Research Proposal

Question:

How accurately can humans distinguish audio clips and images developed by AI versus other humans?

Goal:

Video and image-processing AI continues to get smarter as more data is collected to interpret and replicate media that goes far beyond text. While the general public is largely familiar with AI tools like ChatGPT, media attention has not been placed on models that process data like images and videos. Therefore, I intend to design a study to gauge how familiar people are viewing with AI-generated photos and videos.

This study measures how well humans are able to distinguish various types of media developed by artificial intelligence instead of real people. I plan to gather students from Cal Poly and present them with either two audio clips or two images, and record if they were able to identify the developer of each media correctly. I will continue to keep track of the overall success rate and test for any differences in success rates between both media types.

Explanatory Variables:

Developer of audio/image (Human / AI)
Type of media (Audio clip / Image)

Response Variable:

Proportion of media correctly identified as authored by AI instead of a human

Design/Materials:

This is a 2x2 factorial design, where the first factor is the developer of the media at 2 levels (Human, AI) and the second factor is the type of media at 2 levels (Audio Clip, Image).

The experimental unit is 1 full-time Cal Poly student. Part-time students and students from other campuses are to be excluded from the experiment.

A completely randomized design structure will be used, where participants are randomly assigned one of two types of media - images or audio clips.

AI-generated audio clips will be generated from the resemble.ai platform. An original audio clip will be extracted as a .mp4 file and passed through the resemble.ai software, which will create an artificially generated voice-over from the clip. This will be

done for all audio clips tested so that each original audio clip has an AI-generated version.

Images will be sourced from the Which Face is Real project, a website that generates images and asks the user to identify which image is real or developed by AI.

Steps:

1) Source 5 audio clips (15-20 seconds each) of celebrities or public figures from a video streaming platform like YouTube and download them as .mp4 files. Using the resemble.ai platform (https://www.resemble.ai/), upload your .mp4 files and generate AI-generated versions of these audio clips.

2) Source 5 pairs of images from the Which Face is Real project (https://www.whichfaceisreal.com/index.php). Print out these pairs of images in color, on separate one-sided pieces of paper.

3) Obtain a list of 50 full-time Cal Poly students that have agreed to participate in the study. Randomly assign either the audio clip or image treatment to each participant, so that 25 students listen to one of the five audio clip pairs and 25 students see one of the five image pairs.

4) For each participant in the audio clip treatment group, play the first audio clip and then wait 10 seconds. Then, play the second audio clip. Ask the participant which audio clip they thought was real and which was generated by AI. Record their response.

5) For each participant in the image treatment group, show them the printout of both images side-by-side. Give them 30 seconds to study each image and decide which one is AI-generated. After 30 seconds, ask the participant which image they thought was real and which was generated by AI. Record their response.

6) Calculate the accuracy rate by summing up how many image and audio pairs were correctly identified, and dividing that by the 50 total responses gathered. Multiply that number by 100 to get the accuracy in a percentage format.