DAY – 92

12 December 2025

This project successfully demonstrates the design and implementation of an intelligent chatbot system that provides automated, accurate, and user-friendly responses using both structured and unstructured data sources. The developed chatbot integrates a web-based interface with a Flask backend and employs multiple Natural Language Processing (NLP) and Machine Learning (ML) techniques to deliver reliable answers to user queries.

The system combines a **static dataset**, **PDF-based knowledge extraction**, and **hybrid matching techniques** to enhance the chatbot's understanding capability. The dataset file stores predefined question–answer pairs for frequently asked questions, while the PDF document serves as an official and authentic source of detailed information. By extracting and preprocessing content from the PDF, the chatbot is able to answer complex and procedural queries that are not explicitly defined in the dataset.

Advanced preprocessing techniques such as text normalization, stop-word removal, tokenization, and lightweight stemming are applied to improve query matching accuracy. The chatbot uses **TF-IDF vectorization**, **cosine similarity**, and **machine learning classification** to identify the most relevant response. Additionally, a rule-based fallback mechanism ensures that meaningful answers are provided even when confidence scores are low. For more complex or contextual queries, a **Retrieval Augmented Generation (RAG)** approach is incorporated, enabling document-based conversational responses.

Another significant achievement of this project is **session-based conversation handling**. The chatbot maintains chat history for each user session, allowing it to understand follow-up questions such as "this," "above," or "brief explanation." This makes the interaction more natural and conversational. The system also supports **short and summarized answers** when requested by the user, enhancing usability.

The chatbot interface is designed with a **modern floating chat popup**, providing an interactive and responsive user experience. Features such as typing indicators, streaming-style responses, language selection, and stop controls make the system more engaging and user-friendly. The backend efficiently manages requests, processes user inputs, and returns responses in real time.

Name- Jasleen Kaur   **Branch-D4 CSE (C2)**   **URN-2302723**

Overall, this project achieves its objective of building a **scalable, intelligent, and document-aware chatbot system**. It reduces manual effort, improves information accessibility, and demonstrates practical applications of NLP, ML, and web technologies. The modular architecture allows future enhancements such as multilingual expansion, voice-based interaction, and integration with live databases or government portals. Hence, the project proves to be a valuable and effective solution for automated information dissemination and user support systems.

- The project successfully implements an intelligent web-based chatbot using **Flask, NLP, and Machine Learning techniques**.
- The chatbot provides automated responses using both **dataset-based Q/A** and **PDF document knowledge extraction**.
- **TF-IDF vectorization and cosine similarity** are used to match user queries with the most relevant answers.
- **Rule-based matching** and **machine learning classification** improve accuracy and reliability of responses.
- The system supports **session-based conversation history**, enabling follow-up question handling.
- A **hybrid approach (Dataset + PDF + RAG)** ensures better coverage of user queries.
- Text preprocessing techniques such as **normalization, stop-word removal, and tokenization** enhance response quality.
- The chatbot can generate **short or brief answers** when requested by the user.
- **Multilingual support (English and Punjabi)** increases accessibility for regional users.
- The frontend provides a **responsive floating chat popup** with typing indicators and real-time interaction.
- The project reduces **manual workload** and improves information availability.
- The modular architecture allows **future scalability and enhancements** such as voice input, database integration, and additional languages.

Name- Jasleen Kaur   Branch-D4 CSE (C2)   URN-2302723