Are they deep fakes? Or an observation mistake? With today's riveting advancement in technologies such as Artificial intelligence the cyberspace landscape is becoming a new battle ground for cyber threat actors. When technology advances in the Information Technology realm so does cybercrime. In the article *Testing human ability to detect 'deepfake' images of human faces* the author aims to distinguish the difference between deep fake images created by A.I generators and genuine authentic images. From the information gathered from this document I will discuss the researchers questions or concerns, the type of research methods used, the type of data and analysis done, how the challenges and concerns relates to marginalized groups , the overall contribution on how this study helps our society and how it relates to the principles of social sciences.

Research questions

The first question researchers wanted an answer to was can participants in the study be able to differentiate between deepfakes and images of authentic real humans above chance levels? The second question was do basic interventions improve participants' deepfake detection accuracy? Lastly does a participant's self-reported level of confidence in their response match with their accuracy at detecting deepfakes? Obtaining answers to these questions can help give society a baseline of how easy and or hard it is to detect deepfake images.

Research Methods, Analysis & Concepts

The Empirical research method was utilized by the researchers in this study. They recruited participants from an online platform that offered them compensation at

an hourly rate for their participation to give them motivation. The study had a few different control conditions. The first being a baseline test of showing the participants 20 deep fake images and informing them of that. The second control was showing them a list of “tell-tale features” that deep fake images contain that can be used to discern between the two. They essentially trained them briefly to familiarize the participants and to keep a close enough baseline to have more precise qualitative results. The participants were later showed multiple images of deep fakes and authentic images and were asked their confidence from a scale of 1-10 if the image was an original or not.

Challenges and concern related to Marginalized groups.

There is a plethora of challenges and concerns when it comes to A.I deep fakes and marginalized groups. Deep fake images can be used to target a specific type of victim, a victim that may not even know that they are being targeted. An example could be Immigrants from a third world country moving to a 1st world country. Some immigrants may have not had access to the internet to now having access to the internet in abundance. People of this marginalized group may not be aware of such capabilities since they are not as familiar with the internet as others, and this makes them a sought-out target.

Principles of Social Science

The topic of study in the article relates to Empiricism. Within this study the researchers surveyed a group of participants and analyzed their responses to a controlled set of familiar scenarios of the A.I images. They effectively obtained the

information needed to validate their hypotheses by studying real data from the responses given by the willing participants.

Conclusion

In conclusion this study helped the researchers understand the challenges for the average person to spot a deep fake and their level of confidence in discernment. According to research “The findings of this study suggest that people are better than random but imperfect at detecting deepfakes and that the simple interventions tested do not help.” (Bray et al., 2023). This is very concerning and shows that there needs to be intervention and or public training to keep internet users aware of the deep fakes, so they won’t fall victim to cybercrime.

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