Goh Jia Yi A0259400L
Lin Jiaying Melinda A0281863W
Nathanael Lim Jun Wei A0254652A

# DSA4213 Group 21 Project Proposal

## 1. Introduction

Problem Statement

Travelers face information overload and stress when choosing restaurants abroad. Current recommendation tools provide generic, static suggestions without accounting for the user's emotional state or contextual preferences. This leads to suboptimal choices and heightened stress during trips.

Rationale

As avid travellers, we have personally experienced the frustration of deciding where to eat on the spot, often having to comb through multiple reviews under time pressure. This common pain point demonstrates a clear market gap for a tool that can synthesize vast amounts of data into actionable, personalized advice.

In this project, we are interested in developing a sentiment-aware chatbot that simplifies trip planning by evaluating and recommending restaurants to provide users with concise information that is readily available at their fingertips. We aim to analyse restaurant reviews and provide personalised restaurant recommendations based on existing reviews, locations and other factors to help travellers make quick and informed dining decisions.

## 2. Model

Dataset

The dataset used was the Yelp Academic Dataset from Hugging Face:

- Business dataset (yelp_academic_dataset_business.json) contains metadata about various restaurants including the restaurant name, location, categories, ratings and opening hours
- Review dataset (yelp_academic_dataset_review.json) contains user reviews with ratings and text content

Both review and business data would be merged together before being filtered for restaurant businesses. Thereafter, the data would be cleaned and tokenized. Data would then be split: 80% training, 20% testing for both sentiment analysis and chatbot evaluation.

Methods

Firstly, the sentiment analysis component would utilise a fine-tuned lightweight transformer model trained on Yelp review data in order to classify the user's emotions into positive, neutral or negative. This component would then extract sentiment scores and emotional indicators from user queries to inform subsequent processing stages.

The Retrieval-Augmented Generation (RAG) system would employ vector databases for storing and retrieving restaurant embeddings, implementing semantic search strategies to combine user queries with location preferences and restaurant features to identify relevant establishments and their associated reviews.

Lastly, response generation would leverage on large language models that have been designed with sentiment-aware prompts that shape responses based on the user's emotions.

## 3. Expected Outcomes & Evaluation

Outcomes:

- A functional prototype of a restaurant chatbot that provides accurate, sentiment aware responses
- Evidence that integrating sentiment analysis into RAG improves user satisfaction and perceived empathy
- Insights into the feasibility of using large-scale review data for conversational AI

Data Availability & Feasibility

- The Yelp dataset (approximately 6M reviews) is publicly available and ready to use
- Preprocessing and fine tuning are feasible within the project timeline since the dataset is well-structured and widely used in research
- (If time permits): Beyond the static Yelp dataset, we plan to experiment with scraping live Yelp reviews for selected restaurants. This would allow us to:
    - Incorporate the latest reviews, making the chatbot's knowledge more up-to-date
    - Evaluate whether fresh data improves response relevance and accuracy
    - Compare chatbot performance on static vs dynamic datasets

Evaluation

1. Sentiment Analysis
    - Metrics: Accuracy, F1-score, Precision, Recall
    - Compare with zero-shot LLM performance
2. Chatbot System (Human Evaluation)
    - Test with 20-50 user queries across positive, neutral, and negative sentiment
    - Human evaluators rate responses on Accuracy, Empathy, and Helpfulness (Likert 1-5)
3. Ablation Study
    - Compare full system (RAG + sentiment-aware generation) against baseline RAG only chatbot
    - Conduct A/B testing to see if users consistently prefers sentiment-aware responses