UB

University at Buffalo

Statement for the Record

Dr. David Doermann

Professor, SUNY Empire Innovation

Director, Artificial Intelligence Institute

University at Buffalo, Buffalo, NY

FOR A HEARING ON

The National Security Challenge of
Artificial Intelligence, Manipulated Media, and "Deepfakes"

BEFORE THE

PERMANENT SELECT COMMITTEE ON INTELLIGENCE

U.S. HOUSE OF REPRESENTATIVES

Thursday, 13 June 2019

Chairman Schiff, Ranking Member Nunes, distinguished members of the Committee, thank you for the opportunity to be here to discuss the challenges of countering media manipulation at scale.

For more than five centuries authors have been using variations of the phrase “Seeing is believing”, but in the past half-decade, we have come to realize that this is no longer always true.

With the convergence of mobile devices that have put cameras in everyone’s hands 24/7, the ability to instantly share content on social media, and the proliferation of content editing tools, we are being exposed to a growing volume of edited and manipulated content. While much of the manipulation is benign, and much of the content is being created for entertainment purposes, there is so much of it that people are being desensitized to the harm of misconception and false content.

If we get to the point where someone can point to any audio or video content, claim that it is false, and we believe him or her unless it is proven to be true, then we will be facing a serious threat to democracy.

In late 2013, I was given the opportunity join the Defense Advanced Research Projects Agency (DARPA) as a Program Manger where I was able to address a variety of challenges facing our military and intelligence communities. One of these challenges included our limited ability to analyze, detect and address manipulated media that is being used with increased frequency by our adversaries. At the time, tools for photo and manipulation, many of which had noble purposes, were being developed at a rapid pace, yet we were not actively addressing automated tools to detect and counter this threat. It was clear that our manual processes, despite being carried out by exceptionally competent image analysts, could not deal with the problem at the scale new manipulated content was being proliferated

DARPA’s media forensics program, MediFor, was created to address this problem, and it is doing so with some of the most talented researchers in the world. They are developing a comprehensive suite of tools to address the technical issues of detecting and localizing manipulation in images and video. In typical DARPA fashion, the government got ahead of the problem, knowing that this was a marathon, not a sprint, and the program was designed to address both current and evolving capabilities.

The creation of synthetic image and video content, and manipulation of real content is nothing new. Hollywood has been doing this for years, but it used to be a highly manual process that involved specialized software, expensive hardware, expert operators and days or weeks if not months to create. Things have changed however. Today, the process of content creation and media manipulation can be automated. Software can be downloaded for free from online repositories, it can be run on your average desktop computer with a GPU card by a high school student and it can produce personalized, high-quality video and audio overnight that is either completely synthetic, or looks or sounds like a given person. We can make people dance in ways they have never danced, and put words into people's mouths that they have never said.

The technology that is the basis for these recent developments is "deep learning". Deep learning is a way of realizing machine learning in either a supervised or unsupervised fashion. Over the past decade, deep learning has revolutionized the way many machine learning problems are approached. After initial signficant impact in computer vision, speech recognition and a variety of human language applications, deep learning is now being widely applied to all data intensive domains for classification and prediction. With the convergence of better computer power, more memory and massive amounts of data, we have tools which can learn more accurately than before, from existing data.

In 2014, a paper was published popularizing the concept of a Generative Adversarial Networks (GANs) which allowed systems to automate the process of creating and manipulating an underlying distribution. In a matter of years, we moved from having a plethora of tools for **manually** manipulating images and video, to tools that can both manipulate and create image and video content, completely **automatically**. Now our adversaries can take the human out of the loop and literally bombard us online with deceptive content. Since 2014, we have seen an exponential increase in the quality of the resulting media, from content that was just ho-hum and nowhere near what could be done manually, to content that is practically indistinguishable from real content to the untrained eye. Soon even the trained eye will no longer be able to distinguish between what is real and what is not.

There is nothing fundamentally wrong or evil about this technology. Like basic image and video desktop editor it is only a tool. In fact, without machine learning and in particular deep learning methods, the community would not have progressed as far as it has in being able to detect trace evidence of manipulation.

I need to make it clear however, combating synthetic and manipulated media at scale is not just a technological issue. It is a social one as well. There have been a number of high-profile manipulations and creations of false content, yet even after being sufficiently discredited, the content continued to be shared repeatedly, and often not for entertainment or demonstrative purposes.

I believe this is due in part to a convergence of many factors occurring in society. We now have a public that is willing to accept less than truthful information. Many people are now getting news from the internet and from social media sites, rather than reputable organizations that do their best to vet content using journalistic best practices. There is a signficant difference between factual errors in the media, and media that is being created to purposefully deceive. People are becoming desensitized to the quality of their news and in many cases no longer care if it is fundamentally true or not. They do not see themselves are part of the problem if they spread information that is inaccurate or downright false, yet they are. Until we can change the culture to one where people and organizations do their best to ensure what they are sharing is true, it will be difficult to make signficant progress in combating this problem.

There is no easy solution, and it will likely get much worse before it gets better. Yet we have to continue to do what we can.

- We need to get tools and processes in the hands of individuals, rather than relying completely on the government or social media platforms to police content. If individuals perform the “sniff test” and media fails, they should have ways to verify or prove it.
- We need to continue to work toward being able to apply “automated” detection and filtering capabilities at scale. It is not sufficient to only analyze questioned content after the fact. We need to be able to apply detection capabilities at the front end of the distribution pipeline. And even if we don’t prevent it from appearing, it should come with the appropriate warnings that suggest that it is not real or not authentic.
- We need to continue to put pressure on our social media companies so that they realize that the way their platforms are being abused is not acceptable. And that they must do all they can to address todays issues, and not allow things to get any worse.

Let there be no question that this is a race. The better the manipulators get, the better the detectors need to be. And there are certainly orders of magnitude more manipulators in the race than detectors.

It is also a race that may never end, and may never be won, but one where we must close the gap and continue to make it less attractive (financially, socially, politically) to propagate false information. Like spam and malware, it may always be a problem, but we can make progress toward leveling the playing field. While it is very difficult in today’s political climate, we have to be able to look to colleagues, to our role models and to our leaders to set an example of what is acceptable and demand change.

When the MediaFor program was first conceived at DARPA, one thing that kept me up at night was the concern that someday our adversaries would be able to synthesize entire events with minimal effort. These events might include images of scenes from different angles, video content that appears to be from different devices, and text delivered through various medium providing overwhelming evidence that would lead to social unrest or retaliation before it could be countered. If the past five years are any indication, that someday is not be very far in the future.