GUVI Multilingual GPT Chabot using Streamlit – Integrated Translation and Domain-Specific Model

Deployment

# 1. Project Title

GUVI Multilingual GPT Chatbot using Streamlit – Integrated Translation and Domain-Specific Model Deployment

# 2. Objective

The objective of this project is to build a multilingual AI-powered chatbot using a fine-tuned GPT-2 model that supports real-time interaction in various languages. The chatbot is designed to assist GUVI learners by answering questions in their native languages through integrated translation pipelines.

# 3. Domain

• Artificial Intelligence (AI)  
• Natural Language Processing (NLP)  
• Web Development (Frontend with Streamlit)

# 4. Problem Statement

To create an intelligent, multilingual chatbot that:  
• Accepts user input in multiple Indian and international languages.  
• Translates the input to English using NLP translation models.  
• Generates a response using a fine-tuned GPT-2 model specific to GUVI context.  
• Translates the response back into the user’s original language.  
• Displays it via a responsive and user-friendly Streamlit interface.

# 5. Business Use Cases

1. Customer Support Automation  
2. E-Learning Accessibility  
3. Career Guidance and Mentorship  
4. Course Recommendation System

# 6. Key Features Implemented

• Language detection using langdetect  
• Translation (Multilingual ↔ English) via Hugging Face’s Helsinki-NLP models  
• GPT-2 based response generation using a fine-tuned domain-specific model  
• Dynamic FAQ interface in English, Hindi, Urdu, Russian, and German  
• Chat history management using Streamlit session state  
• Clean UI/UX built with Streamlit widgets and layouts

# 7. Tools and Technologies Used

• Python 3.11  
• Hugging Face Transformers & Pipelines  
• Streamlit  
• Langdetect  
• Git and GitHub  
• Jupyter Notebooks / VS Code

# 8. Technical Architecture

1. User Input Layer  
2. Language Detection  
3. Translation Layer  
4. GPT-2 Response Generation  
5. Streamlit UI

# 9. Model Fine-Tuning Process

1. Data Collection  
2. Preprocessing  
3. Tokenization  
4. Fine-tuning  
5. Evaluation

# 10. Results

• Fully functional multilingual chatbot deployed using Streamlit.  
• Supports real-time query resolution in 10+ languages.  
• Dynamic FAQs with auto-translation and contextual GPT-based answers.  
• Easily extensible for more languages or enterprise use cases.

# 11. Evaluation Metrics

• Modularity of code  
• Multilingual capability  
• Clarity of responses  
• UI responsiveness  
• GitHub documentation

# 12. Dataset Information

• GUVI-based textual data  
• Hugging Face datasets for translation  
• Approx. token count: 13521

# 13. Deliverables

• app.py  
• requirements.txt  
• README.md  
• GitHub repo  
• (Optional) Demo video

# 14. Guidelines Followed

• PEP-8 compliance  
• Modular code  
• Clear documentation  
• GitHub version control

# 15. Suggested Improvements

• Add speech-to-text and text-to-speech  
• Add analytics dashboard  
• Feedback buttons  
• Mobile interface  
• Support more languages

# 16. Deployment

• Hugging Face Spaces   
• Secure token handling  
• Cached model loading for performance

# 17. Conclusion

This project demonstrates the development and deployment of a domain-specific multilingual chatbot tailored to GUVI's educational platform. It showcases practical integration of NLP, translation, and deep learning models in a scalable, user-friendly application that enhances learner engagement and accessibility.