Article Review 2

Elizabeth Simpkins

3/31/25

The Automated Creation of Deepfakes to Try to Trick Society

Research question, methods used, and data analysis used

The research article "Testing human ability to detect 'deepfake' images of human faces" highlights the question of whether or not people are able to identify deepfake images over real images, and whether providing the information to identify them would be beneficial. The research consisted of one control group and three test groups. The first test group were made familiar with deepfakes, the second group was given leading signs to detect deepfakes before the experiment started, and the last group got the signs before and during the experiment. Participants were randomly chosen and were also made to answer a confidence scale on their answers for each image. The same AI-deepfake generating tool was used as well as the same real face generator for the experiment. The data was analyzed by including all of the data collected from whether or not each group answered correctly, had high or low confidence, and reasoning behind their answers using one or more of the signs provided to them. The overall analysis of the data provided led to the conclusion that even though the participants who had the information provided to them did slightly better than the control group that did not, these provided identifiers did not help as much as expected. The provided identifiers to finding deepfakes actually raised the confidence levels of participants in test groups 2 and 3, but did not help them identify the images correctly. The findings of this research article conclude that deepfakes are becoming even more hard to detect and knowing some identifiers can only slightly help if people are untrained to find them.

Connection to the principles of social science

This research first connects to the principle of parsimony due to its simple explanations. This can be seen in the statement that "participants tended to be confident in their ability to differentiate real and deepfake images but their confidence was misplaced" (Bray, 2023). This concludes that participants are not aware of how sophisticated deepfakes are because even when given the confidence in their ability to correctly answer, with identifying factors provided, they still chose wrong most of the time. The principle of empiricism can also be applied to this article. Empiricism is shown because the research question itself of whether providing materials to help people identify deepfakes may be assumed upon based on the relatable factors of it. This research abides by the empiricism principle because the researchers do not assume anything and base their results upon the facts provided. This article finally connects to the principle of ethical neutrality because not only does the experiment keep the participants anonymous, but the research question itself can have underlying ethical questions. The research covers a topic of controversy, AI. Some questions may arise that have to do with ethics such as: should AI be limited in order to prevent more AI related cybercrimes; how much should we allow AI to do? The research question connects with these questions and society as a whole because anyone could be affected by AI's use and deepfakes' misconceptions.

Relating to concepts from class

The first concept covered in this research article that was covered in class is the process of a social science research study. The study itself observes the responses of participants who must try to identify which images are deepfakes and which are not. This research tests the behavior of a group representing society, that will make real-life decisions. These observations help us determine that deepfakes can sometimes be too hard to identify, and identifying factors can only somewhat improve our ability to make the right decision. The topic of psychology is also discussed when the researchers include the testing factor of how confident the participant was in choosing different options and what they saw that influenced their decision. This can help the researchers understand what is going on in the participant's mind when making decisions. By understanding the participant's choices based on what they were given, the researchers can implement psyber security in order to provide information and tips on how to prevent more people from falling for deepfake scams. Another concept from class that connects to this article would be the topic of classical experiments. This is a type of classical cybersecurity experiment because the test groups were randomly assigned, there was a control group, pre-training, extra training for experimental groups, a post test of their behaviors, and finally a comparison of results. The final concept that this article covered was social engineering. Deepfakes are covered under the topic of social engineering because they are a way to manipulate and convince their target into believing what the image states or getting them to follow through with a certain set of commands.

Connections to marginalized groups and societal contributions

The fact that this study was only done with men and women between the ages of 20-29 limits the groups that could be marginalized by deepfakes. One example of this problem would be an older generation of people that may not know or have any indication that certain images can be deepfakes. The article does mention older populations at one point, because they tend to fall for video deepfakes more due to the presence of them on tv and YouTube. Another marginalized group for this topic could be disabled people. There are many programs for computers that can help people who are blind, for example, find their way around sites, but what about deepfakes? When these come up, how would a blind person be able to decipher between a deepfake video and a real one? Would providing identifying factors to disabled people be able to help them if they are surfing the internet and receiving false information from deepfakes? Societal contributions from this article could be the drive to find ways to help society as a whole by providing ways to find deepfakes and avoid falling for them. Another societal contribution would be understanding how people think when making decisions about deepfakes, which can help guide research into the development of deepfakes and how society can grow and learn with them. Both of these concerns can help society by learning about this scam and knowing how to avoid them.

Conclusion

The creation of deepfakes will only get more sophisticated and the use of them by cybercriminals can be dangerous. Therefore, the public needs to be even more aware about where and who they get their information from online. The study of society as a whole about how people interpret and make choices about information, whether it is true or not, determines how researchers and scientists can develop ways to help people make the right decisions. The conclusion of this research study was that by providing identifying factors, only a small percentage of correct answers was added to the scores. Maybe by providing potential consequences to choosing the wrong image, people can be more aware of their choices, because this would resemble a real-life situation. Annual training for people on AI-based deepfakes can also be beneficial to keep people accustomed to the ever-evolving world of AI-cybercrimes.

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