Discerning the Indiscernible: Tackling Deepfake Hoaxes

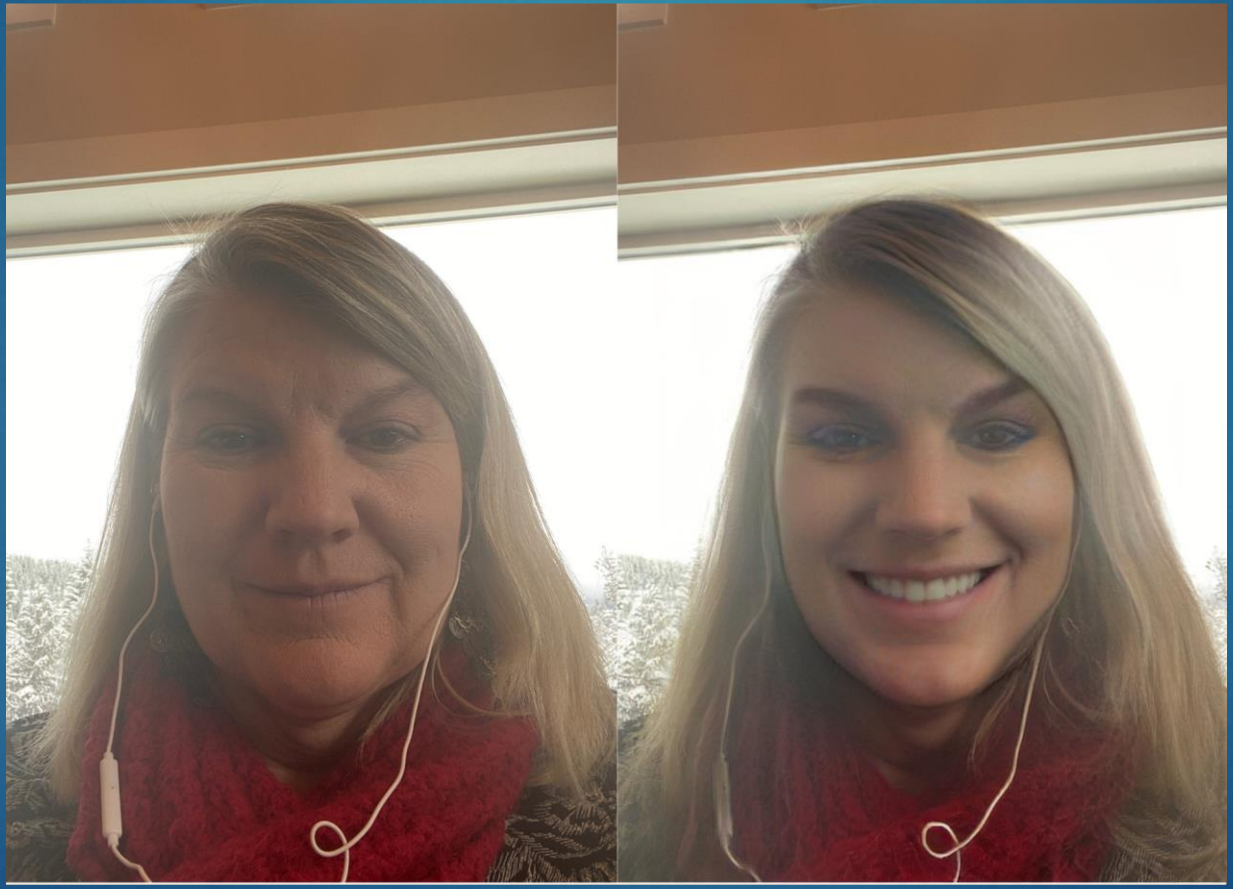
PRESENTER: DEB WELLS

10 JAN 2022

Overview

- Introduction to Deepfakes
- Technology surrounding Deepfakes
- Deepfake Categories and Examples
- Cybersecurity Concerns
- Discerning deepfakes
- Governance and Laws Around Deepfakes

“Seeing is believing”…or is it?



This was done on FaceApp

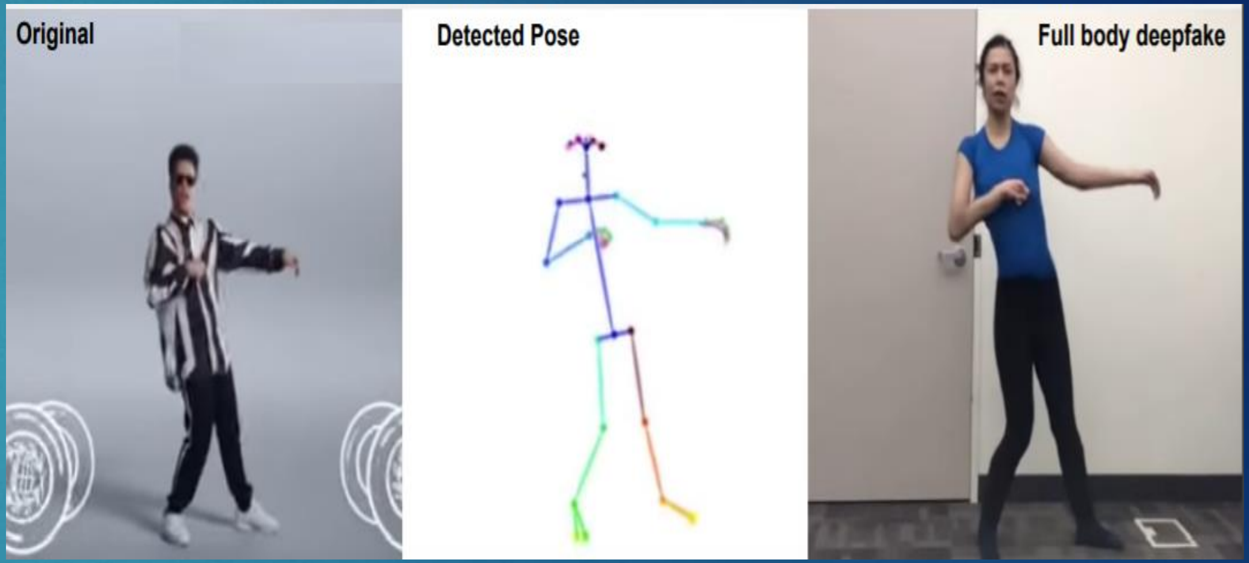
Deepfakes - Defined

- Conjunction – Deep (meaning AI or ML) + Fake (not real)
  - Output = Term ***Deepfake***
- Artificial images and sounds put together with machine-learning algorithms
  - Can create people who do not exist
  - Can impose on real people actions and words they did not really say
- Started in late 2017 – Reddit user began uploading videos of celebrities onto the body of porn star

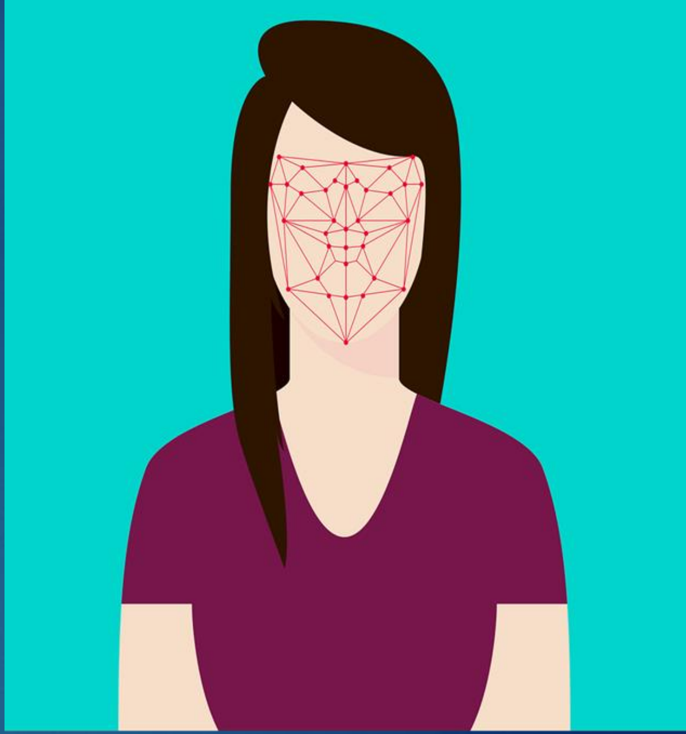
Synthetic Media

Deepfakes

- Traditionally, the better the quality of the deepfake, more images required to make video/audio look and sound better
  - Takes tens of minutes of videos and hundreds of photos
  - Hence the reason by political and celebrities are the main target – today
  - Known as passive information
- **Samsung is perfecting deepfake s/w to allow them with the use of only 1 photo!



Deepfakes are made of…



- Generative Adversarial Network (GAN)
  - Used for face generation
  - It produces faces that otherwise do not exist
- GAN uses two separate neural networks — or a set of algorithms designed to recognize patterns
  - First, network generates the image
  - Second, learns how to distinguish fake from real image
- Output = an algorithm that trains itself using the information generated above to generate fake photos of a real person

Deepfakes also are made of…

- Artificial intelligence (AI) algorithm known as encoder/decorder
  - Used in face-swapping or face-replacement technology
  - First, you run thousands of face shots of two people through the encoder to find similarities between the two images
  - Then, a second AI algorithm, or decoder, retrieves the face images and swaps them
    - End result -- a person's real face can be superimposed on another person's body

Who makes such software?

- Open-source Python software
  - Faceswap and DeepFaceLab
    - Faceswap is free, open-source, multi-platform software
    - Runs on Windows, macOS, and Linux
    - DeepFaceLab is an open-source that also enables face-swapping.
  - FakeApp was developed in 2018
  - FaceApp easily downloaded and used – remember my opening picture!!

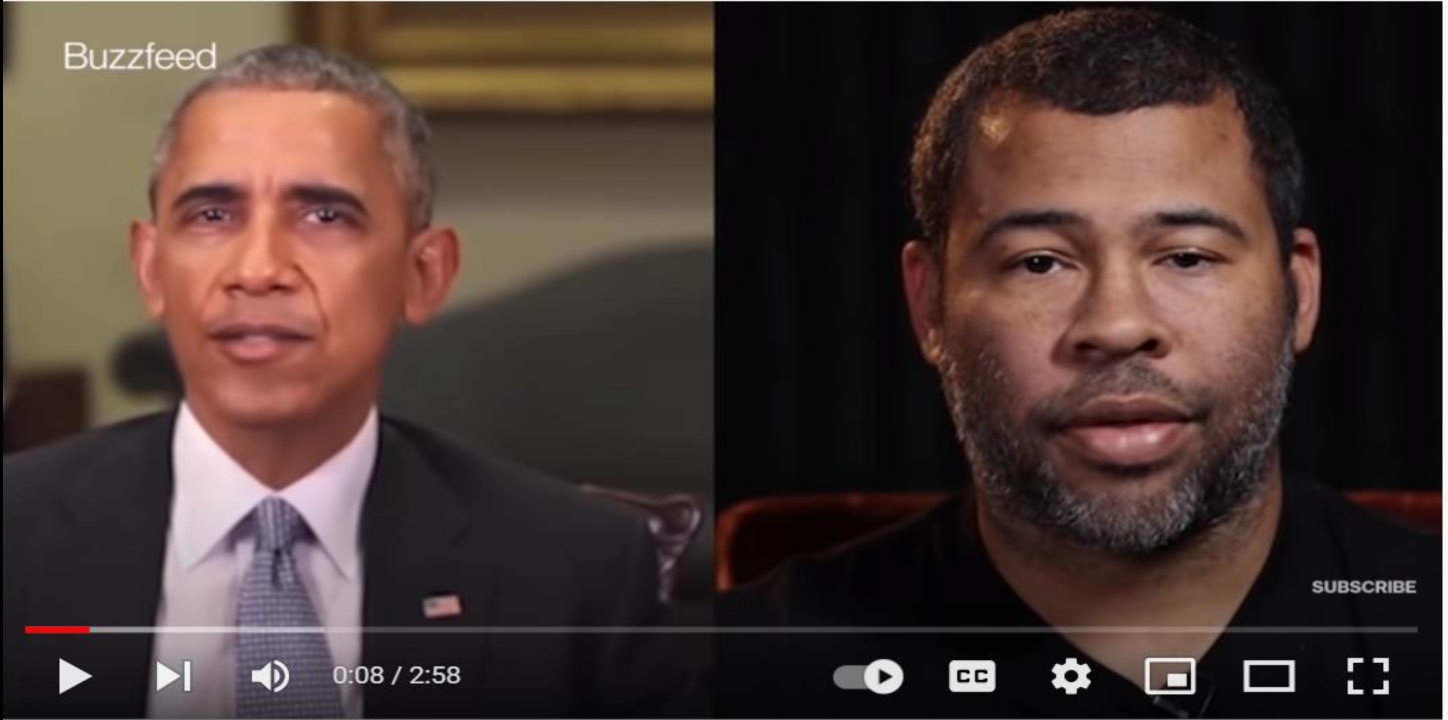
Categories of Deepfakes

- Porn/Revenge Porn
  - Invasion of sexual privacy
- Political campaigns
  - Launching info warfare campaigns: Example -- Gabon, East Africa
- Commercial Uses
  - CereProc uses digital voices for people who have lost the use of their voice
  - Significant cost & time savings for artificial videos (multiple languages)
- Creative Uses
  - Nicholas Cage and face-swapping images
  - Holocaust survivors talking to an audience using holograms, authors reading their own books

SPEAK UP

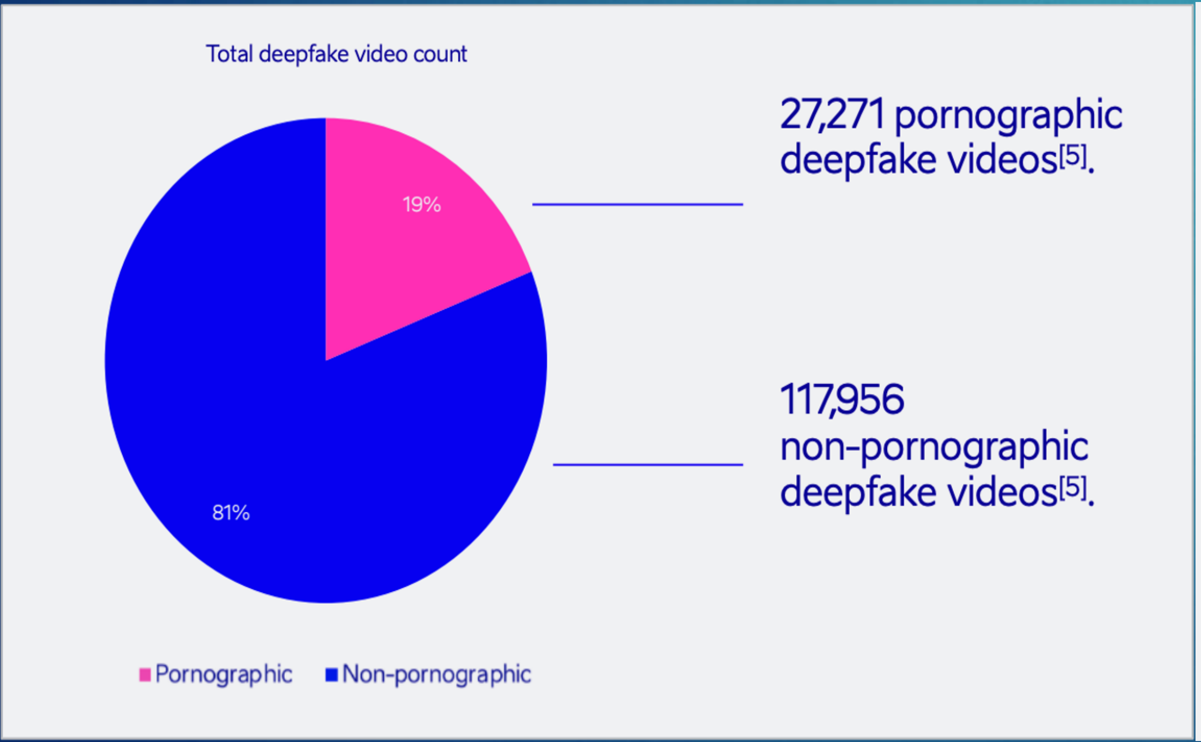
Examples of Deepfakes

- If you watch a Buzzfeed video from 17 April 2018, you will see a video of President Obama making some very outlandish and brazen statements…is it really him?
- Scarlett Johansson’s face was transposed on a porn star back in 2017



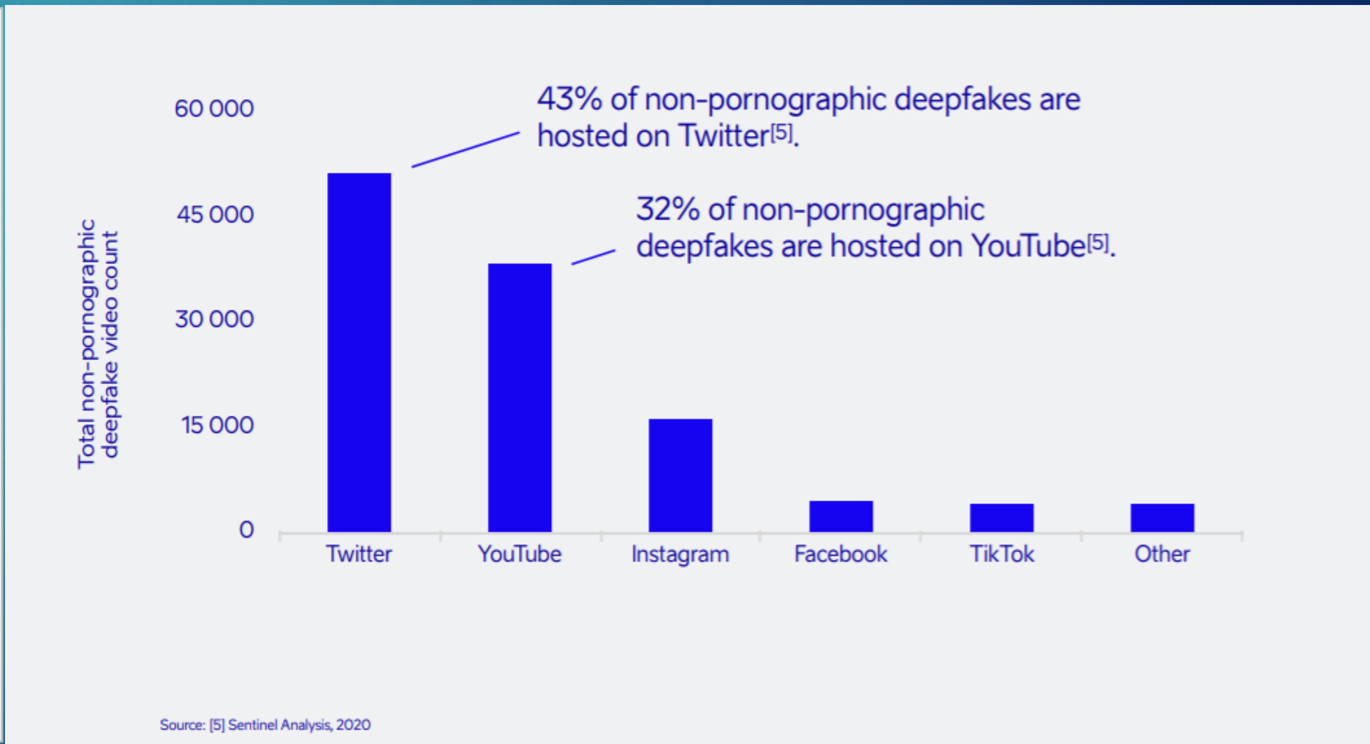
Rise in deepfake volume…

Since 2019, # of deepfakes online has grown at a rate of over 900%!



27,271 pornographic deepfake videos[5].

117,956 non-pornographic deepfake videos[5].



43% of non-pornographic deepfakes are hosted on Twitter[5].

32% of non-pornographic deepfakes are hosted on YouTube[5].

Source: [5] Sentinel Analysis, 2020

https://thesentinel.ai/media/Deepfakes%202020:%20The%20Tipping%20Point,%20Sentinel.pdf

Cybersecurity Concerns

- Phishing scams
- Data breaches
- Hoaxes
- Pornography
- Reputation smearing
- Election manipulation
- Social engineering
- Automated disinformation attacks
- Identity theft
- Financial fraud
- Blackmail

How can I tell?

- Look into their eyes…unnatural eye movement
- Unnatural facial expressions.
- Awkward-looking facial movement or their body does not look right
- Unnatural coloring
- Hair and teeth that look fake
- Blurring or misalignment
- Inconsistent noise or audio
- Images that look unnatural when slowed down
- Hash discrepancies
- Reverse image searches

Countering Deepfakes

- Web browser extensions that help identify a deepfake
- Filtering software – Deeptrace provides this type of protection
- Social Media Rules
- Using soft biometrics to detect
- Deepfake Detection Challenges – like Bug Bounties
- Research Technologies – using watermarks and blockchain to detect a deepfake
- Corporate best practices
- Laws and governance

Governance & Regulations

- National Defense Authorization Act of 2020
- Privacy Act of 1974
- Copyright Laws and Intellectual Property
- Fair Use Doctrine
- Communications Decency Act
- State Laws – Virginia, Texas, and California
- Photoshop law in Israel
- Cyberstalking Law
- General Data Protection Regulation (GDPR) - EU

Is this technology really new?

Deepfakes & Deep Thinking