Testing Human Ability to Detect Deepfake Images of Human Faces
Morris Smith 11-1-2023

Introduction

This article discusses the growing concern around "Deepfakes," these are generated images or videos made by people with the intent to deceive victims into believing something is real when it isn't. These deepfakes pose a threat to cybersecurity and public safety as a whole. Deepfakes can be used for a number of malicious things such as impersonation, identity theft, privacy invasion, frauds and scams along with so much more.

Research Questions/Hypotheses

The research questions were as follows, "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?" All of these questions were taken into account when conducting this experiment.

Collecting the Data

They answered these research questions by conducting three studies, the first of which consisted of showing static images of deepfakes to the participants. This was meant to help them familiarize themselves with the concept of deepfakes and how they can look, they were then asked to label new images as real or deepfake. Different kinds of deepfakes were used and they all had varied levels of accuracy. The second study consisted of participants labeling images without familiarizing themselves with the images first. Some were taught on identifying some aspects of the image that would give it away as a deepfake and the rest received no help. Those who were taught performed slightly better but there wasn't enough of an improvement to make it anything more than chance. In the final study participants were given images in pairs with one being a deepfake and one being real. This was the lowest score in terms of accuracy out of every study.

Contributions

This experiment made many contributions such as highlighting how dangerous this is and how it shouldn't be open to everybody. It shows the decision-making process that happens when presented with a deepfake and highlights the relationship between confidence and accuracy in one's assumption if it's real or fake. These finding expand the harm of this technology and how it can impact millions of people.