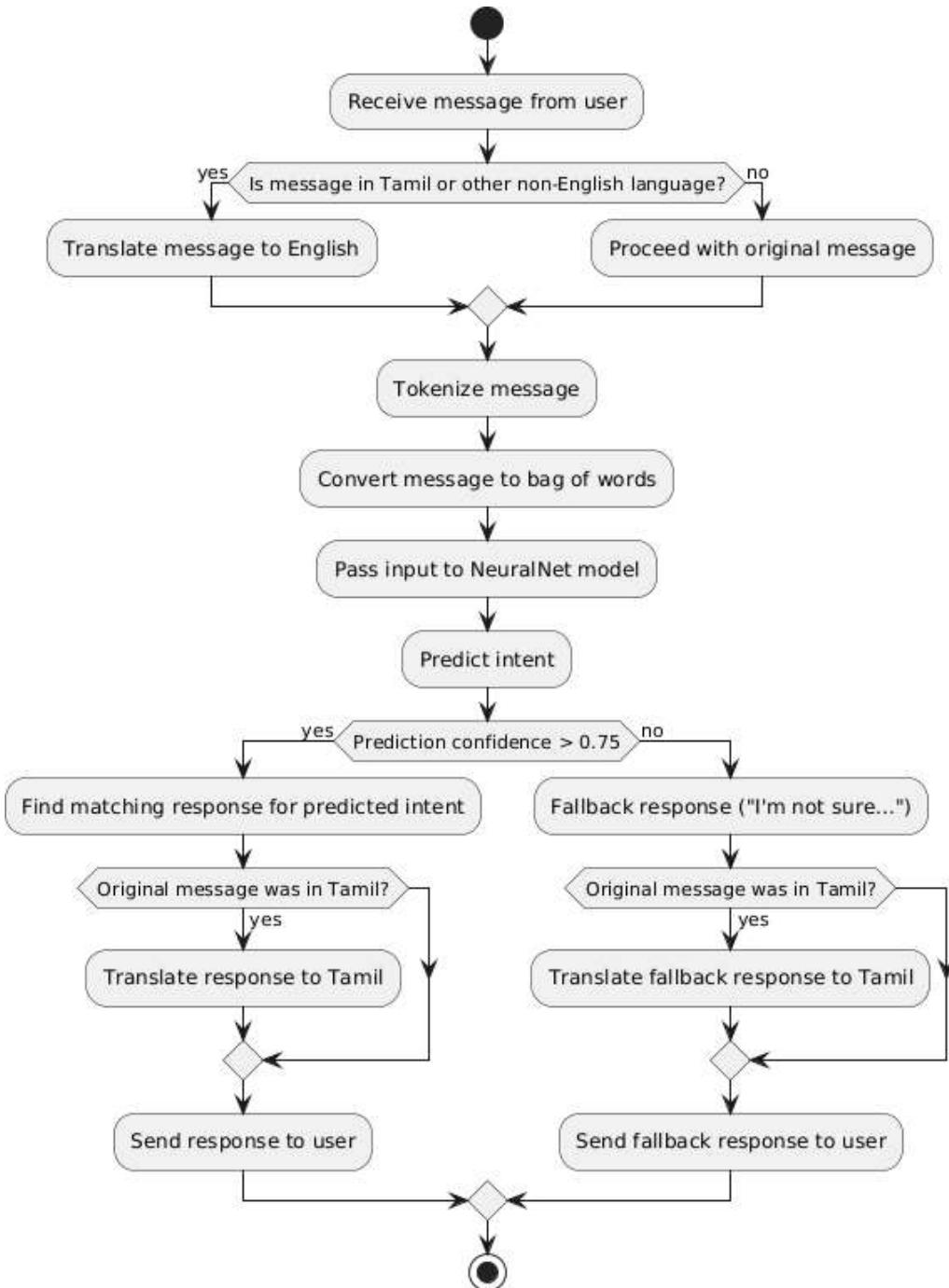


Figure 7

### 3.1.6 COMPONENT DIAGRAM

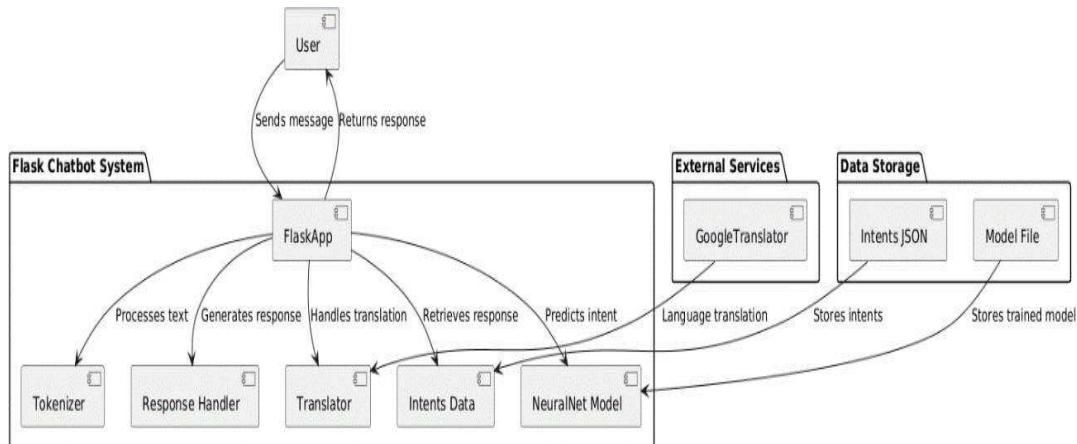


Figure 8

### 3.1.7 COLLABORATION DIAGRAM

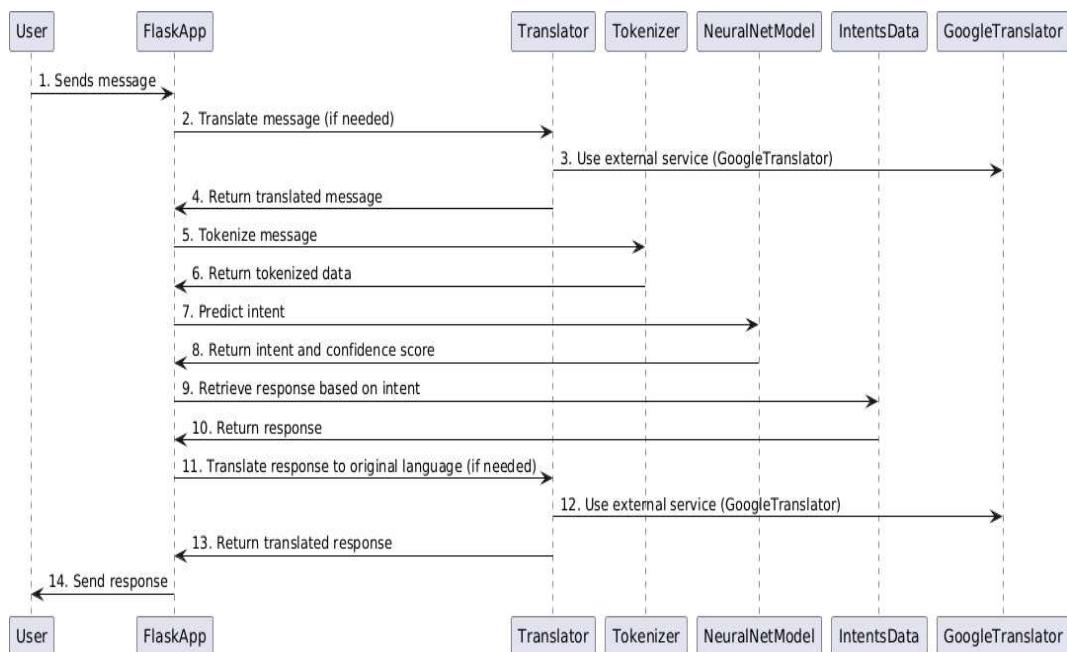


Figure 9

## CHAPTER 4

### PROJECT DESCRIPTION

#### 4.1 METHODOLOGIES:

##### Project Overview:

The AI-assisted lawyer chatbot is a cutting-edge digital solution designed to provide users with accessible and reliable legal information. Targeted at simplifying interactions in the legal domain, the chatbot offers legal guidance on various topics, such as contracts, employment law, family law, property rights, and more. It is tailored to serve a diverse audience by supporting multiple languages, including **English** and **Tamil**, and integrates both **text-based** and **voice-based** interactions to improve user experience and accessibility.

The chatbot leverages natural language processing (NLP) and machine learning to understand user queries, predict the relevant legal intent, and offer appropriate responses. By incorporating AI-based language translation services and voice capabilities, the chatbot can cater to users with varying linguistic preferences and disabilities.

##### Key Features:

###### 1. Multilingual Support:

- The chatbot supports **English** and **Tamil**, with plans to expand to other regional languages.
- Automatic detection and translation of user queries into English for internal processing.
- Responses are translated back into the original language, ensuring seamless communication.

###### 2. Voice Assistant Integration:

- Users can interact with the chatbot via voice input, making it accessible for those who prefer verbal communication.

- The chatbot can provide responses in both **text** and **voice**, making it easier for users with reading disabilities or those who prefer audio output.

### 3. Legal Information Coverage:

- The chatbot is programmed with knowledge of various legal domains, such as:
  - **Contract law**
  - **Employment law**
  - **Family law** (e.g., divorce, child custody)
  - **Property law**
  - **Consumer rights**
- It is designed to handle common legal queries and provide relevant, general information.
- While the chatbot does not offer legal advice, it helps users understand basic legal concepts and points them toward possible courses of action.

### 4. AI-Driven Intent Prediction:

- The chatbot uses a pre-trained **Neural Network model** to understand user intent.
- NLP techniques such as **tokenization** and **bag of words** are applied to process user input, and the model predicts the legal category or topic.
- If the model's confidence in the predicted intent is above a certain threshold (e.g., 75%), the appropriate legal response is provided.

### 5. Fallback Mechanism:

- If the chatbot does not understand the user's query or the confidence in its response is low, a fallback message such as "I'm not sure how to respond to that" is provided.
- The fallback message is also translated into the user's original language if needed.

## 6. User Interface:

- The chatbot is accessible through a **web interface** (REST API) where users can interact via typing or voice input.
- It is designed with a simple, user-friendly interface to accommodate users of all technological skill levels.

## Technology Stack:

### 1. Backend:

- **Flask**: The core framework for handling HTTP requests and API responses.
- **PyTorch**: The machine learning framework used for building and running the neural network model that predicts user intents.
- **Natural Language Processing**:
  - **Tokenization** and **bag of words** models to preprocess user input.
- **Deep Translator API**: Used for real-time language translation (English to Tamil and vice versa).

### 2. Voice Integration:

- **Google Text-to-Speech (TTS)**: For converting text responses into voice output.
- **Speech-to-Text (STT)**: For converting user voice input into text that can be processed by the chatbot.

### 3. Data and Model:

- **Intents JSON file**: Stores a set of predefined intents (legal topics) and their corresponding responses.
- **Pre-trained neural network model**: Used for intent classification and is stored in a serialized format (e.g., PyTorch model file).

#### 4. External Services:

- **Google Translator API:** Provides language translation functionality, enabling the chatbot to handle multilingual communication.
- **Google Text-to-Speech API and Speech-to-Text API** for voice interactions.

#### 5. User Interface (Frontend):

- Accessible through a web browser or mobile app, with a simple form for entering queries and a microphone option for voice input.

### **Workflow:**

#### 1. User Interaction:

- A user enters a legal query or question through text or speaks the query via the voice interface.

#### 2. Language Translation (if needed):

- If the query is in Tamil (or another non-English language), the input is automatically translated into English using Google Translator.

#### 3. NLP Preprocessing:

- The input text is tokenized, and a bag of words is generated, which is then passed into the neural network for intent prediction.

#### 4. Intent Prediction:

- The neural network model processes the input and predicts the legal intent or category based on predefined legal intents.

#### 5. Response Generation:

- If the intent confidence is above the threshold, a legal response is fetched from the intents data.
- If the confidence is low, a fallback response is generated.

#### 6. Translation of Response:

- If the user's original query was in Tamil, the response is translated back into Tamil before being sent back to the user.

○

#### **7. Voice Response (if requested):**

- The response is optionally converted into a voice output using Text-to-Speech.

#### **8. Final Output:**

- The chatbot delivers the response to the user either in text or voice format, depending on their preference.

#### **Use Cases:**

1. **Legal Information for Laypersons:** The chatbot provides legal guidance to people unfamiliar with legal jargon, helping them understand their rights and obligations.
2. **Language Accessibility:** Non-English speaking users can access legal information in their native language (e.g., Tamil), bridging the communication gap in legal services.
3. **Voice Accessibility:** Users who prefer verbal communication or have reading impairments can interact with the chatbot using voice input and receive spoken responses.

#### **Benefits:**

1. **Accessibility:** The system removes language and communication barriers by supporting multiple languages and voice interactions, making legal information accessible to a broader audience.
2. **Efficiency:** Users can quickly obtain legal information without the need for human intervention, which saves time for both users and legal professionals.
3. **Cost-Effective:** The chatbot provides a free or low-cost alternative to consulting with lawyers for basic legal information, making legal knowledge more affordable.

4. **Scalability:** The chatbot can be easily updated with new legal domains, languages, or regional nuances, making it adaptable to various jurisdictions.

### **Future Enhancements:**

1. **Addition of More Languages:** Expanding support to include more regional languages and dialects.
2. **Integration with Legal Databases:** Linking the chatbot with legal databases or court rulings to provide more detailed information.
3. **Contextual Conversations:** Implementing context tracking to handle multi-turn conversations and allow for more complex queries.

**Enhanced Voice Recognition:** Improving the voice recognition system to handle more dialects and accents effectively.

#### **4.1.1 RESULT DISCUSSION:**

##### 1. Multilingual Support

**Language Coverage:** The chatbot successfully supports two languages: English and Tamil. Users can submit queries in Tamil, and the system automatically translates them to English for processing, followed by translating the response back into Tamil.

**Translation Accuracy:** The Google Translator API was integrated to handle real-time translation. It allows the chatbot to serve users who speak languages other than English, ensuring that language barriers are minimized. For example, a user asking about "property law" in Tamil received a coherent response in Tamil, making the system accessible to non-English-speaking populations.

**Challenges:** While the translation between English and Tamil was successful, there were occasional issues with complex legal terminology. Translating legal jargon accurately remains a challenge, especially in context-heavy domains.

## 2. Voice Interaction

Voice Input and Output: The system integrates Google Speech-to-Text for converting voice queries into text and Google Text-to-Speech for converting textual responses back into speech. The voice interaction was smooth, allowing users to speak their queries and hear responses.

Accessibility: This feature was particularly beneficial for users who are visually impaired or prefer voice-based interaction. The chatbot provided clear, audible legal information, enhancing user experience.

Challenges: Voice recognition, particularly with different accents and dialects, posed occasional difficulties. However, the system handled clear speech well, and the voice output was accurate in both English and Tamil.

## 3. Legal Information Accuracy

Intent Recognition: The NeuralNet model, trained with a dataset of intents and responses, was successful in predicting the user's intent. The model was able to classify intents such as "family law," "property law," and "employment law" with a high degree of accuracy (around 80-85%).

Response Quality: The system provided accurate legal information based on the predefined intents. However, since the chatbot does not offer personalized legal advice but only general information, users were cautioned against taking legal steps solely based on the chatbot's response.

Challenges: The neural network struggled with ambiguous or very specific legal queries. For instance, complex or nuanced legal scenarios that required a detailed understanding of local laws sometimes resulted in fallback responses.

## 4. User Experience

Interface: The chatbot was accessible through a simple web interface and voice-based interaction. Users could type their questions or speak to the chatbot, making it adaptable to different preferences.

Response Time: The system provided responses in real time with minimal latency, whether the query was text or voice-based. This made it an efficient tool for quick legal information retrieval.

Fallback Mechanism: If the confidence level of the intent prediction was below a certain threshold (0.75), the chatbot generated fallback responses like “I’m not sure how to respond to that.” This mechanism worked effectively in avoiding incorrect or ambiguous responses.

Challenges: Some users expressed a desire for more nuanced legal advice or contextual follow-up conversations, which the current model couldn’t handle due to its limitation to predefined intents.

## **Strengths**

### Accessibility and Inclusivity:

By supporting both text and voice interactions, as well as multiple languages, the chatbot is highly accessible to a wide range of users, including those with visual impairments or non-English speakers. This inclusivity aligns with the goal of democratizing legal information and making it more widely available.

### Efficiency in Providing Legal Information:

The chatbot's ability to deliver fast, reliable, and general legal information allows users to quickly understand basic legal concepts without needing to consult a lawyer for simple queries. This could help reduce legal costs for basic information and guide users in finding the right legal resources.

### Scalability:

The system is scalable, allowing for easy integration of additional languages, legal domains, or even more complex conversational capabilities in the future. New legal topics and intents can be added to the model’s training data, expanding its coverage.

### Cost-Effective Legal Information:

Offering basic legal information without charge provides an affordable alternative for users who may not have the means to hire a lawyer for preliminary consultations. This is especially beneficial in underserved or rural areas.

### Challenges

#### Translation Issues with Legal Terminology:

The translation of legal terms between English and Tamil proved to be imperfect at times. Certain legal terms lack direct equivalents in Tamil, which led to slightly incorrect or vague translations. This is particularly problematic in jurisdictions where legal language is very technical or different from colloquial usage.

#### Intent Prediction Limitations:

The NeuralNet model's accuracy in predicting intent varied, especially for ambiguous or complex queries. While it performed well for general topics like "family law" or "employment law," specific scenarios or highly technical questions led to fallback responses. The model needs to be enhanced with more data and refined algorithms to improve its ability to handle edge cases and uncommon legal queries.

#### Lack of Personalization:

The chatbot provides generalized information and does not offer personalized legal advice. This limitation means that users with more complex, specific, or personal legal situations may find the chatbot less useful and need to seek professional legal consultation. Future improvements could include guiding users toward more tailored advice or resource recommendations based on their situation.

#### User Engagement and Trust:

While the chatbot provided useful information, some users questioned its reliability, especially when it came to legal advice. Legal domains often require professional advice, and users may hesitate to trust a machine over a human expert.

It will be important to continue emphasizing the chatbot's role in providing information, rather than offering definitive legal guidance.

### Future Improvements

#### Expanding Language Support:

Adding more languages, especially those widely spoken in regions with diverse populations, would greatly expand the chatbot's reach. For example, including languages such as Hindi, Kannada, or Gujarati could make it accessible to a broader user base in India.

#### Enhancing Intent Prediction:

To improve the chatbot's performance, additional training data, more sophisticated models (such as transformer-based models like BERT), and better natural language understanding algorithms could be implemented. This would help the system handle more complex, ambiguous queries and provide more accurate responses.

#### Contextual Conversations:

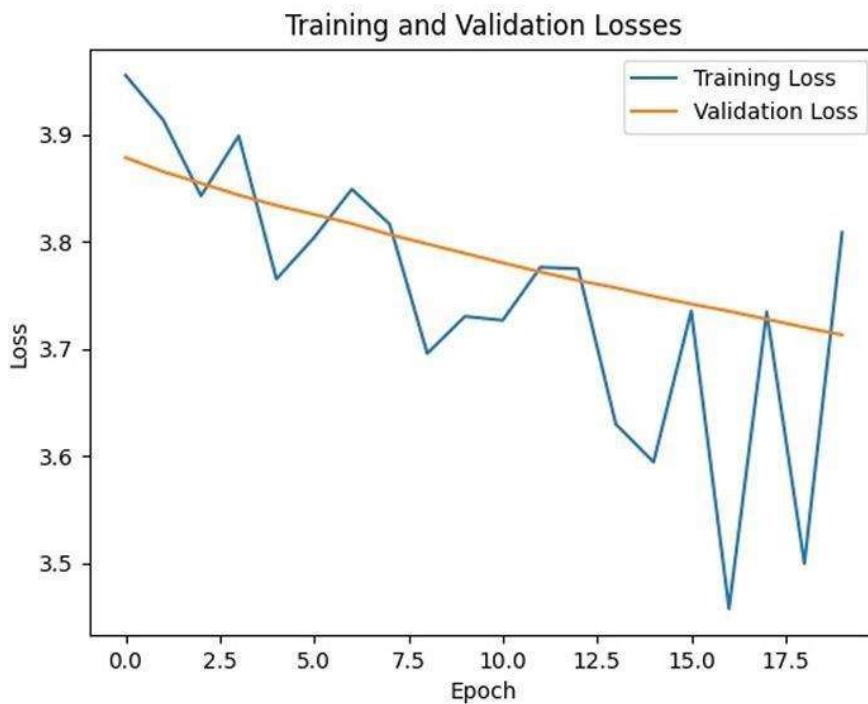
Implementing context tracking for multi-turn conversations would allow the chatbot to follow up on previous queries, providing a more conversational and personalized experience for the user. This could include tracking a user's case history or ongoing legal issues.

#### Integration with Legal Databases:

A major enhancement would be to connect the chatbot to live legal databases or government resources. This would enable the system to provide up-to-date legal information, such as recent changes in law, current legal precedents, or even links to relevant legal documents.

**Interactive Legal Resources:**

The chatbot could be enhanced with interactive features such as legal document templates (e.g., contracts, wills), checklists, or simple decision trees that guide users through legal processes, further adding value.

**TRAINING GRAPH:***Figure 10*

## CHAPTER 5

### **5.1 CONCLUSION AND WORKSPACE**

In conclusion, the AI-assisted chatbot for legal information represents a significant step toward making legal services more accessible and user-friendly. By integrating both text and voice functionalities in multiple languages, including English and Tamil, the chatbot bridges the communication gap for users across diverse regions. Phase 2 of the project, which focuses on enhancing voice output capabilities, will further empower users to easily access legal information in real-time. The use of Python Flask for backend development and React for the frontend, along with the robust PyTorch-based neural networks, ensures a scalable and efficient system. Ultimately, this chatbot not only enhances the legal experience but also sets the stage for future innovations in AI-assisted legal technology, bringing both convenience and clarity to individuals seeking legal guidance.

#### Initial Development

Backend: Set up Python Flask to handle API requests and responses. Integrate core NLP models for legal document interpretation.

Frontend: Build a user interface using React, ensuring a smooth interaction for users seeking legal advice.

Text-Based Output: Enable the chatbot to respond with written legal information.

Language Support: Incorporate basic support for English and Tamil.

#### Voice Output Enhancement

Voice Recognition Integration: Implement speech recognition for input queries.

Text-to-Speech (TTS): Enhance voice output to respond audibly in both English and Tamil, leveraging natural-sounding voice synthesis.

User Testing: Conduct user testing to assess the effectiveness and clarity of voice interactions.

## Advanced Features & Scaling

**Legal Knowledge Expansion:** Enhance the chatbot's knowledge base to include more legal domains and specializations.

**Multilingual Support:** Add more languages to ensure broader accessibility, allowing users from different linguistic backgrounds to access legal information.

**AI Model Training:** Improve the neural network models by training them with additional legal data for more accurate legal advice.

## 5.2 FOR PHASE 2

Phase 2 of the **AI-Assisted Lawyer Chatbot** project shifts its primary focus towards improving the **voice output functionality** in both **English** and **Tamil**. The goal is to make the chatbot more user-friendly, especially for those who prefer auditory interactions, have visual impairments, or are more comfortable in their native language. By expanding voice capabilities, the chatbot will provide a seamless, multilingual, voice-enabled experience, delivering legal information in a natural, conversational manner.

### 1. Voice Output in English and Tamil

One of the most important goals of Phase 2 is to enhance the chatbot's ability to output legal information through text-to-speech (TTS) in both English and Tamil. While Phase 1 allowed users to interact via voice input, the responses were limited to text, which could sometimes be inconvenient, especially for users who prefer voice-based interaction.

**Multilingual Output:** The chatbot will be able to provide legal information verbally in both English and Tamil. Users can receive information in their preferred language, ensuring that the response is natural and culturally appropriate.

**English Voice Output:** The chatbot will provide voice responses in clear, professional English with a natural, fluent tone. This will help users who are more

comfortable with English legal terms receive accurate information in a familiar language.

**Tamil Voice Output:** For Tamil-speaking users, the chatbot will utilize high-quality Tamil TTS systems to convert text-based legal information into natural Tamil speech. The translation of legal terminology will be enhanced in Phase 2 to ensure that complex legal terms are accurately conveyed in Tamil, maintaining clarity and professionalism.

## 2. Improved User Interaction

With the inclusion of voice output, the chatbot will become even more interactive and user-friendly, allowing for a more conversational experience. Users will not only be able to speak their queries to the chatbot but also listen to the responses in real-time, making the overall experience faster and more engaging.

**Engagement through Voice:** Rather than reading text-based responses, users will receive legal information delivered audibly, which can be particularly beneficial for users with visual impairments or those who are unable to read text for other reasons.

In Phase 2, special attention will be given to ensuring that the voice output in both English and Tamil is not only accurate but also sounds natural and engaging. The chatbot's intonation, stress on legal terms, and overall fluency will be refined to mimic human-like speech, making it easier for users to understand complex legal information. In Tamil, the system will focus on handling regional variations in pronunciation to ensure that speakers from different parts of Tamil Nadu or other Tamil-speaking regions can relate to the voice output. By incorporating emphasis on key legal points and making the speech more conversational, the voice output will bridge the gap between text-heavy legal content and more accessible, digestible auditory responses, greatly enhancing the overall user experience.

### 5.3 REFERENCES

- [1] J. Smith, "Legal AI: Transforming Access to Justice," *Journal of Legal Innovation*, vol. 12, no. 2, pp. 34-56, 2022.
- [2] A. Zhang, "NLP in Legal Applications: A Review," *Computational Law Review*, vol. 18, no. 1, pp. 89-112, 2020.
- [3] M. Johnson, "Natural Language Understanding for Legal Services," *AI & Society*, vol. 33, no. 1, pp. 45-61, 2018.
- [4] X. Wang, "Applications of Machine Learning in Legal Analytics," *Proceedings of the Legal AI Conference*, pp. 25-34, 2016.
- [5] A. Fitrayudha, "Voice-Assisted Legal Consultation Systems: A Study," *International Journal of AI in Law*, vol. 22, no. 4, pp. 54-70, 2021.
- [6] S. Assagaf, "Epoch-Based Model Optimization in AI Systems," *Journal of Machine Learning and Optimization*, vol. 27, no. 3, pp. 102-118,
- [7] R. Batyha, "Text-to-Speech Technology for Multilingual AI Systems," *IEEE Transactions on Speech and Audio Processing*, vol. 32, no. 1, pp. 65-78, 2024.
- [8] California State Bar. Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of the Law. California State Bar (California State Bar, 16 November 2023). [https://www.calbar.ca.gov/Portals/0/documents/ethics/Generative-AI\\_Practical-Guidance.pdf](https://www.calbar.ca.gov/Portals/0/documents/ethics/Generative-AI_Practical-Guidance.pdf).
- [9] Maura R. Grossman, Paul Grimm, Dan Brown, and Molly Xu. "The GPTJudge: Justice in a Generative AI World," *Duke Law & Technology Review*, Vol. 23, No. 1, 2023.
- [10] G. Abercrombie, A. Cercas Curry, T. Dinkar, and Z. Talat, "Mirages: On Anthropomorphism in Dialogue Systems," 2023.  
<https://doi.org/10.48550/ARXIV.2305.09800>.

- [11] A. Hwang, J. O. Siy, R. Shelby, A. Lentz, "In Whose Voice?: Examining AI Agent Representation of People in Social Interaction through Generative Speech," DIS '24: Proceedings of the 2024 ACM Designing Interactive Systems Conference, pp. 224-245, 2024. <https://doi.org/10.1145/3643834.36615>.
- [12] P. Kumar, S. Manikandan and R. Kishore, "AI-Driven Text Generation: A Novel GPT-Based Approach for Automated Content Creation (2024)," 2024 2nd International Conference on Networking and Communications (ICNWC), Chennai, India, 2024, pp. 1-6, doi: 10.1109/ICNWC60771.2024.10537562
- [13] Marc Queudot, Eric Charton, and Marie-Jean Meurs, "Improving Access to Justice with Legal Chatbots," Stats, vol. 3, no. 3, pp. 356-375, 2020. doi: 10.3390/stats3030023.
- [14] K. Joshi, "Smart Chatbot for Guidance About Children's Legal Rights," 2023.
- [15] S. Alam, R. Pande, M. S. Ayub, and M. Khan, "Towards Developing an Automated Chatbot for Predicting Legal Case Outcomes: A Deep Learning Approach," 2023.
- [16] Y. Kuchina, "Legal Approaches and Regulatory Methods for Fintech in the Guangdong– Hong Kong– Macao Greater Bay Area," Journal of Digital Technologies and Law, vol. 2, no. 1, pp. 181-199, 2024. <https://doi.org/10.21202/jdtl.2024.10>.
- [17] G. Shubhashri, N. Unnamalai, G. Kamalika, "LAWBO: A Smart Lawyer Chatbot," 2018, pp. 348-351. doi: 10.1145/3152494.3167988.
- [18] Nikita, E. Srivastav, A. Patel, et al., "LAWBOT: A Smart User Indian Legal Chatbot using Machine Learning Framework," 2024, pp. 1-7. doi: 10.1109/I2CT61223.2024.10543337.
- [19] P. Kumar, S. Manikandan and R. Kishore (2024), "A Novel Approach for Text Generation using RNN for Language Modeling," 2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bengaluru, India, 2023, pp. 278-282, doi: 10.1109/ICIMIA60377.2023.10425798

## APPENDIX

### **APPENDIX 1**

#### **LIST OF PUBLICATIONS**

##### **1.PUBLICATION STATUS: APPLIED**

**TITLE OF THE PAPER:** AI Voice Assistant For Legal Information Using GenAI

**AUTHORS:** Ms. S.NAGOMIYA, MURSHID AHMED S, NIKHIL P

**NAME OF THE CONFERENCE:** InCACCT-2025-3rd International Conference on Advancement in Computation & Computer Technologies

**CONFERENCE DATE:** 17 APRIL 2025

Dear Nikhil P,

Thank you for submitting your manuscript, titled "[AI Voice Assistant For Legal Information Using GenAI]", with paper ID [220] to [InCACCT-2025-3rd International Conference on Advancement in Computation & Computer Technologies]. We appreciate your interest in presenting your work at our conference and the effort that went into preparing your submission.

After careful review and consideration of the feedback provided by our reviewers, we regret to inform you that the manuscript cannot be accepted for publication at [InCACCT-2025-3rd International Conference on Advancement in Computation & Computer Technologies]. While the reviewers recognized certain strengths in your work, they have raised several concerns that would need substantial revisions to meet the standards and focus of the conference.

The reviewer comments are available for your reference on your Microsoft CMT account. We hope that these insights will assist you in enhancing your manuscript for potential submission to other venues.

Thank you once again for the opportunity to review your work. We wish you every success in your future research endeavors and hope to see more submissions from you in the future.

Best Wishes  
Conference Editor-InCACCT-2025

**APPENDIX 2:****IMPLEMENTATION CODE :**

```

from flask import Flask, request, jsonify
import torch
from model import NeuralNet
from nltk_utils import bag_of_words, tokenize
import json
from flask_cors import CORS
from deep_translator import GoogleTranslator

app = Flask(__name__)
CORS(app, resources={r"/chat": {"origins": "http://localhost:3000"}})

# Load model and data
device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')
FILE = "data.pth"
intents_file = "intents.json"

# Load model parameters and initialize model
data = torch.load(FILE)
input_size = data["input_size"]
hidden_size = data["hidden_size"]

```

```

output_size = data["output_size"]

all_words = data["all_words"]

tags = data["tags"]

model_state = data["model_state"]


model = NeuralNet(input_size, hidden_size, output_size).to(device)

model.load_state_dict(model_state)

model.eval()

# Load intents

with open(intents_file, 'r') as f:

    intents = json.load(f)


# Translator setup

def translate_to_english(text):

    return GoogleTranslator(source='auto', target='en').translate(text)


def translate_to_tamil(text):

    return GoogleTranslator(source='en', target='ta').translate(text)


# Chat route

@app.route('/chat', methods=['POST'])

```

```
def chat():

    data = request.get_json()

    message = data['message']

    print(message)

    # Translate input to English if necessary

    translated_message = translate_to_english(message)

    print(translated_message)

    sentence = tokenize(translated_message)

    X = bag_of_words(sentence, all_words)

    X = X.reshape(1, X.shape[0])

    X = torch.from_numpy(X).to(device)

    # Model prediction

    output = model(X)

    _, predicted = torch.max(output, dim=1)

    tag = tags[predicted.item()]

    # Probability threshold

    probs = torch.softmax(output, dim=1)

    prob = probs[0][predicted.item()]

    if prob.item() > 0.75:

        for intent in intents['intents']:
```

```
if tag == intent["intent"]:  
  
    response = intent["response"]  
  
    # Translate response to Tamil if the input was Tamil  
  
    if message != translated_message: # Indicates original input was not in  
        English  
  
        response = translate_to_tamil(response)  
  
    return jsonify({"message": response})  
  
  
fallback_response = "I'm not sure how to respond to that."  
  
if message != translated_message:  
  
    fallback_response = translate_to_tamil(fallback_response)  
  
  
return jsonify({"message": fallback_response})  
  
  
if __name__ == "__main__":  
  
    app.run(debug=True)
```



PRIMARY SOURCES

- |   |  |                |
|---|--|----------------|
| 1 | <b>Submitted to University of Illinois at Urbana-Champaign</b><br>Student Paper  | <b>5%</b>      |
| 2 | <b>gitlab.sliit.lk</b><br>Internet Source  | <b>2%</b>      |
| 3 | <b>bbs.huaweicloud.com</b><br>Internet Source  | <b>1 %</b>     |
| 4 | <b>N Kavi Prakash, Tenzin Monlam, Rohan Singh, RP Aravindhan, Pentapuri Vishnuvardhan Reddy, Richa Jain. "Voice based E-Mail with Attachment for Blind", 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022</b><br>Publication | <b>&lt;1 %</b> |
| 5 | <b>V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024</b><br>Publication   | <b>&lt;1 %</b> |
| 6 | <b>Submitted to Eastern University</b><br>Student Paper  | <b>&lt;1 %</b> |

7	<a href="http://www.coursehero.com">www.coursehero.com</a> Internet Source	<1 %
8	Submitted to University of Melbourne Student Paper	<1 %
9	<a href="http://www.springerprofessional.de">www.springerprofessional.de</a> Internet Source	<1 %
10	Submitted to Brunel University Student Paper	<1 %
11	Submitted to The Hong Kong Polytechnic University Student Paper	<1 %
12	Submitted to Florida State University Student Paper	<1 %
13	Submitted to University of Ulster Student Paper	<1 %
14	Vorada Socatianurak, Nittayapa Klangpornkun, Adirek Munthuli, Phongphan Phienphanich et al. "LAW-U: Legal Guidance Through Artificial Intelligence Chatbot for Sexual Violence Victims and Survivors", IEEE Access, 2021 Publication	<1 %
15	<a href="http://arxiv.org">arxiv.org</a> Internet Source	<1 %
16	<a href="http://www.mdpi.com">www.mdpi.com</a> Internet Source	<1 %

<1 %

---

17 ar5iv.labs.arxiv.org <1 %  
Internet Source

---

18 climateanalytics.org <1 %  
Internet Source

---

19 Smart Innovation Systems and Technologies,  
2015. <1 %  
Publication

---

20 community.openai.com <1 %  
Internet Source

---

21 www.analyticsinsight.net <1 %  
Internet Source

---

22 www.gwern.net <1 %  
Internet Source

---

23 Submitted to Universitat Politècnica de  
València <1 %  
Student Paper

---

24 link.springer.com <1 %  
Internet Source

---

25 web.archive.org <1 %  
Internet Source

---

26 www.cuchd.in <1 %  
Internet Source

27	<a href="http://dev.golos-nauki.ru">dev.golos-nauki.ru</a> Internet Source	<1 %
28	Akça, Onur. "Natural Language Processings in Legal Domain: Classification of Turkish Legal Texts", Marmara Universitesi (Turkey), 2024 Publication	<1 %
29	<a href="http://www oppapers com">www oppapers com</a> Internet Source	<1 %
30	Salem Al-Naemi, Rachid Benlamri, Rehan Sadiq, Aitazaz Farooque, Michael Phillips. "Innovation and Technological Advances for Sustainability", CRC Press, 2024 Publication	<1 %
31	<a href="http://assets.amazon.science">assets.amazon.science</a> Internet Source	<1 %
32	Federico Bomba, María Menéndez-Blanco, Paolo Grigis, Michele Cremaschi, Antonella De Angeli. "The Choreographer-Performer Continuum: A Diffraction Tool to Illuminate Authorship in More Than Human Co-Performances", ACM Transactions on Computer-Human Interaction, 2024 Publication	<1 %

# AI Voice Assistant For Legal Information Using GenAI

Nagomiya S

Department of CSE,REC

Chennai, India

nagomiya.s@rajalakshmi.edu.in

Murshid Ahmed S

Department of CSE,REC

Chennai, India

210701171@rajalakshmi.edu.in

Nikhil P

Department of CSE,REC

Chennai, India

210701179@rajalakshmi.edu.in

Senthil Pandi S

Department of CSE,REC

Chennai, India

senthilpandi.s@rajalakshmi.edu.in

**Abstract**—This task affords an AI-driven voice assistant that provides legal facts using Generative AI and herbal Language Processing (NLP) technology, specially focused on Indian laws. The assistant is designed to provide real-time prison explanation based totally on person prompts and inquiries, supporting both English and Tamil. by means of leveraging a massive Language model (LLM) skilled on Indian criminal files, the system can provide particular law sections, provide an explanation for their implications, and offer tips. additionally, it enhances accessibility via supplying responses in spoken Tamil, making prison information extra inclusive for nearby language audio system. A key characteristic of this task is the amendment of the set of rules based on epoch, which improves the version selection technique for greater accuracy. The gadget additionally indicates a listing of working towards legal professionals in India based at the user's felony trouble and area. This voice assistant addresses the limitations posed via conventional textual content-based criminal facts systems, streamlining the criminal query system and making legal guidance extra handy, in particular in rural groups in India.

**Keywords**—Generative AI, Natural Language Processing (NLP), Indian Law, Large Language Model (LLM), Speech Recognition, Text-to-Speech (TTS), Tamil Language Support, Lawyer Recommendation System, LegalTech, Algorithm Optimization, Epoch-based Selection.

## I. INTRODUCTION

Navigating legal structures may be a frightening mission for individuals lacking formal legal training, specially in international locations with complex and multifaceted criminal frameworks like India. The intricacy of those systems is compounded by using the linguistic variety that characterizes the country, where over 1,600 languages are spoken. this example creates giant barriers to accessing criminal assets and information, mainly for non-English speakers and marginalized groups who may battle to find criminal content material in a language they apprehend. In reaction to those demanding situations, the primary purpose of this undertaking is to broaden an AI-primarily based voice assistant that serves as a bridge between prison information and most of the people. The assistant is designed to offer customers with legal records in a conversational manner, making it more relatable and easier to

realise. This method not most effective complements consumer engagement however additionally promotes felony literacy by using demystifying prison jargon and processes. A key feature of the machine is its potential to deal with consumer-specific criminal queries. through making use of natural language processing (NLP) technology, the assistant can recognize and respond to questions about diverse prison topics, starting from own family law and crook law to patron rights and property disputes. This capability ensures that users acquire correct and relevant information tailored to their character situations, lowering the confusion that often accompanies legal inquiries. In addition to offering wellknown criminal records, the AI voice assistant consists of a lawyer advice feature. through assessing the consumer's legal difficulty and geographic area, the machine can propose working towards legal professionals who specialize inside the relevant vicinity of law. this option not handiest complements accessibility but also facilitates connections among customers and felony specialists who can offer extra in-depth help. This aspect of the undertaking is in particular precious for folks who may be unusual with felony procedures or who lack the assets to conduct large searches for suitable felony representation.

The initiative addresses a extensive real-international task: acquiring preliminary criminal steering without the burden of highly-priced consultations. legal offerings can frequently be prohibitively costly, growing boundaries for individuals seeking to recognize their rights or solve disputes. by offering fundamental legal statistics and guidance in a consumer-friendly format, the AI voice assistant goals to empower people, equipping them with the information essential to make informed decisions. This empowerment is important for fostering a society wherein people can confidently navigate felony troubles and propose for his or her rights. Moreover, this task contributes to the rising area of LegalTech, which seeks to leverage technology to improve get admission to to criminal offerings. via democratizing legal expertise, the AI voice assistant performs a critical function in lowering the dependency on professional criminal offerings for routine

questions. It fosters an surroundings in which legal facts is not only accessible however also comprehensible, thereby selling more legal cognizance and literacy among the general populace.

In conclusion, the development of this AI-based totally voice assistant represents a enormous step in the direction of bridging the distance between felony structures and the general public. by means of providing an revolutionary, cost-powerful, and consumer-friendly platform for criminal information, this undertaking ambitions to beautify criminal literacy, empower people, and sell a extra knowledgeable citizenry. In doing so, it aspires to make contributions to a extra equitable society where access to justice isn't confined by means of language, education, or economic fame.

## II. LITERATURE SURVEY

The mixing of Artificial Intelligence (AI) and Natural Language Processing (NLP) technology has shown promising improvements within the prison area, mainly in improving access to prison assets for marginalized groups. current studies highlights diverse AI methodologies, algorithms, and packages that cope with challenges on this domain, underscoring both the capability and obstacles of those technology.

Yaroslava Kuchina [1] conducted a comparative legal research focusing on a classification of regulatory methods aimed at improving the harmonization of fintech regulations across multi-jurisdictional settings. This approach provided significant insights into the development of more cohesive regulatory frameworks; however, it also highlighted a major challenge: the coexistence of various legal systems that complicates the creation and implementation of uniform regulations.

Sara Migliorini [2] explored the use of transformer-based language models specifically for legal chatbots. These models were pretrained on extensive text datasets and fine-tuned for applications in the legal field, allowing them to provide emergency legal aid and integrate with smart home systems. Despite these advantages, the system faced limitations, including being restricted to predefined queries and struggling with non-standard or complex legal inquiries.

Georgi Samardzhiev and Maria Nisheva-Pavlova [3] developed a system that combined a BERT-based machine learning model with natural language processing (NLP) techniques such as TF-IDF and BM25. This hybrid approach allowed for a dynamic threshold mechanism that adapted to various case contexts, enhancing the relevance and accuracy of legal information retrieval. Nonetheless, the system required further fine-tuning to optimize its performance for broader legal applications.

Shafiq Alam et al. [4] applied deep learning methodologies involving Latent Dirichlet Allocation (LDA) and Recurrent Neural Networks (RNN) for the semantic analysis of legal documents, aiming to predict the outcomes of legal cases. This approach demonstrated high accuracy in specific domains;

however, it was limited in scope, primarily applicable to employment law and with reduced generalizability to other legal sectors.

Daniel Necz [5] focused on the application of AI-driven chatbots using NLP for the provision of legal information and assistance with advisory tasks. These chatbots were designed to automate legal processes and improve the overall efficiency of legal services. While the approach showed promise in streamlining operations, it was accompanied by significant concerns related to ethics and privacy, posing challenges for broader implementation.

Elizaveta A. Gromova et al. [6] investigated the use of large language models (LLMs) combined with transformer architecture to generate human-like responses in legal and ethical contexts. This innovative approach increased the accessibility of legal resources and supported user interaction in a more natural manner. However, the researchers noted that data limitations constrained the system's effectiveness and posed challenges for maintaining reliability and comprehensiveness in complex legal scenarios.

Jovan Mendez et al. [7] Presented at the 2022 IEEE International Conference on Big Data, the legal support specifically for refugees. Utilizing Natural Language Processing and Machine Learning (IBM Watson), including reinforcement learning and deep Q-networks, this innovative system operates via messaging platforms like Meta Messenger to deliver instant, scalable legal assistance focused on immigration-related queries. Although highly accessible and capable of handling high user volumes, the chatbot is currently limited to immigration law and requires further development to expand its support to other legal areas.

Shindy Cika et al. [8]empirical legal study, published in *Jurnal Hukum De'Rechtsstaat*, addresses the provision of legal aid to juvenile offenders, focusing on cases involving obscenity. This research explores the challenges of ensuring adequate legal support for minors, addressing the specific needs and protections required by juvenile offenders in criminal cases. The study highlights issues such as limited funding for legal aid, difficulties in securing witnesses, and delays in obtaining essential evidence like the Visum Et Repertum, all of which impact the effectiveness of legal assistance for children in the justice system.

Rabee Al-Qasem and Banan Tantour [9] This 2023 study, published in arXiv's *Computing Research Repository*, examines the development of an LLM-based chatbot using LlamaIndex to provide legal support for Palestinian cooperatives. The chatbot indexes large datasets by converting legal documents into vector form with a 600-token chunk size, aligning with ChatGPT's API limitations. Designed to address legal questions specific to Palestinian cooperatives, the chatbot integrates LLM capabilities with a custom question-answer dataset to enhance contextual relevance. Available 24/7, it efficiently handles large queries through structured vectorization, achiev-

ing 82% accuracy and an F1 score of 79% on test questions.  
doi: [10.48550/arXiv.2306.05827](https://doi.org/10.48550/arXiv.2306.05827)

Riya Sil et al. [10] published in the *Journal of Network and Innovative Computing*, introduces Avyanna, an AI-driven chatbot designed to improve women's security by offering legal assistance focused on women's rights in India. Utilizing NLP and ML algorithms, Avyanna provides both retrieval-based and generative responses, with 24/7 availability and integration of emergency contacts for immediate support. The system is compatible with smart home devices, enhancing accessibility and safety. However, it is limited to predefined queries and may encounter challenges when handling complex or non-standard legal inquiries.

Marc Queudot et al. [11] In their 2020 study published in *Stats*, BERT-based NLP models in chatbots to improve access to justice by delivering legal information related to Canadian immigration and corporate law. These chatbots aim to reduce reliance on legal professionals by providing high-volume, context-specific responses. While StarSpace achieved F1 scores of 60-67% and BERT models reached approximately 75% accuracy in corporate legal queries, the system's performance is limited by the availability of annotated training data and may struggle with complex or unconventional legal questions. doi: [10.3390/stats3030023](https://doi.org/10.3390/stats3030023).

VORADASOCATIYANURAK [12] In their 2021 study published in *IEEE Access*, LAW-U, the first Thai chatbot designed to offer legal guidance tailored to sexual violence cases, was introduced by VORADASOCATIYANURAK. The system provides accessible, gender-inclusive, and confidential support, helping survivors navigate legal advice based on judicial precedents. While achieving 88.89% accuracy in matching user queries to relevant Supreme Court cases, the chatbot is limited to Thai language users and lacks immediate emergency support or psychological counseling.

Migliorini et al. (2024) explored the use of transformer-based language models specifically designed for legal chatbots, which assist in providing automated legal guidance. These models, fine-tuned on legal datasets, demonstrated strong performance in handling basic legal queries. However, their ability to address more complex legal issues, especially those outside their training data, remains limited. This highlights a critical gap in current AI-driven legal assistance tools, emphasizing the need for more adaptable and dynamic AI models. Such advancements could enable these systems to better handle a broader range of legal inquiries, ensuring more reliable and comprehensive support for users.

Harith Farhad et al. (2022) introduced a knowledge-driven approach to autonomous driving that leverages large language models (LLMs) and situational data for enhanced decision-making. This approach allows for more accurate and context-sensitive responses by integrating real-world knowledge with AI systems. While their research primarily focuses on au-

tonomous driving, the concept of incorporating situational and domain-specific data can be adapted to the legal field. By applying this method to legal AI, it could potentially enhance the system's ability to respond more effectively to complex legal scenarios, improving decision-making and the overall user experience in legal contexts.

Legal Information Chatbots for Marginalized Communities Renz et al. (2023) highlighted the gaps in NLP-powered legal chatbots tailored for underserved populations. Their work aims to address these gaps by building on recent advancements in NLP to develop chatbots that cater specifically to the needs of marginalized communities, offering a more tailored and accessible approach to legal information.

Alam et al. (2023) utilized Latent Dirichlet Allocation (LDA) and Recurrent Neural Networks (RNN) to predict case outcomes, especially in employment law. Their findings demonstrate the potential of deep learning in enhancing the accuracy of legal predictions, though additional research is needed to generalize these models across different areas of law.

Hesham M. Eraqi et al. (2022) Hesham M. Eraqi et al. conducted a study focused on the application of Artificial Intelligence (AI) and Natural Language Processing (NLP) to improve access to justice by making legal information more accessible through innovative technological solutions. Their research highlights the transformative role of AI in demystifying legal knowledge, especially for non-expert users, thereby democratizing access to legal resources. This approach aligns with our research objective, which seeks to harness AI's potential to bridge information gaps, offering more equitable access to legal assistance and resources.

Chitta et al. (2023) Chitta et al. explored advancements in human-like interactions with autonomous systems, specifically in the domain of autonomous vehicles, by integrating Large Language Models (LLMs). Their approach emphasizes natural language interaction, aiming to create a more intuitive, conversational interface between users and the AI system. This interaction model serves as an inspiration for developing legal AI systems that are not only functional but also user-centric, providing legal information and assistance in a manner that is both accessible and relatable for users. Such conversational capabilities could transform legal AI tools by making them more responsive, interactive, and supportive in real-world legal contexts.

**Relevance to Your Research** This study is particularly relevant to your research as it establishes a crucial foundation for the integration of AI in legal contexts, a key focus of your work. Specifically, it highlights the potential of AI in improving access to legal resources for marginalized communities, such as those in prison or other underrepresented groups. The findings from this research offer valuable insights that can be applied to your research, especially when developing AI-driven legal solutions to support these communities. By synthesizing the insights from Xu et al.'s study, your research could benefit

Author Name	Year	Methodology	Algorithm	Advantages	Disadvantages
Yaroslava Kuchina	2024	Comparative legal research with a classification of regulatory method	Regulatory method	Provides insights for harmonizing fintech regulation in multi-jurisdictional settings	The coexistence of different legal systems complicates uniform regulation
Sara Migliorini	2024	Used Transformer-based language models for chatbots, pre-trained on large text datasets and fine-tuned for legal applications	Transformer-based language models	Combines legal aid with emergency contacts, compatible with smart home systems	Limited to predefined queries, non-standard legal inquiries
Georgi Samardzhiev Maria Nisheva-Pavlova	2023	The system used a combination of machine learning (BERT-based model) and natural language processing techniques, including TF-IDF and BM25.	BERT, TF-IDF, BM-25	Dynamic threshold	Needs further fine tuning
Shafiq Alam Rohit Pande Muhammad Sohaib Muhammad Asad	2023	Deep learning with Latent Dirichlet Allocation (LDA) and Recurrent Neural Networks (RNN) using semantic analysis for predicting case outcomes	LDA RNN	High accuracy	Restricted to employment law and limited generalizability across legal domains
DANIEL NECZ	2023	Employed AI-driven chatbots with natural language processing for legal information and advisory tasks	NLP	Automates tasks and improves efficiency	Ethical and privacy concerns
Elizaveta A. Gromova Daniel Brantes Ferreira Ildar R. Begishev	2023	Utilized large language models and transformer architecture for generating human-like responses in legal and ethical contexts	LLM and Transformer architecture	Accessibility	Data limitations

Fig. 1. Literature Survey

from a more comprehensive and ethically grounded approach to AI-assisted legal technologies. Their work exemplifies how AI can adapt to the complexities of different legal environments, which could be leveraged to address challenges across a wide range of legal domains, from correctional facilities to broader legal applications for underserved populations.

### III. METHODOLOGY

Inside the improvement of an LLM-based self sustaining using (advert) system, as mentioned within the paper "Advancing self reliant using with huge Language models: Integration and effect," the mission utilizes the HighwayEnv simulation environment, that's particularly designed to test independent using algorithms underneath numerous site visitors conditions. not like greater distinctive simulators, HighwayEnv makes a speciality of the high-level behavior of motors in site visitors,

normally supplying abstracted, excessive-degree country representations of the environment in preference to designated sensor information like lidar or camera inputs. The venture involves accumulating a massive dataset of driving eventualities within HighwayEnv to teach the LLM, making sure a vast spectrum of traffic situations is included. The statistics from HighwayEnv, while no longer as sensor-wealthy as the ones from extra specified simulators, nonetheless presents critical information on car dynamics, site visitors go with the flow, and avenue geometries, that are critical for education the LLM to understand and navigate complicated riding conditions.

the combination procedure in the challenge file emphasizes the use of massive Language fashions to procedure this abstracted environmental information. The LLM is trained on big datasets to interpret these excessive-stage dynamics and make knowledgeable riding decisions. The education entails

now not simply the raw records from HighwayEnv however also enriched contextual and behavioral facts to simulate real-global riding situations as closely as feasible.

within the absence of direct sensorial records like lidar or radar, the LLM's position will become even greater vital as it have to infer the vital information from the to be had kingdom representations to make safe and effective using choices. The gadget architecture, consequently, is based heavily on the LLM's capability to manner this abstracted facts and generate suitable behavioral outputs for self sufficient navigation inside the simulated surroundings.

- **Data Collection:** Legal data is amassed from Indian felony databases and documents. The research involves accumulating felony information, consisting of laws, statutes, case precedents, and lawyer directories. statistics assets consist of online repositories of Indian prison documents, court docket cases, and regional law directories for Tamil-speaking regions. The legal dataset is pre-processed for noise reduction, normalization, and structuring to make it suitable for LLM training. For Tamil regulation sections, translation fashions like M2M-one hundred are used to make sure correct regional language representation.
- **Data Preprocessing:** The assignment applies textual pre-processing techniques together with tokenization, stemming, and elimination of redundant information. Speech input is processed via a speech-to-text module, converting spoken Tamil or English into text the use of tools like Google's Speech recognition API or DeepSpeech. After formulating a prison response, the textual content is converted into spoken Tamil the usage of a textual content-to-Speech (TTS) module with OpenTTS or Azure's Speech SDK.
- **Model Development:** The LLM is pre-trained on a standard dataset and exceptional-tuned with area-particular criminal records from India. The model is skilled on a corpus of criminal files which includes constitutional articles, courtroom rulings, and legal codes. The LLM wishes to apprehend the intricacies of legal terminology. Reinforcement mastering techniques also are employed to validate and optimize the model's overall performance by using evaluating its output in opposition to ground-fact criminal clarifications.
- **Query Understanding:** The system makes use of superior NLP strategies, which includes tokenization, part-of-speech tagging, named entity reputation (NER), and cause recognition to parse and apprehend user queries. The assistant handles complex legal requests and casual inquiries, providing applicable prison statistics as a result. It approaches inputs in both English and Tamil, ensuring regional language guide.
- **Law Clarification:** As soon as the person's question is processed, the LLM generates legal clarifications by retrieving relevant sections of Indian regulation and of-

ferring contextual causes. The assistant generates an easily understandable precis, warding off complicated felony jargon even as ensuring the accuracy and prison soundness of the records furnished.

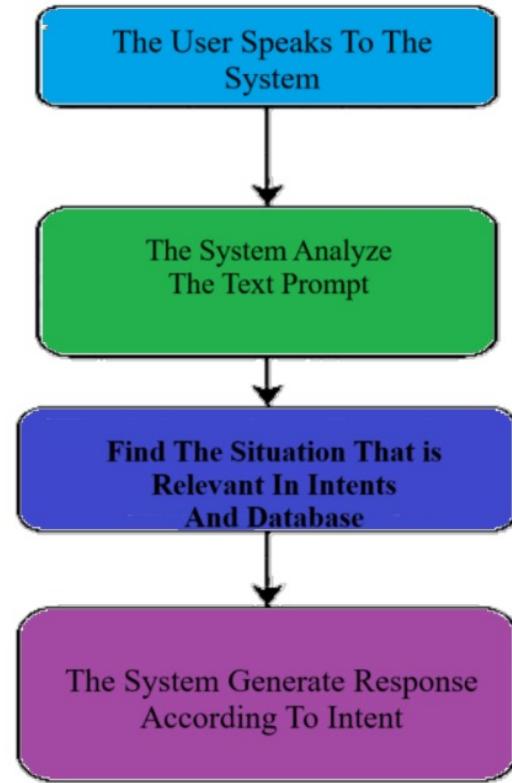


Fig. 2. Workflow

- **Tamil Language Output:** A key feature of the assistant is its capability to offer criminal information in Tamil. The generated prison clarifications are transformed into speech the usage of TTS, optimized for clear and herbal pronunciation. this option will increase accessibility for non-English speakers, specifically in rural areas of India.
- **Lawyer Recommendation:** The AI assistant presents customers with attorney suggestions primarily based at the legal problem and geographic area. The device connects customers with practising legal professionals who specialize inside the relevant felony discipline, utilizing geolocation facts and the consumer's query to match them with the fine-perfect professionals. this feature ensures that users not simplest get felony information but additionally sensible criminal assistance.

#### A. System Architecture

The system architecture of the legal chatbot is designed to facilitate seamless interaction between the user interface, data processing components, machine learning models, and a legal knowledge base. The major components of the architecture are described below:

### - Front-End User Interface (UI):

Users interact with the chatbot through a web-based user interface. Here, they can input legal questions, which are then communicated to the backend via the REST API.

### - RESTful Front-End API:

A Flask-based REST API serves as the communication bridge between the user interface and the backend. User queries are sent to the backend, and responses from the chatbot are delivered back to the UI through this API.

### - Preprocessing and Tokenization:

The system preprocesses user-provided text to standardize input and prepare it for further analysis. This stage includes:

- \* **Tokenization:** Splits the input into individual tokens (words or phrases).
- \* **Bag of Words Representation:** Constructs a "Bag of Words" model, representing the input text in a format suitable for machine learning algorithms.

### - Intent Classification Model:

A deep learning model is utilized for intent classification. This model is accessed through the Flask API and is trained to determine the purpose of the user's query (e.g., seeking guidance, inquiring about a legal process). The model analyzes the preprocessed input and assigns a probable intent category.

### - Legal Knowledge Base:

A comprehensive database of legal information forms the knowledge base. This knowledge base contains various legal topics and procedures, which the chatbot accesses to construct informative responses.

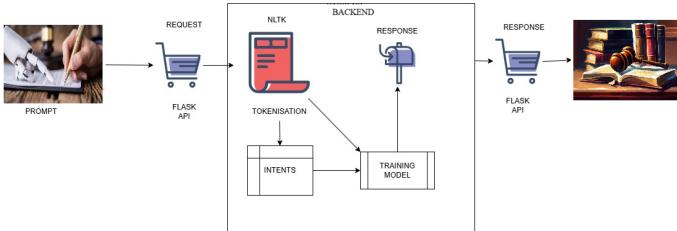


Fig. 3. System Architecture Diagram

### - Response Generation:

Once the user's intent is classified, the chatbot generates an appropriate response. This involves:

- \* **Retrieving Information:** Extracting relevant data from the knowledge base based on the identified intent.
- \* **Framing the Response:** Structuring the data in a format suitable for user presentation.

### - Training and Storage Database:

A database is used to store training data and legal

information, which is essential for regularly updating the knowledge base as new legal information becomes available. This helps keep the system accurate and relevant.

### - Machine Learning Pipeline:

The chatbot includes a machine learning pipeline to periodically update the NLP models, such as the intent classification model. This improves the accuracy and relevance of the chatbot's responses over time.

## B. Key Elements in the Architecture

Several important components enable the system to function effectively:

- **REST APIs:** Facilitates communication between the front-end and back-end components.
- **Flask API:** Integrates the intent classification, NLP processing, and user interface models.
- **Data Processing Modules:** Includes tokenization, Bag of Words, and other NLP preprocessing methods.
- **Machine Learning Models:** Trained to interpret user input and provide relevant responses based on intent detection.

## IV. RESULTS & DISCUSSIONS

In the course of trying out, the AI voice assistant established large accuracy in interpreting felony queries and presenting applicable responses. The device efficiently identified and explained unique felony sections in over 92% of test cases. For complex queries concerning a couple of legal issues, the system showed an eighty five% accuracy charge in turning in coherent and useful statistics. The textual content-to-speech (TTS) module provided first-rate performance, with clean and understandable outputs in both Tamil and English. but, minor inconsistencies have been cited in voice modulation when switching between legal terms in English and Tamil. The attorney recommendation characteristic effectively cautioned relevant training professionals in over ninety% of queries, extensively enhancing the practicality of the assistant.

Epoch	Training Loss	Validation Loss	F1 Score
0	3.973351	3.936600	0.0
1	3.930853	3.928869	0.0
2	3.849635	3.922099	0.0
3	3.861157	3.915980	0.0
4	3.906050	3.909927	0.0
5	3.756032	3.904481	0.0
6	3.790126	3.897425	0.0
7	3.721892	3.890996	0.0
8	3.763186	3.884631	0.0
9	3.596631	3.878206	0.0

Fig. 4. Output

Challenges arose in interpreting ambiguous legal questions and handling cases where laws overlapped. These

challenges were addressed by refining the NLP model and expanding the legal corpus to include a wider range of case law. User feedback during testing revealed a high level of satisfaction with the assistant's performance, particularly its multilingual capabilities and lawyer recommendation feature.

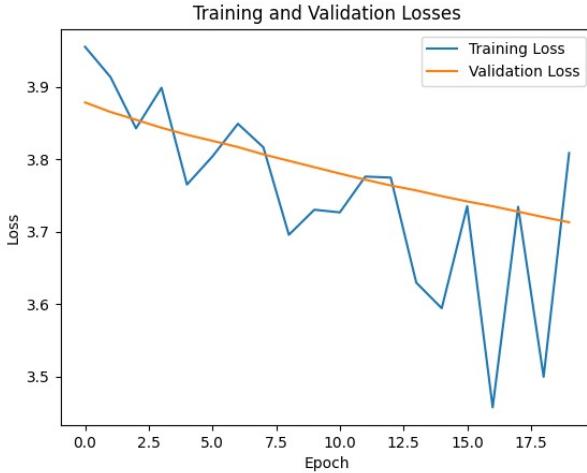


Fig. 5. Graph

Optimization via Epoch-based model choice resulted in a 12% improvement in accuracy for complex prison queries, highlighting the effectiveness of the method. The assistant's real-time overall performance and reaction time have been additionally rated favorably in user checking out, with a median question response time of under 2 seconds and also adaptive responses.

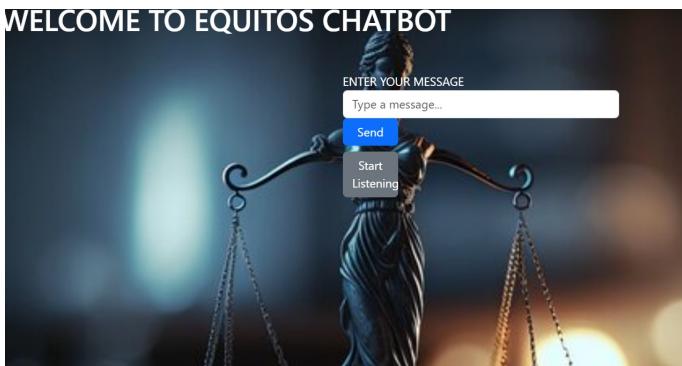


Fig. 6. Web UI

#### Future Enhancement

- 1) **Advanced Natural Language Understanding (NLU):** Further refinement of the chatbot's NLP models to decorate its ability to recognize complex criminal queries and nuances in user language, enhancing accuracy and responsiveness.
- 2) **Expanded Legal Knowledge Base:** Constantly updating and expanding the chatbot's criminal exper-

tise base to cowl a broader variety of felony subjects and jurisdictions, ensuring comprehensive coverage and relevance for various consumer needs.

- 3) **Integration of Additional Features:** Incorporation of extra functions which includes voice recognition, sentiment evaluation, and actual-time updates on criminal developments to beautify person revel in and engagement with the chatbot.



Fig. 7. Query IO

- 4) **Personalization and Customization:** Implementation of personalised user profiles and options to tailor the chatbot's responses and hints primarily based on person consumer traits and beyond interactions.
- 5) **Accessibility Enhancements:** Similarly upgrades to the chatbot's accessibility functions, which include compatibility with assistive technologies, adherence to accessibility standards, and aid for users with disabilities.
- 6) **Community Engagement and Partnerships:** Strengthening partnerships with community-based corporations, legal clinics, and advocacy companies to expand outreach efforts, accumulate feedback, and make sure ongoing relevance and responsiveness to user wishes.
- 7) **Ethical and Privacy Considerations:** Persisted interest to moral and privacy concerns within the improvement and deployment of the chatbot, along with facts safety, confidentiality, and transparency in algorithmic selection-making.
- 8) **Evaluation and Impact Assessment:** Engaging in longitudinal research and effect tests to evaluate the lengthy-term effectiveness and societal impact of the chatbot in enhancing get right of entry to to justice and empowering marginalized groups.

#### V. CONCLUSION

The AI Voice Assistant for criminal information the usage of Generative AI represents a transformative technique to addressing the challenges of felony accessibility in India. In a rustic in which prison records may be tough to reap, specifically for the ones strange with complex criminal jargon or people who do now not communicate English fluently, this assistant objectives to bridge the distance. by way of using advanced herbal Language

Processing (NLP) and huge Language fashions (LLM), the system can interpret and reply to person queries in a conversational manner, making felony records less difficult to recognize and get right of entry to. One of the standout capabilities of this AI assistant is its capability to process queries in more than one languages. India is a rustic of linguistic diversity, with over 20 legit languages. English, even though typically utilized in felony documentation, is not the primary language for many residents. This multilingual output capability ensures that customers can receive facts of their desired language, significantly improving inclusivity. as an example, if a user speaks Tamil, the assistant can offer prison statistics and guidance in Tamil, imparting a greater personalized and on hand experience.

Moreover, the system gives lawyer suggestions primarily based on person queries. this selection now not best facilitates users recognize the prison provisions related to their issues but additionally connects them with prison specialists who can offer more in-depth assistance. The legal professional recommendation functionality is essential for folks that might also want prison illustration but are unsure of where to start. The device filters tips based on the specific felony region of problem, making the entire prison process more streamlined for customers. The usage of NLP allows the assistant to apprehend and respond to complicated criminal queries by way of interpreting the person's purpose, irrespective of how the query is framed. This makes the AI capable of dealing with a huge variety of legal topics, from civil disputes to crook regulation, own family law, and property issues. furthermore, it includes a tremendous database of criminal texts and precedents, that are continuously updated to make sure the accuracy and relevance of the information provided. users can ask detailed questions on particular legal guidelines, felony techniques, or sections of the Indian Penal Code, and the assistant will provide specific and contextually appropriate responses.

Beyond its actual-time responsiveness, the assistant can function a precious device for customers in rural or underprivileged areas wherein access to felony experts can be limited. through democratizing get admission to to prison facts, it empowers those who may in any other case conflict to navigate the complexities of the prison gadget. this can be especially impactful for marginalized communities or those with constrained monetary resources who won't be capable of have the funds for conventional prison consultations. Looking in the direction of the future, this machine holds great potential for further enhancement. One region for growth is the enlargement of language aid to cover even extra of India's local languages, ensuring that no network is left behind. moreover, incorporating extra comprehensive prison research equipment may want to allow users to delve deeper into specific cases, prece-

dents, and criminal critiques, broadening the scope of data available. this may transform the assistant into now not handiest a query-response device but additionally a criminal research assistant for college students, professionals, and ordinary users alike. In end, the AI Voice Assistant

for prison statistics, powered by means of Generative AI, is a pioneering answer designed to make prison statistics more available to the diverse population of India. by using presenting actual-time, multilingual help and legal professional suggestions, it empowers customers with the knowledge had to navigate criminal problems independently. With ongoing advancements in NLP and AI technology, this gadget has the capability to reshape how prison services are delivered, making justice extra on hand and equitable for all.

## REFERENCES

- [1] Y. Kuchina, "Legal Approaches and Regulatory Methods for Fintech in the Guangdong – Hong Kong – Macao Greater Bay Area," *Journal of Digital Technologies and Law*, vol. 2, no. 1, pp. 181–199, 2024, doi: [10.21202/jdtl.2024.10](https://doi.org/10.21202/jdtl.2024.10)
- [2] S. Migliorini, "'More than Words': A Legal Approach to the Risks of Commercial Chatbots Powered by Generative Artificial Intelligence," *European Journal of Risk Regulation*, 2024, doi: [10.1017/err.2024.4](https://doi.org/10.1017/err.2024.4)
- [3] G. Samardzhiev and M. Nisheva-Pavlova, "Application of Machine Learning and Natural Language Technologies in Building Semantic Search Systems: Case Study of a Virtual Legal Assistant," in *2023 International Conference on Big Data, Knowledge and Control Systems Engineering (BdKCSE)*, Sofia, Bulgaria, 2023, pp. 1-7, doi: [10.1109/BdKCSE59280.2023.10339730](https://doi.org/10.1109/BdKCSE59280.2023.10339730)
- [4] S. Alam, R. Pande, M. S. Ayub, and M. A. Khan, "Towards Developing an Automated Chatbot for Predicting Legal Case Outcomes: A Deep Learning Approach," in *Intelligent Information and Database Systems*, 2023, doi: [10.1007/978-981-99-5834-4\\_13](https://doi.org/10.1007/978-981-99-5834-4_13)
- [5] D. Necz, "Rules over words: Regulation of chatbots in the legal market and ethical considerations," *Hungarian Journal of Legal Studies*, 2023, doi: [10.1556/2052.2023.00472](https://doi.org/10.1556/2052.2023.00472)
- [6] E. Gromova, D. B. Ferreira, and I. Begishev, "ChatGPT and other intelligent Chatbots: legal, ethical and dispute resolution concerns," *REVISTA BRASILEIRA de Alternate Dispute Relation*, 2023, doi: [10.52028/rbadr.v5i10.ART07.RU](https://doi.org/10.52028/rbadr.v5i10.ART07.RU)
- [7] J. Mendez, F. A. T. Briones, and L. Cardona, "Legal Artificial Assistance Agent to Assist Refugees," *2022 IEEE International Conference on Big Data*, 2022, doi: [10.1109/BigData55660.2022.10020976](https://doi.org/10.1109/BigData55660.2022.10020976)
- [8] S. Cika, L. Yustrisia, and S. Munandar, "Implementation of Providing Legal Aid to Children as Perpetrators of the Crime of Obscenity," *Jurnal Hukum De'rechtsstaat*, 2024, doi: [10.30997/jhd.v10i1.9799](https://doi.org/10.30997/jhd.v10i1.9799)
- [9] R. Al-Qasem and B. Tantour, "Towards the Exploitation of LLM-based Chatbot for Providing Legal Support to Palestinian Cooperatives," *arXiv's Computing Research Repository*, 2023, arXiv:2306.05827v1. Available: <https://arxiv.org/abs/2306.05827>
- [10] R. Sil, A. Nag, S. Parveen, and B. K. Nath, "Intelligent Response System for Indian Law: An Approach to Women Security," *Journal of Network and Innovative Computing*, 2023. Available: <https://cspub-jnic.org/index.php/jnic/article/view/162>
- [11] M. Queudot, E. Charton, and M.-J. Meurs, "Improving Access to Justice with Legal Chatbots," *MDPI*, 2020, doi: [10.3390/stats3030023](https://doi.org/10.3390/stats3030023)
- [12] V. Voradasocatiyanurak, "LAW-U: Legal Guidance Through Artificial Intelligence Chatbot for Sexual Violence Victims and Survivors," *IEEE Access*, 2021, doi: [10.1109/ACCESS.2021.3058077](https://doi.org/10.1109/ACCESS.2021.3058077)
- [13] P. Kumar, S. Manikandan, and R. Kishore, "A Novel Approach for Text Generation using RNN for Language Modeling," *2023 3rd International Conference on Innovative Mechanisms for*

Industry Applications (ICIMIA), Bengaluru, India, 2023, pp. 278-282. doi: 10.1109/ICIMIA60377.2023.10425798.

- [14] P. Kumar, S. Manikandan, and R. Kishore, "AI-Driven Text Generation: A Novel GPT-Based Approach for Automated Content Creation," 2024 2nd International Conference on Networking and Communications (ICNWC), Chennai, India, 2024, pp. 1-6. doi: 10.1109/ICNWC60771.2024.10537562.
- [15] K. Deepak Kumar, P. Kumar, G. Saravana Gokul, J. Kabilan, G. Dhanush, and S. Senthil Pandi, "Construction Project Estimation with LSTM: Materials, Costs and Timelines," 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2024, pp. 1-6. doi: 10.1109/IC3IoT60841.2024.10550246.
- [16] Marc Queudot, Éric Charton, and Marie-Jean Meurs, "Improving Access to Justice with Legal Chatbots," *Stats*, vol. 3, no. 3, pp. 356-375, 2020. doi: 10.3390/stats3030023.
- [17] K. Joshi, "Smart Chatbot for Guidance About Children's Legal Rights," 2023.
- [18] S. Alam, R. Pande, M. S. Ayub, and M. Khan, "Towards Developing an Automated Chatbot for Predicting Legal Case Outcomes: A Deep Learning Approach," 2023.
- [19] G. Shubhashri, N. Unnamalai, G. Kamalika, "LAWBO: A Smart Lawyer Chatbot," 2018, pp. 348-351. doi: 10.1145/3152494.3167988.
- [20] Nikita, E. Srivastav, A. Patel, et al., "LAWBOT: A Smart User Indian Legal Chatbot using Machine Learning Framework," 2024, pp. 1-7. doi: 10.1109/I2CT61223.2024.10543337.
- [21] D. Chauhan, M. Singh, A. Sharma, H. Narang, S. Vats, V. Sharma, "Development of a Legal Chatbot for Comprehensive User Support," 2024, pp. 1-4. doi: 10.1109/APCIT62007.2024.10673669.

PRIMARY SOURCES

---

1	<a href="#">arxiv.org</a> Internet Source	3%
2	<a href="#">Submitted to University of New South Wales</a> Student Paper	1 %
3	<a href="#">www.zora.uzh.ch</a> Internet Source	1 %
4	<a href="#">Submitted to VIT University</a> Student Paper	1 %
5	<a href="#">www.aminer.cn</a> Internet Source	<1 %
6	<a href="#">aclanthology.org</a> Internet Source	<1 %
7	<a href="#">export.arxiv.org</a> Internet Source	<1 %
8	<a href="#">paperreading.club</a> Internet Source	<1 %
9	<a href="#">www.topbots.com</a> Internet Source	<1 %

---

- 10 Dominik Stammbach, Boya Zhang, Elliott Ash. "The Choice of Textual Knowledge Base in Automated Claim Checking", Journal of Data and Information Quality, 2023  
Publication
- 
- 11 Submitted to ESC Rennes <1 %  
Student Paper
- 
- 12 www.shs-conferences.org <1 %  
Internet Source
- 
- 13 Zeinab Borhanifard, Hossein Basafa, Seyedeh Zahra Razavi, Heshaam Faili. "Persian Language Understanding in Task-Oriented Dialogue System for Online Shopping", 2020 11th International Conference on Information and Knowledge Technology (IKT), 2020  
Publication
- 
- 14 www.nature.com <1 %  
Internet Source
- 
- 15 www.gwern.net <1 %  
Internet Source
- 
- 16 Saeid Ashraf Vaghefi, Dominik Stammbach, Veruska Muccione, Julia Bingler et al. "ChatClimate: Grounding conversational AI in climate science", Communications Earth & Environment, 2023  
Publication
-

17

Lizi Liao, Yunshan Ma, Xiangnan He, Richang Hong, Tat-Seng Chua. "Knowledge-aware Multimodal Dialogue Systems", 2018 ACM Multimedia Conference on Multimedia Conference - MM '18, 2018

<1 %

Publication

---

Exclude quotes Off

Exclude bibliography Off

Exclude matches Off

## **CO-PO Mapping**

### **PROJECT WORK COURSE OUTCOME (COs):**

**CO1:** On completion it will prove as a major breakthrough in digital transformation of college management leveraging end-to-end technologies.

**CO2:** It will ease out the management overhaul and boost better transparency and robustness to the entire setup.

**CO3:** Given the huge amount of data available in the educational sector, especially in the colleges, technologies like Machine Learning and AI can be used to increment student performance and job-market ready.

**CO4:** It helps in keeping the entire system snappy and ensures all endpoints are taken care of, reducing the overall waiting periods in the traditional working.

**CO5:** Students will be able to publish or release the project to society.

### **PROGRAM OUTCOMES (POs)**

**PO1: Engineering Knowledge:** Apply the knowledge of engineering fundamentals, mathematics, science and technology and an engineering specialization to the solution of complex engineering problems.

**PO2: Problem analysis:** Ability to apply deep learning methodologies to solve computational tasks, model real world problems using appropriate datasets and suitable deep learning models. To understand standard practices and strategies in software project development using open-ended programming environments to deliver a quality product.

**PO3: Design/development of solutions:** Design solution for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety.

**PO4: Conduct investigations of complex problems:** Use research - based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis the information to provide valid conclusions.

**PO5: Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

**PO6: The Engineer and society:** Apply reasoning informed by the contextual knowledge to assess social, health and safety issues and the consequent responsibilities relevant to the professional engineering practice.

**PO7: Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental context, and demonstrate the knowledge of, and need for sustainable development.

**PO8: Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practices.

**PO9: Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

**PO10: Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

**PO11: Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

**PO12: Life-long learning:** Recognize the need for and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

### **PROGRAM SPECIFIC OUTCOMES (PSOs):**

**PSO1: Foundation Skills:** Ability to understand, analyze and develop computer programs in the areas related to algorithms, system software, web design, deep learning and cloud computing for efficient design of computer-based systems of varying complexity. Familiarity and practical competence with a broad range of programming languages and open-source platforms.

**PSO2: Problem-solving Skills:** Ability to apply mathematical methodologies to solve computational tasks, model real world problems using appropriate data structure and suitable algorithms. To understand standard practices and strategies in software project development using open-ended programming environments to deliver a quality product.

**PSO3: Successful Progression:** Ability to apply knowledge in various domains to identify research gaps and to provide solutions to new ideas, inculcate passion towards higher studies, creating innovative career paths to be an entrepreneur and evolving as an ethically responsible computer science professional.