

Article Review #2: Testing Human Ability to Detect Deepfake Images of Human Faces

Student Name: Zack Fok

School of Cybersecurity, Old Dominion University

CYSE 201S: Cybersecurity and the Social Sciences

Instructor Name: Diwakar Yalpi

Date: 4/14/2026

Introduction/BLUF

The publication "Evaluating Human Ability to Detect Deepfake Images of Human Subjects" looks at how well people can identify the difference between a live human and an AI deepfake produced digitally. Artificial Intelligence has advanced at such a rate that the use of deepfake software to make another person appear to look like someone else, even in videos. There are many serious security hazards posed by these Deepfake technology have raised issues about digital security, false information, and trust in technology. The purpose of the investigation is to assess the ability of humans to detect a deepfake photo or a digitally produced image and to explore whether quick intervention can enhance or lessen the ability of a person to correctly identify a deepfaked picture. BLUF: People struggle to detect deepfakes, performing near chance level while remaining overconfident in their judgments.

Relation/Connection to Social Science Principles

From the results, it can be concluded that the ability of people to correctly identify digital images produced by a deepfake was almost 50%. The confidence of the participants was not matched by their real accuracy, which identifies serious weaknesses in human perception abilities. This research reflects several principles of social science. First, it is based on empiricism, which collects data from participants in an experiment and uses the scientific method by developing hypotheses before testing them through experimentation under controlled conditions. The study also reflects ethical neutrality, as the researchers do not judge or blame participants for their inability to accurately detect deepfakes, instead focusing on understanding human cognitive limitations rather than blaming them for poor recognition. Finally, the findings point to evidence of determinism, meaning that people have limitations in their ability to act or decide and thus determine the outcomes of the study.

Research Question /Hypothesis/ Independent Variable/Dependent Variable

The central research questions are: can humans accurately detect deepfakes? Do interventions increase accuracy in detecting deepfakes? And do humans' levels of certainty correspond to their actual performance? The hypothesis is that participants' accuracy in detecting deepfakes will be just above chance level but will improve when they receive guidance or additional help. The independent variables in the study were: type of intervention; and type of image. The dependent variables were the accuracy in detecting deepfakes and whether confidence indicates how accurate participants were.

Types of Research Methods used

In the current study, the quantitative experimental method was used to gather data about deepfakes, as around 280 participants were surveyed using an online survey. Measurements of accuracy and confidence in detecting deepfakes were measured after participants classified images as being either real or fake. An experimental design was selected to conduct the study using a quantitative approach.

Types of Data Analysis used

Researchers collected data from participants using an anonymous online survey that contained images, either true images or deepfakes, and asked them to classify the image into one of the two categories and report their confidence level. The structure of the data collection process enabled the researchers to assess accuracy and compare the findings' results from each condition.

Connections to other Course Concepts and Connections to the Concerns or contributions of

Marginalized Groups

This article is relevant to topics studied in this course through reasons such as the importance of humans in the identification of fake people. A specific example is the vulnerabilities of people who can be affected by social engineering and the risks of overconfidence. The results of this article are especially significant to members of vulnerable groups, like the elderly, who may be more prone to being victims of deepfake abuse. These groups have a limited ability to identify digital technology fraud, so using the research they could learn tactics to properly spot scams and fakes in the future.

Overall societal contributions of the study/Conclusion

In conclusion, this paper illustrates the poor ability of humans to identify AI deception or deepfakes, and the need for better tools and training, public education and awareness towards these as they're becoming more common in our everyday lives, and as well as our understanding of the intersection of social science and cybersecurity.

Reference

Testing Human Ability to Detect 'Deepfake' Images of Human Faces | Journal of Cybersecurity |

Oxford Academic, academic.oup.com/cybersecurity/article/9/1/tyad011/7205694.

Accessed 16 Apr. 2026.

Article Link: <https://doi.org/10.1093/cybsec/tyad011>