Article Review #2: A Person's Ability to Perceive Deepfake Images

Student Name: Matthew Fox

School of Cybersecurity, Old Dominion University

CYSE 201S: Cybersecurity and the Social Sciences

Instructor Name: Diwakar Yalpi

Date: 11 / 12 / 2025

What is the Article about?

The term “Deepfake” refers to an Artificial Intelligence that mimics real-life videos. They can create videos, images, and audio that closely represent how a person would normally create one of these. They are a concern because they can be used for malicious reasons, one huge one being for propaganda and fraud. This article attempts to understand if people are able to tell the difference between a Deepfake and an actual video.

How does this article relate to the Social Sciences?

The most impactful relation to the Social sciences would be Skepticism. In this experiment, they are expecting people to doubt that the video is fake, and in turn, realize the differences between the deepfake and a real video. Even with the real videos, they should take the time to consider the fact that it may possibly be fake. Relativism is another Social Science that is applicable here. Most people have a reason for creating a deepfake, whether it's out of curiosity or for their own gain. The realism of the deepfakes can help sway how people may feel about someone, which helps the creator of the deepfake to push their wants and beliefs onto other people. Finally, Empiricism helps simplify how we should identify and react to the deepfakes. Instead of having a gut feeling about a video being a deepfake, we should take the time to look over and find the flaws in case the video is a deepfake.

Research Question, Hypothesis, And Variables

Their research questions are as follows: “Are participants able to tell the difference between deepfakes and images of real people?”, “Do simple interventions improve a participant’s deepfake detection accuracy”?, and “Does the self-reported level of confidence a participant has align with their accuracy at detecting deep fakes?”. The authors hypothesize that because deepfakes are being integrated into our society, fewer people can identify and deal with

them appropriately. In this experiment, the independent variable is the alternating real-life person image and the Deepfake image, while the dependent variable is the accuracy with which people are able to correctly guess if it is a Deepfake or a real image.

Research Method Used

The researchers gathered 280 participants and put them into 4 groups, with one of them being the control group, with no intervention, and the others having an assistance intervention of different levels: Familiarization, One-time advice, and Advice with reminders. They presented the participant with 20 random images from a pool of 100 images, half of them real, the other half Deepfake. After the participants guessed, they were tasked with reporting their confidence levels and explaining why they felt that way.

Results of the experiment

From this experiment, they concluded that in all of the groups, there is a 60% accuracy rate. Depending on the group the participants were in, there was a slight increase, with the control group having a 59.65% accuracy rate, and Advice with Reminders having a 64.26% accuracy rate. In terms of confidence levels, it was discovered that the mean of the rate of confidence was high across all groups; the participants given advice had different types of it, the highest one being that “accessories do not make sense”. They presented all of this data in a simple chart, except for the Advice list, as that was presented in a bar graph.

Connections to the Concerns of Marginalized Groups

There is a concern that people outside of the experiment will have a harder time differentiating a Deepfake image from a real one. Bray et al. (2023) state that “ This means that participants would have likely viewed the images in a much larger format than such images would commonly be viewed in real life.”. To explain this, normally you would be viewing these

images in a smaller format, and they give an example of Instagram: “Instagram has a profile picture limit of 110 by 110 pixels. The effect of different image sizes upon participant accuracy is an area for future work, but it seems reasonable to suggest that certain details (e.g., the absence of symmetry for earrings) that would affect a participant’s decision would be hard or near impossible to see in smaller images” (Bray et al., 2023). Not only will the images viewed be smaller, but now it is harder to clearly tell if there is a flaw that could help identify if the image is an anomaly or not. The authors are also concerned about their results, stating that they believe this is a huge concern.

What do we gain from this? / Conclusion

This experiment shows the desperate need for a solution against Deepfakes. Even with the results being over half, there is still huge room for someone to improve a Deepfake and use it, and be able to manipulate people into thinking one way. This article also helps show what types of Deepfakes are effective and where they may commonly be, like a Social media app or dating site.. The authors suggest that a change in how individuals behave may be a solution for mitigating the damage that Deepfakes cause.

References

Bray, S. D., Johnson, S. D., & Kleinberg, B. (2023). Testing human ability to detect "deepfake" images of human faces. Journal of Cybersecurity, 9(1). https://doi.org/10.1093/cybsec/tyad011

Article Link: Testing human ability to detect 'deepfake' images of human faces | Journal of Cybersecurity | Oxford Academic