

Appendix

In total there are 4523 chunks, we mentioned that each chunk consisted of around 500 tokens, for this calculation we consider some additional space and suggest 700 tokens per chunk. To store embeddings for 4,523 chunks of approximately 700 tokens each in a FAISS vector database locally, the following hardware requirements are considered:

RAM: The memory requirement primarily depends on the dimensionality of the embeddings. Using the text-embedding-ada-002 embedding model with 1,536 dimensions, each vector would require 6,144 bytes (1,536 dimensions * 4 bytes per float32). For 4,523 vectors, the total memory needed would be approximately 27.8 MB. For efficient querying, it is recommended to have at least 3 times this amount available in RAM which equates to a minimum of 84 MB. Furthermore, other processes in the RAG system also fill up space in the RAM memory, therefore we consider a minimum of 4GB of RAM which is present in most modern-day hardware systems.

CPU: For a dataset of this size, a modern multi-core CPU is sufficient. A standard desktop or laptop CPU with 4-8 cores will be adequate to run FAISS efficiently.

Storage: The storage requirements will be similar to the RAM usage for the vectors themselves, approximately 28-30 MB. However, to account for storage of the other datasets and taking today's modern systems into account we suggest that 4GB of storage is more than sufficient.

GPU: For a dataset of this size a GPU is not necessary. FAISS can operate efficiently on a CPU for smaller datasets. This makes a GPU optional unless dealing with significantly larger datasets.

B Evaluation Criteria

B.1 Informativeness

Measures the chatbot's ability to come up with relevant and accurate information.

- 1: Provides mostly inaccurate or irrelevant information.
- 2: Occasionally provides accurate and relevant information, but often misses the mark.
- 3: Provides somewhat accurate and relevant information, with room for improvement.
- 4: Provides mostly accurate and relevant information.
- 5: Consistently provides highly accurate and relevant information.

B.2 Guidance

Evaluates the chatbot's capability in giving clear, actionable instructions and guidance to the patient.

- 1: Provides unclear or unhelpful guidance.
- 2: Provides occasionally clear but often unhelpful guidance.
- 3: Provides somewhat clear and occasionally helpful guidance.
- 4: Provides mostly clear and helpful guidance.
- 5: Consistently provides clear, actionable, and helpful guidance.

B.3 Empathy

Reflects the extent to which the chatbot demonstrates compassion, understanding, and support for the patient's circumstances.

- 1: Shows little to no empathy or understanding.
- 2: Occasionally shows empathy, but often lacks depth and understanding.

- 3: Shows a moderate level of empathy and understanding.
- 4: Shows a high level of empathy and understanding most of the time.
- 5: Consistently demonstrates deep empathy, compassion, and understanding.

B.4 Relevance

Measures the chatbot's effectiveness in generating responses that pertain to the patient's specific problems.

- 1: Responses are mostly irrelevant to the patient's specific problems.
- 2: Occasionally generates relevant responses, but often misses the patient's specific problems.
- 3: Generates somewhat relevant responses, with room for improvement.
- 4: Generates mostly relevant responses to the patient's specific problems.
- 5: Consistently generates highly relevant responses tailored to the patient's specific problems.

B.5 Understanding

The level of accurately grasping and reflecting the patient's experiences and emotions from the chatbot.

- 1: Shows little to no understanding of the patient's experiences and emotions.
- 2: Occasionally grasps the patient's experiences and emotions, but often inaccurately.
- 3: Somewhat accurately grasps and reflects the patient's experiences and emotions.
- 4: Mostly accurately grasps and reflects the patient's experiences and emotions.
- 5: Consistently and accurately grasps and reflects the patient's experiences and emotions.

B.6 Exploration

Assesses the degree to which the chatbot delves further into topics discussed during the interaction. Exploration can uncover underlying issues that the patient might not immediately disclose, which can help identify the root cause of distress and aid in the therapeutic process.

- 1: Rarely explores topics beyond the surface level.
- 2: Occasionally explores topics further, but often remains superficial.
- 3: Somewhat explores topics further, but with limited depth.
- 4: Mostly delves deeper into topics, uncovering underlying issues.
- 5: Consistently and effectively explores topics in depth, identifying root causes.

B.7 Coherence

Evaluates the chatbot's ability to maintain a logical and focused conversation flow and to transition smoothly between topics.

- 1: Conversations are often illogical and disjointed.
- 2: Occasionally maintains coherence, but often loses focus and clarity.
- 3: Somewhat coherent, but with occasional lapses in logic and focus.
- 4: Mostly maintains a logical and focused conversation flow.
- 5: Consistently maintains a coherent, logical, and focused conversation flow.

B.8 Reflectiveness

Measures the chatbot's proficiency in echoing back what the patient says, rephrasing, and demonstrating active listening.

- 1: Rarely reflects back or rephrases what the patient says.
- 2: Occasionally reflects back, but often inaccurately or without understanding.
- 3: Somewhat proficient in reflecting back, with room for improvement.
- 4: Mostly proficient in reflecting back and demonstrating active listening.
- 5: Consistently and accurately reflects back, rephrases, and demonstrates active listening.

B.9 Engagement

The ability of the chatbot to maintain longer and engaging conversations with the patient.

- 1: Rarely engages the patient, leading to short and unproductive conversations.
- 2: Occasionally engages the patient, but often fails to maintain interest.
- 3: Somewhat engaging, but with limited ability to maintain longer conversations.
- 4: Mostly engaging, maintaining longer and more productive conversations.
- 5: Consistently maintains high engagement, leading to long and productive interactions.

C Simulated patient script for depression

C.1 Instruction for Role-Playing

Welcome, and thank you for participating in this important study. This study regards developing chatbots for mental health support. You will be asked to interact with two different chatbots while role-playing or simulating having a mental health condition. It's important to clarify that this is not a test of your acting skills. The purpose is to interact with a mental health counselor chatbot, providing us with valuable insights into the chatbot's responses to different expressions of mental health symptoms.

You will be asked to have a conversation of around three minutes with each chatbot. After each conversation, you will be asked to fill in a questionnaire to measure the effectiveness of the chatbot as a support tool. Thus, while having the conversation, pay close attention to the responses of the chatbot and how you would feel towards these responses if you had the mental health condition.

C.2 Depression

Description of the Mental Health Condition

Depression is a common but serious mood disorder that affects how you feel, think, and handle daily activities. It is characterized by persistent feelings of sadness, hopelessness, and a lack of interest or pleasure in previously enjoyed activities. These symptoms must last for at least two weeks for a diagnosis of depression.

Key Symptoms to Portray

- **Persistent Sadness:** You often feel melancholic, tearful, and empty for most of the day, nearly every day.
- **Loss of Interest:** Activities that were once pleasurable, including hobbies, social activities, or sex, no longer interest you.
- **Fatigue or Decreased Energy:** You feel tired nearly every day, even without physical exertion. Simple tasks require more effort than usual.
- **Feelings of Worthlessness or Excessive Guilt:** You are overly critical of yourself for perceived faults and mistakes. You may dwell on past failures.
- **Difficulty Concentrating:** You find it hard to focus, make decisions, or remember details.
- **Changes in Appetite and Weight:** Significant weight loss or gain unrelated to dieting.
- **Sleep Disturbances:** This includes insomnia or sleeping too much.

- **Physical Symptoms:** Unexplained aches and pains, headaches, cramps, or digestive problems that do not ease even with treatment.

Impact on Daily Functioning

Depression can significantly impair your daily functioning. Work, school, family life, and personal relationships suffer. Routine tasks become burdensome, and maintaining productivity and social connections becomes increasingly difficult.

Emotional Experience and Attitude Towards Life

You may feel overwhelmed by despair, making it difficult to see any hope or positivity in life. There can be a pervasive feeling of being stuck or believing that things will never get better. Thoughts about life being unworthy of living can occur in severe cases. Social withdrawal is common, as you may feel that nobody understands your struggle, or you don't want to burden others with your feelings.

How to Interact with Others

In interactions, you might display irritability or frustration more easily, especially in situations that require emotional response or engagement. There may be a tendency to isolate yourself, preferring to be alone, and feeling that it's easier not to have to pretend to feel okay around others.

C.3 Role-Playing Guidance

Cause of Mental Health Problem

You have been grappling with depression for several years, but recent events have exacerbated the condition. The primary trigger was the sudden loss of a close friend in an accident six months ago. Additionally, high stress at work due to an ongoing demanding project and lack of support from superiors have contributed to a decline in your mental health.

Daily Feelings

You wake up feeling overwhelmed and exhausted most days, despite sleeping for more than eight hours. There is a persistent sense of sadness and a lack of interest in activities that once brought joy. You struggle with feelings of worthlessness and often blame yourself for minor mistakes at work or misunderstandings in personal relationships. Concentration difficulties are common, making it hard to stay focused during meetings or while completing tasks.

Impact on Life

Your social interactions have significantly diminished. You have started to isolate yourself, declining invitations to social gatherings and only responding to messages sporadically. This behavior has strained friendships and caused a rift in your relationship with your partner, who has expressed feelings of helplessness regarding your condition. The sense of disconnect from people and the things you once loved has left you feeling lonely and misunderstood.

Conversation Starter

You can start the conversation like this: "I've been feeling really overwhelmed and disconnected from everything lately. I'm not sure where to start, but I know I need help. I get a feeling that I am really depressed. Can we talk about where this feeling might be coming from and what I can do about it?"

C.4 Ethical Considerations and Support

We want to emphasize that your participation in this study is entirely voluntary. You have the right to withdraw at any time, for any reason, without any consequences. We are committed to ensuring your well-being throughout the study. If at any point you feel uncomfortable or need support, please let us know immediately.