### **Chatbot for DoJ website**

#### A PROJECT REPORT

Submitted by,

Mr. Abi Roshan - 20211CAI0033 Mr. Pranav Pradeep - 20211CAI0073 Ms. Huzaifa Fathima - 20211CAI0178

Under the guidance of,
Ms. Josephine R.

in partial fulfillment for the award of the degree of BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING
ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

At



# PRESIDENCY UNIVERSITY BENGALURU

MAY 2025
SCHOOL OF COMPUTER SCIENCE ENGINEERING

#### **CERTIFICATE**

This is to certify that the Project report "Chatbot for DoJ website" being submitted by Abi Roshan, Pranav Pradeep & Huzaifa Fathima bearing roll number(s) 20211CAI0033, 20211CAI0073 & 20211CAI0178 in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering Artificial Intelligence and Machine Learning is a bonafide work carried out under my supervision.

Ms. Josephine R.

**Assistant Professor** 

School of CSE

Presidency University

Dr. Zafar Ali Khan

Professor & HoD

School of CSE

**Presidency University** 

Dr. L. SHAKKEERA

Dr. MYDHILI NAIR

Dr. SAMEERUDDIN KHAN

Associate Dean

School of CSE

School of CSE

Associate Dean

Pro-Vc School of Engineering

Dean -School of CSE & IS

Presidency University P

Presidency University

Presidency University

# PRESIDENCY UNIVERSITY SCHOOL OF COMPUTER SCIENCE ENGINEERING

#### **DECLARATION**

We hereby declare that the work, which is being presented in the project report entitled CHATBOT FOR DOJ in partial fulfillment for the award of Degree of Bachelor of Technology in Computer Science and Engineering Artificial Intelligence and Machine Learning, is a record of our own investigations carried under the guidance of Josephine R, Assistant Professor, School of Computer Science Engineering, Presidency University, Bengaluru.

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

Name	Roll No.	Signature
Abi Roshan	20211CAI0033	
Pranav Pradeep	20211CAI0073	
Huzaifa Fathima	20211CAI0178	

#### **ABSTRACT**

This project presents a dynamic AI-powered chatbot system developed for the Department of Justice (DoJ) website. The chatbot aims to streamline user interactions and improve access to judicial services by offering real-time information on various aspects such as case statuses, judicial appointments, traffic fines, and eFiling procedures. Leveraging natural language processing (NLP), generative AI, and cloud deployment, the chatbot understands user queries, provides accurate responses, and learns over time to enhance interaction quality. This system is designed to be scalable, multilingual, and user-friendly, thereby making legal resources more accessible and understandable for citizens across India.

#### **ACKNOWLEDGEMENT**

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr. Shakkeera** L and Dr. Mydhili Nair, School of Computer Science & Engineering, Presidency University, and Dr. Zafar Ali Khan, Head of the Department, School of Computer Science & Engineering, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide. Ms. Josephine R, Assistant Professor and Reviewer Mr. John Bennet, Assistant Professor, School of Computer Science & Engineering, Presidency University for his/her inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators **Dr. Sampath A K, Dr. Abdul Khadar A and Mr. Md Zia Ur Rahman,** department Project Coordinators **Dr. Afroz Pasha** and Git hub coordinator **Mr. Muthuraj.** 

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Abi Roshan

**Pranav Pradeep** 

Huzaifa Fathima

#### TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO
	ABSTRACT ACKNOWLEDGMENT	i ii
1	INTRODUCTION	1
	<ul><li>1.1 Challenges in Accessing Legal Information</li><li>1.2 Need for an AI-based Chatbot</li><li>1.3 Objectives</li></ul>	1 1 1
2	LITERATURE REVIEW	2
3	RESEARCH GAPS	4
4	PROPOSED METHODOLOGY	5
5	SYSTEM DESIGN & IMPLEMENTATION	7
6	OUTCOMES	9
7	CONCLUSION	10
	REFERENCES	11
	APPENDIX-A PSEUDOCODE	12
	APPENDIX-B SCREENSHOT	14
	APPENDIX-C ENCLOSURES	15

#### **CHAPTER 1:**

#### INTRODUCTION

#### 1.1 Challenges in Accessing Legal Information

In India, access to justice remains a significant issue due to the complexity of legal terminology, lack of digital literacy, and fragmented sources of legal information. Government websites often lack user-centric designs, making it difficult for individuals—especially those from rural backgrounds—to understand or access judicial processes. Many citizens are unaware of their rights or how to avail government legal services, resulting in underutilization of crucial facilities.

#### 1.2 Need for an AI-based Chatbot

To overcome these hurdles, an interactive chatbot powered by AI technologies offers a transformative solution. A chatbot can interpret user queries in natural language, guide users to relevant legal resources, and perform functions such as tracking case status or providing eFiling instructions. The chatbot simplifies the user interface by providing step-by-step conversational assistance and makes complex legal processes comprehensible for the general public.

#### 1.3 Objectives

- Develop a fully functional AI-based chatbot integrated into the DoJ website.
- Address diverse user needs by providing support across judicial divisions, appointments, and digital court procedures.
- Reduce digital and linguistic barriers through multilingual support and simple conversational design.
- Ensure robust security and confidentiality while handling legal and personal data.
- Leverage continuous learning mechanisms to enhance chatbot performance over time.

#### **CHAPTER 2:**

#### LITERATURE REVIEW

The integration of artificial intelligence (AI) into chatbot systems has evolved considerably, particularly in sectors such as customer service, education, and healthcare. AI-powered chatbots are capable of simulating human conversation using natural language processing (NLP) and machine learning algorithms. The key benefit of chatbots lies in their availability, scalability, and the ability to provide fast, consistent, and contextual responses. However, the adaptation of such systems to the legal domain is still an emerging field.

One of the widely used frameworks in chatbot development is TensorFlow, which provides the infrastructure for training deep learning models that can understand context and semantics in user input. Additionally, GPT-based models (developed by OpenAI) have shown remarkable performance in generating human-like responses in natural conversations. These models use transformer architecture to process and predict the next word or phrase based on prior context, making them suitable for dialogue-based applications.

RASA, an open-source machine learning framework, is another prominent tool used for building contextual chatbots. It supports intent classification, entity extraction, and dialogue management, which are crucial in building reliable chatbot systems for complex domains like legal or governmental sectors. RASA also allows easy integration with APIs and external databases, making it viable for our project.

Microsoft Bot Framework and Google Dialogflow are also extensively used tools for chatbot development. They offer features such as multilingual support, voice interface compatibility, and AI model training through graphical interfaces. However, these tools sometimes present limitations in terms of customizability and fine-grained control over the NLP pipeline.

Several research papers highlight the deployment of chatbots in legal and judicial environments. For instance, some systems are designed to assist users in understanding legal terminologies or guiding them through document submission processes. These systems often rely on static datasets and are limited to predefined interactions, which hinders their effectiveness in dynamic real-world applications. Moreover, many legal chatbots lack adaptability, resulting in subpar performance when exposed to unfamiliar queries or case-specific language.

Another major limitation is the absence of real-time data integration. Legal services often require up-to-date information, such as hearing dates, case statuses, and judicial appointments. Chatbots built without access to live databases fail to deliver timely and relevant assistance to users. Therefore, integrating live sources like the National Judicial Data Grid (NJDG) into a chatbot system is essential for maintaining credibility and usability.

Language support is another critical area where existing systems fall short. In a diverse country like India, it is important for legal chatbots to support multiple regional languages to ensure equitable access to justice. Although some frameworks provide rudimentary multilingual

support, they often fail to capture the contextual meaning behind regional phrases or cultural expressions.

In summary, while chatbot technologies have matured in commercial domains, their implementation in legal services remains limited. Most systems lack domain-specific training, real-time database connectivity, and multilingual capabilities. The proposed chatbot for the Department of Justice addresses these gaps by incorporating robust NLP, real-time integration, multilingual support, and a feedback loop that enables continuous learning and enhancement.

Most reviewed systems lack real-time data access, limiting their utility in time-sensitive applications like legal support. This project proposes a novel integration of live judicial data sources like NJDG into the chatbot's backend.

#### **CHAPTER 3:**

#### RESEARCH GAPS

Despite the remarkable advancements in AI and chatbot development, several specific limitations still hinder the effectiveness and scalability of these technologies in the legal domain. The following research gaps have been identified as major barriers in the development and deployment of intelligent legal assistants:

- Lack of Domain-Specific Chatbots: Most existing chatbot frameworks are developed for general customer support tasks and lack the specialization required for legal services. Legal language involves domain-specific terminology, complex procedural logic, and hierarchical structures that general-purpose bots cannot easily comprehend. This limits their ability to respond meaningfully to queries about court procedures, case law, or legal statutes.
- Absence of Real-Time Data Syncing: Chatbots that operate on static data repositories
  cannot provide accurate, up-to-date responses about ongoing legal cases, judicial
  vacancies, or changes in filing protocols. The lack of integration with live databases
  like the National Judicial Data Grid (NJDG) severely reduces the utility of such systems
  in real-world legal assistance scenarios where data is frequently updated.
- Limited Learning Capabilities: Traditional rule-based or static bots operate on a fixed set of responses and cannot adapt to new user queries or evolving legal terminology. Without reinforcement learning or user feedback loops, these bots fail to evolve and improve over time, resulting in repetitive and sometimes irrelevant responses. This stagnation diminishes user trust and engagement.
- Inadequate Multilingual Support: India is a multilingual nation with users accessing legal services in various regional languages. However, most bots are built with English as the primary language, and even when regional language support is added, it is often superficial. The inability to capture cultural and linguistic nuances in vernacular languages leads to communication gaps and poor user experience for non-English speakers.
- Security Concerns: Chatbots dealing with legal data must maintain high standards of
  privacy and security. Many current implementations overlook the importance of secure
  data transmission, encryption, and compliance with government data protection
  guidelines. Given the sensitive nature of legal queries and personal information shared
  through these systems, robust security protocols are a necessity that is often lacking in
  existing frameworks.

Addressing these gaps is crucial to building a chatbot system that is reliable, inclusive, adaptable, and legally compliant. Our project directly targets these challenges through real-time data integration, continuous learning models, multilingual interfaces, and privacy-first design architecture.

#### **CHAPTER 4:**

#### PROPOSED METHODOLOGY

The proposed methodology outlines a comprehensive pipeline for developing an AI-based chatbot system capable of delivering real-time, accurate, and user-friendly assistance on legal matters. The methodology is broken down into modular phases to ensure seamless functionality and scalability.

#### 1. User Input Acquisition

Users interact with the chatbot through either text-based or voice-based input on the Department of Justice website. The chatbot interface is embedded on the homepage to facilitate ease of access, and it supports both English and regional languages to accommodate users across different regions of India.

#### 2. Preprocessing and Language Normalization

Once the user submits a query, the system performs natural language preprocessing. This involves removing punctuation, correcting common spelling errors, lowercasing all text, and eliminating stop words. For voice inputs, speech-to-text conversion is handled using libraries such as Google Speech Recognition API. The preprocessing module ensures that the query is clean and standardized before analysis.

#### 3. Intent Recognition

Using pretrained models such as BERT or RASA NLU, the chatbot classifies the user's query into specific intent categories such as "Case Status", "Judicial Appointment Info", "eFiling Process", "Fast-track Court Details", or "Traffic Fine Payment Instructions." Intent classification is central to identifying the purpose of the user's message, allowing the bot to respond appropriately.

#### 4. Entity Extraction

The chatbot then extracts specific keywords and phrases (known as entities) from the query. These entities could be case numbers, district names, court names, dates, or keywords such as "appointment" or "pending." Named Entity Recognition (NER) is used to dynamically identify and tag important details that help narrow down the context for data retrieval.

#### 5. Query Routing to Pre-trained Knowledge Sources

Once the intent and relevant entities are recognized, the system routes the query to the appropriate data source. For example, a question about "case status" is routed to the NJDG API or DoJ's internal database for real-time information. The system has access to pre-trained datasets and connected portals, allowing it to fetch data in response to specific user queries. This modular access ensures that only the most relevant information is retrieved.

#### 6. Response Generation

The chatbot utilizes generative AI models or response templates to generate a human-like, informative answer. Depending on the query, the response could include:

- Direct answers with case status or appointment schedules.
- Step-by-step guidance for legal processes (like eFiling).
- Hyperlinks or attachments (PDFs or external resources).
- Suggestions for further reading or action.

The responses are tailored to be concise, accurate, and understandable for users with no legal background.

#### 7. Output Display & Multi-modal Support

The generated response is displayed to the user via the interface. For voice-enabled users, the answer is read aloud using text-to-speech (TTS) technology. The chatbot also supports multimedia such as clickable buttons and embedded links to allow users to explore related content without typing further queries.

#### 8. Feedback Capture & Continuous Learning

At the end of each interaction, users are prompted to provide feedback via thumbs-up/down buttons or a short rating. This data is stored and used to retrain the model periodically, allowing it to improve its accuracy, adapt to new user intents, and better serve public expectations.

This methodological pipeline ensures the chatbot operates with precision, efficiency, and user-centric design, providing valuable assistance in a legally sensitive environment.

#### **CHAPTER 5:**

#### SYSTEM DESIGN & IMPLEMENTATION

The chatbot system is structured into modular layers for better maintenance, scalability, and performance. This section provides a detailed breakdown of how the system was architected and implemented.

#### **5.1** Architecture Overview

The architecture of the chatbot consists of the following layers:

- **Presentation Layer (Frontend)**: Built with HTML, CSS, and JavaScript, this layer handles all interactions with the user. It supports a text box for input, a voice assistant icon for speech input, and an animated avatar to create a dynamic experience.
- Logic Layer (Middleware): Python-based backend logic (using Flask/Django) connects the interface with AI services. It controls query flow, tracks session history, and manages logic trees based on user interactions.
- AI Layer (NLP and ML Engine): This is the brain of the system. It processes input using pre-trained models for intent classification, entity recognition, and text generation. TensorFlow Lite and NLTK are used for lightweight local processing.
- Data Access Layer: This connects with external APIs like NJDG, DoJ public services, and internal knowledge bases. All responses and data fetch operations are executed securely using REST APIs.

#### 5.2 Tools & Technologies

- **Python**: Core language for backend scripting and ML integration.
- TensorFlow Lite & NLTK: Used for machine learning and NLP.
- Spacy & RASA: Intent detection and contextual understanding.
- SpeechRecognition API & gTTS: For voice-based input/output processing.
- **AWS / Google Cloud Platform**: Cloud hosting to ensure uptime, scalability, and data backup.
- **GitHub**: For version control, collaboration, and code maintenance.

#### **5.3 Implementation Workflow**

1. **User Interaction Interface**: Designed using responsive web development techniques to support both desktop and mobile browsers. Special attention was paid to accessibility.

- 2. **Input Handling**: Accepts and converts voice/text input into standard NLP-friendly format.
- 3. **Query Routing Module**: Forwards preprocessed queries to relevant NLP modules or APIs.
- 4. **Model Inference Engine**: Interprets the intent and fetches data from the correct database or pre-trained model.
- 5. **Response Generator**: Generates user-readable answers based on templates, natural generation, or database entries.
- 6. **Feedback Logger**: Logs user satisfaction and response accuracy to improve future performance.

#### **5.4 Key Features**

- **Multilingual Support**: Built-in translation APIs allow real-time support for regional languages.
- Security & Compliance: Secure endpoints using HTTPS and token authentication. Sensitive data is encrypted during transit.
- Scalability: Modular structure enables new services (like appointment booking or document upload) to be added without major rework.
- **Custom Knowledge Base**: Designed to be updated regularly by DoJ staff for evolving services and FAQs.

Together, the system's design ensures real-time responsiveness, adaptability to new topics, and high usability for a wide audience accessing justice-related services.

#### **CHAPTER 6:**

#### **OUTCOMES**

The successful implementation of this chatbot is expected to deliver multiple social and technical benefits:

- **Broadened Access to Legal Services**: Citizens can access accurate legal information without needing legal representation.
- **Minimized Bottlenecks on Government Sites**: Reduces the pressure on support staff and helpdesks.
- **User-Centric Experience**: By mimicking natural conversation, the system creates a welcoming experience, particularly for first-time users.
- **Continuous Improvement**: The feedback mechanism enables the system to adapt to emerging user needs and update its knowledge base accordingly.
- **Inclusive Governance**: By supporting regional languages, the chatbot bridges the gap between legal institutions and linguistically diverse citizens.

#### **CHAPTER 7:**

#### **CONCLUSION**

This project highlights how artificial intelligence and natural language processing can revolutionize public access to legal systems in India. By integrating real-time judicial data with a responsive, user-friendly chatbot interface, the project offers an innovative solution to longstanding accessibility and communication challenges.

The developed system paves the way for future enhancements, including voice-based access, biometric security, and broader integration with national legal portals. The chatbot is a scalable, efficient, and inclusive tool that can be a game-changer in the digital transformation of legal services in India.

#### REFERENCES

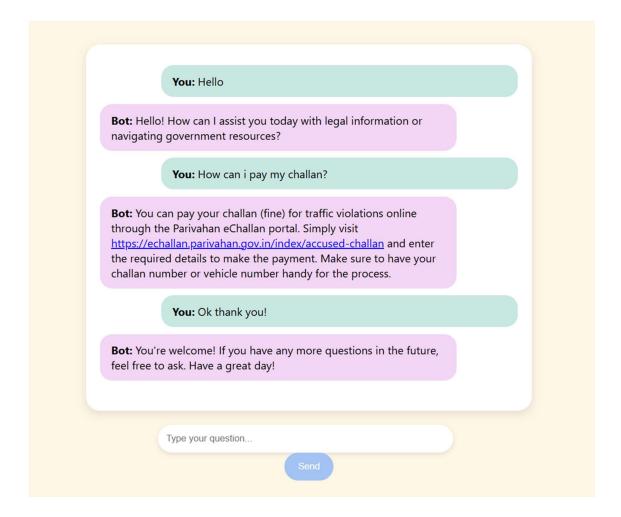
- [1] Amir Vakili and Azadeh Shakery. "Enriching Conversation Context in Retrieval-based Chatbots," *Proceedings of the Asian Federation of Natural Language Processing (AFNLP)*, 2019, pp. 1–8. arXiv preprint arXiv:1911.02290.
- [2] Pragaash Ponnusamy, Alireza Roshan Ghias, Chenlei Guo, and Ruhi Sarikaya. "Feedback-Based Self-Learning in Large-Scale Conversational AI Agents," *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020. arXiv preprint arXiv:1911.02557.
- [3] Chokri Kooli. "Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions," *Sustainability*, vol. 15, no. 7, 2023, article 5614. DOI: 10.3390/su15075614
- [5] Sophia Keyner, Vadim Savenkov, and Svitlana Vakulenko. "Open Data Chatbot," arXiv preprint, arXiv:1909.03653, 2019.
- [6] Lasha Labadze, Maya Grigolia, and Lela Machaidze. "Role of AI Chatbots in Education: Systematic Literature Review," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 56, 2023. DOI: 10.1186/s41239-023-00426-1
- [7] Aishwarya Gupta, Divya Hathwar, and Anupama Vijayakumar. "Introduction to AI Chatbots," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 7, July 2020. ISSN: 2278-0181.
- [8] Rohit Tamrakar and Niraj Wani. "Design and Development of CHATBOT: A Review," *Conference Paper*, Sardar Vallabhbhai National Institute of Technology, April 2021. Available via ResearchGate.
- [9] Vajinepalli Sai Harsha Vardhan, Parsi Anurag, and Richa Sharma. "Research Paper on Rule Based Chatbot," *International Research Journal of Modernization in Engineering, Technology and Science*, vol. 4, no. 5, May 2022, pp. 254–255. ISSN: 2582-5208.
- [10] Chiara Valentina Misischia, Flora Poecze, and Christine Strauss. "Chatbots in Customer Service: Their Relevance and Impact on Service Quality," *Procedia Computer Science*, vol. 201, pp. 421–428, 2022. DOI: 10.1016/j.procs.2022.03.055

#### APPENDIX -A

#### **PSEUDOCODE**

```
app.py >  chatbot_response
        from flask import Flask, request, jsonify
        from flask_cors import CORS
       from chatbot import chat # import your function from chatbot.py
       app = Flask( name )
       CORS(app) # Enables cross-origin requests for frontend
        Comment Code
       @app.route('/chat', methods=['POST'])
        def chatbot response():
           data = request.get_json()
            user_message = data.get("message", "")
11
12
            response = chat(user message)
            return jsonify({"response": response})
       if __name__ == "__main__":
           app.run(debug=True)
17 +
```

# APPENDIX – B SCREENSHOT



### APPENDIX-C ENCLOSURES

1. Journal publication/Conference Paper Presented Certificates of all students.









### CERTIFICATE

OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

#### PRANAV PRADEEP

Student, Department of Computer Science and Engineering, Presidency University, Bengaluru, India

in Recognition of Publication of the Paper Entitled

"Chatbot for DoJ Website"

in IJIRCCE, Volume 13, Issue 5, May 2025









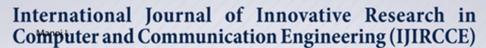


e-ISSN: 2320-9801 p-ISSN: 2320-9798





⊕ www.ijircce.com 
□ ijircce@gmail.com



(A Monthly, Peer Reviewed, Refereed, Multidisciplinary, Scholarly Indexed, High Impact Factor, Open Access Journal since 2013)





### CERTIFICATE

OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

#### **HUZAIFA FATHIMA**

Student, Department of Computer Science and Engineering, Presidency University, Bengaluru, India

in Recognition of Publication of the Paper Entitled

"Chatbot for DoJ Website"

in IJIRCCE, Volume 13, Issue 5, May 2025









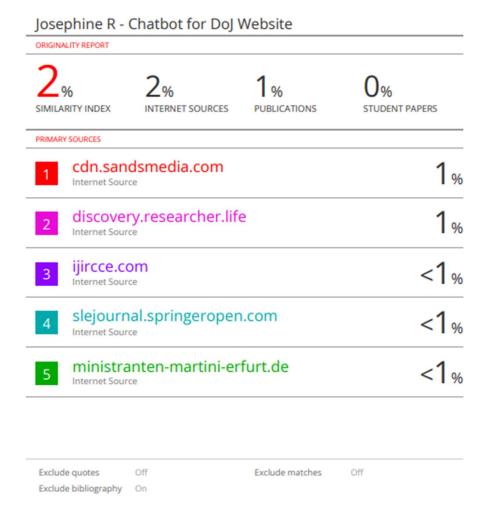


e-ISSN: 2320-9801 p-ISSN: 2320-9798

Editor-in-Chief

⊕ www.ijircce.com ijircce@gmail.com

2. Similarity Index / Plagiarism Check report clearly showing the Percentage (%). No need for a page-wise explanation.



# 4. Details of mapping the project with the Sustainable Development Goals (SDGs).



#### **SDG 16: Peace, Justice and Strong Institutions**

- The project **promotes access to justice** by making legal information more accessible and understandable to the public.
- It **strengthens institutions** by digitizing services, improving transparency, and reducing the complexity of legal procedures.
- By using AI and multilingual support, it ensures **inclusivity and efficiency**, helping citizens navigate the justice system more easily.