RSAConference2020
Asia Pacific & Japan

A Virtual Learning Experience | 15–17 July

HUMAN ELEMENT

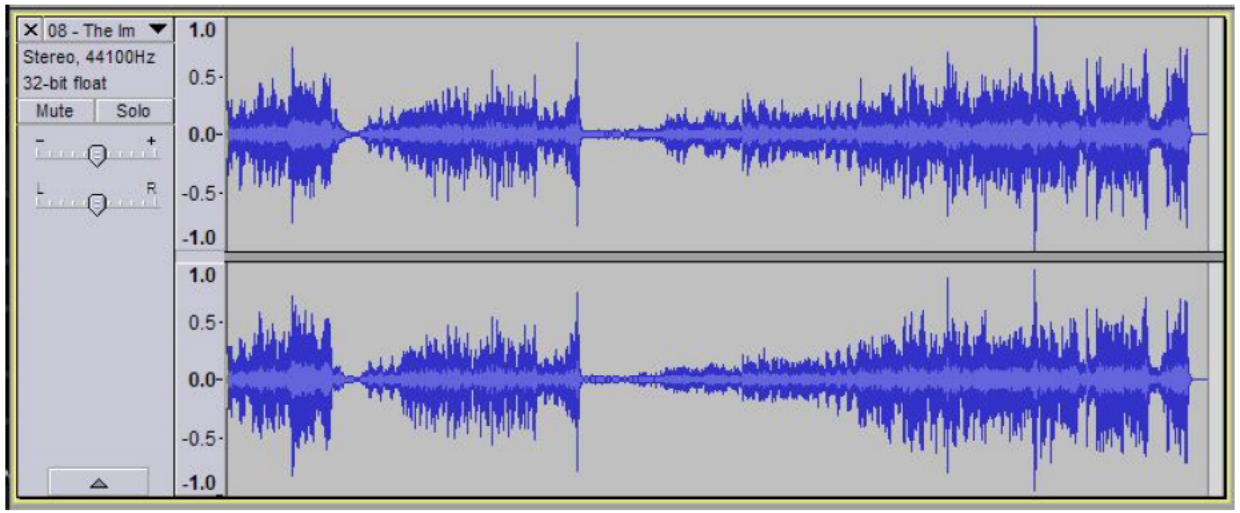
SESSION ID: **RMG-R05V**

Reality Under Attack: Just How Serious Is the Deep Fake Threat?

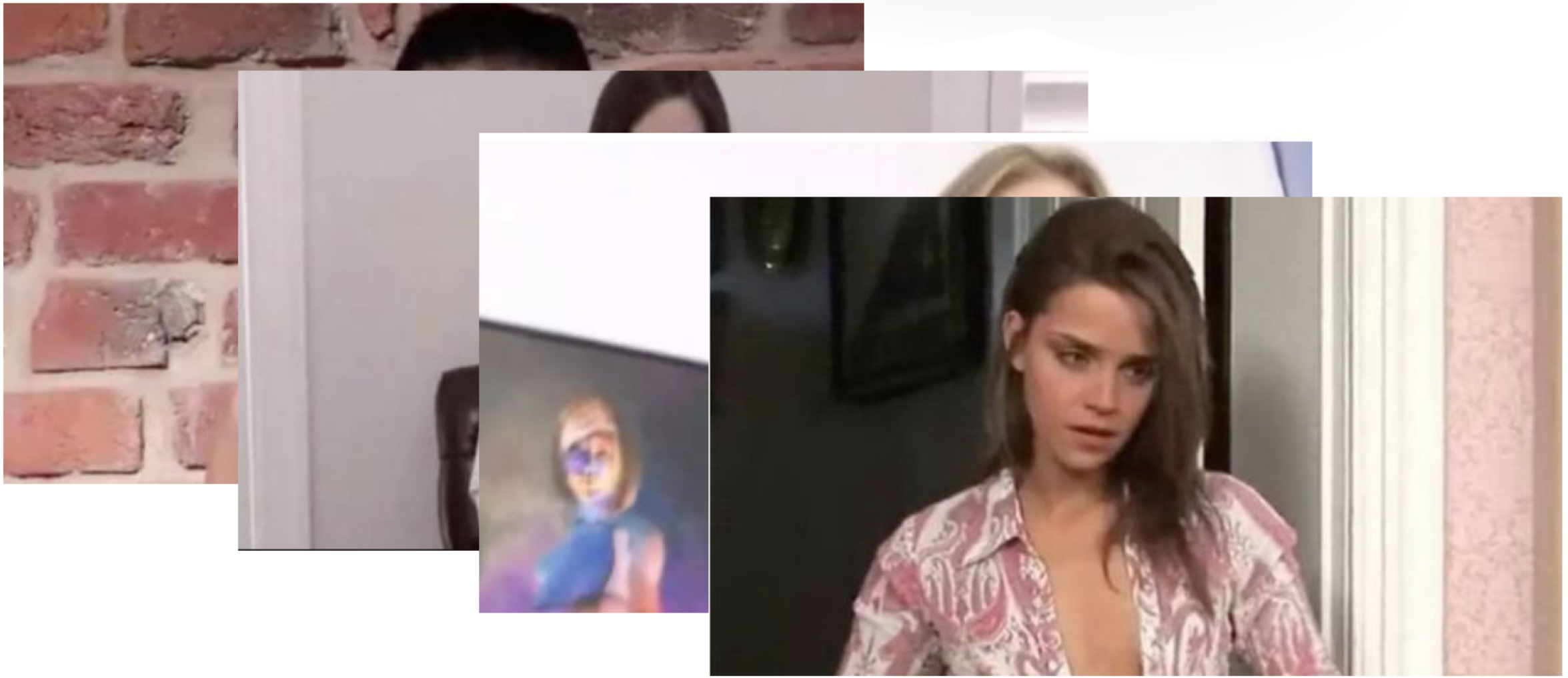
Alyssa Miller

Application Security Advocate
Snyk Ltd.
@AlyssaM_InfoSec

What Are Deepfakes?



A quick history on the rise of deepfakes



Turning Political

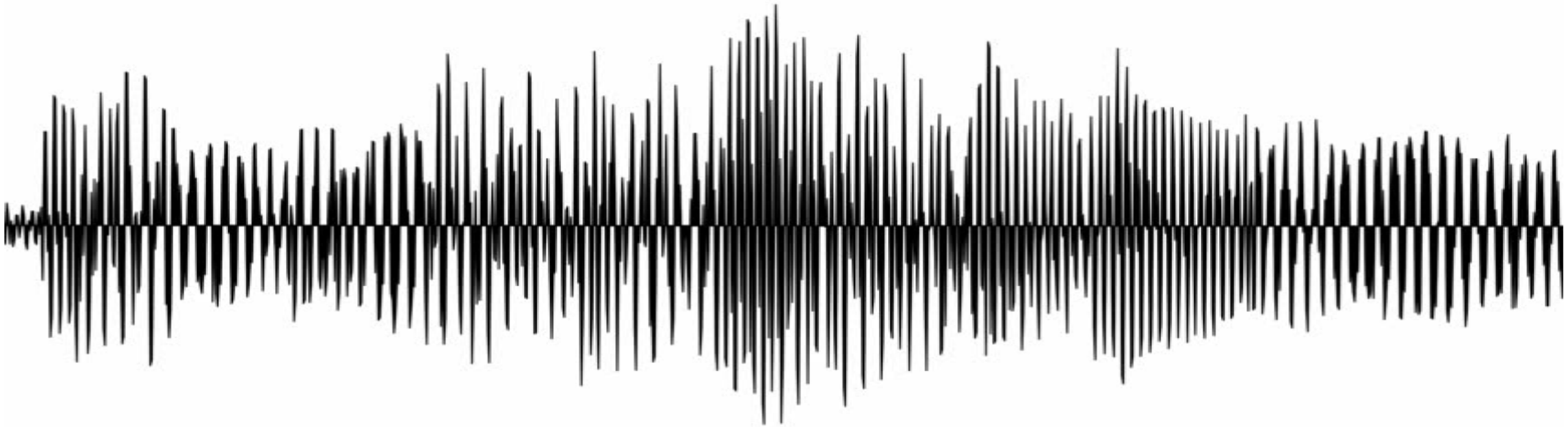
Business leaders

RSAConference2020 APJ

A Virtual Learning Experience

The Untold Business Impact

Threats to business



SOCIAL ENGINEERING

Threats to business

EXTORTION

Threats to business

“OUTSIDER” TRADING

Threats to business

FINANCIAL/MARKET MANIPULATION

RSAConference2020 APJ

A Virtual Learning Experience

How Serious is this Really?

Understanding GANs

Target Media

Generator

Training Set

Discriminator

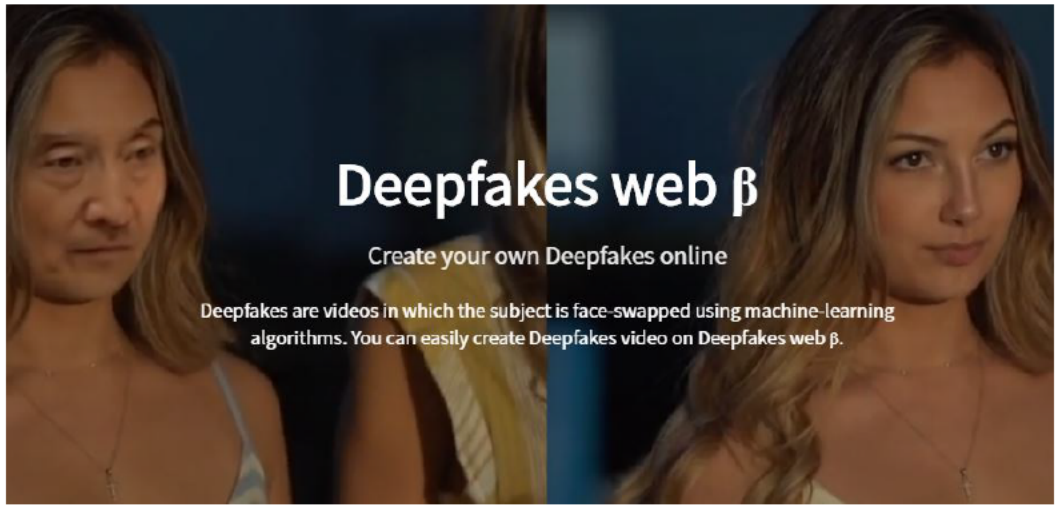
Fake

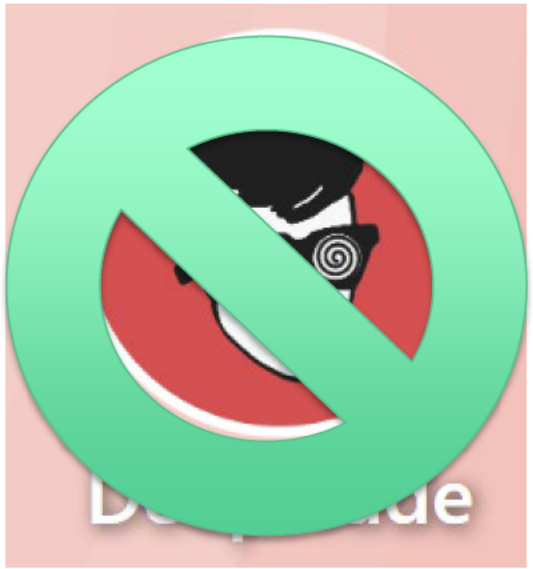
Real

Creating Deepfakes

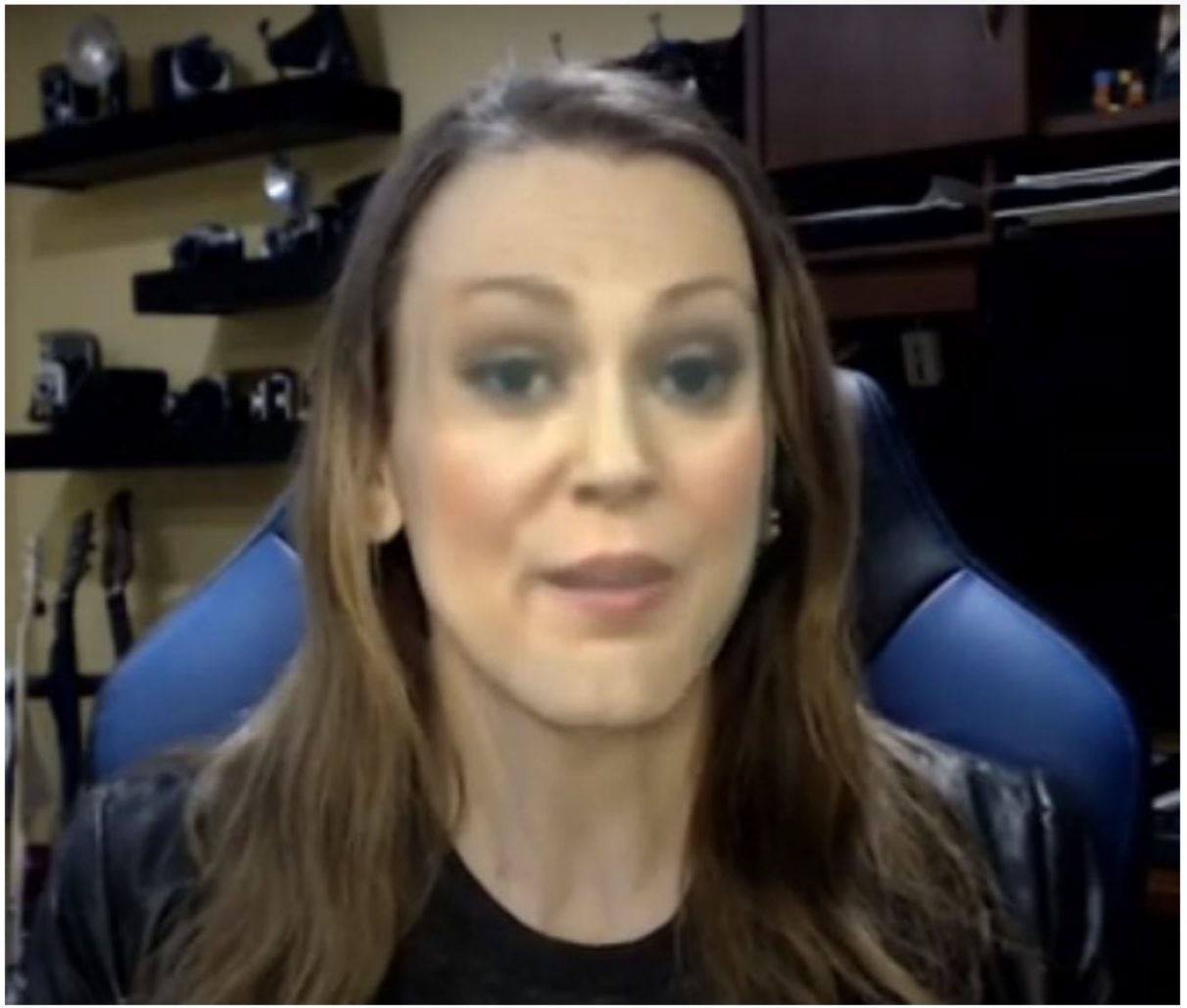
FAKEAPP

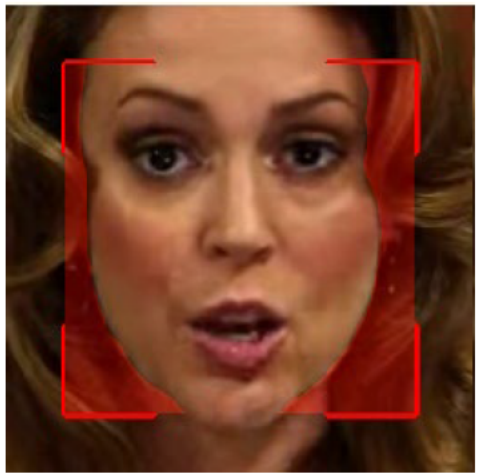
faceswap





Not So Easy Afterall…





Detecting Deepfakes

Image: Phys.org

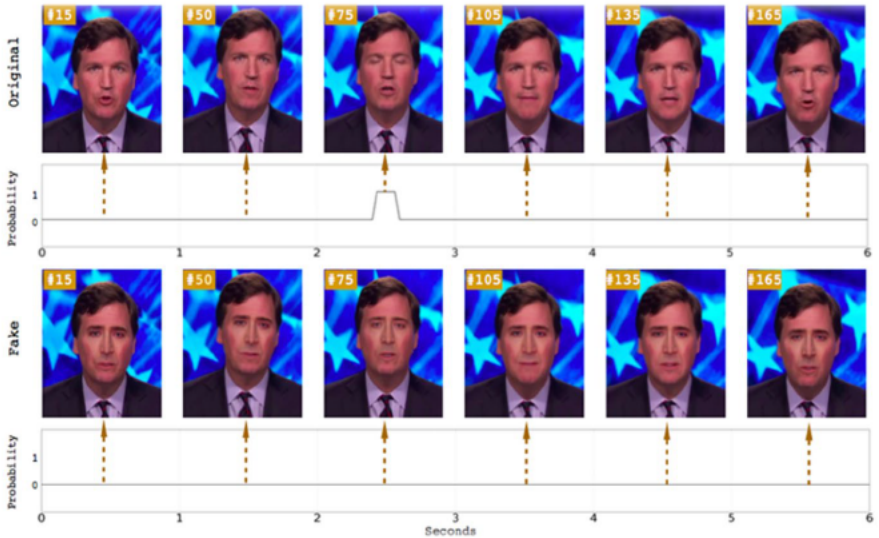
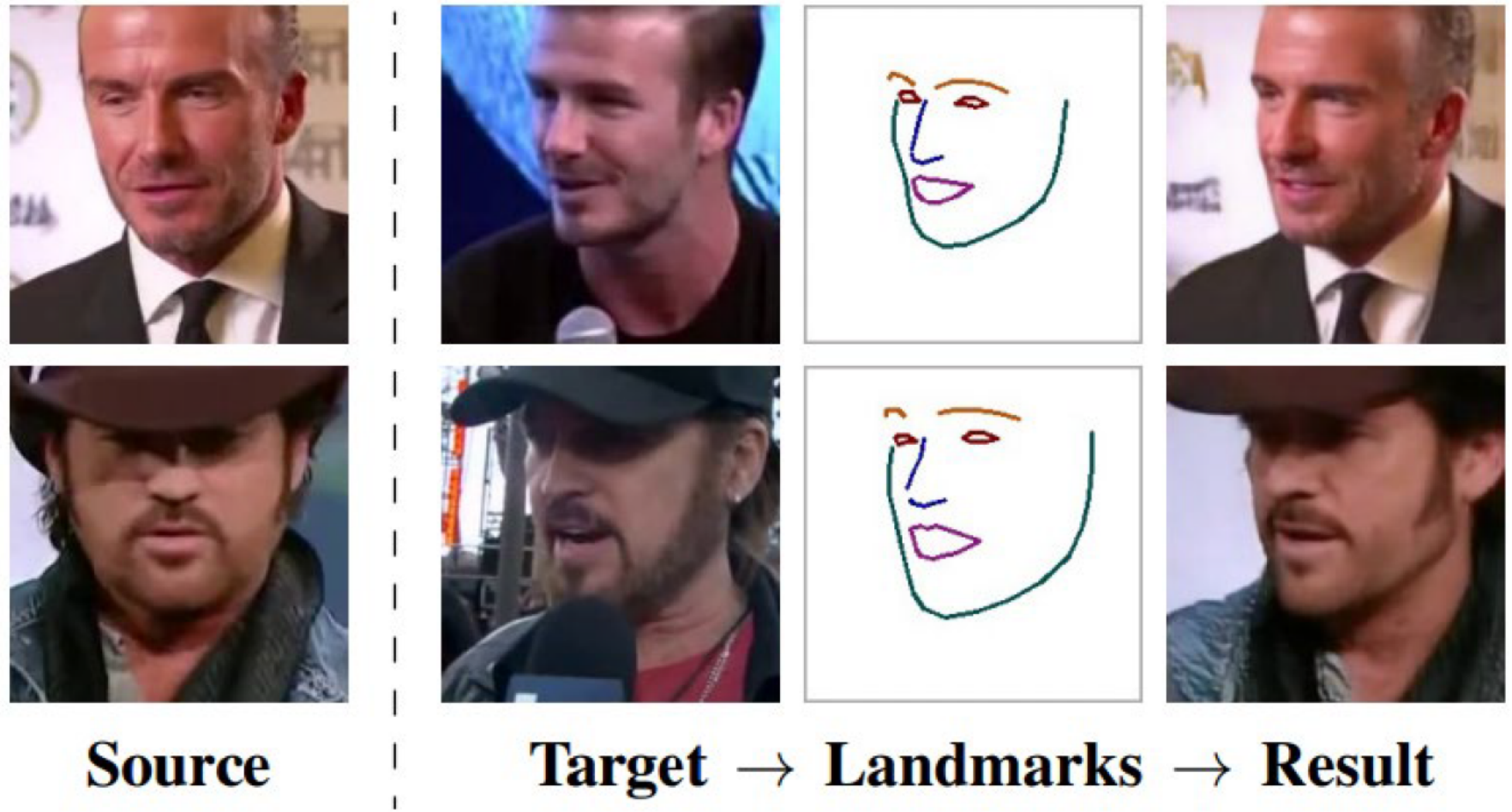


Image: University at Albany, SUNY

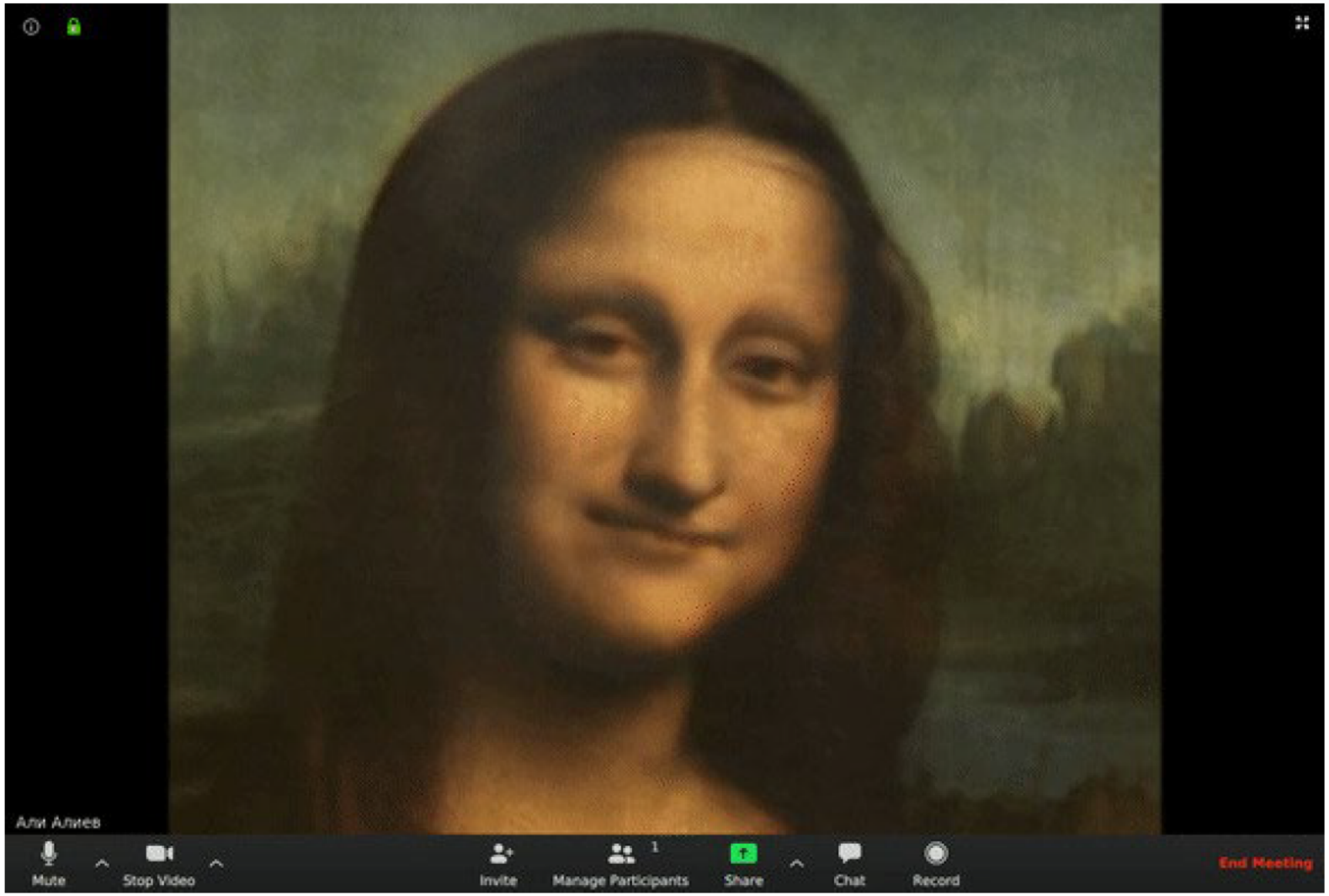
Image: Berkeley & USC

A New Approach



Current Limitations

Avatarify (Plugin for Video Conferencing)



Can we prevent deepfakes from being created?

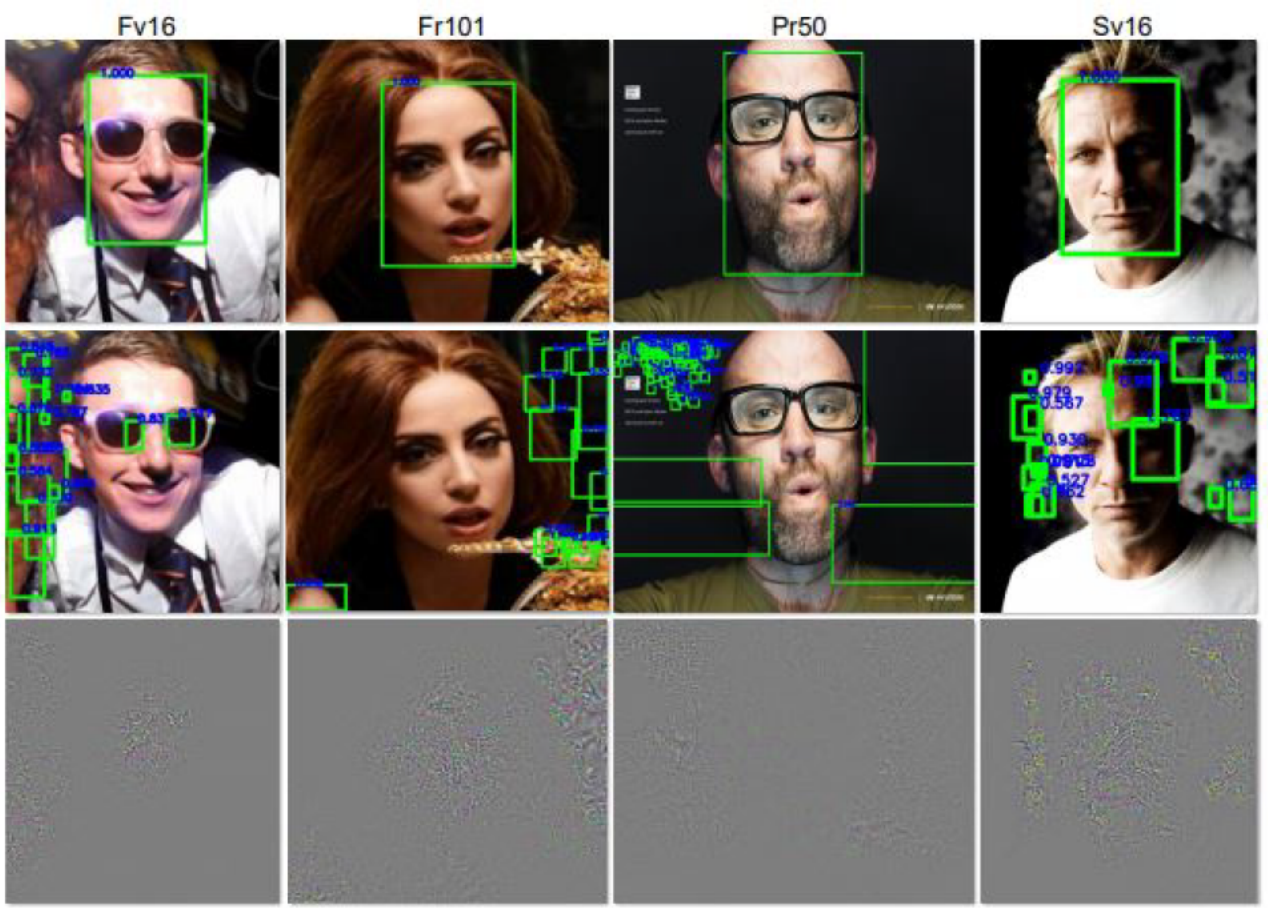


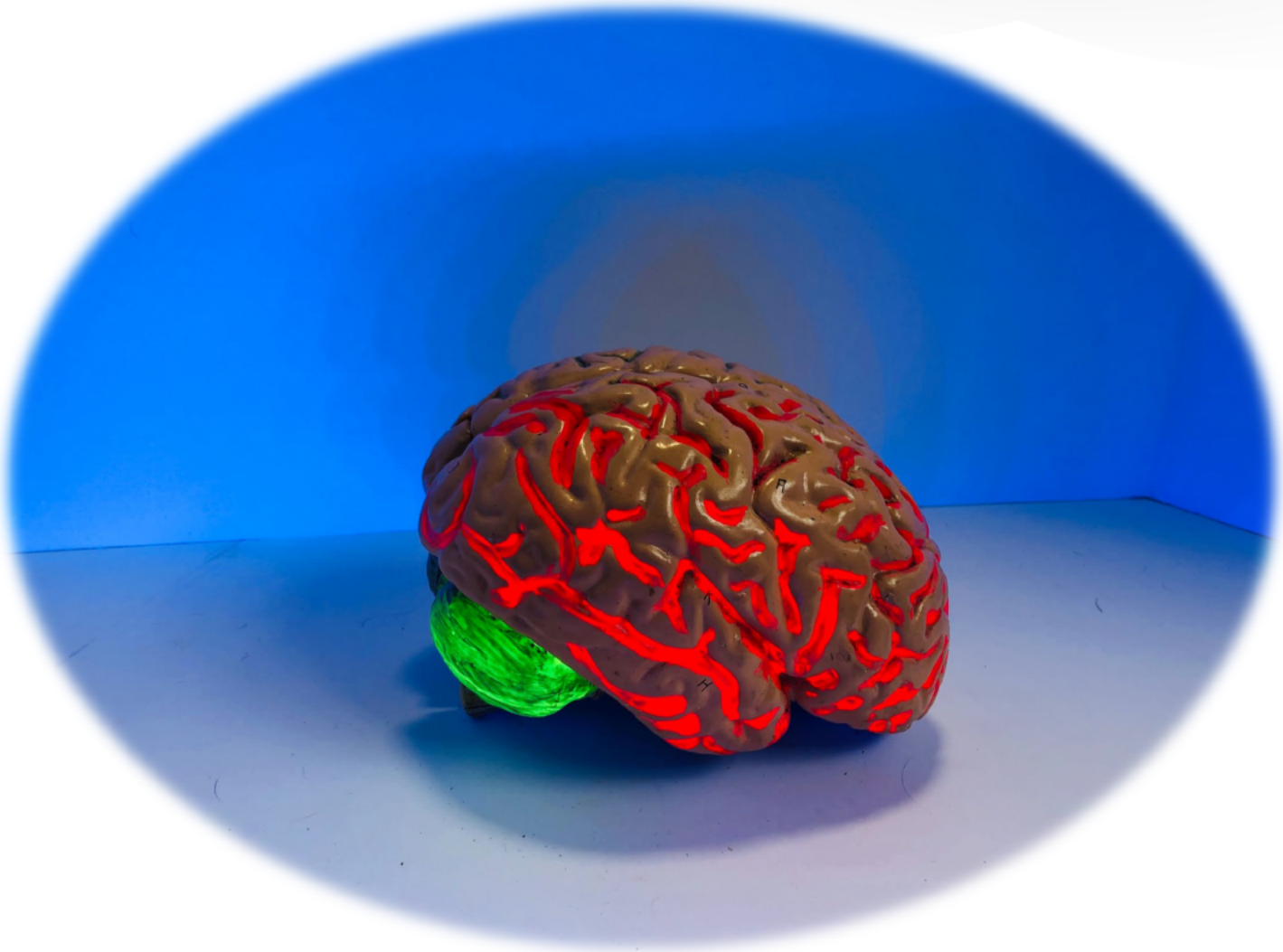
Image: Cornell University

RSAConference2020 APJ

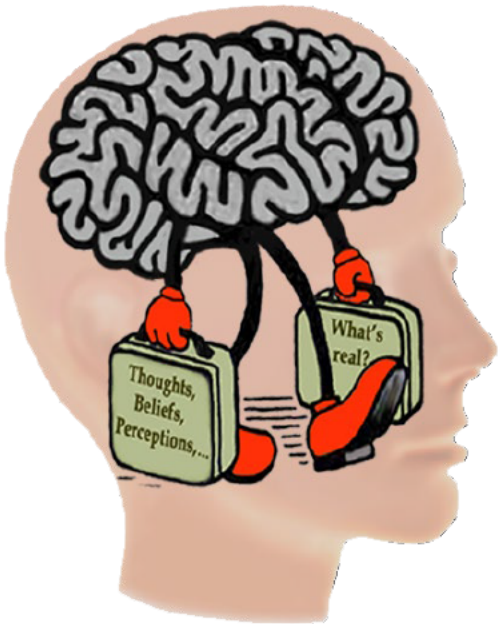
A Virtual Learning Experience

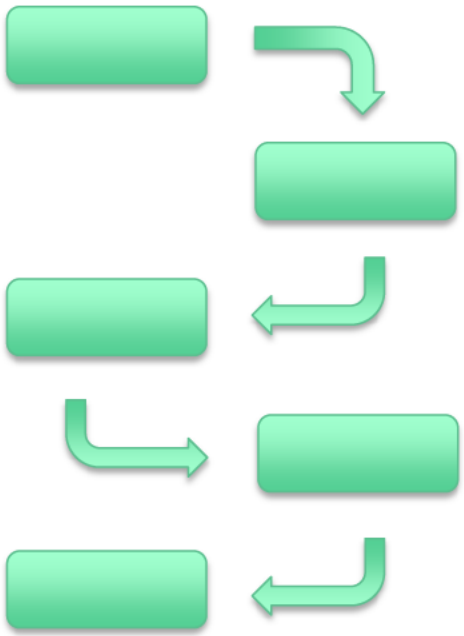
Combating Deepfakes

Misinformation is a Human Problem

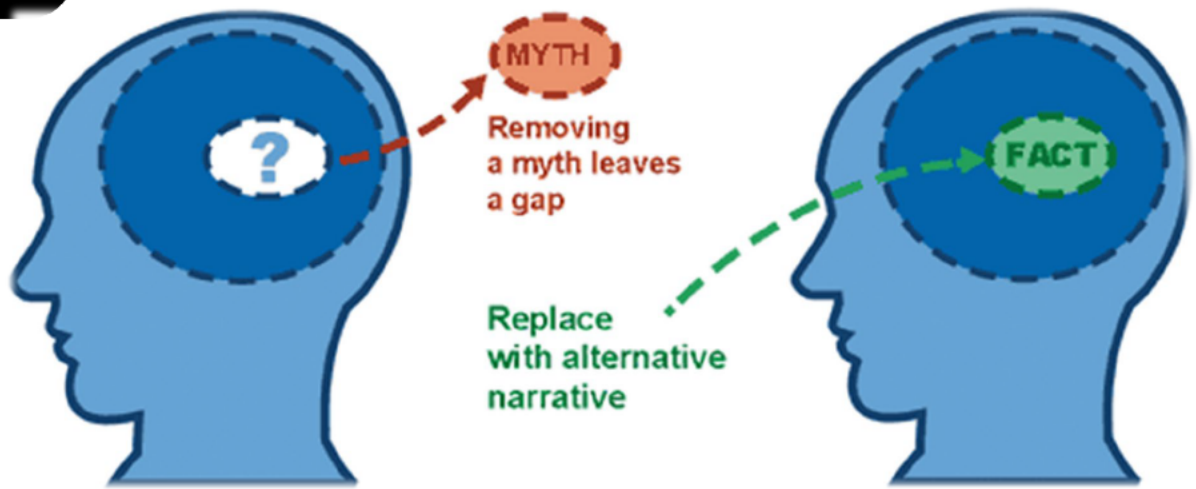


The Disinformation Problem





Combating Disinformation



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

What Should Businesses Do?

- 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