DEEPFAKES: A GROWING CYBERSECURITY CONCERN
What Are Deepfakes?

A combination of "deep learning" and "fake", deepfakes are hyper-realistic videos digitally manipulated to depict people saying and doing things that never actually happened.

Deepfakes rely on neural networks that analyze large sets of data samples to learn to mimic a person's facial expressions, mannerisms, voice, and inflections.

The process involves feeding footage of two people into a deep learning algorithm to train it to swap faces. In other words, deepfakes use facial mapping technology and AI that swaps the face of a person on a video into the face of another person.
What Are Deepfakes?

Deepfakes are difficult to detect, as they use real footage, can have authentic-sounding audio, and are optimized to spread on social media quickly. Thus, many viewers assume that the video they are looking at is genuine.

Deepfakes target social media platforms, where conspiracies, rumors, and misinformation spread easily, as users tend to go with the crowd.

At the same time, an ongoing ‘infopocalypse’ pushes people to think they cannot trust any information unless it comes from their social networks, including family members, close friends or relatives, and supports the opinions they already hold.

In fact, many people are open to anything that confirms their existing views even if they suspect it may be fake.
An Autoencoder is a special type of neural network that is trained to copy its input to its output.

For example, given an image of a handwritten digit, an autoencoder first encodes the image into a lower dimensional latent representation, then decodes the latent representation back to an image.

An autoencoder learns to compress the data while minimizing the reconstruction error.
How Do Deepfakes Work?

The result is that the two autoencoders have a shared encoder that can "read" either a Mark Zuckerberg face or a Mr. Data face.

The goal is for the encoder to use the same representation for things like head angle or eyebrow position whether it's given a photo of Mark Zuckerberg or a photo of Mr. Data.

And that, in turn, means that once you've compressed a face with the encoder, you can expand it using either decoder.
Generative Adversarial Networks (GANs) are one of the most interesting ideas in computer science today. Two models are trained simultaneously by an adversarial process. A generator ("the artist") learns to create images that look real, while a discriminator ("the art critic") learns to tell real images apart from fakes.
How Do Deepfakes Work?

During training, the generator progressively becomes better at creating images that look real, while the discriminator becomes better at telling them apart.

The process reaches equilibrium when the discriminator can no longer distinguish real images from fakes.
Deepfake Examples

+ Tech Timeline
Platforms Banning Deepfakes:
• Several websites, including Discord, Gfycat, and Twitter, ban deepfakes with varying degrees of success.

Obama Deepfake Video:
• Deepfake video of former US president Barack Obama raises mainstream awareness.
  
  https://www.youtube.com/watch?v=cQ54GDm1eL0

US Senator Voices His Concern About Deepfakes:
• US Senator Marco Rubio voices his concerns about deepfakes at the Senate Intelligence Committee nomination hearing

Adobe #VoCo is an audio manipulator that allows you to change words in a voiceover simply by typing new words. Presented live during the Adobe MAX 2016.

https://www.youtube.com/watch?v=I3l4XLZ59iw
“#VoCo. Adobe’s Audio Manipulator” Demo

https://www.youtube.com/watch?v=I3I4XLZ59iw
Obama Deepfake Video

https://www.youtube.com/watch?v=cQ54GDm1eL0
David Beckham Deepfake Video:
- The recent global campaign showing Malaria survivors speaking through David Beckham to help raise awareness around the Malaria Must Die initiative spooked a lot of people.

House holds hearing on "deepfakes" and artificial intelligence amid national security concerns:
- The House Intelligence Committee heard from experts on the threats that so-called "deep fake" videos and other types of artificial intelligence-generated synthetic data pose to the U.S. election system and national security at large.

FaceApp:
- AI photo editor FaceApp goes viral after adding AI-based age filter.

DARPA Is Taking On the Deepfake Problem:
- The Defense Department is looking to build tools that can quickly detect deepfakes and other manipulated media amid the growing threat of “large-scale, automated disinformation attacks.”
David Beckham Deepfake Video

https://www.youtube.com/watch?v=QiiSAvKJIHo
CEO Deepfake Scam:
- Criminals used artificial intelligence-based software to impersonate a chief executive’s voice and demand a fraudulent transfer of €220,000 ($243,000) in what cybercrime experts described as an unusual case of artificial intelligence being used in hacking.

California and Texas ban political deepfake videos:
- California and Texas have passed a law meant to prevent altered “deepfake” videos from influencing elections in a plan that has raised free speech concerns.

Deepfake Queen: 2020 Alternative Christmas Message:
- An alternative Christmas message for a very alternative year.

DeepFaceLive:
- Real-time face swap for PC streaming or video calls.

https://www.youtube.com/watch?v=IvY-Abd2FfM
https://github.com/iperov/DeepFaceLive
“Deepfake Queen: 2020” Video

https://www.youtube.com/watch?v=lvY-Abd2FfM
“This is not Morgan Freeman” Video

https://www.youtube.com/watch?v=oxXpB9pSETo
Making A Deepfake With DeepFaceLab

DeepFaceLab is the leading software for creating deepfakes.

https://github.com/iperov/DeepFaceLab
DeepFaceLab | Extract Faces from SRC
DeepFaceLab | Extract Faces... 27 hours later...

Choose one or several GPU idxs (separated by comma).

[CPU]: CPU
[0]: NVIDIA GeForce RTX 3080 Laptop GPU

[0] Which GPU indexes to choose?:

[xf] Face type ( f/wf/head :help ) :
wf

[8] Max number of faces from image ( :help ) :
8

[512] Image size ( 256-2048 :help ) :
512

[90] Jpeg quality ( 1-100 :help ) :
90

[n] Write debug images to aligned_debug? ( y/n ) :

n
Extracting faces...

Running on NVIDIA GeForce RTX 3080 Laptop GPU

100% [================================================] 123676/123676 [27:20:53<00:00, 1.26it/s]

Images found: 123676
Faces detected: 125003

Done.
Press any key to continue...
DeepFaceLab | Sorting Faces

Running sort tool.

Choose sorting method:
[0] blur
[1] motion_blur
[2] face yaw direction
[3] face pitch direction
[4] face rect size in source image
[5] histogram similarity
[6] histogram dissimilarity
[7] brightness
[8] hue
[9] amount of black pixels
[10] original filename
[11] one face in image
[12] absolute pixel difference
[14] best faces faster

[5] : 4
4

Sorting by face rect size...

Loading: 12%#####8

OWASP FOUNDATION  owasp.org
Model Training

SRC Real  SRC Gen  DST Real  DST Gen  SRC + DST Mix
DeepFaceLab | Preprocessing Target Video
DeepFaceLab | Extracting Faces from DEST
DeepFaceLab | Preprocessing DEST
DeepFaceLab | Preprocessing DEST
Detection
Deepfake Detection

The basic idea is to look for inconsistencies between “visemes,” or mouth formations, and “phonemes,” the phonetic sounds.

Specifically, the researchers looked at the person’s mouth when making the sounds of a “B,” “M,” or “P,” because it’s almost impossible to make those sounds without firmly closing the lips.

Researchers at Stanford have said that their approach is merely part of a “cat-and-mouse” game. As deep-fake techniques improve, they will leave even fewer clues behind.

In the long run, the real challenge is less about fighting deep-fake videos than about fighting disinformation. To reduce disinformation, we need to increase media literacy and develop systems of accountability. For example, laws against deliberately producing disinformation and consequences for breaking them, as well as mechanisms to repair the harms caused as a result.
Positive Benefits
Benefits of Deepfakes?

Deepfake technology also has positive uses in many industries, including movies, educational media and digital communications, games and entertainment, social media and healthcare, material science, and various business fields, such as fashion and e-commerce.

The film industry can benefit from deepfake technology in multiple ways. For example, it can help in making digital voices for actors who lost theirs due to disease, or for updating film footage instead of reshooting it.
Questions?