DeepFakes & Fake News: AI challenges in the media industry

Dr. Marina Arvanitidou

IRT – Institut für Rundfunktechnik

BLM
Bayerische Landeszentrale für neue Medien

IRT

Introduction

Deepfakes

Open Source Intelligence

Outlook

Introduction

Deepfakes

Open Source Intelligence

Outlook

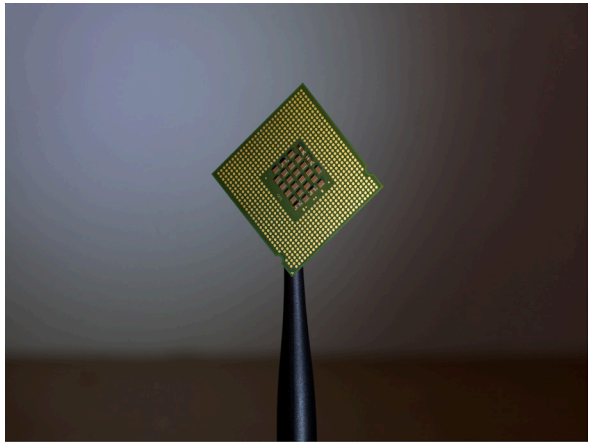
Manipulating video is not new!

Forrest Gump, 1994: filmmakers digitally inserted archival footages of JFK into this scene and manipulated his mouth movements.

Why now?

Algorithms

Resources



Data

Introduction

Deepfakes

Open Source Intelligence

Outlook

Deep Learning Fake News

How

Threats

Solutions

Increasingly realistic synthetic faces

Deepfake technology - ***Generative Adversarial Networks (GANs)***

maliciousaireport.com

Which person is real?

a

b

- ❑ a and c ?
- ❑ a and d ?
- ❑ b and c ?
- ❑ b and d?

VOTE time !

c

d

http://www.whichfaceisreal.com

Which person is real?

Play again.

a

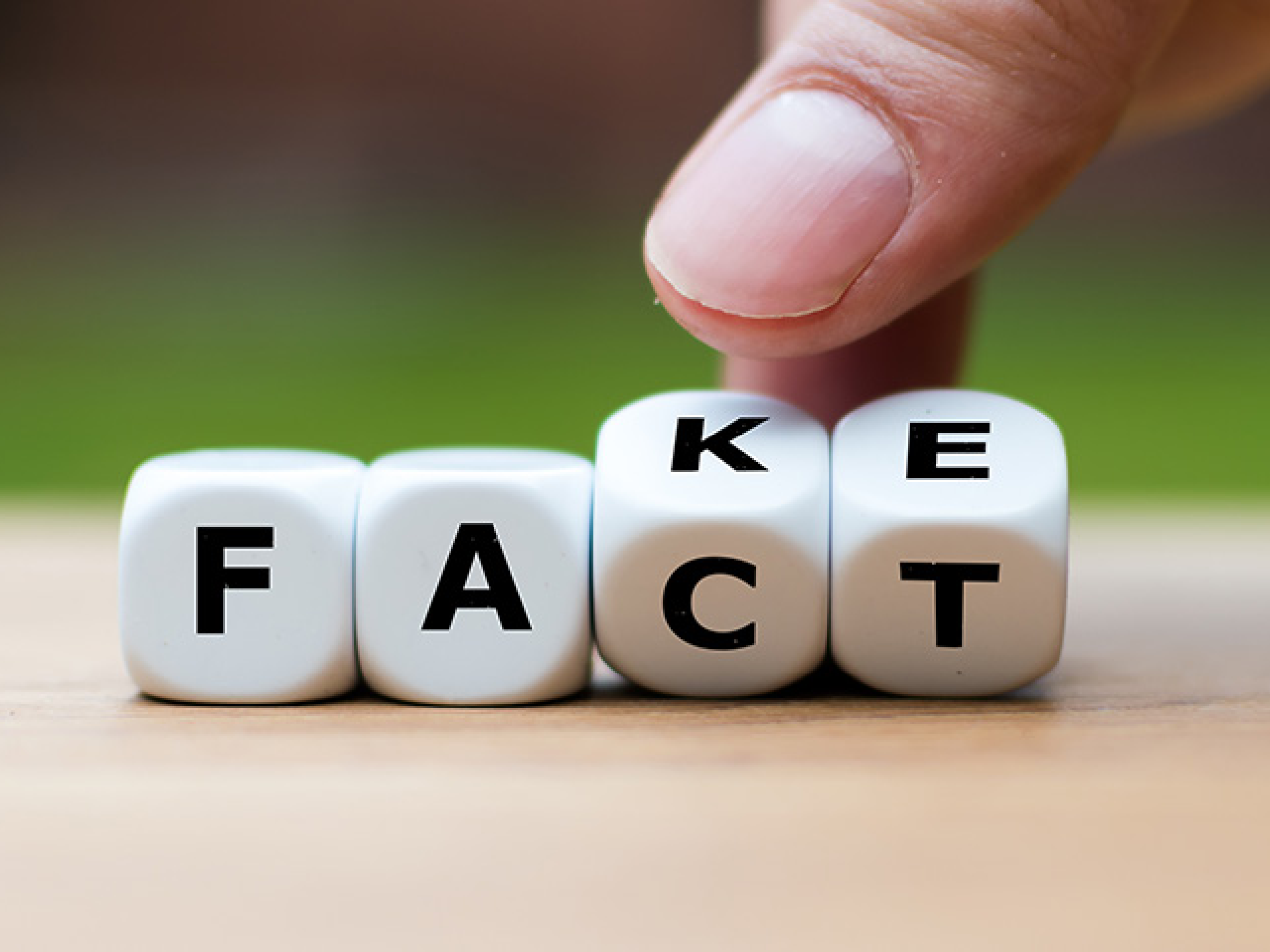
b

- ❑ a and c ?
- ❑ a and d ?
- ❑ b and c ?
- ❑ b and d?

c

d

http://www.whichfaceisreal.com


FAKE
FACT

Deepfakes are only the tip of the iceberg...

>2.1B

FAKE Facebook accounts disabled in 1Q2019

<40%

of deepfake images can be identified by an average person

3.6T

FAKE Youtube views per year

>50%

of new PhD's reported are FAKE

$1.3BN

Cost of Fake Instagram followers

>300M

FAKE Twitter accounts removed each year

$17BN

Lost due to financial misinformation

$9BN

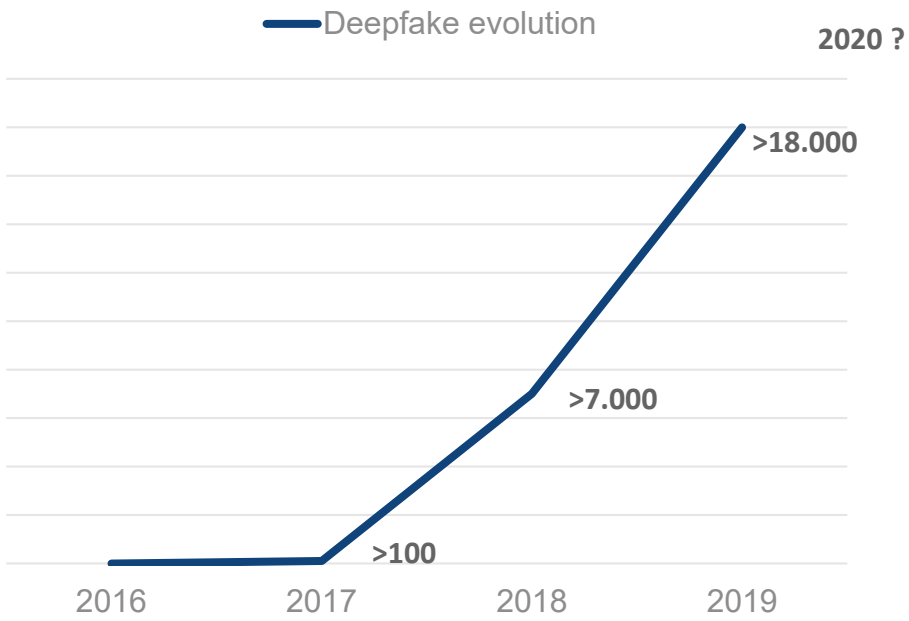
Spent repairing corporate reputation

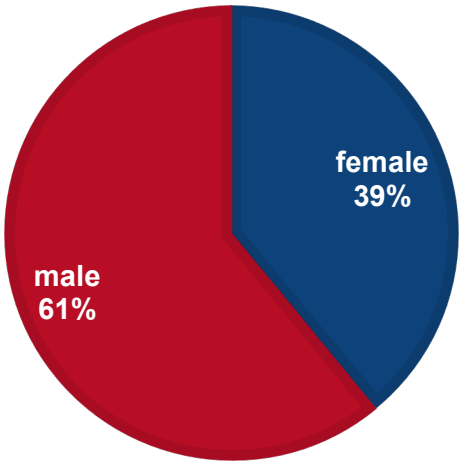
$3BN

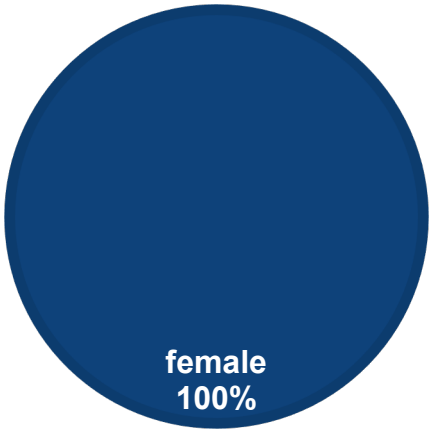
Spent by platforms to combat misinformation

https://www.deeptrustalliance.org/

Threats







DeepTracelabs.com

Deepfakes: Now what?

Solutions

Invest on technology	Policies and guidelines to be used by tech platforms to define what is harmful	Educate
Governments Academia Private	Policy makers	PSM

How to detect DFs?

Deepfakes' "flaws"
Image :
- Assymetries
- Blob artifacts
- Blurred background / regions
- Artifacts in face boundaries/edges

Audio:
- Lip Sync

Video:
- Real person's movements (head tilt, eyebrows, chin, "soft biometrics")

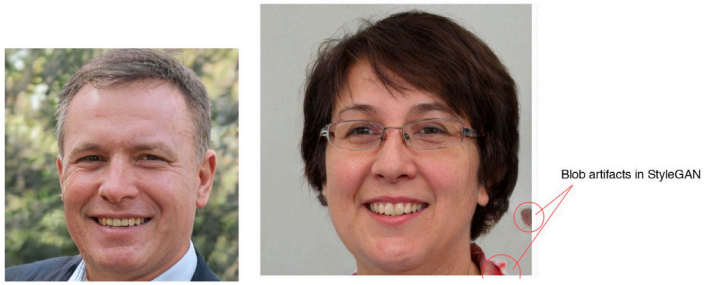
https://medium.com/@jonathan_hui/detect-ai-generated-images-deepfakes-part-1-b518ed5075f4

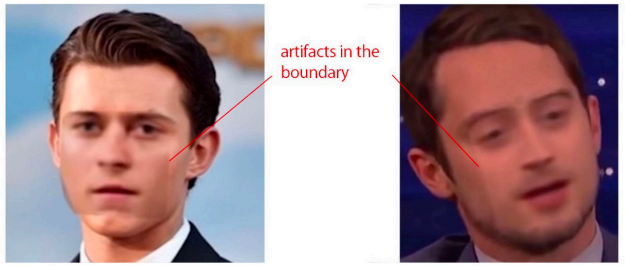
How to detect DFs?

Deepfakes' "flaws"
Image :
- Assymetries
- Blob artifacts
- Blurred background / regions
- Artifacts in face boundaries/edges

Audio:
- Lip Sync

Video:
- Real person's movements (head tilt, eyebrows, chin, "soft biometrics")



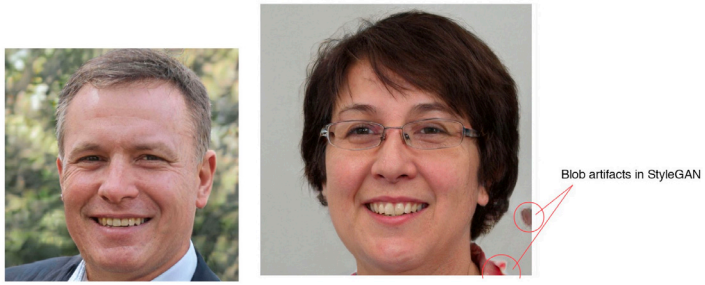


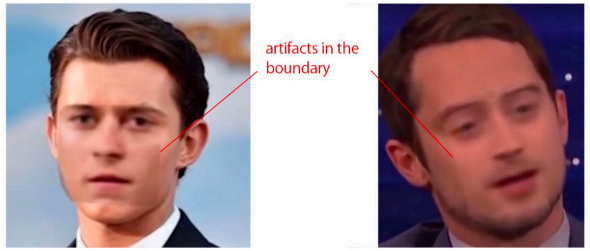
How to detect DFs?

Deepfakes' "flaws"
Image :
- Assymetries
- Blob artifacts
- Blurred background / regions
- Artifacts in face boundaries/edges

Audio:
- Lip Sync

Video:
- Real person's movements (head tilt, eyebrows, chin, "soft biometrics")





"Spotting Deepfakes is like spotting viruses: an ever changing challenge"

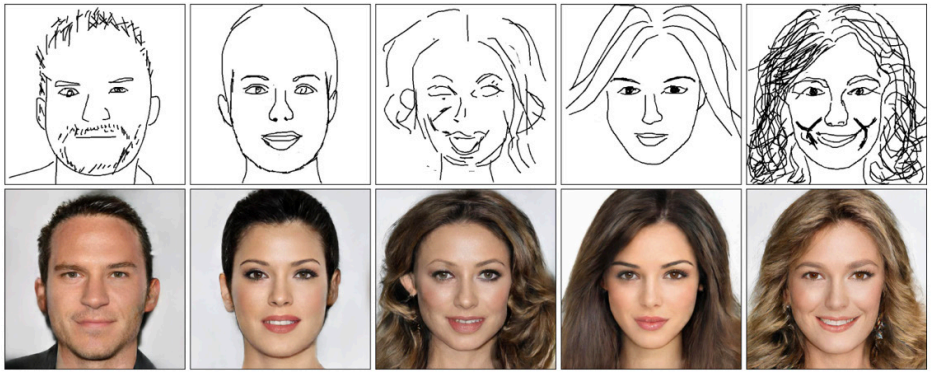
Solutions: Education

Education
- Professionals in the newsroom
  - Learn, use, create tools
  - Modify workflows
- Public
  - Trustworthy resources
  - How to distinguish fake/real

Collaboration (e.g. in EU level, EBU)
- Trainings
- Catalogue “digital forensics”

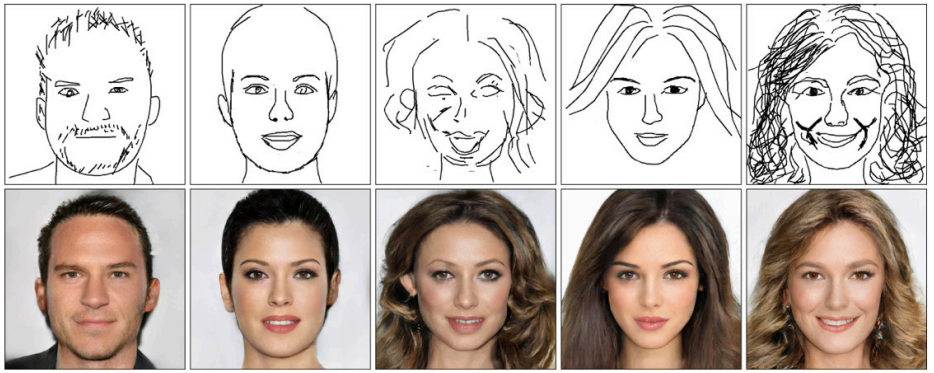
More applications… or business opportunities!

- generating realistic face images from sketches (Deepface Drawing)



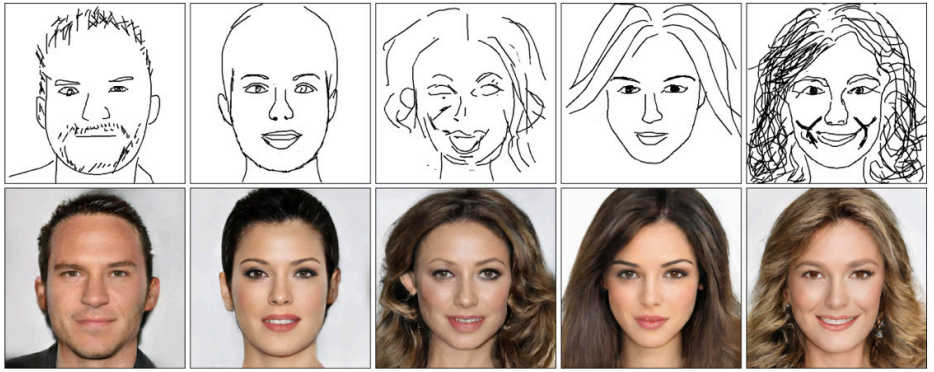
More applications… or business opportunities!

- generating realistic face images from sketches (Deepface Drawing)
- ..or even from text



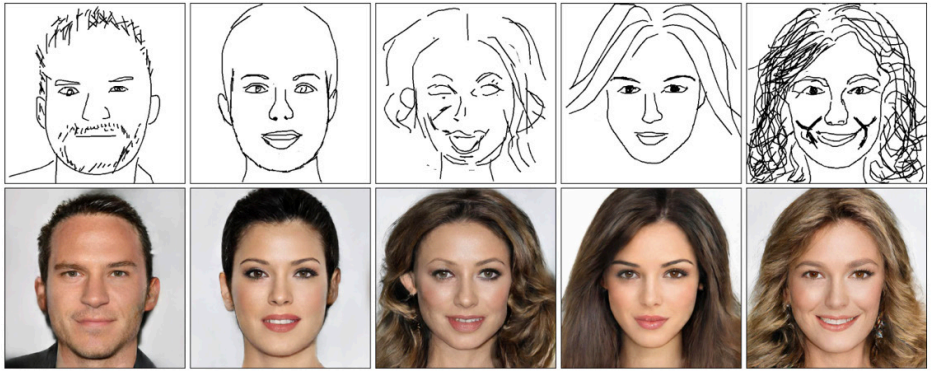
More applications… or business opportunities!

- generating realistic face images from sketches (Deepface Drawing)
- ..or even from text
- dubbing foreign-language films



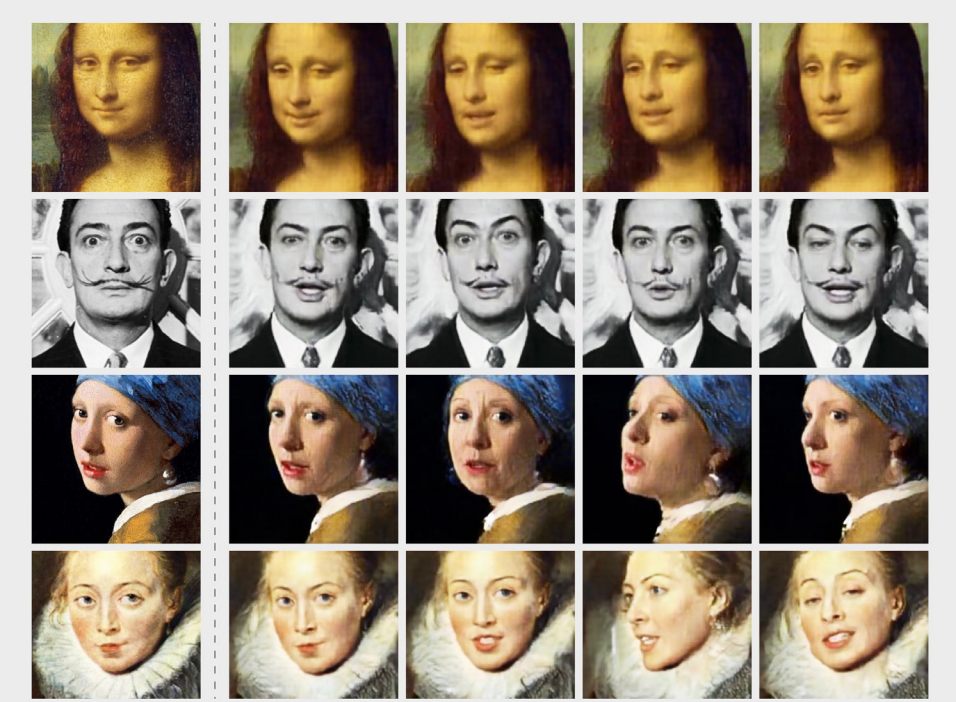
More applications… or business opportunities!

- generating realistic face images from sketches (Deepface Drawing)
- ..or even from text
- dubbing foreign-language films
- creating personalized synthetic advertising



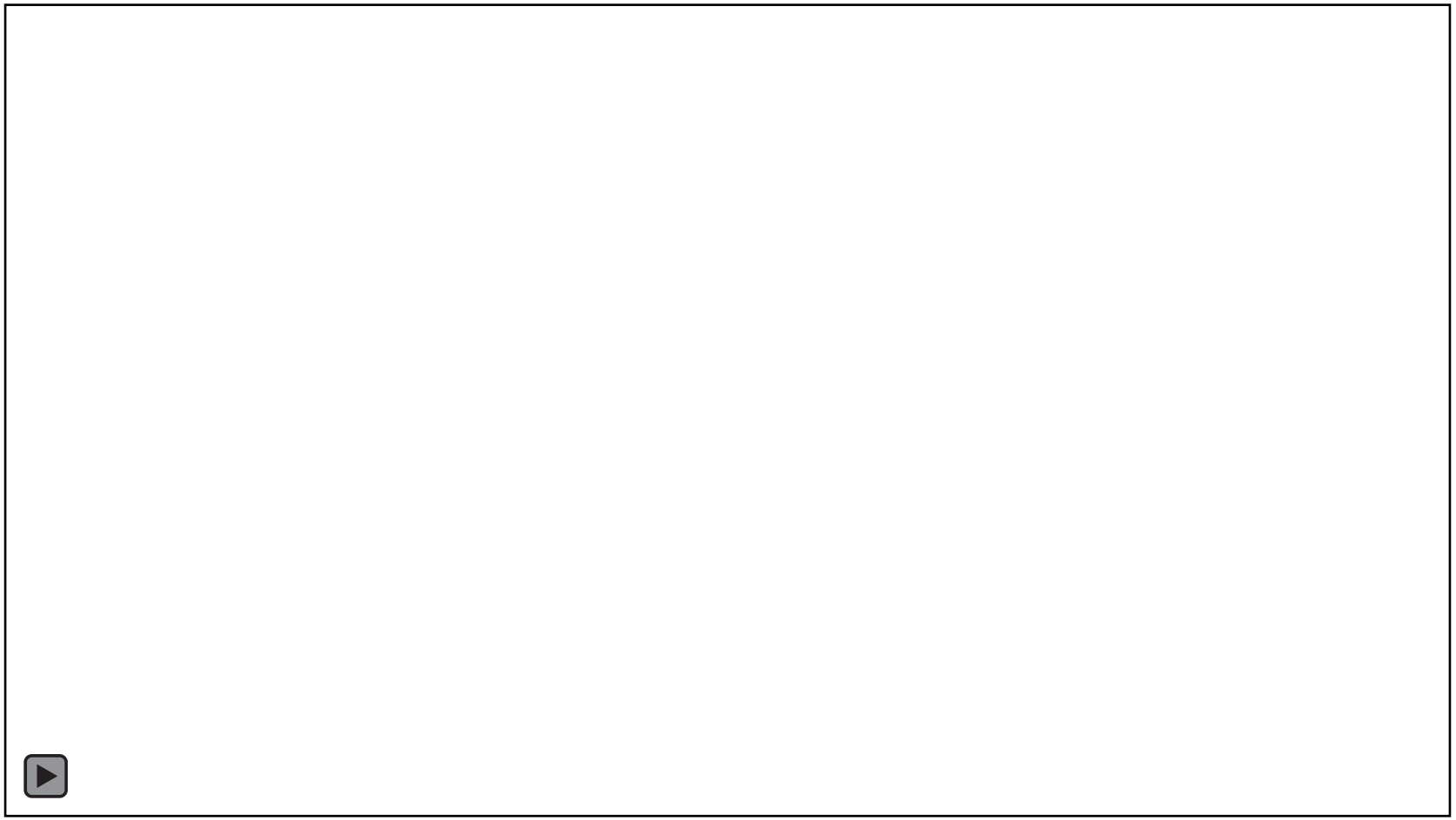
More applications… or business opportunities!

- generating realistic face images from sketches (Deepface Drawing)
- ..or even from text
- dubbing foreign-language films
- creating personalized synthetic advertising
- creating living prortaits of favourite persons for education



Living portraits… or Mona Lisa smiled!
https://www.wired.com/story/deepfakes-getting-better-theyre-easy-spot/

Dalí once said… “If I die, I won’t completely die.”



“Dali Lives,” Dali Museum in St. Petersburg, Florida

Introduction

Deepfakes

Open Source Intelligence

Outlook

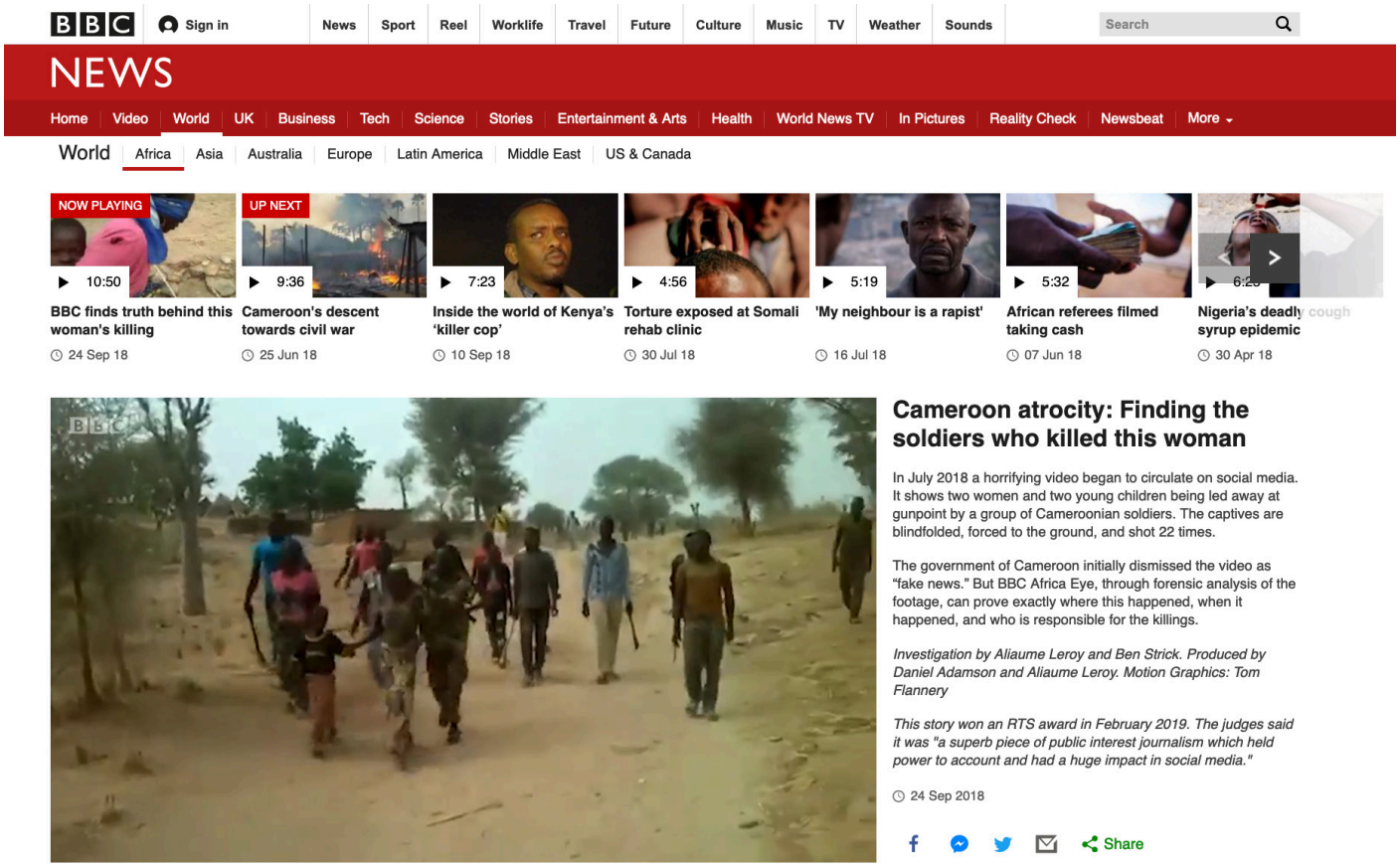
Open Source INTelligence

Investigative Journalists

OSINT

Photo credits: franki chamaki

Investigative journalism – BBC Cameroon atrocity



https://www.bbc.com/news/av/world-africa-45599973/cameroon-atrocity-finding-the-soldiers-who-killed-this-woman

Investigative journalism – BBC Cameroon atrocity

Where?

Topography match: mountain ranges

Tracks/roads, buildings, trees

When?

Shadow analysis

Historical satelite data

Who?

Identify objects: gun models, soldier uniforms

Analyze background sounds and voices

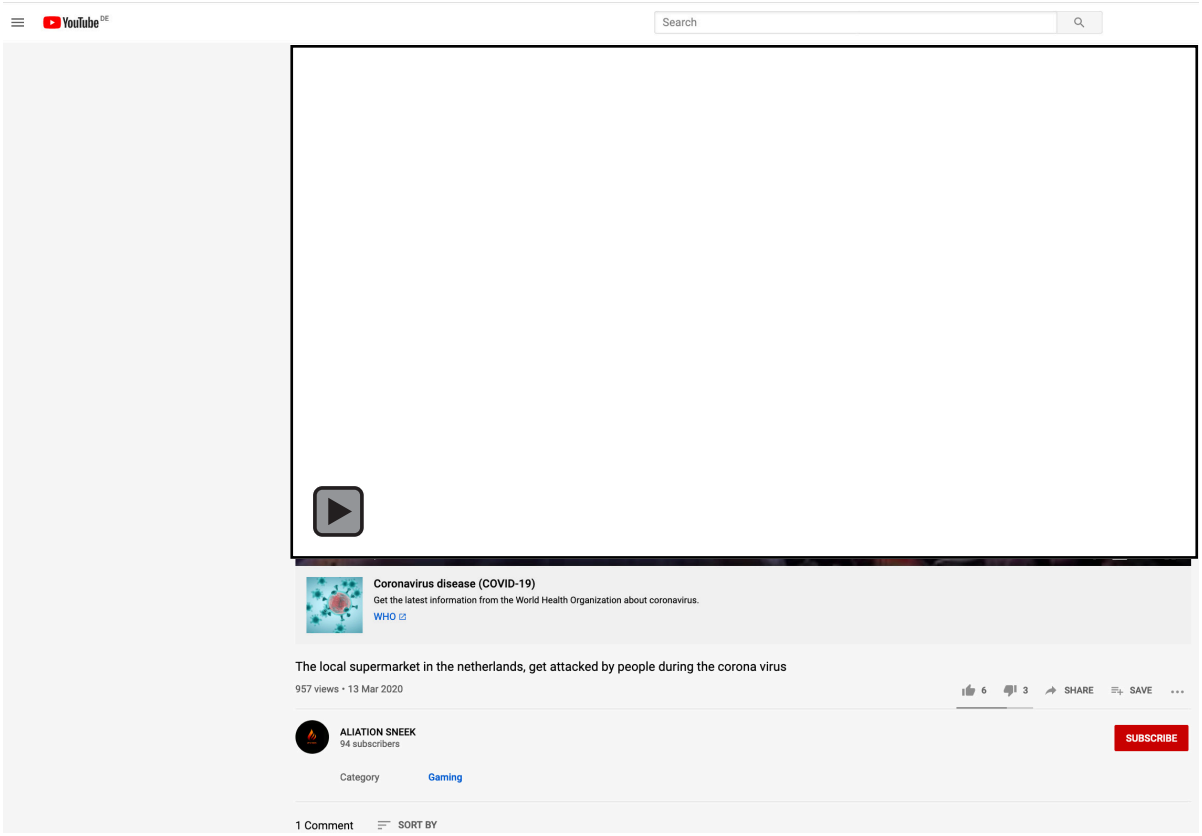
uniforms used b
addition, and st
equipment of th
heavy helmets,
the video;

The weapons di
are not those us

One of the men,
indicating his ra
conformity with

As you can see.

COVID-19 Panic: The “Haarlem AlDI” by Bellingcat



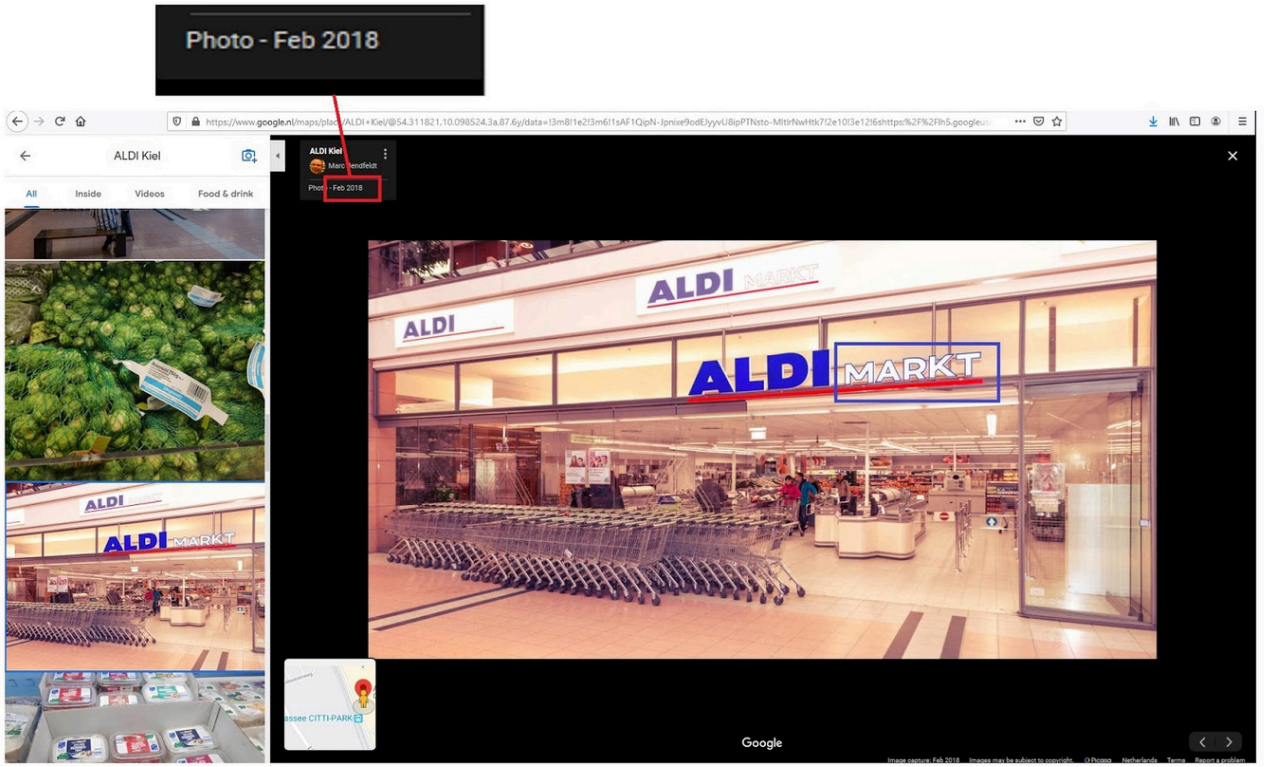
https://www.bellingcat.com/news/2020/03/13/monitoring-and-debunking-covid-19-panic-the-haarlem-aldi-hoax/

COVID-19 Panic: The “Haarlem AlDI” by Bellingcat

1. Sings in German NOT in dutch

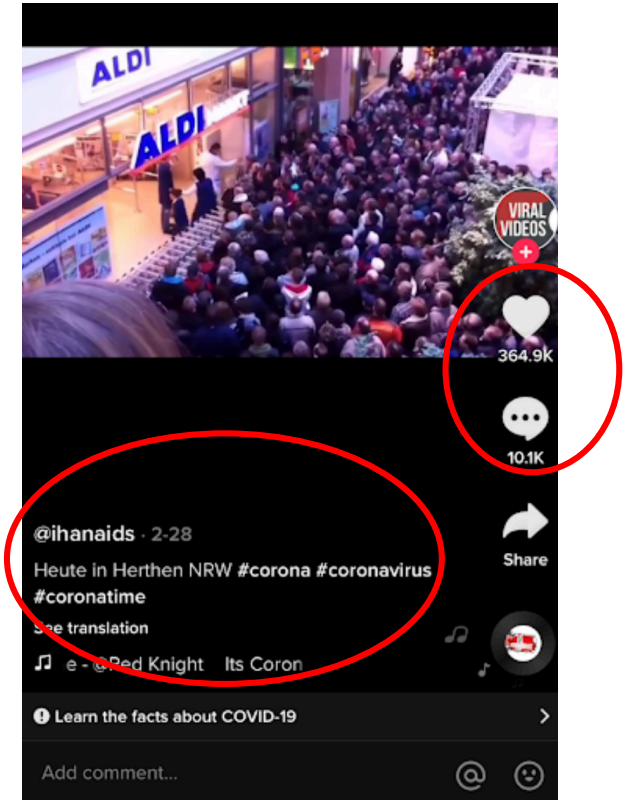
COVID-19 Panic: The “Haarlem AlDI” by Bellingcat

1. Sings in German NOT in dutch
2. Track back: logo not up-todate



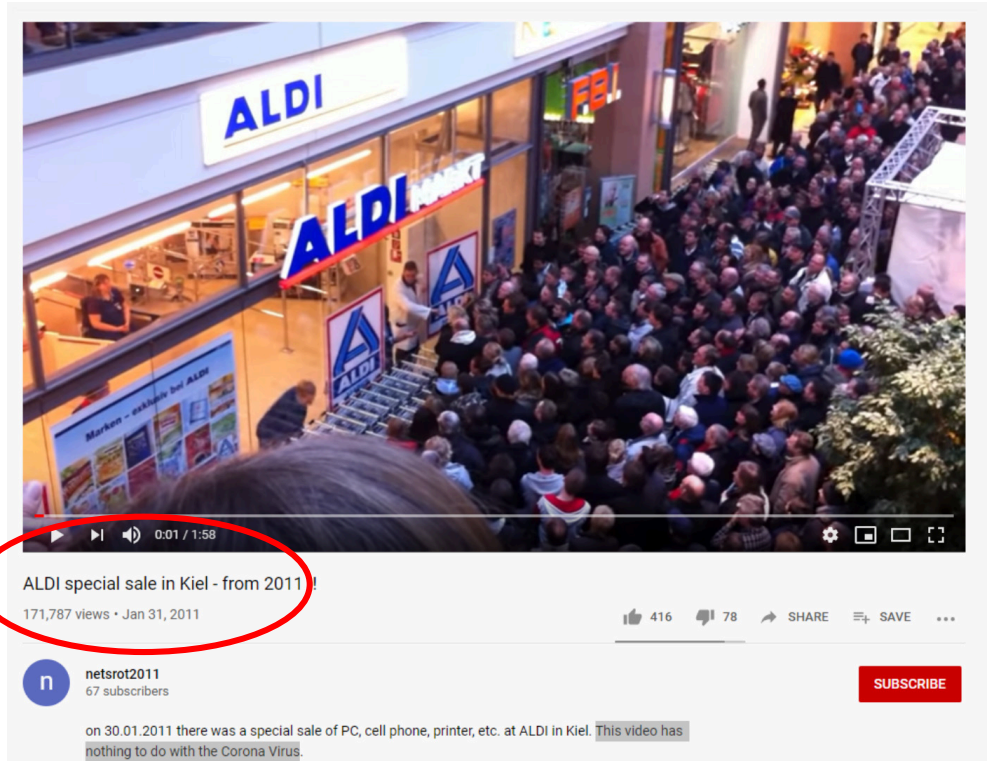
COVID-19 Panic: The “Haarlem AlDI” by Bellingcat

1. Sings in German NOT in dutch
2. Track back: logo not up-todate
3. Track in social media: twitter, facebook and TikTok (@Herthen) <- viral



COVID-19 Panic: The “Haarlem AlDI” by Bellingcat

1. Sings in German NOT in dutch
2. Track back: logo not up-todate
3. Track in social media: twitter, facebook and TikTok (@Herthen) <- viral
4. Find original video: Kiel 2011!



Introduction

Deepfakes

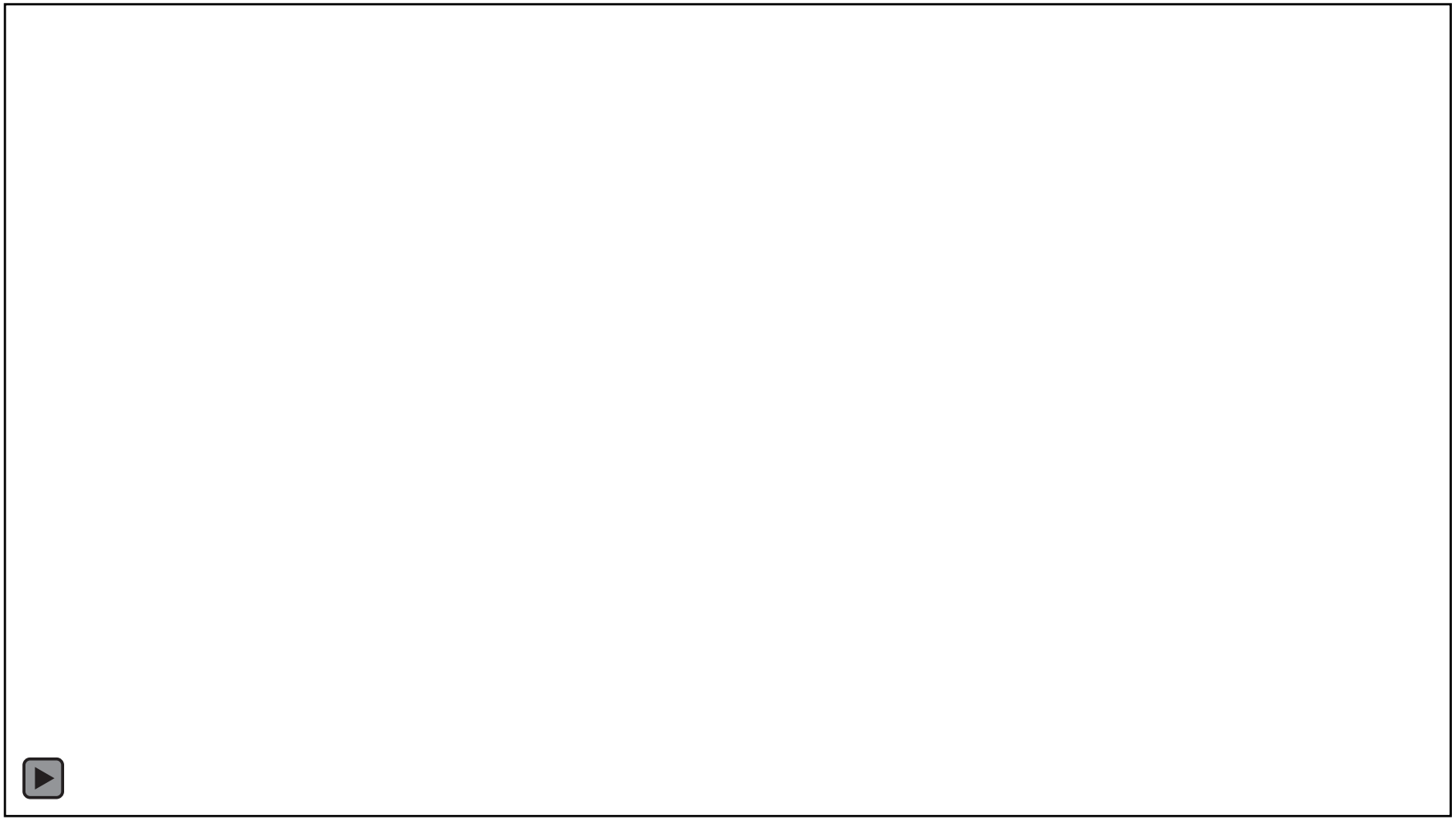
Open Source Intelligence

Outlook

Outlook

- AI changes media (both opportunities & risks)
- Synthesized media: Deepfakes - a major threat
- Technology: changes the way journalists work – trustworthy reporting
- Fight misinformation

On the way out.. ☺



Thank you for your attention!

Experts in audio-visual media

Dr. Marina Arvanitidou
Lead R&D Engineer Artificial Intelligence

Floriansmuehlstrasse 60
80939 München
www.irt.de
marina.arvanitidou@irt.de

Bayerische Landeszentrale für neue Medien

All rights reserved. All text, images, graphics and charts are protected by copyright.
Reproduction or use of the content is not permitted without the express consent of the author.