Legal & Policy Explainer Bot

Project Overview

The Legal & Policy Explainer Bot is an AI-powered system designed to make Indian legal and policy documents easier to understand for students, researchers, and general users. Legal texts such as Acts, constitutional provisions, and policy documents are often lengthy, complex, and written in highly technical language. This project bridges that gap by converting raw legal PDFs into a searchable knowledge base and providing simplified explanations in natural language.

The system allows users to ask legal or policy-related questions in plain English and receive context-aware explanations derived directly from authoritative source documents.

Problem Statement

Access to legal information does not always translate into understanding. Most legal documents:

- Use dense and formal language
- It is difficult for non-lawyers to interpret
- Require navigating long documents to find relevant sections

As a result, students and citizens struggle to understand their rights, procedures, and obligations despite information being publicly available.

Solution Approach

This project builds a **Retrieval-Augmented Generation (RAG)** pipeline tailored for legal text:

1. **Document Ingestion**
   Legal and policy PDFs are loaded directly from Google Drive and converted into plain text.
2. **Legal Structure Awareness**
   The text is cleaned and segmented based on legal units such as *Articles* and *Sections* to preserve meaning.
3. **Chunking & Indexing**
   Relevant chunks are embedded using a SentenceTransformer model and indexed using FAISS for efficient semantic search.
4. **Query Understanding**
   User queries are analyzed to detect intent (conceptual, procedural, penalty-related).

5. **Contextual Explanation**
   Retrieved legal text is passed to a large language model (Gemini) to generate simplified explanations without altering legal meaning.

Key Features

- Semantic search over legal documents
- Article/Section-level retrieval
- Intent-aware explanations
- Source document citation
- Confidence scoring for responses
- Interactive command-line interface

Technologies Used

- Python
- Sentence Transformers (all-MiniLM-L6-v2)
- FAISS (Vector Search)
- Google Gemini API
- PyPDF2
- NumPy

Intended Use

This project is intended for **educational and research purposes only**. It does not provide legal advice and should not be used as a substitute for professional legal consultation.

Future Improvements

- Support for more Indian laws and policies
- Multilingual explanations
- Web-based interface
- Improved citation accuracy
- User feedback integration