

Communication Analysis Tool for Human-AI Interaction Driving Simulator Experiments

Video to Audio Conversion (File 1):

This script uses the `mediapy` library to extract audio from a video file. The audio is divided into **5-second chunks**, with the last chunk being **2 seconds** if the total duration is not divisible by 5. Each chunk is saved as an individual audio file.

Audio to Transcription using Whisper (File 2):

The second script uses OpenAI's `whisper` model to transcribe the audio files into text. If an audio file is empty or cannot be transcribed, the transcription is recorded as `null` to handle edge cases gracefully.

Transcription to Sentiment Analysis (File 3):

The third script applies Hugging Face's `sentiment-analysis` pipeline to each transcription. It creates a structured dictionary containing:

- **Filename**
- **Transcription**
- **Sentiment Label** (POSITIVE, NEGATIVE)
- **Confidence Score** (model's prediction score)
Transcriptions with no valid audio are assigned `NA` for both sentiment and confidence.

Data Visualization (File 4):

The final script leverages `pandas`, `matplotlib`, and `seaborn` displays the count of each sentiment category—**POSITIVE**, **NEGATIVE**, and **NEUTRAL** (representing segments with null or non-analyzable transcriptions). This count-based bar chart provides insight into the overall emotional distribution across all audio segments, including periods of silence or ambiguity. It helps assess how often the model was able to detect emotional tone versus instances where speech was absent or unclear.