

NAME:jayda mohamed abd el-rehim

ID:2305284

NAME:jana kareem ismail

ID:2305283

NAME:jana hossam

ID:2305233

# Travel Guide Chatbot Project Report

## Overview

This report documents the development of a Travel Guide Chatbot that provides information about travel tips and hotel experiences based on real reviews. The chatbot leverages natural language processing and retrieval-augmented generation to answer user queries.

## Design Decisions

### 1. Data Processing Pipeline

- **Dataset Selection:** Used a travel/hotel review dataset (500 samples for demo purposes)
- **Preprocessing:**
  - Removed HTML tags and newline characters
  - Cleaned text for better embedding quality
- **Chunking:**
  - Used RecursiveCharacterTextSplitter with 500-character chunks
  - 50-character overlap to maintain context

## 2. Vector Store Implementation

- **Embedding Model:** Chose "paraphrase-MiniLM-L6-v2" for its balance between performance and resource requirements
- **Vector Database:** Selected FAISS for efficient similarity search
- **Indexing:** Created persistent FAISS index for quick reloading

## 3. Language Model Selection

- **Flan-T5-small:** Chosen for its:
  - Small footprint suitable for local deployment
  - Good performance on question-answering tasks
  - Open weights and commercial usability

## 4. Retrieval-Augmented Generation

- **RetrievalQA Chain:**
  - Combines document retrieval with generation
  - "stuff" chain type for simplicity
  - Retrieves top 3 most relevant documents

## 5. User Interface

- **Streamlit:**
  - Rapid UI development
  - Built-in chat components
  - Easy deployment options
- **Session State:** Maintains conversation history
- **Visual Design:** Clean, travel-themed interface

## Implementation Challenges

1. **Memory Constraints:** Had to limit dataset size for Colab demo
2. **Model Performance:** Trade-off between Flan-T5-small's speed and answer quality
3. **Safety Considerations:** Needed `allow_dangerous_deserialization` for FAISS loading
4. **Deployment:** Ngrok configuration for Colab-to-web access

## Future Improvements

1. **Enhanced Data Processing**
  - a. Incorporate more comprehensive travel datasets
  - b. Add entity recognition for locations/amenities
  - c. Implement sentiment analysis for review summaries
2. **Model Upgrades**
  - a. Experiment with larger Flan-T5 variants
  - b. Try fine-tuning on travel-specific data
  - c. Implement caching for frequent queries
3. **User Experience**
  - a. Add multilingual support
  - b. Incorporate visual elements (maps, hotel images)
  - c. Implement follow-up question handling
4. **Deployment Architecture**
  - a. Containerize with Docker for easier deployment
  - b. Set up proper CI/CD pipeline
  - c. Add monitoring for user queries/feedback
5. **Advanced Features**

- a. Personalization based on user preferences
- b. Integration with booking APIs
- c. Seasonal/temporal awareness for recommendations

## Conclusion

This project demonstrates how retrieval-augmented generation can be applied to create a specialized travel assistant. The combination of semantic search with a capable language model provides relevant, context-aware answers while keeping computational requirements manageable. Future work could significantly enhance the system's capabilities and user experience.