

# Formal Study Protocol

*What's real and what's fake? Analyzing AI beyond text*

---

## Research Question

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

## Objectives

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 non-textual forms of data like images and videos. Therefore, I intend to design a study that gauges how familiar people are with identifying AI-generated photos and videos, as they may unconsciously see them online.

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.

## Variables

The explanatory variable is the type of media, presented to each participant as either an audio clip or an image. As a result, the response variable is the proportion of media correctly identified as authored by AI instead of a human. These proportions will be calculated separately for the audio clip and image categories, and compared with each other during data analysis.

## Design/Materials

This is a one-way design, with the only factor being 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.

The design structure implements two levels of randomization. First, 10 pairs of images and 10 pairs of audio clips will be sourced. Each participant will be randomly assigned to either the image or audio clip treatment. Then, they will be randomly assigned either 5 pairs of images or 5 pairs of audio clips depending on which category they were assigned to.

AI-generated audio clips will be generated from the resemble.ai platform. An original audio clip will be extracted as a .mp4 file and then 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 10 audio clips (10-15 seconds each) of celebrities or public figures speaking 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 store AI-generated versions of these audio clips.
- 2) Source 10 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. Note that each pair should be printed out on one side so that both images can be viewed alongside each other.
- 3) On Dexter Lawn, conduct a convenience sample by recruiting 50 full-time Cal Poly students who have agreed to participate in the study.
- 4) Randomly assign either the audio clip or image treatment to each participant, so that 25 students listen to five audio clip pairs and 25 students see five image pairs.
- 5) For each participant in the audio clip treatment group, start by randomly selecting 5 pairs of audio clips from the 10 pairs of audio clips previously gathered. Make sure to randomize the order in which the AI clip and original clip are played within each pair.

Play each audio clip, waiting 10 seconds between each clip before playing the next. Ask the participant which audio clip they thought was real and which was generated by AI, recording their response. Repeat this process for all 5 pairs of audio clips.

- 6) For each participant in the image treatment group, start by randomly assigning 5 image pairs from the 10 image pairs previously gathered.

Show them the printout of an image pair, giving them 30 seconds to study each image. After 30 seconds has elapsed, ask the participant which image they thought was real and which was generated by AI, recording their response. Repeat this process for all 5 pairs of images.

- 7) Keep track of the results in a table similar to the format shown below.

<b>Participant</b>	<b>Treatment</b>	<b># Correct</b>	<b># Total</b>
1	Image	3	5
2	Audio	2	5

- 8) For each participant, calculate the proportion of images or audio clips identified correctly, by dividing correct responses by total responses. Then, calculate the average proportion of responses identified correctly within the image group; do the same for the audio clip group.