The Report Committee for Lu Jin
Certifies that this is the approved version of the following Report:

Cheated by Deepfakes?
Deepfake Detection Ability, People's Reactions, and Ethical Implications

APPROVED BY
SUPERVISING COMMITTEE:

Dr. Kenneth R. Fleischmann, Supervisor

Dr. Danna Gurari, Member

Cheated by Deepfakes?
Deepfake Detection Ability, People's Reactions, and Ethical Implications

by

Lu Jin

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

The University of Texas at Austin

May 2020

Acknowledgements

I would like to thank Dr. Kenneth R. Fleischmann for serving as my supervisor and Dr. Danna Gurari for serving as my second reader for this interesting and inspiring master's report. I also appreciate the guidance and technical support provided by Nitin Verma. Given the challenges of this unexpected special situation of a pandemic shutdown, completing this report is particularly special and memorable.

Abstract

Cheated by Deepfakes?
Deepfake Detection Ability, People's Reactions, and Ethical Implications

Lu Jin, MS

The University of Texas at Austin, 2020

Supervisor: Dr. Kenneth R. Fleischmann

Recent dramatic developments in the fields of computer vision and deep learning technology have opened up a range of possibilities not previously imagined. The applications of computer vision technology include manipulating any face in any video and changing the environment of photos, just to name a couple of the new applications. However, these applications are already having impacts on our everyday lives. Given these recent advances in computer vision technology, people may not be able to trust images and videos we see on any media channel. These videos and images have the potential to deceive us.

Throughout the history of technology development, the pros and cons of new technology are often in dispute. New technology is often sensationalized in terms of the benefits for people, which may go beyond anyone's control and imagination. For example, the internet was started with a goal of developing a decentralized network. However, due to how it was commercialized in use, the Internet actually became more centralized than had been intended. Since a centralized platform has the advantage of

controlling all users' data and information, these can be sold to companies to help them engage in targeted marketing. Thus, the Internet fell short of its expectations and hype. Now, the focus and hype has largely shifted to artificial intelligence. In which direction will these new technologies go? What are humans' relationships with these emerging technologies? How can we use this technology safely and ensure that it leads to a future that we want? This goal is the starting point for this report.

In this report, I will use the latest FaceForensics++ dataset as a base for an experiment to answer three research questions: First, how well do people detect deepfakes, and what factors affect their ability to detect deepfakes? Second, what are their reactions when deepfakes are revealed ? Third, what do they see as the ethical implications of deepfakes, and how deepfakes could be used or abused?

For RQ1, I explore the elements that can help people detect deepfakes. For RQ2, I evaluate their reactions. For RQ3, I explore how they perceive the ethical implications of deepfakes. More generally, my findings offer guidance for thinking about how to rebuild trust in video data in an era of deepfakes?