

# Dynamic, Multilingual AI Chatbot

**Intern Name:** Kaushal Sahu

**Internship Duration:** 2/7/2025 – 5/5/2025

**Organization:** NullClass

## 1. Introduction

This report outlines the work done during my internship, focused on developing an intelligent, multilingual chatbot capable of dynamically expanding its knowledge. The project utilized modern AI tools and NLP models, aiming to build a system that can understand and respond accurately to user queries in various languages, while learning and updating over time.

## 2. Background

The growing demand for smart assistants across industries requires chatbots that are both intelligent and adaptable, understanding multiple language and answer accordingly. Most static bots fail to grow post-deployment. This project was designed to address that gap using open-source tools like Meta's LLaMA model and FAISS for efficient vector-based knowledge retrieval.

## 3. Learning Objectives

- Understand how to implement a vector database using FAISS.
- Learn to integrate large language models (LLMs) like LLaMA for response generation.
- Develop a dynamic knowledge base that expands over time.
- Enable chatbot to learn multilingual capabilities for global user interaction.
- Design an interactive interface using Streamlit.

## 4. Activities and Tasks

- Set up FAISS for storing and retrieving document embeddings.
- Connected the vector database with LLaMA.
- Implemented a script to dynamically update the FAISS database from .txt sources.
- Integrated a language detection module to support English, Hindi, Spanish, and French.

- Designed a Streamlit app for user interaction with real-time responses.
- Successfully tested multiple knowledge topics, including "Lifecycle of a Frog."

## 5. Skills and Competencies

- Python (file handling, NLP, APIs)
- Natural Language Processing (SentenceTransformer, summarization)
- Large Language Models (LLaMA, inference logic)
- Vector Search (FAISS)
- UI/UX using Streamlit
- Data preprocessing and Language Detection
- Problem-solving and debugging

## 6. Feedback and Evidence

Feedback was gathered through consistent testing, showing high accuracy in multilingual understanding and contextual response generation. All components performed efficiently, and responses adapted to newly added information without retraining.

The sample pictures were taken as an evidence and uploaded on the folder named "sample".

## 7. Challenges and Solutions

**1. Limited access to commercial APIs (e.g., OpenAI GPT due to subscription limits).**

**SOL-** Switched to LLaMA with local embedding and FAISS for full control.

Note:- LLaMA doesn't have the capability to detect the language and change responses accordingly (but as it was free and made for Task 3). This is been coded so that the chatbot can do it.

**2. Making the bot multilingual without complicating architecture.**

**SOL-** Used lightweight language detection library called "langdetect" to detect the language and make a custom response in chatbot backend which helps in ChatBot performance.

**3. Updating knowledge without retraining.**

**SOL-** Made a custom script to update knowledge. It can be customized according to the developer needs and can be updated in the future pretty easily.

## **8. Outcomes and Impact**

- Built a fully functional, multilingual AI chatbot with dynamic learning capability.
- Demonstrated practical use of vector databases and LLMs in real-time applications.
- Gained hands-on experience with production-level chatbot architecture.
- The project is scalable and can be extended into voice, analytics, or domain-specific bots.

## **9. Conclusion**

This internship significantly enhanced my understanding of real-world AI systems, from embedding techniques to interactive user-facing apps. The chatbot developed is not only technically robust but also adaptable and user-friendly. It stands as a strong foundation for future AI-based assistants in education, business, or research.

The chatbot can be modified to learn more different languages and answer according to the custom data provided. The more vast the data, the more smart the model becomes.