Look at the input for visual as well as audio.

Multimodal model with video and images.

11 labs. text to speech api paid.

Synchronizing the audio with the movement of the bot.

See who is talking during a speech.

Real time audio speech management.

playcanvas.com

first prototype in web xr.

MELD dataset

affective-meld.github.io

EEG dataset for emotion recognition.

Email Kien about audio-visual multimodal emotion recognition

* Model will recognize emotion.
* teach the agent about how to react to clam down the participant.
* Hostage scenario, psychiatrist training
* How many emotions do we use?
* More use cases of scenarios where we teach people about reacting to people.
* Use case, lab tours, informative questions.
* safety agent. clams people down in situations.