Committee on Science, Space, and Technology
Online Imposters and Disinformation

Hany Farid, Ph.D.

Background

Rumors quickly spread in Trent, Italy that members of the Jewish community murdered a young boy and drained and drank his blood to celebrate Passover. Before long, the city's entire Jewish community is arrested and tortured, fifteen of which are found guilty and executed. The year was 1475.

Fast forward to 2018. Rumors quickly spread in Athimoor-Kaliyam, India that roving gangs are kidnapping children. Over a period of several months, nearly two dozen innocent people are dragged from their vehicles and killed. The rumors this time spread through WhatsApp instead of word of mouth.

Disinformation is not new, nor are its deadly consequences. What is new, thanks to the internet and social media, is its reach and frequency. Today, disinformation propagates around the world at the speed of light. From small- to large-scale fraud, to sowing civil unrest, interfering with democratic elections, and inciting violence, disinformation campaigns today are leading to dangerous and deadly outcomes.

Add to this phenomenon the ability to create increasingly more compelling and sophisticated fake videos of anybody saying and doing anything – popularly referred to as deep fakes – and the threat only increases. This is the landscape that awaits us in 2019 and beyond.

Creating Deep Fakes

Advances in machine learning and access to large and diverse data sets have led to computer systems that are able to synthesize images of people who don't exist, videos of people doing things they never did, and audio recordings of them saying things they never said. These deep fakes are a dangerous

addition to an already volatile on-line world in which rumors, conspiracies, and disinformation spread often and quickly.

By providing millions of images of people to a machine-learning system, the system can learn to synthesize realistic images of people who don't exist. Similar technologies can, in live-stream videos, convert an adult face into a child's face, raising concerns that this technology will be used by child predators.

With just hundreds of images of someone, a machine-learning system can learn to insert them into a video. This face-swap deep fake can be highly entertaining, as in its use to insert Nic Cage into movies in which he never appeared. The same technology, however, can also be used to create non-consensual pornography or to impersonate a world leader. Similar technologies can also be used to alter a video to make a person's mouth consistent with a new audio recording of them saying something that they never said. When paired with highly realistic voice synthesis technologies that can synthesize speech in a particular person's voice, these lip-sync deep fakes can make a CEO announce that their profits are down, leading to global stock manipulation, a world-leader announce military action, leading to global conflict, or a presidential candidate confess complicity in a crime, leading to the disruption of an election.

What is perhaps most alarming about these deep-fake technologies is that they are not only in the hands of sophisticated Hollywood studios. Software to generate fake content is widely and freely available on-line, putting in the hands of many the ability to create increasingly compelling and sophisticated fakes. Coupled with the speed and reach of social media, convincing fake content can instantaneously reach millions.

How do we manage a digital landscape when it becomes increasingly more difficult to believe not just what we read, but also what we see and hear with our own eyes and ears? How do we manage a digital landscape where if anything can be fake, then everyone has plausible deniability to claim that any digital evidence is fake?

Detecting Deep Fakes

Despite efforts by digital forensic researchers, no current technology exists that can contend with the vast array of different types of deep fakes at a speed and accuracy that can be deployed at internet-scale.

There are several challenges that the digital forensic community is facing.

Deep fakes are relatively new and have developed in sophistication much faster than expected. There are significantly more researchers working to develop techniques for synthesizing increasingly more realistic audio, images, and video, than there are those of us trying to detect this content. This means that the nature and quality of deep fakes is developing at an unprecedented rate that is difficult to keep pace with. In addition, the scale and speed of the internet makes deploying effective technology incredibly challenging: Facebook, for examples, sees some one billion daily uploads and YouTube sees some 500 hundred hours of video uploaded every minute. The sheer amount of information uploaded everyday makes any filtering technology incredibly difficult.

There is, however, a family of technologies that could be considered for wide deployment. Control-capture technologies can authenticate content at the point of recording by extracting, at the time of recording, a unique digital signature from any recorded digital content, cryptographically signing this signature, and then placing it on a secure central server or a distributed immutable ledger like the blockchain.[1] This signature can then be compared to any version of the same content found online to determine if the content has been altered from the time of recording. Although this approach tackles disinformation differently than forensic techniques – by telling us what is real instead of what is fake – these technologies are available today and can operate at internet-scale.

We should be exploring the further development and deployment of both control-capture and forensic technologies.

The Future

Despite the challenges, I propose several calls to action.

1. Funding agencies have to invest at least as much financial support to programs that seek to build systems to detect fake content as they do to programs in computer vision and computer graphics that are giving rise to the sophisticated synthesis technologies described above.

2. Researchers that are developing technologies that we now know can be weaponized should give more thought to how they can put proper

[1]For full disclosure, I am a paid advisor to a company, Truepic, that develops control-capture technology.

safeguards in place so that their technologies are not misused.

3. No matter how quickly forensic technology advances, it will be useless without the collaboration of the giants of the technology sector. The major technology companies (including, Facebook, Google/YouTube, and Twitter) must more aggressively and proactively deploy technologies to combat disinformation campaigns, and more aggressively and consistently enforce their policies. For example, Facebook's terms of service state that users may not use their products to share anything that is "unlawful, misleading, discriminatory or fraudulent". This is a sensible policy — Facebook should enforce their rules.

4. Lastly, we should not ignore the non-technological component to the issue of disinformation: us the users. We need to educate the public on how to consume trusted information, we need to educate the public on how to be better digital citizens, and we need to educate the public on how not to fall victim to scams, fraud, and disinformation.

Conclusions

I will end where I began. Disinformation is not new. Deep fakes is only the latest incarnation. We should not lose sight of the fact that more traditional human-generated disinformation campaigns are still highly effective, and we will undoubtedly be contending with yet another technological innovation a few years from now. In responding to deep fakes, therefore, we should make every effort to consider the past, present and future as we try to navigate the complex interplay of technology, policy, regulation, and human nature.

Lastly, I would be remiss in not mentioning that although there are serious issues of on-line privacy, moves by some of the technology giants to transform their platform to an end-to-end encrypted system will only make the problem of disinformation worse. Such end-to-end encrypted systems will make it even more difficult to understand and slow or stop the spread of disinformation. We should make every effort to consider the balance between privacy and safety and how these can be best accomplished.