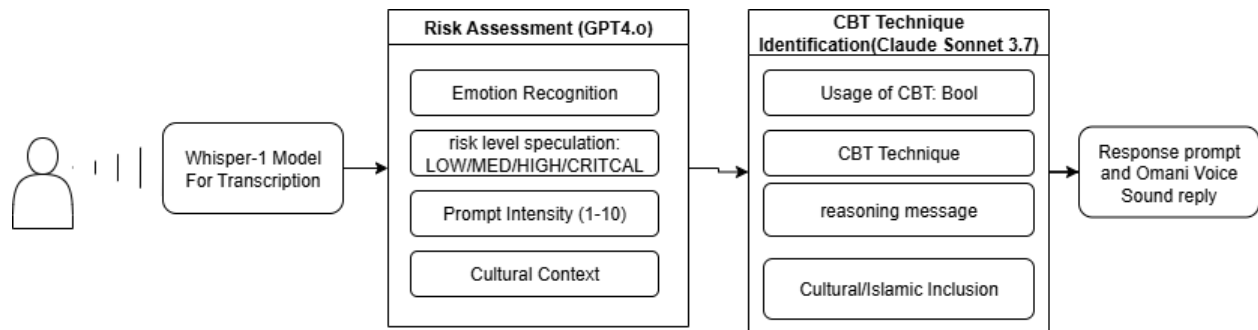


Omani-Voice-Therapist

Objective:

Develop a culturally-sensitive, voice-only Omani Arabic mental health chatbot that delivers therapeutic-grade support through real-time speech processing, emotional understanding, and secure, empathetic interaction—optimized for low latency, dialectal authenticity, and clinical safety standards.

Architecture/ System Design:



1. User records an audio in real time.
2. Risk Assessor Works to identify the following
 - a. Detect Emotion of User
 - b. Provide a risk level about the user record (and inform authorities if risk is critical)
 - c. Identify prompt intensity of the user
 - d. Generate a prompt about user cultural omani background
3. CBT Technique Identification
 - a. Identify whether to use Cognitive Behavioral techniques
 - b. Identity possible Cogonitve Behaviour technique to use based on a vector database
 - c. Create a short reasoning message to be used to influence the user
 - d. Include cultural and Islamic verses if needed
4. Response generated is sent to Azure Speech endpoint to generate an authentic Omani voice

Other Features.

- Professional Referrals for mental health hospitals for Oman
- Addition of Emergency contacts

- Error recovery system by relying only on GPT4.0 response if error arises and falling back to claude sonnet 3.7 if GPT4.o fails.

Deployment:

The system was deployed on an azure app service that is connected to a github repo to ensure CI/CD. A docker container and image were created and pushed to azure container registry autonomously after any changes in the main branch.

FASTAPI was used to create an endpoint on the azure app service that allows users to access the web page UI

A Simple Gradio App was created to be the User Interface of the system.

No database or user authentication were added, so it is a simple static website where changes are visible across users having the link.

Link to the Omani-Therapist : <https://omani-voice-therapist-evapawacaphzgfw.westeurope-01.azurewebsites.net/>

Analysis:

Dual Model analysis

The enhanced prompt is the output of Claude which will be fed into GPT4.o for response generation. Here are some examples for it.

Enhanced prompt 1:

المستخدم: السلام عليكم ورحمة الله وبركاته أنا أعاني من اكتئاب أنا جالس في البيت منذ فترة ولا أجد عملاً يناسبني بماذا تنصحنني؟

الحالة: sadness (6/10) - MEDIUM

CBT: True

Technique to help user: إعادة البناء المعرفي

مناسبة CBT المستخدم يعاني من اكتئاب مرتبط بالبطالة وعدم إيجاد عمل مناسب. تقنيات Reasoning for this technique: هنا لمساعدته على تحديد وتغيير الأفكار السلبية المرتبطة بالبطالة، وتطوير استراتيجيات للبحث عن عمل، وبناء روتين إيجابي. إعادة البناء المعرفي ستساعد في تحدي الأفكار السلبية حول قدراته وفرص العمل.

Reply from GPT4.o:

وعليكم السلام ورحمة الله وبركاته. الله يفرج عنك ويوفقك لكل خير. اكتئابك بسبب عدم إيجاد العمل المناسب مفهوم، ولكن مهم نبدأ بتغيير بعض الأفكار السلبية التي مرتبطة بهالوضع. قد يكون من المفيد تسأل نفسك بعض الأسئلة، مثل: "هل أنا فعلاً غير قادر على إيجاد عمل، أم أن هناك فرص لم ألاحظها بعد؟" و"هل هناك مهارات جديدة يمكنني تعلمها لتحسين فرصتي في العمل؟"

جرب تكتب أفكارك السلبية وناقشها مع نفسك. حاول استبدالها بأفكار أكثر إيجابية وواقعية. مثلاً، بدلاً من التفكير "ما في شغل يناسبني"، "حاول تقول "يمكنني البحث بشكل أوسع واستغلال مهاراتي بطرق مختلفة

نصيحتي لك أنك تخصص جزء من يومك للبحث عن عمل أو تطوير مهاراتك، حتى لو كان بسيط. ويمكنك أيضاً استشارة أصدقاء أو زملاء قد يكون لديهم معلومات عن فرص عمل جديدة.

والله يقول في كتابه الكريم: "فَإِنَّ مَعَ الْعُسْرِ يُسْرًا"، تذكر دائماً أن الأمور قد تتحسن بإذن الله مع المثابرة والدعاء.

Enhanced prompt 2:

المستخدم: انا ايضا اعاني من تقلبات شديدة في المزاج بماذا تتصحي كي استطيع ان اسيطر على هذه التقلبات

الحالة: distress (6/10) - MEDIUM

CBT: True

Technique to help user: مراقبة المزاج وإعادة البناء المعرفي

النص يشير إلى معاناة من تقلبات مزاجية شديدة مع رغبة واضحة في السيطرة عليها. Reasoning for this technique: مناسبة جداً لهذه الحالة لأنها تساعد في تحديد الأفكار السلبية التي تسبب تقلبات المزاج وتعديلها، كما تساعد في تطوير CBT تقنيات الاستراتيجيات للتعامل مع المشاعر المتقلبة.

Reply From GPT4.o:

Key Performance Indicators:

There is no methodology yet to test the whole of the pipeline due to lack of a complete omani Arabic word dataset. So I relied only on the accuracy metrics of whisper-1 GPT4.0 and Claude 3.7.I Need more time.

Reliability:

The Whisper-1 model exhibits sometimes wrong audio capturing, mishearing sometimes one word or all of the sentence. Hence, the user may require to record again

Emotion recognition: Most of the time the model predicts the correct emotion of the user with few wrong predications, mainly relying on correctly a transcribed audio

CBT Technique: The only verification that was made was a manual verification. The cbt technique generated was compared online with mental health websites (mayo clinc mental health). For an initial manual inspection, the prediction is approximately accurate.

Latency:

The model average response time is about 17 ± 8 seconds. Most delays occur from very long recorded audios, long replies by GPT4.o, claude API takes as well some time. Usually, a 5 second recorded audio takes about 15 – 18 seconds to finish processing.

Some rare cases were found where claude API took too long to respond. Taking up to a minute. This happens in the prod environment. The red arrow is trials to ping claude's api

```
Connected!
2025-07-21T15:04:55 Welcome, you are now connected to log-streaming service.Starting Log Tail -n 10 of existing logs ---
-/appsvctmp/volatile/logs/runtime/container.log
2025-07-21T15:04:14.1802540Z 2025-07-21 15:04:14,180 INFO | Retrying request to /v1/messages in 0.413740 seconds
2025-07-21T15:04:29.6745064Z 2025-07-21 15:04:29,674 INFO | Retrying request to /v1/messages in 0.756440 seconds
2025-07-21T15:04:35.0382197Z 2025-07-21 15:04:35,037 INFO | HTTP Request: POST https://api.anthropic.com/v1/messages "HTTP/1.1 200 OK"
2025-07-21T15:04:36.4964636Z 2025-07-21 15:04:36,496 INFO | HTTP Request: POST https://api.openai.com/v1/chat/completions "HTTP/1.1 200 OK"
2025-07-21T15:04:36.5042066Z 2025-07-21 15:04:36,503 INFO | ??? GPT response received in 35.59s
2025-07-21T15:04:36.5048640Z 2025-07-21 15:04:36,504 INFO | ??? Total processing time: 38.93s
2025-07-21T15:04:36.5054102Z 2025-07-21 15:04:36,505 INFO | ??? Speaking with Azure TTS...
2025-07-21T15:04:40.3250168Z 2025-07-21 15:04:40,324 INFO | ??? Speech synthesized.
```

Scalability:

As mentioned before the azure app service hosts a static website, so any affects is seen by every users that has the link to the azure app service. It was planned to connect the project to an Azure Cosmos DB where users' login details, old conversations, and emergency contacts will be recorded, but time was not sufficient.

Future improvements:

- Instead of calling whisper api, download it manually
- Use an azure cosmos DB
- Modularize the code instead of using a single monolithic code
- Rely on azure to handle all requests from future. Upgrading to premium will help accommodating many users