AI, or Human?

Alex G. Bretana

Old Dominion University

CYSE201: Cybersecurity and the Social Sciences

Professor Yalpi

9/30/2025

After reading the article “Testing Human Ability to Detect “Deepfake” Images of Human Faces,” I found the results both surprising and concerning. The study looks at a topic that is very important in today’s digital world; how well people can tell the difference between real photos of human faces and fake ones created by artificial intelligence. The main takeaway is that we are not very good at it, and worse, we are often confident even when we are wrong.

This topic connects strongly to the principles of social sciences we learned in class. The study focuses on human behavior and perception, which is a core part of psychology. It also looks at how this technology can be used for crime, which relates to criminology. The researchers used an experimental method, which is one of the main research strategies in social science. They applied principles like empiricism by gathering real data, and skepticism by testing common-sense ideas about training people to spot fakes.

The study had three clear research questions. First, can people tell deepfakes from real images better than random guessing? Second, do simple interventions like training or advice help people get better at spotting fakes? Third, is there a connection between how confident people feel about their choices and how accurate they actually are?

The researchers set up an online experiment with 280 participants. The independent variable was the type of help participants received. They were randomly split into four groups: a control group that got no special help, a group that saw examples of deepfakes beforehand, a group that received advice about what to look for, and a group that got the same advice plus reminders during the test. The dependent variables were whether participants correctly identified each image as real or fake, and how confident they felt about each decision.

For data collection, each participant was shown 20 images—half real and half AI-generated—and had to label each one. They also rated their confidence for each choice and

explained their reasoning. The researchers then analyzed this data using statistical tests to compare accuracy between groups and to check if confidence levels matched actual performance.

The findings were clear and troubling. Overall, people were only correct about 62% of the time, which is just slightly better than guessing. None of the interventions—familiarization, advice, or reminders—significantly improved this overall accuracy. Even more concerning was the relationship between confidence and accuracy. When people reported high confidence in their choices, this confidence had almost no connection to whether they were actually right. This means someone could be completely wrong but still feel certain they were correct.

This research matters because it shows how deepfakes can particularly harm marginalized groups. As the article notes, fake content can be used to harass or silence people, especially women and minorities. If people can't tell what's real online, it can lead to reputational damage, spread of misinformation, and make certain groups less likely to participate in public life. The fact that people are both bad at detecting fakes and overconfident makes this threat even more dangerous.

The study's main contribution to society is showing that we can't rely on simple solutions like public education to solve the deepfake problem. The researchers tested common-sense interventions and found they didn't work. This tells policymakers and technology companies that we need better solutions, whether through more advanced detection technology, new laws, or different approaches to digital literacy. The study serves as an important warning about our collective vulnerability to this emerging technology.

In conclusion, this research reveals a serious gap in our defenses against digital deception. We're not good at spotting AI-generated faces, and simple training doesn't help much.

Worse yet, we're often confident when we're *wrong*. As deepfake technology becomes more common and convincing, these findings show we need better protection strategies that don't rely solely on human judgment. Without them, our world is quickly going to become a very dangerous and deceptive place.

Works Cited

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