Minghua Zhang (6636852110)
(I just registered for the course so I do not have a team right now…)

Brainstorming possible domains:
1. Mental Health Support Group
Available Datasets: Reddit posts talking about mental health : https://zenodo.org/records/3941387
WHO publications related to mental health: https://www.who.int/publications/i

….

Reasoning: Communication/messaging with others is an important part of mental health support groups. Such an application can help people to reach for support from others with similar conditions and seek for help from professional resources through AI agent.

2. Healthcare for Patients in Hospital
Available Datasets: clinical database (in Beth Israel Deaconess Medical Center):
https://physionet.org/content/mimiciii/1.4/

….

Reasoning: patients can connect with others with similar conditions or talk to their doctor since they often need guidance. Besides, AI agent provides service like making an appointment or explanation of lab result.

Continuing with **Mental Health Support Group** since psychology is always a field I wish to explore more. Plus, I have found a lot of interesting and useful datasets.

1. CSV/Excel

A large CSV of counselor-client professional conversations covering full spectrum of mental health topics. It trains our chatbot how to respond to users when they wish to start a conversation with chatbot and talk about their feelings. (especially when other people in the group are not around)

Kaggle : https://www.kaggle.com/datasets/nguyenletruongthien/mental-health

Sample:
Q: I guess I tried to be present and supported, but it seemed like it was never enough.
A: Remember, small efforts can make a big difference. How do you handle your anger when it arises in co...

2. PDF

WHO publications related to Mental Health topic. Those PDF publications include various research/survey/manual/ report/ …. about mental health

An example of publication: https://www.who.int/publications/i/item/9789240114487

Sample:
'The Mental Health Atlas initiative was established in 2000, and the first global report was published in 2001…..'

3. HTML - Forum posts

7cups has a online forum for people sharing mental health issues and caring for each other. It contains tons of daily conversations related to mental health topic.

Here an example: https://www.7cups.com/forum/games/General_2573/Asimpleyesorno_359506/

Sample:
"Just say yes or no and leave a question for the next person?

Are you in a good mood today?"

Python Script:

CSV/Excel files: using pandas to save them into a dataframe
Plus, my csv file is extracted through a Kaggle API first

PDF files: using pdfplumber library to read the pdf file and transform them into text
Then, I could save the text into the dataframe

HTML/webpage: most online forum / website will block/gate the Python Request so I am not able to directly scrap them even though I found a lot of mental health forums. Instead, I found some datasets containing crawled data from online mental health forums or blogs
(https://www.dropbox.com/scl/fo/y7p04go4bvvcwojalcwie/AAhiHEVm8jj_mVcGJB7JmPc?dl=0&e=1&rlkey=gvl1q8zgru1drkloa0og27fgq)
Similarly, save the csv file into a dataframe

How our datasest improve exisitng chatbot:
By fine-tuning on those datasets,
Counselor-client conversation -> learn how to talk like a supportive/professional/sympathetic counselor and how to respond a user's respond
WHO publications -> learn about mental health in-depth research and knowledge
Mental Health Forum -> learn to identify mood signal of the conversation
We can make the daily / generic chatbot into a supportive and professional mental health assistant.
The goal is to make people to believe that they could talk to a chatbot when a real-person support is not available
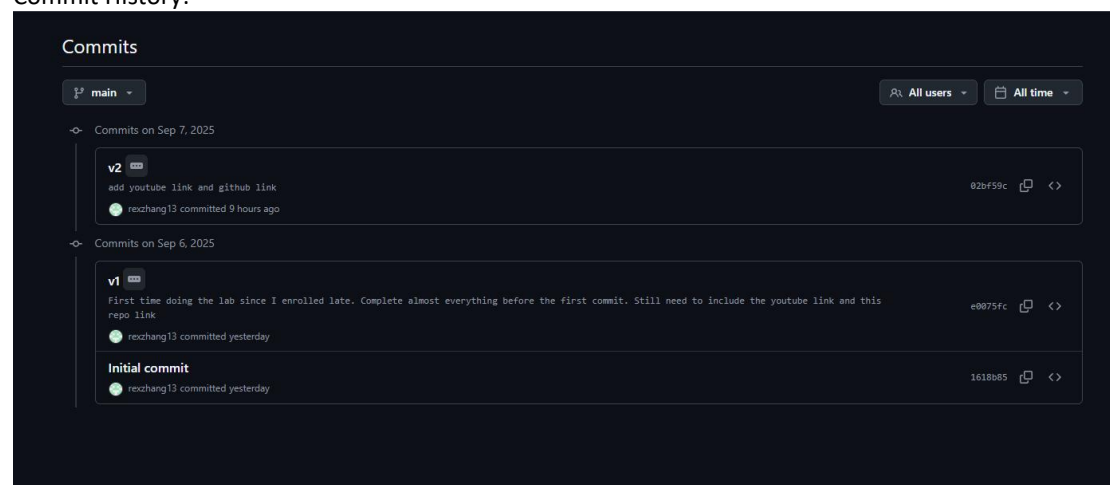
Later, we can also include many services into our application such as calling a crisis line or making a therapist appointment

Minghua Zhang (6636852110)
Youtube Demo: https://youtu.be/06XiMjOlSPI
Github Repo: https://github.com/rexzhang13/DSCI560_lab2
Commit History: