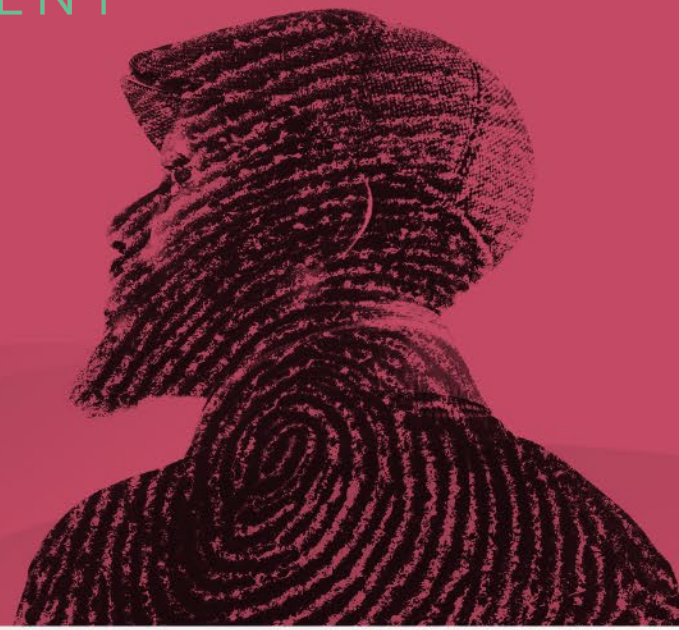
RSAC Sandbox

RSAConference2020

HUMAN ELEMENT



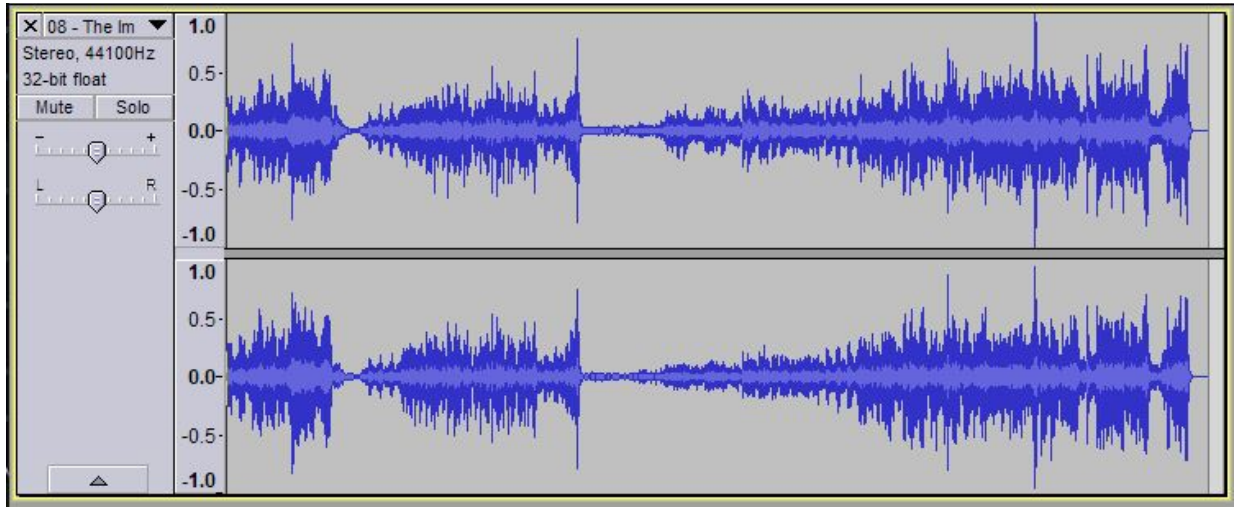
SESSION ID: **SBX1-W5**

Losing Our Reality: How Deepfakes Threaten Businesses and Global Markets

Alyssa Miller

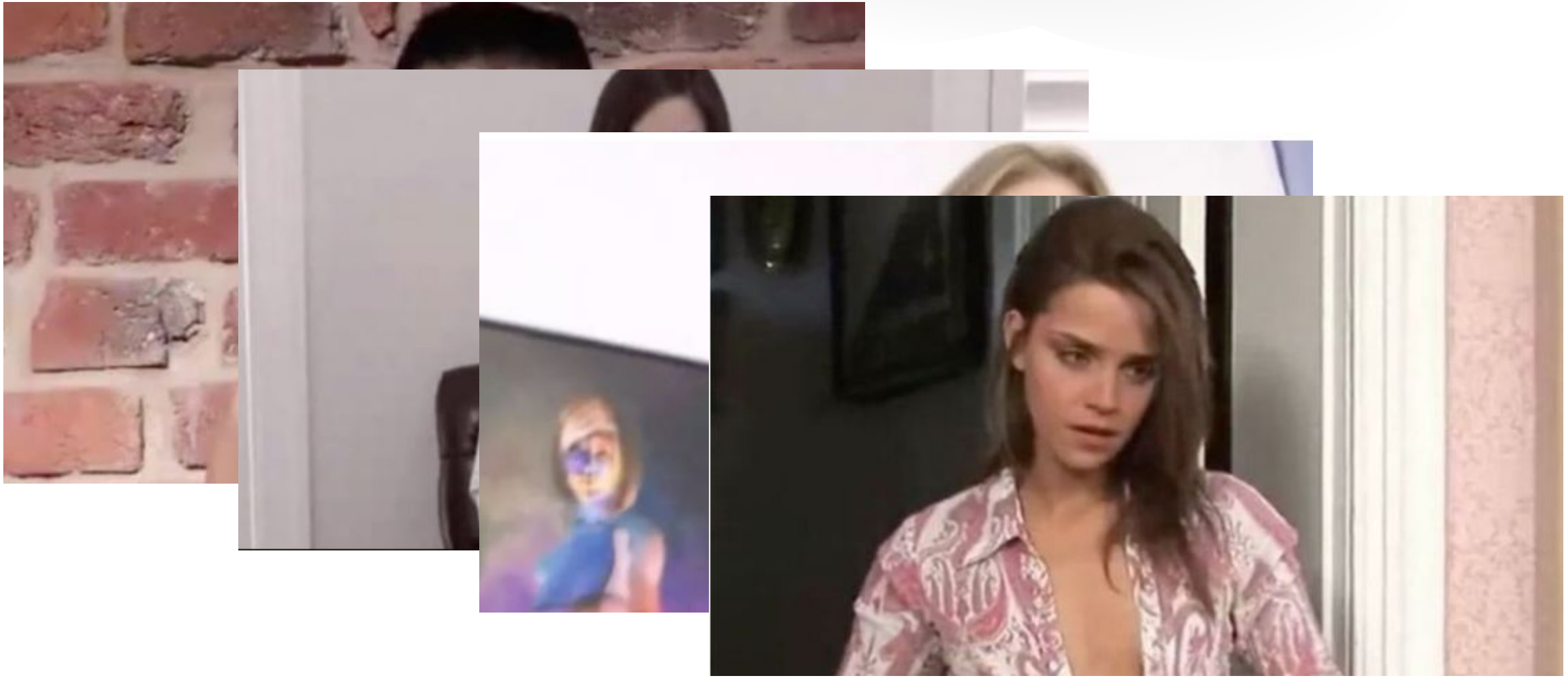
Application Security Advocate
Snyk Ltd.
@AlyssaM_Infosec

What Are Deepfakes?



A quick history on the rise of deepfakes

RSAC Sandbox



Turning Political

RSAC Sandbox

Business leaders

PLACEHOLDER: Deepfake Video Here

Understanding GANs

Target Media

Generator

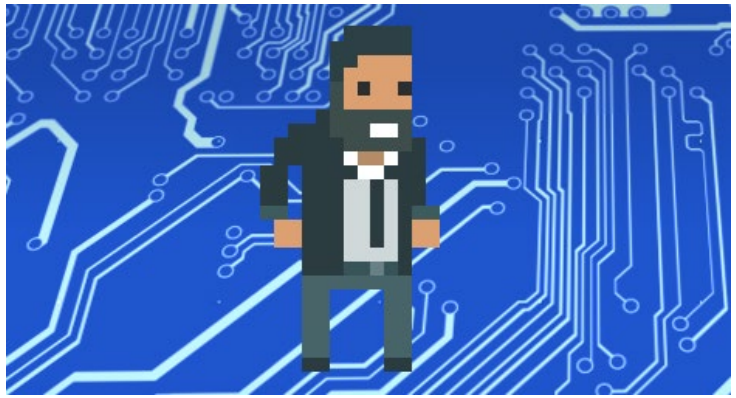
Training Set

Fake

Discriminator

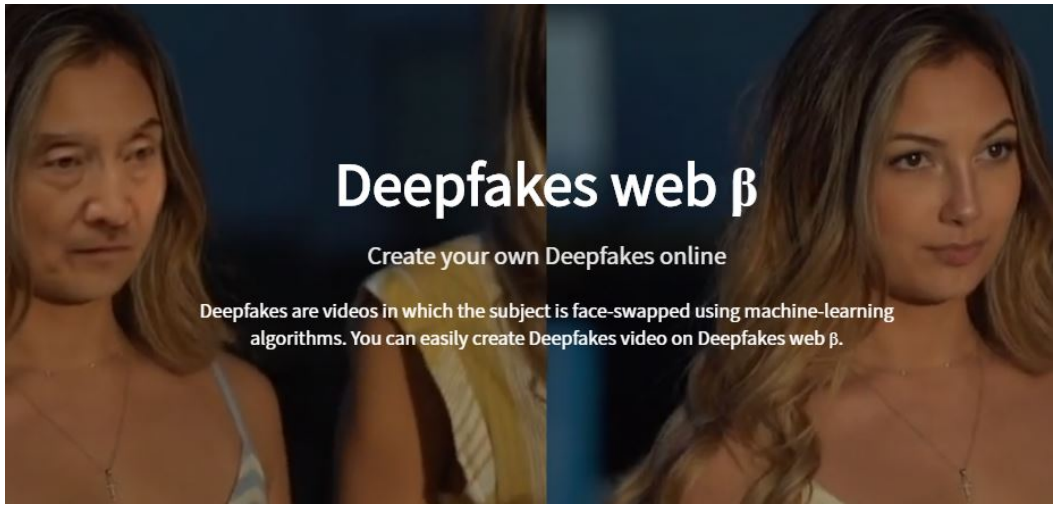
Real

Creating Deepfakes



FAKEAPP

faceswap



DeepNude

Understanding the Threats

How does this impact businesses and markets?

Threats to business

RSAC Sandbox

SOCIAL ENGINEERING

Threats to business

EXTORTION

Threats to business

RSAC Sandbox

“OUTSIDER” TRADING

Threats to business

RSAC Sandbox

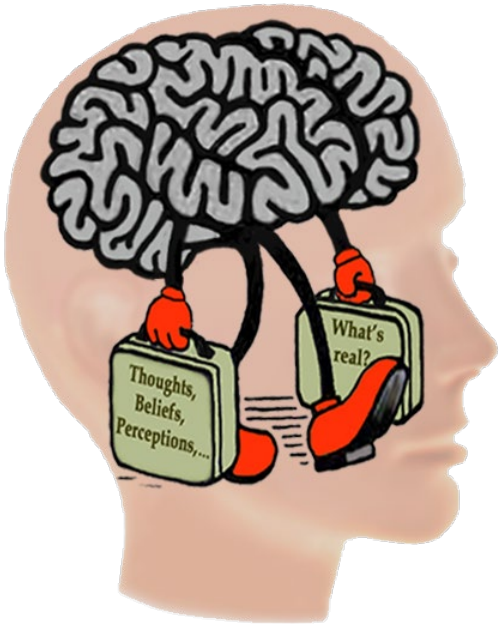
FINANCIAL/MARKET MANIPULATION

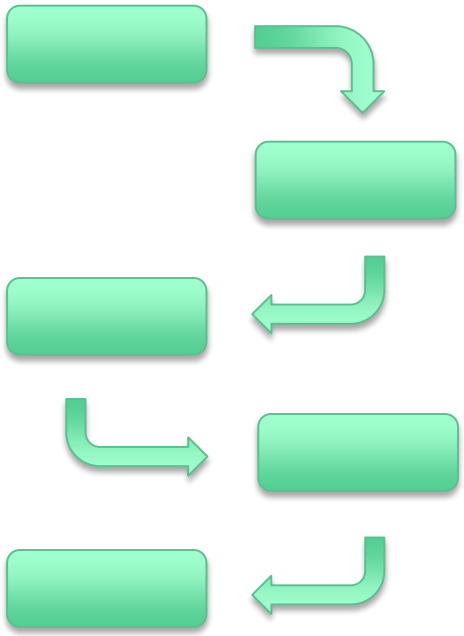
The World of Disinformation

What can be done to combat this threat?

The Disinformation Problem

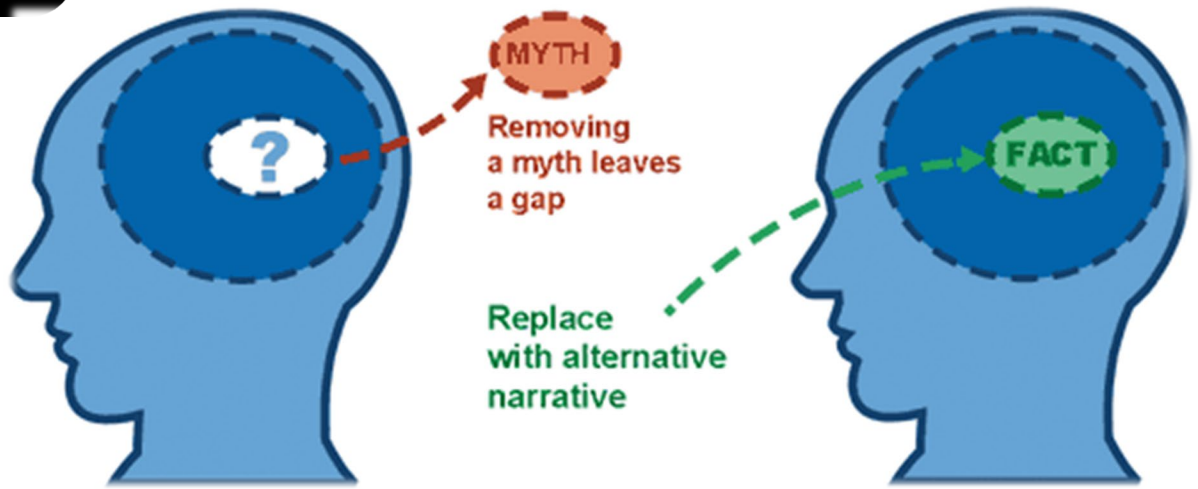
RSAC Sandbox





Combating Disinformation

RSAC Sandbox



Detection, Certification, Prevention

Countermeasures for Deepfake Disinformation

Detecting Deepfakes

RSAC Sandbox

Image: Phys.org

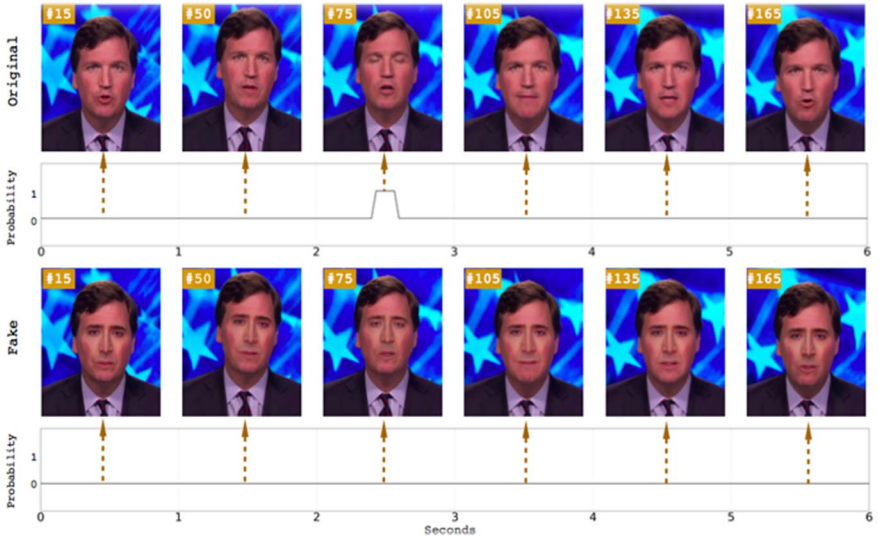


Image: University at Albany, SUNY

Image: Berkeley & USC

Certifying Original Videos

RSAC Sandbox

Photo by Hunter Moranville on Unsplash

Can we prevent deepfakes from being created?

RSAC Sandbox

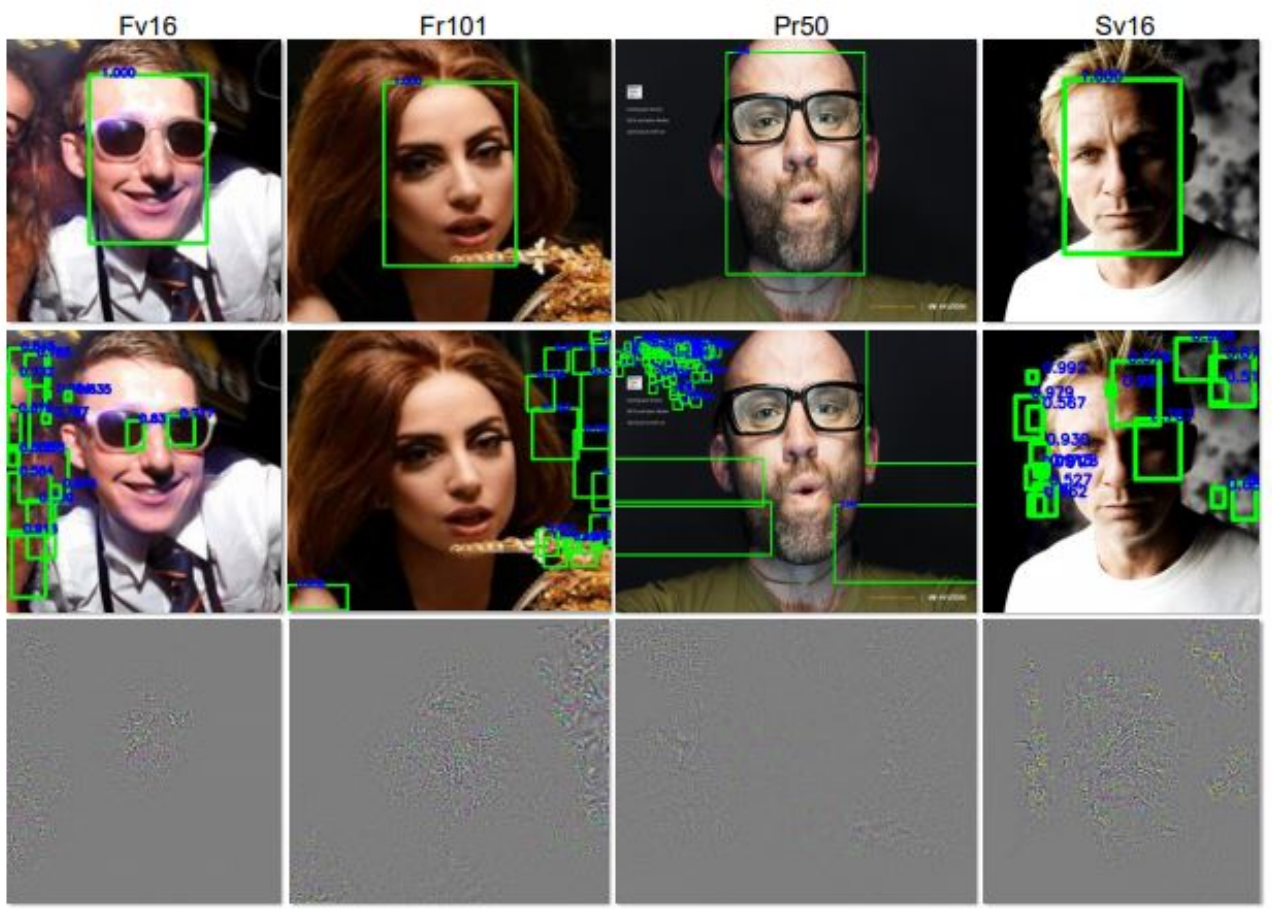


Image: Cornell University

Taking Action

Preparing the enterprise to combat the DeepFake Threat

So what can business do?

RSAC Sandbox

- Immediately
  - Minimize channels for company communications
  - Drive consistent information distribution
- Prepare for the Future
  - Develop a disinformation response plan (treat these as incidents)
  - Organize a centralized monitoring and reporting function
- For the long term
  - Encourage responsible legislation and private sector fact verification
  - Monitor development of detection and prevention countermeasures

Alyssa MILLER

Continue the Conversation

RSAC Sandbox

@AlyssaM_Infosec

Linkedin.com/in/alyssam-infosec

https://alyssasec.com

References…

Misinformation and its Correction
https://www.researchgate.net/publication/277816966_Misinformation_and_its_Correction

Detecting Deepfakes by Looking Closely…
https://phys.org/news/2019-06-deepfakes-reveals.html

In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking
https://arxiv.org/pdf/1806.02877.pdf

Exposing DeepFake Videos By Detecting Face Warping Artifacts
https://arxiv.org/pdf/1811.00656.pdf

Protecting World Leaders Against Deep Fakes
http://openaccess.thecvf.com/content_CVPRW_2019/papers/Media%20Forensics/Agarwal_Protecting_World_Leaders_Against_Deep_Fakes_CVPRW_2019_paper.pdf

Hiding Faces in Plain Sight: Disrupting AI Face Synthesis with Adversarial Perturbations
https://arxiv.org/pdf/1906.09288.pdf