The Human Perspective of Altered Media

Nneka Udeagbala

Computer Information Sciences and Engineering,

Herbert Wertheim College of Engineering, University of Florida,

Gainesville, FL 32611, USE

Abstract. Deep fakes introduce a new level of complexity by creating distrust towards video evidence of events. Our research discusses the utilization of social media and news sources for media authentication, lessening the impact of published altered videos. User studies were used to investigate human perception limitations and dependency on media sources for detecting altered media.

Keywords: Technology, Media, Democracy, Digital Literacy, Human factors, Deepfake

1 Introduction

1.1 Motivation

Deepfake technology introduces a new level of complexity to the fake news phenomenon by adding distrust towards video evidence of events. There is a need to identify the limitations of human perception while viewing the altered media while locating means of mitigating distrust. Though research has been done on edited audio and images and AI detection of altered videos, no research has been done on the human detection of altered videos. Our research discusses the utilization of social media and news sources for media authentication, lessening the impact of

published altered videos. User studies were employed to investigate human perception limitations and dependency on media sources for detecting altered media.

Photoshop allowed humans to create the most rudimentary version of image alteration. The secondary step that lead to the creation of DeepFake technology can be assigned to the software used to identify people made by companies such as Google and Facebook. The last step was the creation of generative adversarial networks (GANs). Generative adversarial networks consist of a "generator" network that creates images and a "discriminator" network that evaluates their authenticity. (note: encoders vs decoders) The term "adversarial" is derived from the relationship between the generator and discriminator: the generator gets better at producing fake images and the discriminator gets better at detecting it. GANs, created by Ian Goodfellow, continue to learn without human supervision. There is a setback, if the generator does not get better, the discriminator will not either. Nvidia used a database of over 200,000 celebrity images to train its GANs which was then used to produce realistic images of individuals who do not exist. (Greenemeier)

The motivation for this research was the wide-spread alarm at the sophistication of these forms of media. As stated above, organizations have begun using GANs to create fake individuals. It should be noted, however, that reactionary articles have been published to highlight ways to detect fake pictures of individuals (McDonald). This leads us to question what types of methods can be applied to prevent individuals from falling victim to the new age falsified media created with machine learning. Understanding perspective and how angles lead to distortion is also vital when considering doctored images. While the machine can take a face and recreate it, there are unrealistic distortions that occur due to the absence of reality. Thus, when taking visual doctoring into account in the conversation of distortion, do we consider the angle at which the image was

taken? These types of questions were analyzed at the beginning of this project. However, to obtain a strong basis of research, it was decided that much simpler questions be asked. Therefore, we settled on approaching the problem by mostly asking whether or not individuals could tell real from fake.

1.2 Approach

Deep fake technology has been an increasing concern in the fight against the dissemination of fake news. This project is centered around the overlap between Technology, Media, and Democracy. To explore the phenomenon of sophisticated altered media, we compiled three main learning goals:

1. Do individuals detect Deepfakes?
2. Do individuals trust news/social media outlets to determine if the media is fake?
3. What context clues help the viewer come to a conclusion?

1.3 Methodology

A selection of real and fake videos were amassed from multiple online sources. These videos were placed into the context of news sources or social media sources either showing the material or alerting the viewer of its lack of authenticity. A survey was created within which the taker would see four videos, real or fake, in any or all of the created contexts. 32 videos were pulled together for the creation of this survey though only 4 are shown to the taker. The environments created are as follows:

- News outlet reports fake video
- Social media outlet reports fake video

- News outlet reports with video (realness neither confirmed nor denied) “this video appeared on…”
- Social media outlet shows video (realness neither confirmed nor denied)

2 Results

2.1 Hypotheses

- Most individuals who state a video is familiar will assume it is real.
- Half of the subjects will default to believing what they are told or simply trust the source on at least one video.
- Lack of audio will result in a “neither agree nor disagree” response in over half of respondents presented with that type of video.
- Most participants will look at the background to determine authenticity.

3 Conclusion

3.1 Challenges and Limitations

The videos were obtained from the internet, therefore the material available was dependent on the interests of those producing Deepfakes and altered videos. This caused some complications because it was not as easy to control the types of media being used as test material. Another challenge was that study requires a large amount of material and participants. Finding a large amount of good quality edited media, then obtaining real video of a similar genre, was moderately difficult to achieve. This contributed to the reasoning behind having both Deepfake media and Photoshopped videos as materials within the survey. Lastly, written out responses are

not guaranteed, therefore much more investigation will be needed to obtain concrete answers to any of the study goals.

3.2 Future Work

The next step for this research would be to conduct interviews with individuals watching Deepfake videos and providing real-time commentary. This type of inquiry would provide more insight into how individuals think through the altered videos they are provided with.

References

Note: many sources were utilized in gaining insight on the topic, though none were used in the development of this project.

Cutting, J.E. Behavior Research Methods, Instruments, & Computers (1997) 29: 27. https://doi.org/10.3758/BF03200563

Cutting, J. E. (1987). Rigidity in cinema seen from the front row, side aisle. Journal of Experimental Psychology: Human Perception and Performance, 13(3), 323-334.

Greenemeier, Lawrence. “Spot the Fake: Artificial Intelligence Can Produce Lifelike Photographs.” Scientific American, Springer Nature, 1 Apr. 2018, www.scientificamerican.com/article/spot-the-fake-artificial-intelligence-can-produce-lifelike-photographs/?redirect=1.

McDonald, Kyle. “How to Recognize Fake AI-Generated Images.” Medium, Medium, 5 Dec. 2018, medium.com/@kcimc/how-to-recognize-fake-ai-generated-images-4d1f6f9a2842.

Sakshi Agarwal, Lav R. Varshney (2019). Limits of Deepfake Detection: A Robust Estimation Viewpoint