

Nigel Adkins
2/26/2026

In the article “Testing human ability to detect ‘deepfake’ images of human faces” by Sergi D Bray, Shane D Johnson, and Bennett Kleinberg, they examine if humans can detect deepfakes on human faces to see the true power of the emerging technology. This topic relates to the principles of social sciences because of how it examines human behavior and interaction and its use of the scientific method and research.

Background

The article gives a background on deepfakes and states that they are ranked as one of the most serious AI threats. The research question aims to find out the human ability to detect deepfakes on Human faces. In the experiment, the researchers conducted an online survey and used a sample size of 280 participants with around 40% being female and 26.24 being the mean age. Each participant completed 20 trials where images would be selected from a group of 50 real and deepfake images. The 50 real images were drawn from Flickr and a high resolution image dataset called FFHQ. The images that were Deepfake were generated by the StyleGAN2 deep learning algorithm which was trained using tens of thousands of Images from the FFHQ dataset.

Procedure

The procedure included the participants reading about the aims of the study and learning about real images and AI generated ones. They were then given instructions on how they would complete the task. All the participants were told how their data would be stored and told that they have the right to withdraw at any time. They were given the author contact details, they provided informed consent, and then allocated into different study conditions. The participants were

shown one image at a time and they were asked to label the image as “AI generated ” or “Real” and rated their confidence on a scale of 1 to 10, collecting both quantitative and qualitative data.

Results

The results of the experiment showed that on average participants were correct about 60% of the time and that there was no statistically significant difference between any of the groups of the participants. There was a slight increased accuracy among those with experimental conditions then in the control group.

Relation to class and challenges/concerns

This Study relates to the topics in the Powerpoint presentations we learned in class because of the complexity of the scientific research and its examination of human social behavior and their ability to detect a difference in something that may look familiar to them. I believe that this study highlights the challenges and concerns about the abuse of new and emerging technologies like AI and deepfake images could lead to misinformation, fraud, and identity theft especially among marginalized groups like the elderly.

Conclusion

Conclusion; This study contributed to society by putting focus on the realistic nature of Deep fake images as shown by how 40% of people in the experiment were incorrect about the images. This study was a call to action for people to be more observant, skeptical, and knowledgeable about the deceptive nature of new technologies.

*Testing Human Ability to Detect 'Deepfake' Images of Human Faces | Journal of
Cybersecurity | Oxford Academic,*

academic.oup.com/cybersecurity/article/9/1/tyad011/7205694

Link to article

<https://academic.oup.com/cybersecurity/article/9/1/tyad011/7205694?searchresult=1#415793212>