VOIDS

Visual Object Identification, Detection, and Speaker tracking.

By,
Karan Vora
(Expert Category)

VOIDS: Visual Object Identification, Detection, and Speaker tracking.

-> Majority of the interviews are being conducted over a virtual medium and in my personal experience 3 in every 10 interview has possible chance of attempt at fraud or straight up fraud.

-> The most common method to cheat is to have a third person to tag along in the process.

-> Successfully flagging the possible attempts at fraud becomes important in preserving the sanctity and integrity of the hiring process.

Overview

VOIDS uses 4 models to provide and intermediate representation of the chunk of video from the whole process.

1) Object Detection and Tracking
2) Object Identification
3) Audio Transcribing
4) Diarization

-> It maintains a global lookup table to keep track of unique and repeated entities of audio and object detected in the video stream

-> It uses purely heuristics to infer if there is an attempt and flag the said chunks for further analysis

-> The whole process is a Zero-Shot analysis process.

Workflow

-> VOIDS works in the backend after the video interview has been completed. This is an automated tool to provide you with insights that you might have skipped during the interview.

-> To make it compute efficient it works on randomize instances of the full video instead of the full-video

-> It provides you with flagged instances for further analysis

Live Demo on a previously recorded interview and usecases

-> Demo

-> Use cases in bot based interview

Future Improvement

-> Integration with Casual and Correlational model to provide detail visual analysis in form of video output

-> Safeguarding against audio based prompt injection attempts.

-> Flagging and classifying the individual channel for detailed analysis

-> Real-Time??