DM22-0547

Script: *A Dive into Deepfakes*
SME(s): *Shannon Gallagher and Dominic Ross*
Interviewer: *Dominic Ross*
Interview Conducted: Thursday, April 28, 2022 11 a.m. ET

<Canned Intro>

Dominic: Welcome to the SEI Podcast Series. My name is Dominic Ross. I am the multimedia design team lead in the SEI's CERT Division.

Today, I am joined by Shannon Gallagher, a data scientist in the CERT Division. Today we are here to talk about our work on DeepFakes, specifically how easy is it to make and detect a deepfake. As we stated in a recently published SEI blog post on this work, there were more than 85,000 harmful deepfake videos detected up to December 2020, with the number doubling every six months since observations began in December 2018. So Deepfakes are definitely a problem for many organizations.
Welcome Shannon.

Shannon: *Thank you.*

1. **Dominic**: We are both new to the podcast series so let's start by telling our audience about ourselves, what brought us to the SEI, and the work that we do here. Shannon, you can start. One thing I think we should touch on is how someone with a background in multimedia design and a data scientists came together to work on this problem.
   - Data science is a supplemental tool

2. **Dominic:** What are deepfakes, and how would you characterize the current state of the technology that is being used to create them?
   - Believable media created by neural nets, typically refers to videos but can also be image, text, audio
   - Fake images have been around for a long time and certainly originated from multimedia experts and visual effects.
   - GAN vs Autoencoder (tom cruise)

3. **Dominic**: How easy is it for people to create deepfakes? How easy are they to detect and what can happen if they are left unchecked.
   - Well it depends. There are some phone apps available and if we have a broad view, snapchat has filters that can be considered deepfakes that are quite amusing. Would people think they're real images? Probably not.
   - There are some tools available online ranging from point and click interfaces to full code-oriented packages. Usually these require being proficient in coding and using GPUs.
   - Your regular citizen probably isn't going to make stellar ones anytime soon

- Similar to how people leave fingerprints or different camera lens leave artifacts in an image, it's been studied how deepfake generators also usually leave digital fingerprints. So right now, if we can get a match to our 'fingerprint' we can have a pretty good guess to what it is.
- A potential problem is that people can start subtracting the fingerprint features from future iterations of the model, sort of by layering images over one another.
- So we must assume generators are going to get better and better, directly from using information from detectors.
- The main concern is that doubt is added to whether you can trust an image/video/text. Trusting a fake video can obviously have bad consequences but doubting a real video can be very bad as well! It's hard to dismiss something as fake that you see/hear.

4. **Dominic**: Now let's talk about the SEI's work in this area. To quote Heilmeier's Catechism, "If we are successful, what difference will it make?"
   - Social media
   - Hopefully real time
   - Difficult area because sometimes simply releasing a video is enough to sew discord

5. **Dominic**: What do current trends suggest about the future of deepfakes and the efforts to detect and combat them?
   - Iterative 'game' of deepfake makers and deepfake detectors
   - With the concept of generative adversarial networks ("GANs") this game can in a way to automated
   - Ultimately it may be easier to have trust through metadata, like 'seals' rather than to detect truth

from the actual content or image/video. However, this could be a problem of equity.

6. **Dominic**: One aspect of our work that we like to highlight in our podcasts is transition. How might our listeners learn more about deepfakes and about the work that the SEI is doing on deepfake-detection models and software frameworks?
    - Check out Dom and Cathy's blog post if you haven't already. And we plan on having one or two more in the series
    - Hany Farid has a great book on Photo Forensics https://www.amazon.com/Photo-Forensics-Press-Hany-Farid/dp/0262035340
    - Keep an eye out for Chris Ume's videos on youtube

Dominic: Shannon, thank you for talking with us today. For our audience, we will include links in the transcript to resources mentioned during this podcast.

The SEI Podcast Series is available on Apple Podcasts, Google Podcasts, Soundcloud, Stitcher, and the SEI's YouTube Channel. If you like what you see and hear, please give us a thumbs up.

Thanks again for joining us.

<Canned Outro>

Resources:

Deepfake images are more trustworthy: https://www.pnas.org/doi/abs/10.1073/pnas.2120481119

Nice overview of deepfakes: Mirsky, Yisroel, and Wenke Lee. "The creation and detection of deepfakes: A survey." *ACM Computing Surveys (CSUR)* 54.1 (2021): 1-41.