Article Review #1

Chandler A. West

Old Dominion University

Course Number: Course Name

Professor Yalpi

10/2/2025

Human Ability to Determine AI Deepfaking

Introduction

The article *"Testing human ability to detect 'deepfake' images of human faces"* looks at how well people can tell the difference between real photos of human faces and fake ones created by artificial intelligence. Deepfakes are a growing problem because they can spread lies, harm reputations, and cause people to lose trust in what they see online. This review explains how the study connects to social science, what research methods were used, what data was collected, and why the findings matter to society.

Relation to Social Science Principles

This study connects to social sciences because it deals with human behavior, judgment, and decision-making. Social sciences often ask how people think, how they interact with technology, and how society reacts to new problems. In this case, the research shows how people try to spot fake media and what mistakes they make. It also relates to bigger social issues, like the spread of misinformation and how people build or lose trust when using digital tools.

Research Question, Hypotheses, and Variables

The main research question was: Can people correctly tell the difference between real face photos and AI-generated deepfake faces, and does training or advice help them do better? There were three thoughts the researchers had before performing this experiment. The researchers thought people would do better than randomly guessing. They thought that advice and training against deepfakes would make them more accurate. They felt people were more confident and would also be more correct. In this situation, the independent variable was the type of help given to participants (no help, practice with fake images, advice before the task, or advice

during the task). The dependent variable was how accurate participants were when deciding if an image was real or fake. Another dependent variable was their confidence level.

Research Methods

The study used an experiment with four different groups. Participants were recruited online through a platform called Prolific. They looked at a mix of 20 real and fake face images and had to say if each one was “real” or “AI-generated.” They also rated how confident they were and explained their reasoning. This method was mostly quantitative because the researchers measured accuracy and confidence in numbers. There was some qualitative data though, which were the participants’ written explanations for what they chose.

Data and Analysis

Some of the data that was included in this experiment were participant confidence ratings, participant answers on whether images were fake or real, and the participant’s reasons for their choices. The researchers compared average accuracy across the groups, checked if people were better than random guessing, and looked at whether confidence matched accuracy. They also examined which images were easier or harder for people to judge.

Class Presentation Relations

From Module 3, the powerpoint explains that social scientists often use experiments to study cybersecurity. In these experiments, researchers test how people behave when faced with certain situations or tools, and whether small changes can affect their choices. The deepfake article is a good example of this approach. In the study, the researchers split participants into different groups: one group had no extra help (control group), one group practiced by looking at fake images, and other groups were given advice about spotting fakes. They then tested everyone to see if these steps helped people do a better job finding deepfake images from real ones. This

shows how experiments can also be used in cybersecurity research to understand how people think, make decisions, and react to digital threats.

Marginalized Groups

Deepfakes are especially harmful to marginalized groups. Women and minorities are often the targets of fake images, such as non-consensual sexual content or manipulated videos meant to spread lies. Since this study shows that most people cannot reliably detect deepfakes, it means these groups are more at risk. The research indirectly highlights the need for stronger protection, because people cannot depend on their own judgment to tell what is real online.

Contributions to Society

This study helps society by showing the limits of human ability to detect deepfakes. It found a couple of important points. First, people are only a little better than chance when spotting fake faces. Second, training and advice did not really help anyone. Third, confidence does not match accuracy, which means people may think they are right even when they are wrong. These findings are important because they show that humans alone cannot solve the deepfaking problem. Instead, we need better technology to detect fakes, stronger policies to protect people, and more education about the risks. This is especially important to protect vulnerable groups and to keep public trust in online media.

Conclusion

This article review shows that people struggle to tell real and fake faces apart. While people can sometimes spot deepfakes, their performance is inconsistent, and advice or training does not make much difference. The study's results highlight the importance of building tools, laws, and awareness to fight against deepfakes. Overall, this research adds valuable knowledge to how society can respond to digital threats and protect individuals from harm.

References

Sergi D Bray, Shane D Johnson, Bennett Kleinberg, Testing human ability to detect 'deepfake' images of human faces, *Journal of Cybersecurity*, Volume 9, Issue 1, 2023, tyad011, https://doi.org/10.1093/cybsec/tyad011