GUVI Multilingual Chatbot – Project Report

# 1. Project Overview

The GUVI Multilingual Chatbot is an AI-powered assistant that answers user queries about GUVI’s platform, courses, and features. The chatbot is designed to understand multiple languages and provide context-aware responses using Retrieval-Augmented Generation (RAG).

## Key Highlights:

* Multilingual support (Indian + global languages)
* Context-based retrieval from a knowledge base
* Translation using NLLB-200
* Text generation with LaMini-Flan-T5-783M
* Simple and interactive Gradio interface

# 2. Objectives

* Build a chatbot capable of understanding and responding in multiple languages
* Provide accurate, context-aware answers using a RAG pipeline
* Integrate translation and generation for seamless multilingual support
* Deploy the chatbot for real-time interaction via Hugging Face Spaces

# 3. Features

|  |  |
| --- | --- |
| Feature | Description |
| Multilingual Support | Handles Indian & global languages including Hindi, Tamil, French, Korean, etc. |
| Context Retrieval | Uses FAISS with SentenceTransformer embeddings to fetch relevant information from text chunks. |
| Translation | NLLB-200 translates user input and answers for accurate multilingual responses. |
| Answer Generation | Lightweight text generation using LaMini-Flan-T5-783M. |
| Interactive Interface | User-friendly interface built with Gradio. |
| Error Handling | Robust handling of translation or generation failures. |

# 4. Tools & Technologies

* Python
* Hugging Face Transformers
* Sentence Transformers
* FAISS (vector search)
* NLLB-200 (translation)
* Gradio (UI)
* NLTK (text preprocessing)

# 5. How the Chatbot Works

## Flow Diagram:

1. User Input (Any Language)
2. Language Detection & Translation to English (if needed)
3. Embedding Generation (SentenceTransformer)
4. Top-K Context Retrieval (FAISS Index)
5. Prompt Creation for Text Generation
6. Answer Generation (LaMini-Flan-T5-783M)
7. Back Translation to User Language (if needed)
8. Display Answer

## Step-by-Step:

1. Detect the user’s language using langdetect.
2. Translate input to English using NLLB-200 if it’s not in English.
3. Convert the input question to embeddings using SentenceTransformer.
4. Search the FAISS index for the top relevant text chunks.
5. Construct a RAG prompt combining context and user question.
6. Generate the answer using LaMini-Flan-T5-783M.
7. Translate the answer back to the user’s language if required.
8. Display the final response.

# 6. Steps Taken to Build the Chatbot

* Collected GUVI-related content from GUVI website, Shiksha, Scribd, ChatGPT, and Perplexity.
* Cleaned the data and split it into chunks using NLTK.
* Generated embeddings for each chunk and built a FAISS index.
* Created a RAG pipeline combining translation, retrieval, and generation.
* Integrated the pipeline with Gradio for an interactive interface.
* Deployed the chatbot on Hugging Face Spaces.

# 7. Use Cases

* Instant Student Support: Provide answers about courses and platform features in real time.
* Course Recommendations: Suggest relevant GUVI courses based on user queries.
* Knowledge Base Access: Quickly retrieve important information without manual searching.
* 24/7 Assistance: Offer multilingual support anytime.
* Accessible Learning: Enable learners worldwide to interact with GUVI content in their language.
* Interactive Learning: Provide contextual explanations and clarifications for better understanding.

# 8. Demo

Try the chatbot live on Hugging Face Spaces:  
👉 <https://huggingface.co/spaces/gaja1995/GUVI_CHATBOT>

# 9. Conclusion

The GUVI Multilingual Chatbot demonstrates a practical application of RAG with multilingual translation. It improves accessibility, speeds up information retrieval, and provides a seamless user experience for students interacting with GUVI’s platform.