Researcher – 1: Naveen:

Participant 1:

Female, 25, Graduate Student

Phase 1,2:

P1 gave brief on both videos by starting with the first video (real) where her focus total on the content spoken and celebrity face (facial expressions/moments) in the video. Explained most of the content spoken by the person in the video. For the second video (deep fake), again, P1 focused on the content and spoken person’s face (facial expressions) in the video. Both faces in the videos are attractive (familiar celebrities) and the videos are single monologue. P1 shared experience about the Phase1 videos that she stated “It’s nice, short and momentary. I could grasp most of the things” and she doesn’t feel boring due to small video duration. P1 felt the videos are normal/natural like social media and didn’t observe any differences like the videos are real or unreal. Persons in the video are talking general terms

Phase 3,4:

P1 clearly brief about the content spoken in the first video (deepfake). P1 understood the content in this video. P1 said about the first video “I was like focusing on what he's saying more than what he's doing”. P1 couldn’t get the content spoken in the second video (real) even though she rewinds the video. P1 said about the second video “I focused on his ring because I couldn't focus on what he was saying”. The characters/persons in this phase are new to P1. P1 focused more on the faces however in the second video, due to content opaqueness, she didn’t focus on the person/content (due to unfamiliarity of person and content) in 2nd video. No substantial change in the perception of watching 2 new set of videos to the warning label. P1 claimed that have forgotten the warning label while watching the second video.

Participant 2:

Male, 22, Graduate Student  
Phase 1,2:  
P2 summarized the first video (real) as a discussion of football-related concussions, highlighting the person’s personal experiences. In the second video (deepfake), P2 described Tom Cruise performing a magic trick, focusing on potential misdirection in the performance. P2 primarily focused on the speaker’s **face** (facial expressions) and **microphone** (mic) that speaker wore in the first video and the **coin** and **face** in the second video, attempting to detect trickery. P2 also mentioned that he focused on the frame behind the Tom Cruise which is in attracting blue color. P2 felt the videos are normal like “we watch in Instagram reels and YouTube shots”. P2 judged the first video as real due to its serious topic and personal narrative, while the second was deemed a deepfake, citing the lack of natural flow in expressions. Initially, P2 doubted on the Tom Cruise (deepfake) video that it was a deepfake, however because of the hair swing, he thought it was a real video. P2 stated about Tom Cruise video that “I think I've seen this video before also, I don't know if it's still a deep fake or not”.

Phase 3,4:  
P2 described the first video (deepfake) as discussing AI and time travel, focusing on the speaker’s **lips** (lip movement), which he found unnatural and inconsistent with the audio and told that it was a deepfake video. P2 stated about first video “there were no lips for in the first 5 seconds to Obama and there was a pause in the vocals and still the lips were moving”. The second video (real) appeared more natural due to fluid hand gestures and speech patterns. P2 stated that the warning label prompted him to scrutinize facial and lip movements closely, which helped identify the first video as a deepfake. He judged the second video as real based on its alignment of speech and gestures, emphasizing the role of natural expressions in his evaluation. This is because the presence of warning label from the first video changes his perception towards the second video where he started focusing on the Obama’s lips in the second video. P2 didn’t focus on the content due to busy in observing any unnatural movements in the video and later focused on the speaker’s watch.

Researcher-2:

Participant 1

Male, Age, Graduate Student  
Phase 1,2:  
P1 summarized the first video (real) as a discussion about a cricket player’s experiences and the second (deepfake) as an explanation about ChatGPT’s latest module. He remembered the content spoken by both speakers. He noted that the first video had clearer textures and appeared natural, while the second video seemed blurry and had unnatural lip movements. P1 focused primarily on the speaker’s face in both videos but observed fuzzy details (right hand side shoulder seemed quite blurred) and inconsistent lip synchronization (lip movements) in the second video, which he identified as a deepfake. He judged the first video as real based on clarity and natural expressions. Due to the short length, he didn’t observe much other than the above-mentioned features. P1 said that “Well, I was looking at the video as a whole”.

Phase 3,4:  
P1 noted that the first video (deepfake) had abnormal textures, particularly in the facial features, while the second video (real) appeared clearer and more expressive. The warning label influenced P1 to actively look for deepfake indicators, particularly focusing on the alignment of facial expressions and audio. He judged the second video as real due to its dynamic range of expressions and natural hand movements, while the first video was identified as a deepfake due to rigidity and lack of expression. P1 also stated that “it was not just because of the warning label, it was just the quality of video and the texture of video”.

Participant 2

Male, Age, Graduate Student  
Phase 1,2:  
P2 summarized the first video (real) as an actress discussing her experiences, while the second video (deepfake) involved Tom Cruise talking about his music and stunts. P2 told about the brevity of video. P2 focused on the speaker’s face in both videos, briefly observing peripheral elements like a painting and guitar in background in the second video. He judged the first video as real due to the speaker’s wider range of facial expressions, while the second was deemed a deepfake due to limited expressions and stiffness from the P2 words “I think the first video was more convincing as the real one because, the range of expressions in the first video was large and in the second were comparatively smaller.”.

Phase 3,4:  
P2 noted rigidity and lack of expression in the first video (deepfake), while the second video (real) had dynamic **facial expressions** and hand movements, which he found convincing. The content spoken by both speakers was general. The warning label heightened his awareness, leading him to actively search for deepfake indicators such as unnatural rigidity which continued to second video as well. He concluded that the second video was real and the first was a deepfake, based on the fluidity and expressiveness of the speakers. P2 stated his explanation “high probability of being defect in the first video, the **camera angle**, I mean the expressions and overall, the moment of the body was very rigid, and the expressions weren't changing that much whereas in the second video there was a lot of changing of expression as well as the **hand moment**”.

Researcher – 3:

Participant 1

Female, 60, Housewife  
Phase 1,2:  
P1 had difficulty recalling the spoken content from both videos, citing the brevity as a limiting factor for the video. In the first video (real), P1 mentioned focusing on the speaker’s face but did not elaborate on expressions or specific observations. For the second video (deepfake), P1 commented on the excessive movement of the speaker's eyes, finding it distracting but didn’t identify it as a deepfake. P1 described the videos as “pleasant” but expressed indifference, mentioning that they didn’t leave a lasting impression with her statement “but I don't know why I don't remember what they were talking about”. When asked to tell which one is real, P1 stated she couldn’t tell which video was real or fake due to her unfamiliarity with AI-generated imagery.

Phase 3,4:  
P1 briefly summarized the first video (deepfake) as a discussion about relativity and time travel, focusing on the speaker’s hat as a visual distraction. In the second video (real), P1 noticed exaggerated eye movements and facial expressions, which she found slightly overwhelming. Despite the presence of the warning label, P1 did not find the label impactful and didn’t change her perception. She mentioned that as long as the content wasn’t controversial or provoking misinformation, it didn’t matter whether the video was real or AI-generated. P1 judged both videos as equally real/normal, claiming that the label didn’t influence her judgment from her lines “I don't think it's provoking any dislikeness in me”.

Participant 2

Male, 63, Professor in Business  
Phase 1,2:  
P2 briefly summarized the first video (real) as a discussion about football concussions and the second (deepfake) as a magic trick involving misdirection. P2 primarily focused on the speaker’s face in the first video, attempting to interpret the emotional context related to sports injuries. In the second video, his focus shifted to the coin and the speaker’s hands, trying to detect sleight-of-hand movements. P2 described the first video as informative and socially relevant, while the second was playful, entertaining and person in the video is attractive. He judged the first video as real due to its serious tone and emotional depth, while the second was deemed a deepfake due to its performative nature.

Phase 3,4:  
P2 summarized the first video (deepfake) as a discussion about robots and AI, focusing on the speaker’s hand and mouth, which he found unnatural and inconsistent with the audio. In the second video (real), P2 observed fluid hand gestures and speech patterns that aligned with his perception of the speaker. The warning label heightened his suspicion, causing him to scrutinize lip movements and gestures more closely. P2 identified the first video as a deepfake due to unnatural lip synchronization (lip movements) and judged the second video as real, supported by familiar and natural expressions. P2 said that “his like accent and like the content caused me to think it wasn't a real”. P2 said that he has knowledge on Obama’s speeches from the lines “when I listened to him and it, it more aligned with how I've seen Obama talk in the past”.

Judgement:

| **Participant** | **Phase** | **Real Video Correct** | **Deepfake Video Correct** | **Overall Result** |
| --- | --- | --- | --- | --- |
| **Naveen P1** | 1 & 2 | ✅ | ❌ | Partially Correct |
|  | 3 & 4 | ✅ | ❌ | Partially Correct |
| **Naveen P2** | 1 & 2 | ✅ | ❌ | Partially Correct |
|  | 3 & 4 | ✅ | ✅ | Fully Correct |
| **Shubhom P1** | 1 & 2 | ✅ | ✅ | Fully Correct |
|  | 3 & 4 | ✅ | ✅ | Fully Correct |
| **Shubhom P2** | 1 & 2 | ✅ | ✅ | Fully Correct |
|  | 3 & 4 | ✅ | ✅ | Fully Correct |
| **Ling P1** | 1 & 2 | ❌ | ❌ | Incorrect |
|  | 3 & 4 | ❌ | ❌ | Incorrect |
| **Ling P2** | 1 & 2 | ✅ | ✅ | Fully Correct |
|  | 3 & 4 | ✅ | ✅ | Fully Correct |

A graph of different colored bars

Description automatically generated

Total Participants: 6 (each with two phases).

Real Videos:

Phase 1 & 2: 5 participants identified the real video correctly.

Phase 3 & 4: 5 participants identified the real video correctly.

Deepfake Videos:

Phase 1 & 2: 3 participants identified the deepfake video correctly.

Phase 3 & 4: 4 participants identified the deepfake video correctly.

A graph of a comparison of a video type

Description automatically generated with medium confidence

A graph of blue rectangular objects

Description automatically generated with medium confidence

Peripherals: ring, mic, blue frame, watch, shoulder texture (said blurry), guitar, painting, hat, coin

A graph of a warning label

Description automatically generated

Descriptive statistics

Inferential statistics

CAI/Security venue

Two phases