# Ziqi:

## Problem:

Mental Health Sentiment Analysis

Sentiment analysis -> model development (finetune BERT, RoBERTa, GPT, etc)

Possible reference:

* MentalBERT: Publicly available pre-trained language model for mental healthcare <https://arxiv.org/abs/2110.15621>
* EmoBERTa: A Robust Pre-trained Model for Emotion Understanding <https://arxiv.org/abs/2108.12009>

## Dataset:

Depression: Reddit Dataset (Cleaned)

<https://www.kaggle.com/datasets/infamouscoder/depression-reddit-cleaned>

Will:

Real Estate Information Query System

**Problem**: Helping real estate buyers and renters navigate available property listings and make informed decisions.

**Solution**: Use **Langchain** to enable dynamic interactions between users and the system, such as filtering based on preferences like budget or location. **RAG** can be used to retrieve current listings from a database or scraped data, and **prompt engineering** is used to phrase user preferences to narrow down results.

**Key Features**: Users can ask questions like "Show me 3-bedroom apartments under $2000 in Toronto," and the bot uses **RAG** to query the relevant listings.

Dataset: 我有code 可以爬zillow的data 并且form一个非常well-formatted 的table，example 放到了drive 里面了。 我可以爬一下 JSQ, Grove street, new port 以及学校周边几个zip-code area 的renting data 来形成一个 大的dataset 然后用llm来retrieve information 回答user的问题

[Zheng Tong](mailto:zt672@nyu.edu)

Problem: Generate Recipe Based on Food Image

Multimodal, fine-tuning LLM

Reference: <https://arxiv.org/pdf/2409.12010v1>

Dataset: RecipeNLG

Zihang:

Privacy for sentiment classification

Existing foundation model, fine tuning

Dataset: RealToxicityPrompts

https://docs.aws.amazon.com/bedrock/latest/userguide/model-evaluation-prompt-datasets-builtin.html

Reference: https://aws.amazon.com/getting-started/hands-on/machine-learning-tutorial-generate-predictions-without-writing-code/

https://docs.aws.amazon.com/sagemaker/latest/dg/jumpstart-solutions.html