Article Review #1 Testing human ability to detect ‘deepfake’ images of human faces

Kaniel Stephens
School of Cybersecurity, Old Dominion University
CYSE 201S: Cybersecurity and the Social Sciences

Diwakar Yalpi
2026/02/20

Introduction/BLUF

In this day and age people’s trust in institutions are at all time low but this seems to be opening up a market for false narratives online that play into base desires that mainstream outlets try to avoid trafficking in. In this environment the use of false digital images like deepfakes is on the rise. In this article the researchers have set out people's ability to detect these sorts of images. The main three questions stated in the article are: “Are participants able to differentiate between deepfake and images of real people above chance levels? Do simple interventions improve participants’ deepfake detection accuracy? Does a participant’s self-reported level of confidence in their answer align with their accuracy at detecting deepfakes?”(Bray et al., 2023)

Relation/Connection to Social Science Principles

The study of deepfakes and the human reaction to false or preternatural images can fall into a number of camps, namely psychology. The effect these images have on society and the way we interact can fall under sociology.” In the real world, when people make the types of decisions that deepfakes may be intended to affect, decision-making is unlikely to be quite so binary. People may have their doubts about an image, which may in turn affect the decisions they make.” (Bray et al., 2023) Says the authors of the paper

and this shows how these images permeate our world. In this study they also mention the dissemination of false images can have a political impact as well.

Research Question /Hypothesis/ Independent Variable/Dependent Variable

The question that the researchers set out to answer is if people find real human faces more trustworthy than deepfakes. The answer to this would help us understand if people can differentiate between deepfakes and real faces. The independent variable is the amount of advice the test groups are given beforehand. The three variables are familiarization, one time advice, and advice with reminders. The dependent variable is the trustworthiness of the pictures.

Types of Research Methods used

The study uses a quantitative mode of research to study and test how many people are able to differentiate between fake and real faces. The data was gathered by testing a pool of people's reactions to a series of images depicting both real and fake faces.

Types of Data Analysis used

The study uses inferential analysis to assess how the population at large reacts to fake images online. By breaking down the data gathered we can make assumptions about larger sizes of people. There is also an effort to study the data through a causal lens to define what impacts people's ability to perceive falsehoods.

Connections to other Course Concepts

This study is a good example of a classical experiment that uses randomization and would stand up to scrutiny. There are aspects of relativism as it discusses the wide ranging impact of deepfakes on society.

Connections to the Concerns or contributions of Marginalized Groups

The study did not involve any minority groups and while this matter impacts us all there might be a need to discover how misinformation might impact the more vulnerable in society. “Of the 273 participants who volunteered their nationality information, 15 (5.5%) were UK nationals, while 229 (84%) were (non-UK) European nationals.” (Bray et al., 2023)

Overall societal contributions of the study/Conclusion

Identifying and addressing deepfakes is a very important step in internet security. The web is a melting pot of fact and fiction where nothing is real and to combat that we need a guiding hand to keep people safe. People should have some reasonable degree of certainty that if their grandson facetimes them it's really them. So, the conclusions found in this study can help people address the failure of our abilities in the field of discerning reality.

Reference

Bray, S. D., Johnson, S. D., & Kleinberg, B. (2023, January). Testing human ability to detect ‘deepfake’ images of Human Faces | Journal of Cybersecurity | Oxford academic. https://academic.oup.com/cybersecurity/article/9/1/tyad011/7205694