DSA4213 Project Proposal
Deng Haoyun A0286459J
Li Xiaoyue A0286700E
See Sze Yui A0282690X

1. Project Title

Travel Assistant Chatbot: A Lightweight Domain-Specific NLP Application

2. Motivation

Travelers often face repetitive questions, such as baggage policies, airport transfers, or short itineraries. Moreover, finding reliable answers requires navigating multiple websites. Building a chatbot that provides concise, accurate, and travel-focused responses demonstrates how NLP methods can solve a real-world problem while staying computationally efficient.

3. Problem Statement and Goals

3.1 We aim to answer:

How can lightweight language models be adapted into a domain-specific travel chatbot that provides accurate responses without requiring large-scale computation?

3.2 Goals:

- Implement a prompt-only baseline chatbot using Flan-T5 small.
- Add a simple retrieval component (RAG-lite) for improved factual grounding.
- Evaluate performance on a small curated dataset of travel FAQs and destination facts.

4. Scope

4.1 In scope:

- FAQs: baggage, check-in rules, airport transfers.
- 5–6 destination cards (basic transport, 2–3 attractions, 1–2 travel tips each).
- Comparison of prompt-only vs retrieval-based chatbot.
- Small-scale evaluation (automatic + human).

4.2 Out of scope:

- Personalized trip planning.
- Live/real-time data (flight status, prices).
- Large ablation studies or complex fine-tuning.

5. Dataset and Knowledge Base(KB)
- 15–20 FAQ pairs (from 2–3 airlines/airports).
- 5–6 destination cards, each with 2–3 attractions + 2 travel tips.
- Evaluation set: 20–25 test queries (mix of FAQ and itinerary-style).
- All stored in JSON/CSV, with a simple FAISS index for retrieval.

6. Methods

- Models: Flan-T5 small (baseline); optionally DistilGPT-2 for comparison.
- Chatbot setups:
  - Prompt-only generation.
  - RAG-lite: query → retrieve top-k KB facts → augment prompt.
- Evaluation:
  - Automatic: semantic similarity (e.g., cosine embeddings) instead of BLEU/ROUGE; perplexity for fluency.
  - Human: 8–10 students rating answers on correctness, clarity, helpfulness.
- Analysis: Focus on error cases (hallucination, incomplete answers).

7. Implementation Plan (6 weeks)

- **W1–2:** Collect FAQs + destination cards; build KB + evaluation set.
- **W3:** Implement prompt-only chatbot with Flan-T5 small.
- **W4:** Add a simple retrieval component (MiniLM + FAISS).
- **W5:** Run evaluation (20–25 queries, automatic + human).
- **W6:** Final analysis, report, and presentation.

8. Deliverables

- Curated dataset & KB (JSON/CSV).
- Prompt-only chatbot (baseline).
- RAG-lite chatbot (improved version).
- Evaluation results (automatic metrics + human ratings).
- Final report (10–12 pages) + presentation slides.

9. Risk and Mitigations

- **Data gaps:** fallback to simple FAQs only.
- **Evaluation weakness:** rely more on semantic similarity + qualitative analysis.
- **Time constraints:** keep scope small (≤ 20 FAQs, ≤ 6 destinations).