Article review #2: Deepfakes and the Growing Population of Fake Users

Daniel BrzezinskiRwockwell

CYSE 201S

Diwakar Yalpi

11 April 2025

Introduction

Artists, since humanity started to look into itself, have tried to transpose the human condition onto a canvas, cave wall, or any hard surface. Vincent Van Gogh, with his impressionist style; Pablo Picasso, with surrealism; and Fransico Goya, with romanticism artwork, were revered and honored throughout history. In today's world, there is a new revolution of artists that has the potential to cause real and permanent damage. This new revolution is generative AI creating *'deepfakes.'* The following article *Testing human ability to detect 'deepfake' images of human faces,* asks 280 individuals to determine if they can tell if the image of a person shown to them was real or fake and how confident they were in their answer.

Research Questions Asked and Research Methods used

The main and probably the most important question this article asks and attempts to answer is whether people are able to discern between authentic and generated images of human faces. This type of deceitfulness has existed for many years, where the easily susceptible are fooled and doped out of their belongings. Another question asked and addressed was whether the person's odds in determining the fake images if they were given a guide or trained beforehand. The researchers then attempted to determine if a person's accuracy correlates with their confidence in guessing.

To experiment, the researchers split the sample size into a control group, a group familiar with deepfakes, people given a one-time training beforehand, and finally, a group that was constantly coached. The test group participated online and was shown 20 out of 100 that were 50/50 split between real and fake images to tell whether they were a deepfake and how confident they were with their answer. At the end of the research, it was noted that "...although participant accuracy was 62% overall, this accuracy across images ranged quite evenly between 85 and 30%, with an accuracy of below 50% for one in every five images." (Bray et al., 2023) It was also noted that users' odds of correctly guessing the right image did not go up with coaching.

Data and Analysis

The expirment perfromed by the researchers collected accuracy rates, confidence ratings, reasonings for picks, and heat maps of what drove the participants to make their decisions. For analyzing the data, they used descriptive statistics and correlation analysis to breakdown the accuracy and correlation to confidence.

Concepts from Lecture Applied in the Article

Though the participants had higher confidence in correctly picking the right images as fake, a heavy case of overconfidence bias was noted. The correlation in whether the participant felt confident in their choice could be lad to the growing distrust in online media and news outlets, "...according to 2020 statistics from Reuters, 56% of people (sampled across 40 countries) are

concerned about the veracity of news found online." (Bray et al., 2023) Even with heightened skepticism for online content, concerns are deep with the growing popularity of cybercrime utilizing deepfakes.

Challenges, Concerns and Contributions of Marginalized Groups

In these scams, many of those affected are women who are targeted with generated explicate images of their likeness. Another group that can easily stir an emotional response is political misinformation specific to promoting the ideals of one side over the other or an attempt from outside actors to disrupt a rival country's political system.

Contribution and Conclusion

Even with a not-as-large sampling size for the experiment, the article has shown a significant concern about the fidelity of deepfakes. The research also shows that new or different methods are needed to help educate people on recognizing deepfakes, as their instruction did not raise guessing probabilities. With a better way to educate people, rules and laws should be put in place to help deter illicit activities with deepfake-generated media. For a follow-up to gain more data, a larger group should be brought together.

References:

Bray, S. D., Johnson, S. D., & Kleinberg, B. (2023). Testing human ability to detect 'deepfake' images of human faces. *Journal of cybersecurity (Oxford)*, *9*(1). https://doi.org/10.1093/cybsec/tyad011