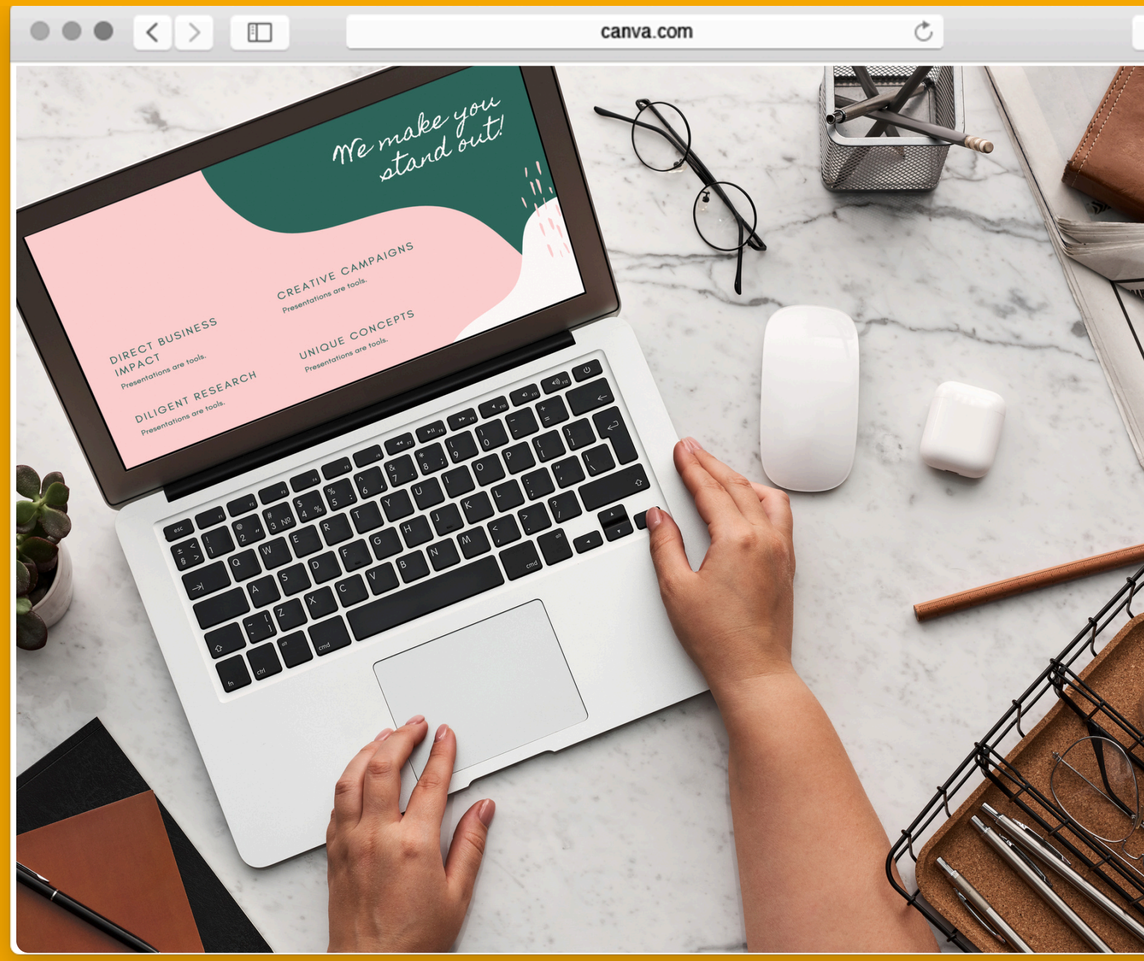
Statement B

Development of an artificial intelligence-based model for conversational use case chatbot in English and scheduled languages of the Constitution of India, 1950, to answer queries about case-related information, summarisation of judgments, court documents, etc.

VIVARAN



Our Progress Towards the Solution →

Introduction

Vivaran, a chatbot driven by advanced Artificial Intelligence technology. It is designed to empower users with comprehensive knowledge of Indian judicial laws, as established by the Indian Constitution and the Supreme Court. We are committed to ensuring that every Indian citizen has the right to access and understand the laws and court rulings that affect them. We believe that every Indian citizen should have easy access to legal information. To make this possible, Vivaran a conversational chatbot provides translations in scheduled languages, ensuring that everyone can understand the laws that impact them.

Introduction

- Building a chatbot for interaction in English and scheduled Indian languages.
- Answer questions about cases, judgments, and court documents.
- Summarize complex legal texts.
- Enable access to court documents summary through file upload in the chatbot.
- Provides accurate, timely information.
- Intuitive and user-friendly design.
- Simplifies complex documents.

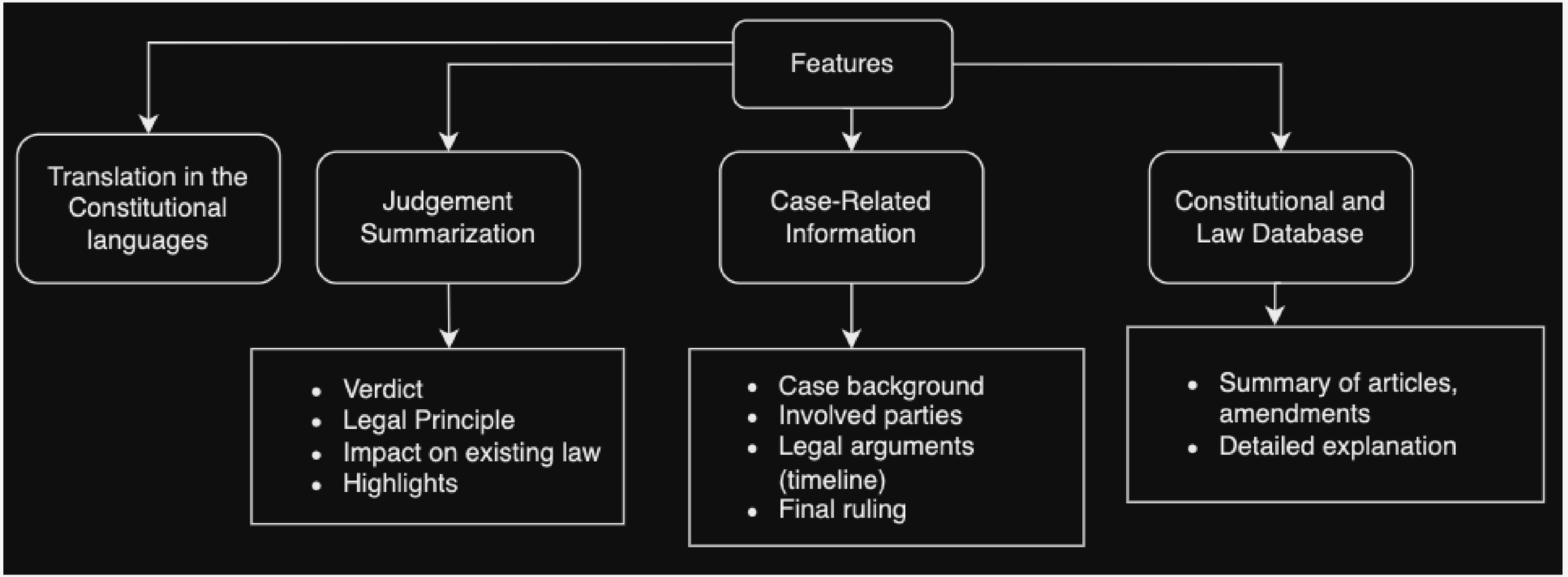
Challenges:

- Ensure up-to-date, correct information.
- Translate legal terms accurately.

Impact:

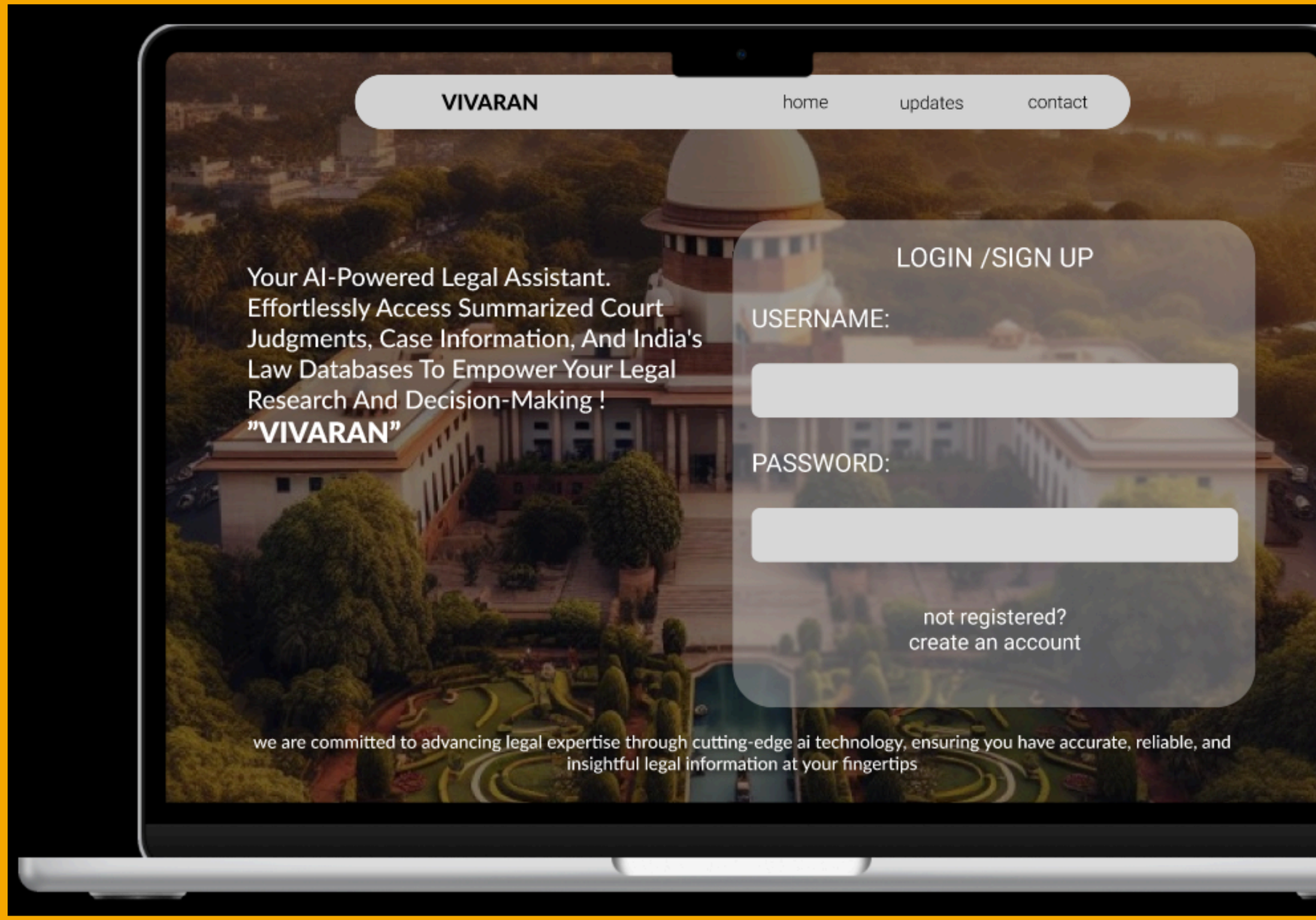
- Enhance access to legal information across languages.
- Provide a resource for understanding legal processes.

Main ChatBot Features



Our Solution

- Landing page: Home, Updates, Contact
- User authentication for secure access
- Logs user activity to protect confidential information - Audit trails and transparency



- Access: Chatbot page available post-login
- UI: Simple interface with essential component
- Bookmarking: Use the star icon to save important chats
- File Sharing: Upload files and images for summaries, highlights, and opinions
- Translation: Convert conversations into constitutional languages
- Voice Features: Input and receive responses via voice (translation support)

- Audience: Especially beneficial for students seeking educational information
- Access: Available without chatbot login
- Function: Provides updates on the latest judicial system changes and policies
- Benefit: Keeps users informed about recent developments without chatbot use

- Purpose: Allows users to reach out to developers
- Function: Share feedback and suggestions
- Access: Direct communication channel for users

User-Friendliness

- **Intuitive Interface:** The chatbot has a clean, user-friendly interface for easy navigation and use.
- **Multilingual Support:** Supports English and scheduled Indian languages for broad accessibility..
- **Interactive Design:** Conversational elements are designed to be engaging and straightforward.
- **Accessibility Features**: Includes options for voice input and output, allowing users to interact with the chatbot using voice commands, which enhances accessibility for those with visual impairments or preference for voice interaction.
- **Bookmarking Functionality:** Users can easily save and revisit important chats using a simple bookmarking feature, improving the efficiency of retrieving previously accessed information.
- **File and Image Upload:** Users can upload files and images to get summaries or opinions, with an easy-to-use upload interface that supports a seamless experience.
- **Real-Time Responses:** Provides instant replies to queries, ensuring that users receive timely and relevant information without delays.
- **Feedback Mechanism:** Includes options for users to provide feedback on their experience, helping to continually refine and enhance the user experience based on real user input.

Understanding the approach

To create an advanced legal chatbot capable of offering case-related information, judicial law summaries, judgment summaries, and access to a constitutional database, while also supporting multi-language translation and document analysis, a robust architecture integrating several AI technologies is essential.

A. Case-Related Information, Judicial Law Summaries, and Judgment Summarization

Data Retrieval and Summarization:

- Natural Language Processing (NLP) Models: Using transformer-based models like BERT, T5, or GPT-3.5 for natural language understanding and summarization. These models can be fine-tuned on Indian legal data to generate precise summaries and explanations of judicial laws, case details, and court judgments.
- Semantic Search: Implementing semantic search techniques using models like Sentence-BERT to provide relevant case-related information by understanding the context and meaning of user queries.

Understanding the approach

Constitutional Database Access:

- Structured Database: Building an indexed database containing the Indian Constitution's articles, amendments, and relevant annotations. The chatbot can query this database based on user input.
- GraphQL or SQL-Based API: Using APIs to enable dynamic querying of the database based on user requirements.

Understanding the approach

B. File and Image Input for Legal Document Analysis

Text Extraction from Files:

- Optical Character Recognition (OCR): Utilizing advanced OCR technology like Google Cloud Vision, Tesseract, or AWS Textract to convert scanned documents and images into machine-readable text. These OCR tools are capable of extracting text with high accuracy from various document formats and image qualities.
- Document Parsing Libraries: Using libraries like PyPDF2, pdfplumber, and python-docx to parse text from common legal document formats (PDFs, DOCX).

Highlight Extraction, Summarization, and Opinion Generation:

- Text Summarization Models: Using models like BART or Pegasus for both extractive and abstractive summarization to highlight key points and generate summaries of extracted text.
- Custom Fine-Tuning: Fine-tuning these models on Indian legal datasets to enhance the accuracy and relevance of summaries and opinions generated by the chatbot.

Understanding the approach

C. Multilingual Translation and Voice Support

Translation Across 22 Indian Languages:

- Multilingual NLP Models: Using multilingual models like mBERT or XLM-R that can handle multiple Indian languages. These models can be fine-tuned on Indian legal texts to improve translation accuracy and contextual understanding.
- Machine Translation APIs: Integrate with machine translation services like Google Cloud Translation API or Microsoft Translator API. These services support multiple Indian languages and offer high-quality translations, leveraging neural machine translation techniques.

Voice-to-Text and Text-to-Voice Capabilities:

- Speech Recognition: Using Google Cloud Speech-to-Text API or Microsoft Azure Speech Service for voice-to-text conversion. These services support numerous Indian languages and provide accurate transcriptions.
- Text-to-Speech (TTS): Using TTS services like Google Cloud Text-to-Speech or Amazon Polly to convert text responses into speech in the user's preferred language.

Optimized Solution

Backend and Data Processing:

- Transformers (Hugging Face): Use for implementing, fine-tuning, and deploying NLP models for tasks like summarization, translation, and semantic search on legal texts.
- Elasticsearch or Solr: Utilize for scalable, real-time semantic search, enabling efficient querying of legal databases.

AI and ML Services:

- Google Cloud AI Platform or AWS SageMaker: Employ cloud services for training, deploying, and scaling AI models, offering robust infrastructure for AI workflows.
- Google Cloud Translation and Speech APIs: Integrate for multilingual translation and voice support, covering extensive Indian language needs.

Frontend and User Interface:

- React.js or Angular: Build the web application’s frontend for a dynamic and responsive user interface, allowing smooth integration with backend services.
- WebSockets or REST APIs: Use for real-time communication or standard data exchange, ensuring instant responses and efficient user interaction handling.

Timeline for Development

1. Planning and Requirements Analysis (2 weeks):

- Define project scope and objectives.
- Gather requirements for NLP, data, translation, and speech.
- Choose tech stack (e.g., Hugging Face, Elasticsearch).
- Design architecture and project plan.

2. Data Collection and Preprocessing (3 weeks):

- Collect and preprocess legal texts and documents.
- Prepare multilingual datasets for NLP and translation models.

3. Model Development and Fine-Tuning (5 weeks):

- NLP Model Development (2 weeks): Fine-tune models for summarization and translation.
- Semantic Search Setup (1 week): Implement Elasticsearch for querying legal databases.
- Translation and Speech Integration (2 weeks): Integrate Google Cloud Translation and Speech APIs.

Timeline for Development

4. Backend Development (4 weeks):
   - API Development (2 weeks): Create APIs for data processing and communication.
   - Database Integration (1 week): Set up and connect to legal databases.
   - Testing and Optimization (1 week): Optimize backend services.
5. Frontend Development (3 weeks):
   - UI/UX Design (1 week): Design interfaces.
   - Implementation (2 weeks): Develop using React.js or Angular.
6. Integration and System Testing (3 weeks):
   - Integrate all components and perform end-to-end testing.
   - Conduct user acceptance testing (UAT).
7. Deployment and Launch (2 weeks):
   - Deploy on a cloud platform, monitor performance.
   - Conduct a soft launch and address issues before full release.

Previous Work Done in This Tech

LegalRobot

An AI-driven platform that simplifies understanding, drafting legal documents by providing clear explanations

AI Lawyer

It is a platform that streamlines legal tasks by providing instant legal research, document drafting, and review

Latch

It is an AI-driven platform that automates legal operations, case management, document drafting, time tracking, billing, and client communication.

Similar Works Executed in Government or Public Sector Enterprise

SUPACE

It is an AI tool that automates legal research and document summarization to aid judges in finding relevant cases and precedents, expediting the decision-making process.

COMPAS

It aid in risk assessment and sentencing by analyzing offender data, while US court chatbots provide public information on procedures and schedules, reducing staff workload and improving accessibility.

SUVAS

It is an AI-powered translation tool that converts judicial documents between English and nine regional languages.

Feasibility

- Accuracy of NLP Models: High accuracy (85-90%) is achievable with Hugging Face Transformers, BERT, and mBERT, contingent on fine-tuning with quality legal data and handling complex legal language.
- Semantic Search Precision: Technologies like Elasticsearch and Solr can achieve high accuracy, aiming to deliver relevant results within the top 3-5 responses with proper tuning and indexing.
- Translation and Speech Recognition Accuracy: Google Cloud APIs offer generally high accuracy, with 85-90% achievable through customization for complex legal terminology and less common dialects.
- Multilingual Support: Ensuring accurate translations and speech recognition across 22 Indian languages is feasible with expert customization and advanced APIs.
- NLP Expertise Needed: Requires data scientists and NLP engineers skilled in fine-tuning models for legal texts and multilingual support, and linguists with legal expertise for contextual accuracy.
- Backend and Frontend Development: Skilled backend developers for managing databases and APIs, and frontend developers/designers for creating intuitive user interfaces.
- Machine Learning and AI Infrastructure: ML engineers for managing AI workflows on cloud platforms, and speech technologists for accurate speech recognition in Indian languages.

Conclusion

This innovative legal chatbot leverages cutting-edge technology to provide comprehensive, user-friendly access to legal information and services. Key features include judgment summarization, case-related insights, constitutional law explanations, and multilingual support for over 22 Indian languages, ensuring inclusivity and accessibility. The chatbot is built using a robust tech stack: Hugging Face Transformers for NLP tasks, Elasticsearch for real-time semantic search, and Google Cloud APIs for translation and speech recognition, ensuring high accuracy and scalability. By integrating these advanced technologies, the solution aims to enhance user experience, provide reliable legal information, and support diverse linguistic needs, making it a valuable tool for both legal professionals and the general public.