Chatbot for Government Welfare Schemes

NLP PROJECT REPORT

Submitted by

1. **Kishan Payadi 01314811621**
2. **Amar Singh 04214811621**

BACHELOR OF TECHNOLOGY
IN

ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

Under the Guidance

of

Mrs. Gunjan Chugh

(Assistant Professor, AIML)

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

Maharaja Agrasen Institute of Technology,
PSP area, Sector – 22, Rohini, New Delhi – 110085
(Affiliated to Guru Gobind Singh Indraprastha, New Delhi)

Abstract

Today, despite the introduction of numerous government welfare schemes aimed at benefiting citizens, a significant portion of eligible individuals remain unaware of these programs, Furthermore, they often struggle to identify the appropriate schemes that address their specific needs. This lack of awareness often hinders beneficiaries from accessing the support they are entitled to. To address this critical issue, our project focuses on leveraging cutting-edge technologies in Natural Language Processing (NLP) and machine learning. We propose the development of a user-friendly chatbot system that serves as an interactive platform for individuals to inquire about government welfare schemes easily and also has multilingual capability, enabling it to provide prompt and accurate responses in any language. This ensures accessibility for a diverse range of users, regardless of their linguistic background. Through intuitive conversational interfaces, users can articulate their concerns and queries to the chatbot, which then provides personalized recommendations on relevant schemes tailored to their circumstances. By offering proactive assistance and guidance, our solution aims to empower individuals to make informed decisions and access the support they need effectively. Through the seamless integration of technology and social welfare, we strive to enhance accessibility and inclusivity in government welfare initiatives.

Introduction

To effectively address inquiries regarding government welfare schemes, our proposed solution centers around the development of a chatbot powered by a Large Language Model (LLM) utilizing advanced machine learning techniques. This chatbot serves as an interactive platform, aimed at simplifying user inquiries about various welfare schemes. In addition to its robust functionality, This chatbot boasts multilingual capabilities, enabling it to understand and respond to queries in different linguistic languages. To achieve this, we integrate the Google Translate API, allowing seamless communication with users from diverse language backgrounds. Leveraging a comprehensive database curated meticulously through web scraping of government websites, the chatbot aims to furnish users with accurate and relevant information of all government initiatives. Additionally, our approach incorporates the utilization of a vector database to store information by converting it into vector embeddings. Through this innovative strategy, we endeavour to streamline the process of accessing crucial welfare support for individuals, offering a user-friendly and informative solution.

1. Chatbot

A chatterbot (chatbot) is a computer program designed to simulate an intelligent conversation with human users in natural language via voice or textual methods.

Chatbots are versatile enough to carry out conversations and even tasks for the users. They can be as simple as programs that answer a simple query with a single-line response, or as complex as digital assistants that learn and evolve to deliver increasing levels of personalization.

Chatbots are primarily of two types,

A. **Rule-based chatbots (Scripted chatbots)** — These types of chatbots are only capable of interacting with users by following pre-programmed rules. Instant responses, quick answers, FAQs, complaint resolution, mostly all the facets of customer support are achievable through these chatbots. This ensures the round-the-clock availability of businesses. These chatbots are less intelligent.

B. **AI chatbots (Virtual assistants)** are self-trained and powered by machine learning. They bank on artificial intelligence to learn and respond based on interactions with users. These are not hardcoded like the rule-based predecessor. These bots learn from the conversational datasets(conversations with the users) and predict the response.

- **Retrieval-based model-based chatbots** are close-domain and sweep through the database of answers and provide the most relevant answer with the highest rank. Usually used when conversation scope is limited to a topic.
- **Generative models-based chatbots** are open-domain and generate answers based on the words in the input. It is complex and prone to error. Used when there is no focused scope and generic. Siri and Google assistant are good examples of AI chatbots.

2. LLM

LLMs are complex AI models trained on massive amounts of text data. This data can include books, articles, code, and even conversations. By analyzing these vast datasets, LLMs learn the patterns and relationships between words, allowing them to:

- Understand the meaning and context of a sentence.
- Generate human-quality text, translate languages, write different kinds of creative content, and answer your questions in an informative way.
- Adapt their responses based on the conversation flow.

Example:- GPT, BERT, Claude, Cohere etc.

3. LLMs in Generative AI Chatbots

LLMs are the driving force behind generative AI chatbots. These chatbots go beyond simple rule-based responses. Here's how LLMs empower them:

- **Natural Language Processing (NLP) on Steroids:** LLMs take NLP to a whole new level. They can grasp the nuances of human language, including sarcasm, humor, and slang. This allows them to have engaging and natural conversations that feel less robotic.

- **Personalized Interactions:** LLMs can analyze past interactions with a user and adapt their responses accordingly. Imagine a chatbot that remembers your preferences and tailors its suggestions or recommendations.
- **Open-Ended Conversations:** Unlike retrieval-based AI chatbots that rely on pre-defined answers, LLM chatbots can handle open ended questions and even generate creative text formats like poems or scripts based on your prompts.

4. Vector Database

A vector database is a type of database specifically designed to store and manipulate vector data efficiently. In the context of natural language processing (NLP) and machine learning, vector databases are often used to store word embeddings, which are numerical representations of words or phrases.

Word embeddings are essential in NLP tasks because they capture the semantic meaning of words and their relationships with other words in a high-dimensional vector space. These embeddings are typically generated using techniques like word2vec, GloVe, or FastText, which map words to dense vectors of real numbers.

A vector database organizes these word embeddings in such a way that they can be efficiently queried and retrieved based on similarity or other criteria. This enables various NLP applications to quickly access and manipulate word embeddings for tasks such as text similarity calculation, information retrieval, document clustering, and more.

Example:- Pinecone, Milvus, Chroma, Faiss etc.

5. Google Translate API

It is a powerful tool provided by Google that enables developers to integrate automatic translation capabilities into their applications. It offers a wide range of features and functionalities for translating text between languages quickly and accurately.

Key features of the Google Translate API include:

- ✓ Language Detection
- ✓ Text Translation
- ✓ Multilingual Support
- ✓ Customization Options
- ✓ Scalability and Reliability
- ✓ Integration Flexibility etc.

Objectives

The objective of this project is to develop a conversational chatbot capable of interacting with users in a multilingual format and effectively communicating information about various government welfare schemes. The chatbot will be designed to engage users in casual or layman language, ensuring accessibility and ease of understanding for individuals from diverse linguistic backgrounds. we aim to empower users to access crucial information about welfare schemes effortlessly, thereby enhancing awareness and utilization of government initiatives.

Methodology

The project was divided into two parts:

1. **Data Collection and Processing**

   - **Data Collection:** We collected information about government welfare schemes from the official website https://www.india.gov.in/my-government/schemes. Using the Selenium web scraping tool, we extracted all scheme names and their corresponding links. This data was saved into a CSV file for further processing.
   - **Scheme Information Extraction:** Using the CSV file containing scheme names and links, we scraped detailed information about each scheme from the website. This information included scheme descriptions, eligibility criteria, application procedures, and other relevant details.
   - **Vector Embedding Generation:** The extracted scheme information was then converted into vector embeddings using the Hugging Face model "sentence-transformers/all-mpnet-base-v2". This model transformed the textual data into dense vector representations, enabling efficient similarity search and retrieval.
   - **Vector Database Integration:** The vector embeddings generated were stored in a vector database using Faiss, an efficient similarity search library. This database facilitated fast and accurate retrieval of scheme information based on user queries.

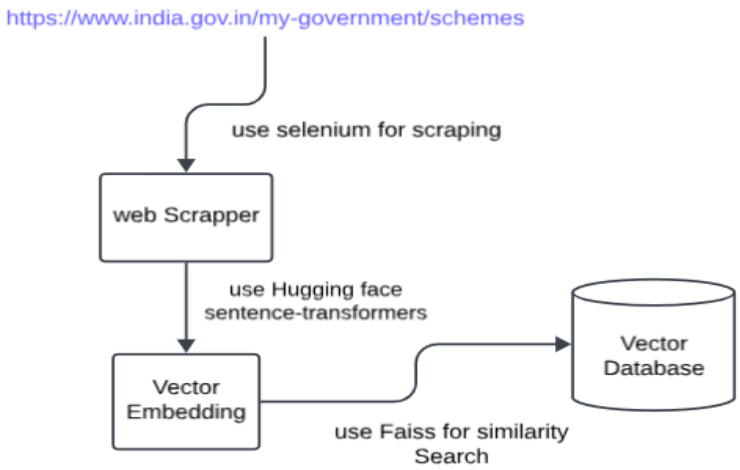

Fig1: Data collection and processing workflow

2. Chatbot implementation and Operation

- **Chatbot User Interface:** For the chatbot user interface, we utilized Streamlit, a Python library for creating interactive web applications. This interface provided users with a user-friendly platform to interact with the chatbot and inquire about government welfare schemes.
- **Language Identification and Translation:** When users entered queries in any language, we utilized the Google Translate API to identify the query language. If the query was not in English, it was automatically translated into English to ensure uniform processing.
- **RAG (Retrieval Augmented Generation):** The entire concept of our chatbot is based on the RAG framework, which combines retrieval-based and generation-based approaches. In the retrieval phase, the chatbot retrieves relevant scheme information from the vector database based on the user's query. In the generation phase, the retrieved information is passed to the Gemini model for response generation, ensuring accurate and informative responses tailored to the user's needs.
- **Query Processing and Response Generation:** The user's query, whether in English or translated, was converted into vector embeddings using the "sentence-transformers/all-mpnet-base-v2" model. These embeddings were then used to perform similarity search in the vector database, retrieving scheme information relevant to the user's query. The retrieved information was passed to the Gemini model for response generation. Finally, the response was translated back into the user's desired language before being displayed on the user interface.

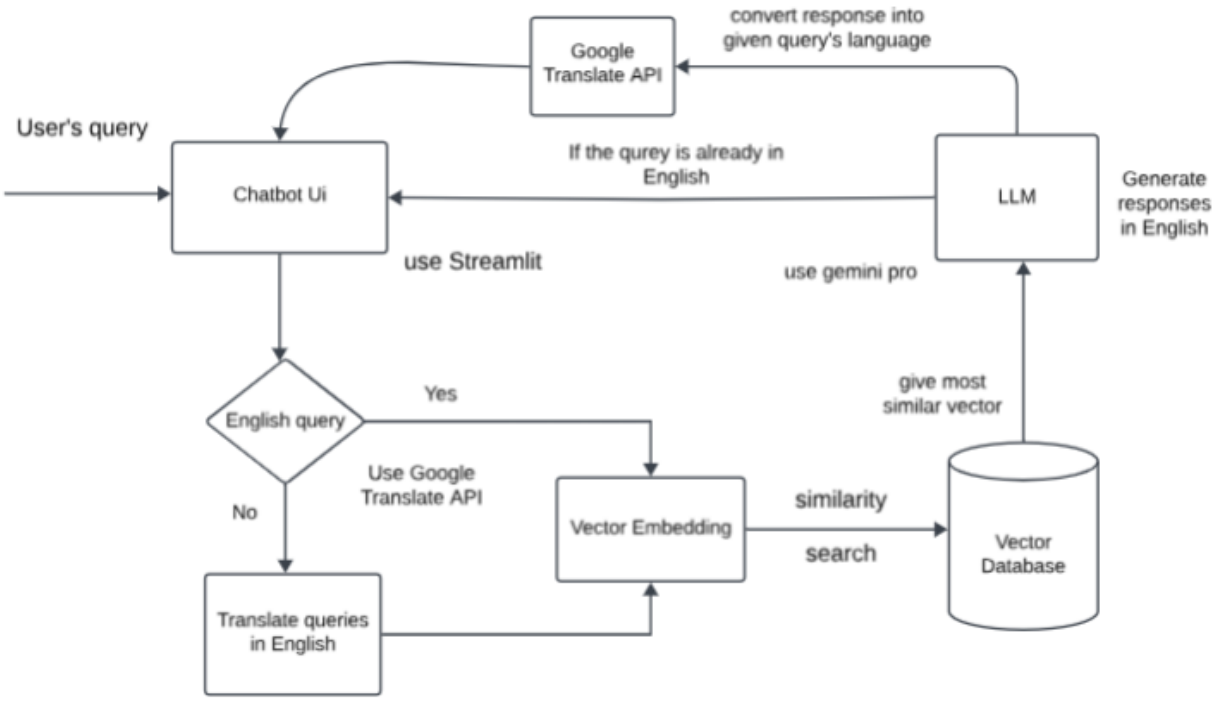

Fig2: Chatbot implementation and Operation workflow

Results

Fig3: Chatbot UI

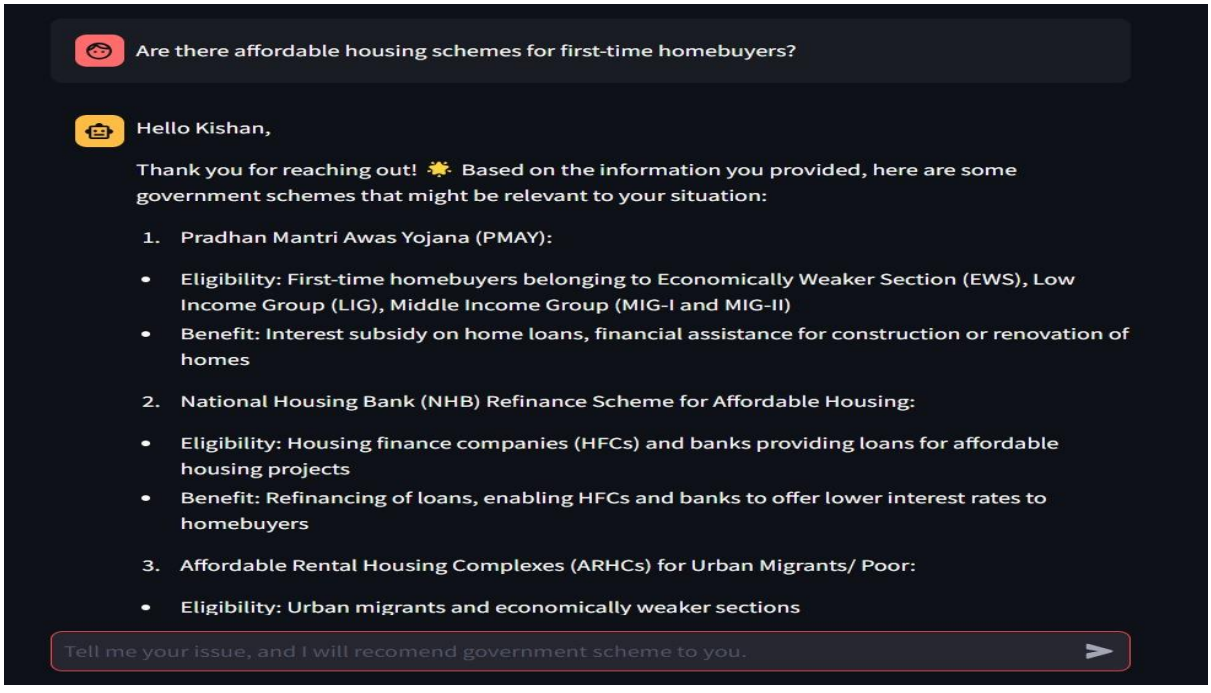

Fig4: Query responses in English

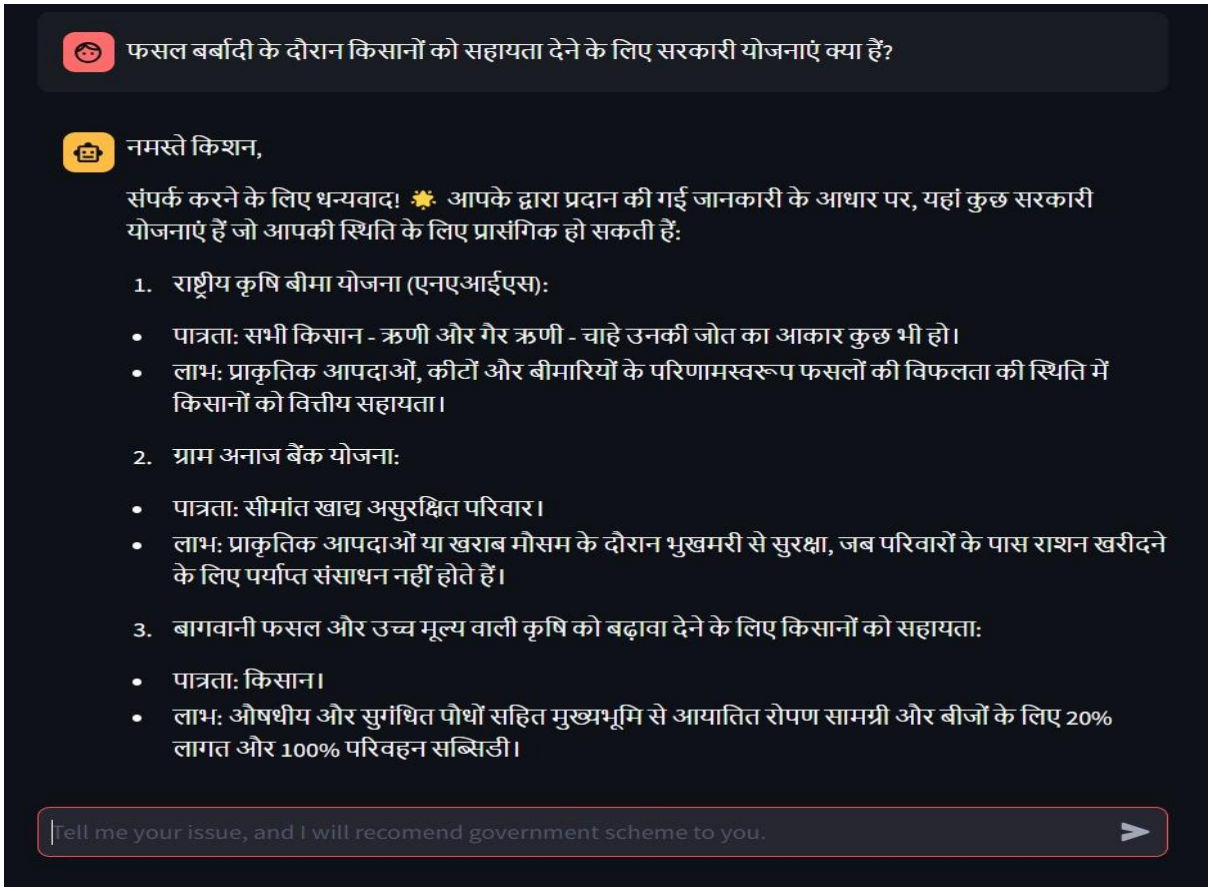

Fig5: Query responses in Hindi

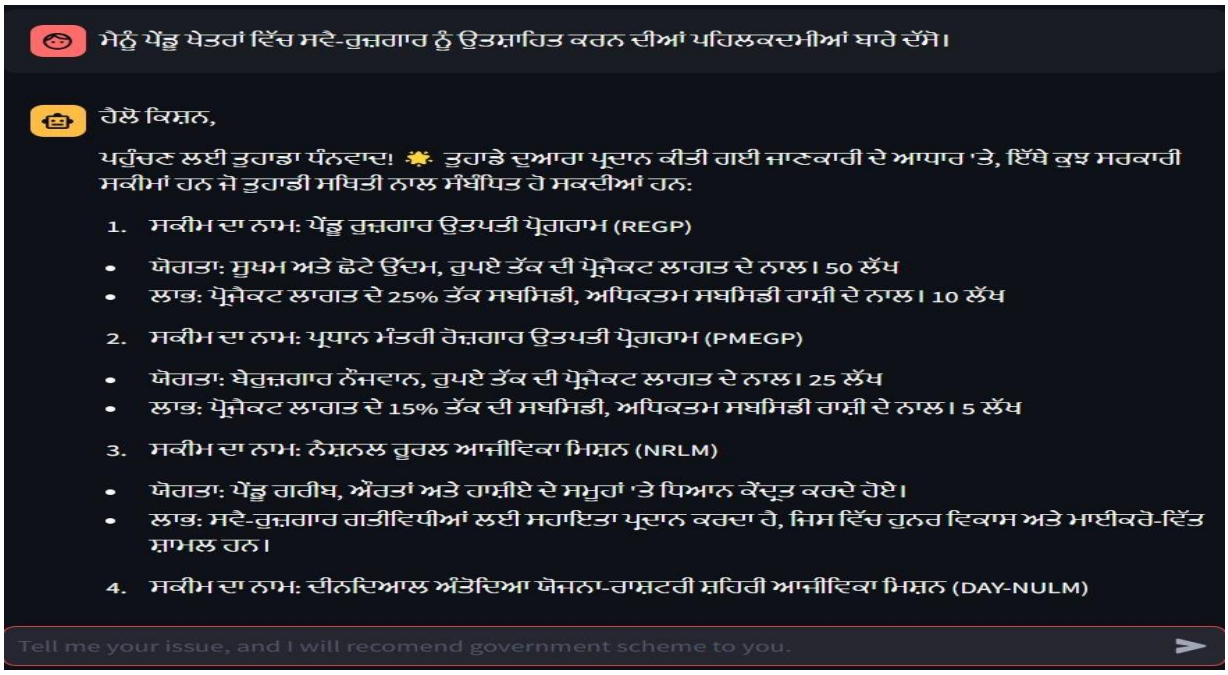

Fig6: Query responses in Punjabi

Conclusion

This project presents a comprehensive solution to address inquiries regarding government welfare schemes through the development of a multilingual chatbot powered by a Large Language Model (LLM). By leveraging advanced machine learning techniques and integrating features such as multilingual conversations and casual language support, we aim to simplify the process of accessing crucial welfare support for individuals from diverse linguistic backgrounds. Additionally, our extensive database and innovative use of vector embeddings ensure that users receive accurate and relevant information about government initiatives. Through this user-friendly and informative solution, we strive to enhance accessibility and inclusivity in government welfare initiatives, ultimately empowering individuals to make informed decisions and access the support they need effectively.

Future Work

Real-Time Updates on Welfare Schemes: Implement functionality to provide users with real-time updates on changes to government welfare schemes, ensuring that they always have access to the most current information.

Enhanced User Personalization: Develop features for personalized information delivery based on individual preferences and circumstances, allowing the chatbot to tailor responses to meet the specific needs of each user.

Integration with Government Portals: Explore opportunities to integrate the chatbot with official government portals and websites, facilitating seamless access to welfare scheme information directly from authoritative sources.

Adoption of Voice-Based Interaction: Investigate the implementation of voice-based interaction capabilities, allowing users to interact with the chatbot through voice commands and natural language processing.

Incorporation of User Feedback Mechanisms: Introduce mechanisms for gathering user feedback and insights to inform iterative improvements and enhancements to the chatbot's functionality and user experience.