577 Improving user’s comprehension of deepfake analysis

P1 -Researcher-2

P1 was able to describe both short video clips, explaining their content and what was being discussed while watching. However, due to the limited time frame, he couldn’t determine the overall relevance of the videos. His focus was on the entire screen rather than just the face or lips. He observed that the quality of the first video was superior, with clearer and more defined textures. In contrast, he found the lip movements in the second video to be unnatural and the right shoulder blurred, which led him to conclude that the first video appeared more authentic than the second.

In the second round, he could describe the first video but struggled to explain the second one, possibly due to the short time provided. Again, his attention was on the video as a whole, but he noticed that the texture in the first video seemed abnormal. The warning label may or may not have influenced his opinion about which video was real and which was a deepfake.

P2- Researcher-2

P2 was able to describe both videos and summarized that they depicted regular day-to-day conversations. In the first video, his attention was on the face and the monologue, while in the second video, his focus shifted from the face to the background elements, such as the painting on the wall and the guitar. He concluded that the first video was real due to the wider range of facial expressions compared to the second one.

In the second phase, P2 was less focused on the content of both videos due to the limited time frame. His attention remained on the faces in both videos, as there wasn’t much else happening in the foreground. The warning label influenced his perception, prompting him to look for signs indicating the video could be AI-generated. He observed that the expressions and body movements in the first video were rigid, with little variation in facial expressions, unlike the second video. Additionally, the hand movements in the second video helped him conclude that the first video was AI-generated, whereas the second one was a deepfake

Researcher-3- P1

P1 struggled to recall details about the two videos during the first phase and was unable to focus on specific elements within them. Her attention was more directed toward the monologues and what the individuals in the videos were saying. Due to limited exposure to AI-generated content, she couldn’t determine which video was real and which was AI-generated.

In the second phase, however, P1 was able to explain both videos and their contexts. She found the first video to be more professional, while in the second video, the eye movements and expressions were distracting. The presence of a warning label did not significantly influence her perception of the videos, as their content was neither controversial nor conveyed messages that she found unconvincing or objectionable.

Researcher-3-P2

P2 was able to describe both videos and their content effectively. In the first video, his focus was primarily on the monologue and the message being conveyed, which revolved around personal experiences with brain injuries and concussions. In contrast, his attention in the second video shifted to the coin and the type of game being played, anticipating some form of misdirection. He concluded that the first video was real due to its serious subject matter, while the second video seemed more entertainment-focused and less credible.

During the second phase, P2 again described the videos well, focusing on their content and context. In the second video, his attention was drawn to the hands and gestures, while in the first video, he focused more on the mouth and speech. He observed that the mouth movements in the first video appeared unnatural and misaligned with the way the individual typically speaks. Based on these observations, he concluded that the second video was fake. The warning label initially influenced his perception, but ultimately, it was the mismatch between the persona’s usual impression and the way they appeared to speak in the first video that solidified his judgment

Naveen -P1

P1 demonstrated a strong ability to explain and describe the content of the videos, noting that they were concise and engaging rather than overly lengthy. In the first video, her focus was on changing facial expressions, highlighting a natural human tendency to pay attention to moving or visually appealing elements. Since only one person was speaking in the frame, her attention centered on the face. She also noted that the topic being discussed was familiar and had already reached the audience. However, she couldn’t determine which video was real and which was a deepfake.

In the second phase, P1 was again able to describe both videos but struggled with the context of the second one, as the character was not properly introduced. She paid attention to the visuals and actively listened to the discussion of an important term related to relativity in the first video. In contrast, in the second video, her focus shifted to the ring rather than the content, possibly due to her lack of background knowledge. The presence of a warning label did not influence her perception of the videos, as she found them believable and reflective of real-life scenarios.

Naveen -P2

Participant 2 provided detailed descriptions of both videos, though he found them not particularly engaging or compelling as experiences. In the first video, his attention was drawn to the microphone below the neck and the background picture, the face, and the hair in the second video. He noted that the background image caught his eye, and he read the text in the image. He also thought he had seen the second video before. He concluded that both videos were unrelated. He perceived the first video as more authentic, as it portrayed a personal experience of trauma, whereas the second video lacked depth, being merely about Tom Cruise laughing.

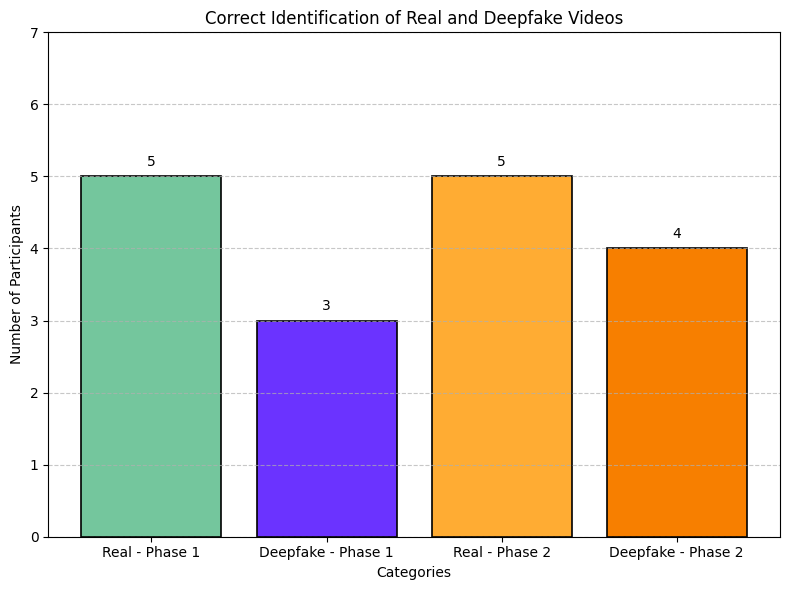
In the second phase, Participant 2 quickly identified the first video as a deepfake due to mismatched lip movements and speech alignment. He noted that the editing quality was poor, as there were moments when the vocalization paused, but the lips continued moving. This observation led him to focus on the lips from the beginning of the first video. Similarly, in the second video, he initially focused on the lips, expecting it to be a deepfake, but later shifted his attention to the watch strap, observing the brand. He also searched for any blurring or irregularities that might indicate a deepfake. After about five seconds of observing the second video, he concluded that it was not AI-generated.

While the warning label did not significantly influence his decision-making, it did prompt him to watch the videos more attentively, particularly focusing on the lips, to determine whether they might be deepfakes.

Interview survey result

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Participant** | **Phase 1 - Real Identified** | **Phase 1 - Deepfake Identified** | **Phase 3 - Real Identified** | **Phase 3 - Deepfake Identified** |
| Researcher-2-P1 | Yes | Yes | Yes | yes |
| Researcher-2-P2 | Yes | Yes | yes | Yes |
| Researcher-3-P1 | No | No | No | No |
| Researcher-3-P2 | Yes | Yes | yes | Yes |
| Naveen-P1 | yes | No | Yes | No |
| Naveen-P2 | yes | No | Yes | yes |

* **Yes**: Indicates the participant correctly identified the video.
* **No**: Indicates the participant did not identify the video correctly.



The parameterS that can be considered for qualitative analysis:

**Comparison of Perception**:

* Analyze the differences in participants' perceptions of real vs. deepfake videos, especially regarding which seemed more convincing and why.

**Impact of Warning Labels**:

* Measure how the presence of a warning label changes the perceived authenticity of videos and compare it to videos without such labels.

**Summarization Accuracy**:

* Quantify how accurately participants summarize the videos to evaluate comprehension.

**Focus Area Distribution**:

* Use visual heatmaps or other methods to evaluate and compare focus areas marked by participants on video screenshots.

**Response Times**:

* Record the time taken by participants to answer questions, especially regarding identifying deepfake vs. real videos.

**Content Recall**:

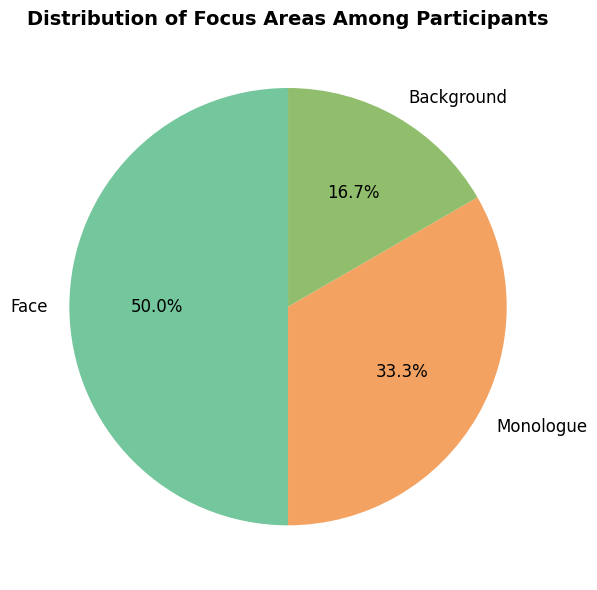
* Measure how much detail participants recall from each type of video (real vs. deepfake, with and without warnings).

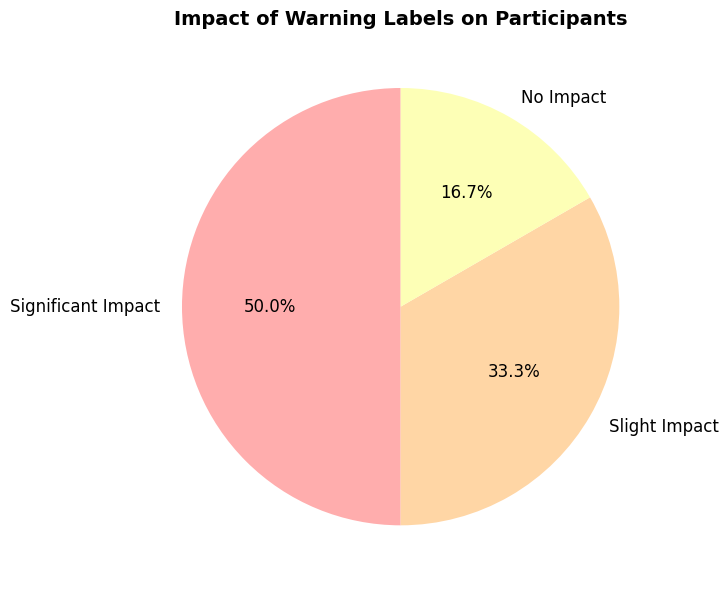
**Emotional or Cognitive Indicators**:

* Use qualitative data from participants' verbal explanations (e.g., "What caught your attention?") and analyze it quantitatively by coding responses into categories.

**Impact of Phase Progression**:

* Compare data from phases 1 and 3 to understand how viewing experiences evolve with increased familiarity or the introduction of warning labels.





**CODEBOOK:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Level of Analytics Code(Theme )** | **Code Name (Subtheme)** | **Code Description** | **Example from the Transcript** |
| **Attention Patterns** | Focus Area: Face | The participant's attention was directed primarily at the face while watching the video. | "His attention was on the face and the monologue." (Researcher-2 - P2, Phase 1) |
|  | Focus Area: Background | The participant shifted focus to the background elements of the video. | "His focus shifted from the face to the background elements, such as the painting on the wall and the guitar." (Researcher-2 - P2, Phase 1) |
|  | Focus on the Entire Screen | The participant's attention was distributed across the whole video frame. | "His focus was on the entire screen rather than just the face or lips." (Researcher-2- P1, Phase 1) |
| **Perception of Realism** | Authenticity of Content | The participant assessed the realism based on video quality, expressions, and natural movements. | "He found the lip movements in the second video to be unnatural and the right shoulder blurred." (Researcher-2 - P1, Phase 1) |
|  | Credibility of Context | The Participant evaluated the believability based on context and subject matter. | "He concluded that the first video was real due to its serious subject matter, while the second video seemed more entertainment-focused and less credible." (Researcher-3 - P2, Phase 1) |
| **Impact of Warning Label** | Increased Attentiveness | The warning label prompted the participant to observe more details. | "The warning label influenced his perception, prompting him to look for signs indicating the video could be AI-generated." (Researcher-2 - P2, Phase 2) |
|  | Negligible Impact | Warning labels had minimal effect on perception due to the non-controversial content. | "The presence of a warning label did not significantly influence her perception of the videos." (Researcher-3 - P1, Phase 2) |
| **Detection of Deepfake Cues** | Facial Movements | Unnatural lip or facial movements were identified as cues for deepfake detection. | "The editing quality was poor, as there were moments when the vocalization paused, but the lips continued moving." (Naveen - P2, Phase 2) |
|  | Visual Blurring or Texture | Texture anomalies and blurring indicated deepfake content. | "He noticed that the texture in the first video seemed abnormal." (Researcher-2 - P1, Phase 2) |
|  | Discrepancy in Persona | The mismatch between typical persona and observed behavior helped identify deepfakes. | "It was the mismatch between the persona’s usual impression and the way they appeared to speak in the first video that solidified his judgment." (Researcher-3 - P2, Phase 2) |
| **Engagement Level** | Limited Time Effect | Time constraints affected participants' ability to evaluate videos comprehensively. | "He struggled to explain the second one, possibly due to the short time provided." (Researcher-2 - P1, Phase 2) |
|  | Attention to Monologue | The Participant concentrated on the speech or discussion in the video. | "Her attention was more directed toward the monologues and what the individuals in the videos were saying." (Researcher-3 - P1, Phase 1) |
| **Content Comprehension** | Description Accuracy | The Participant accurately described the video content or context. | "He provided detailed descriptions of both videos, though he found them not particularly engaging or compelling." (Naveen - P2, Phase 1) |
|  | Misunderstood Context | Participants struggled with or misunderstood the video's intent or context. | "She struggled with the context of the second video, as the character was not properly introduced." (Naveen - P1, Phase 2) |

### Explanation of Columns:

1. **Level of Analytics Code**: The data analysis overarching category (theme).
2. **Code Name (Subtheme)**: Specific subtopics or recurring themes that emerged during coding.
3. **Code Description**: Brief explanation of what the code captures from the data extracted from the interview.
4. **Example from the Transcript**: Direct quotes or summarized insights from participants illustrating the code from the interview.