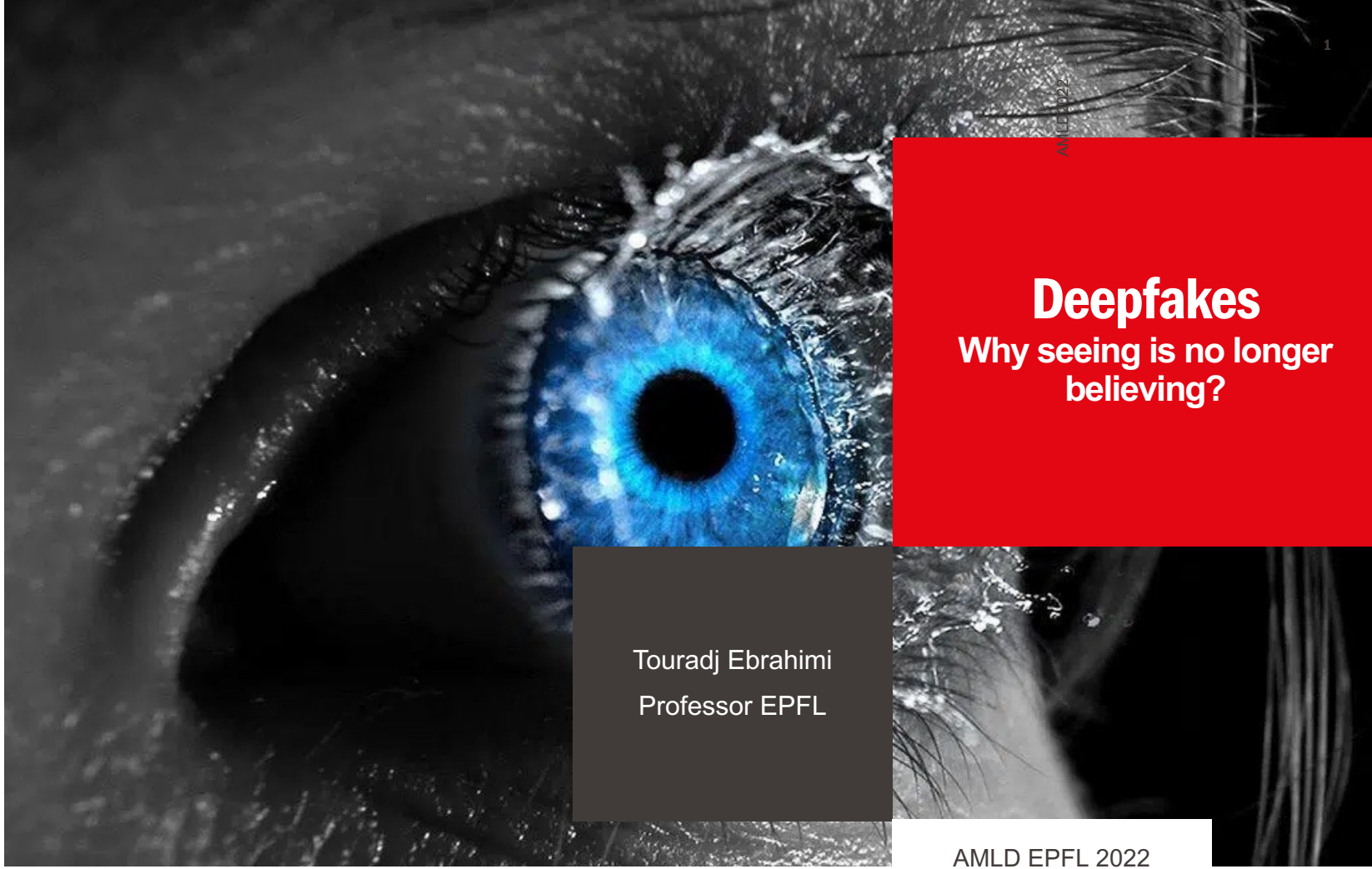


Deepfakes

Why seeing is no longer believing?

Touradj Ebrahimi

Professor EPFL

AMLD EPFL 2022

AMLD 2022

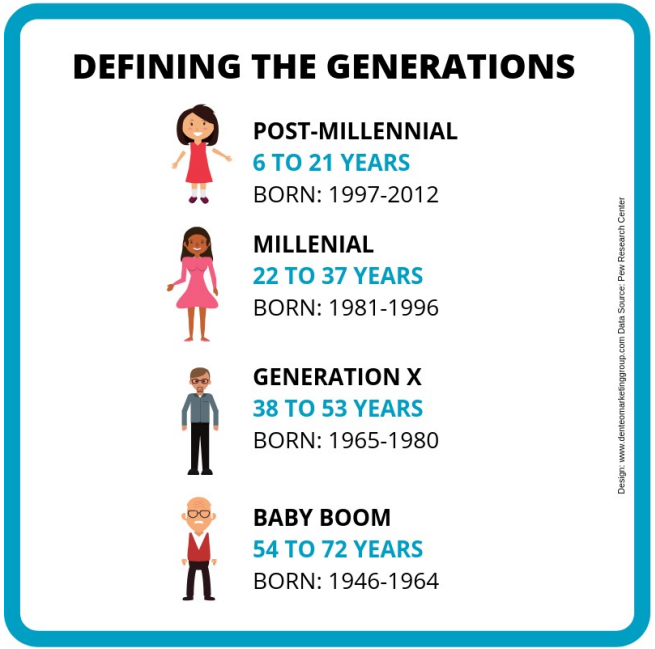
Manipulation of Public Opinion

30 March 2022

Multimedia content is produced by anybody and consumed everywhere

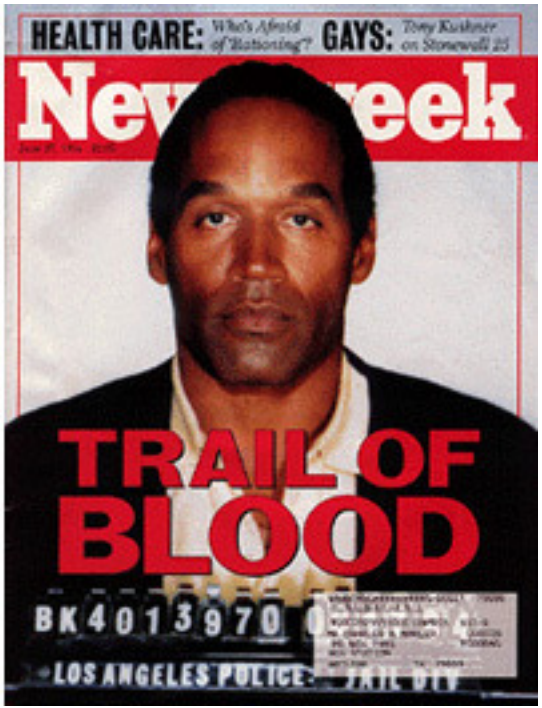
Main drivers behind the era of seamless multimedia content

- Cheap capture and display connected devices
  - smart phones
  - tablets
  - laptops
- Free, effective and easy to use software to edit content
- Cheap or free storage on devices and on the cloud
- Efficient means to distribute content
  - Internet (Mailing lists, WhatsApp interest groups, blogs, …)
  - Social network (Instagram, Snapchat, Tiktok, …)
- Generational change

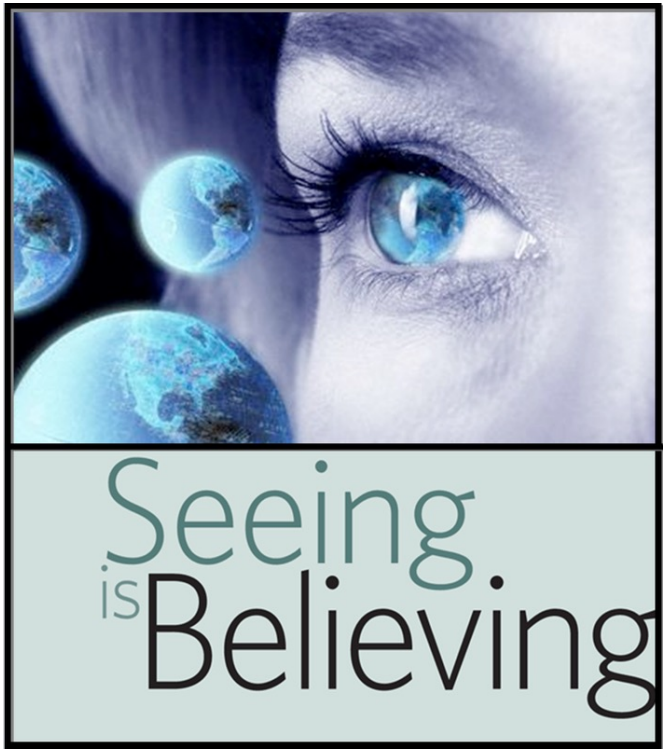


Misleading pictures reinforce disinformation

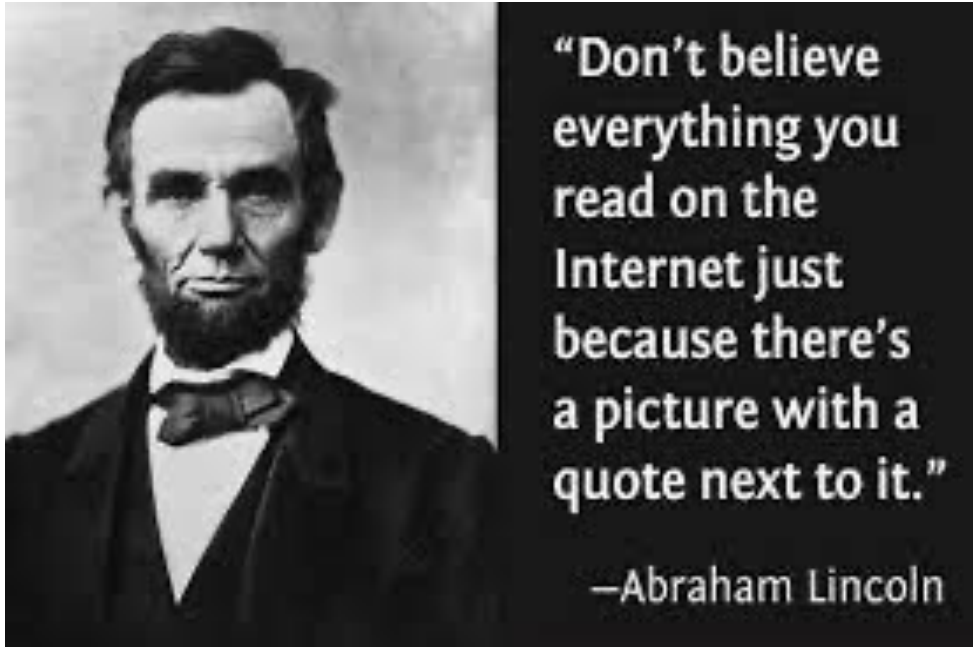
Hidden messages through pictures…



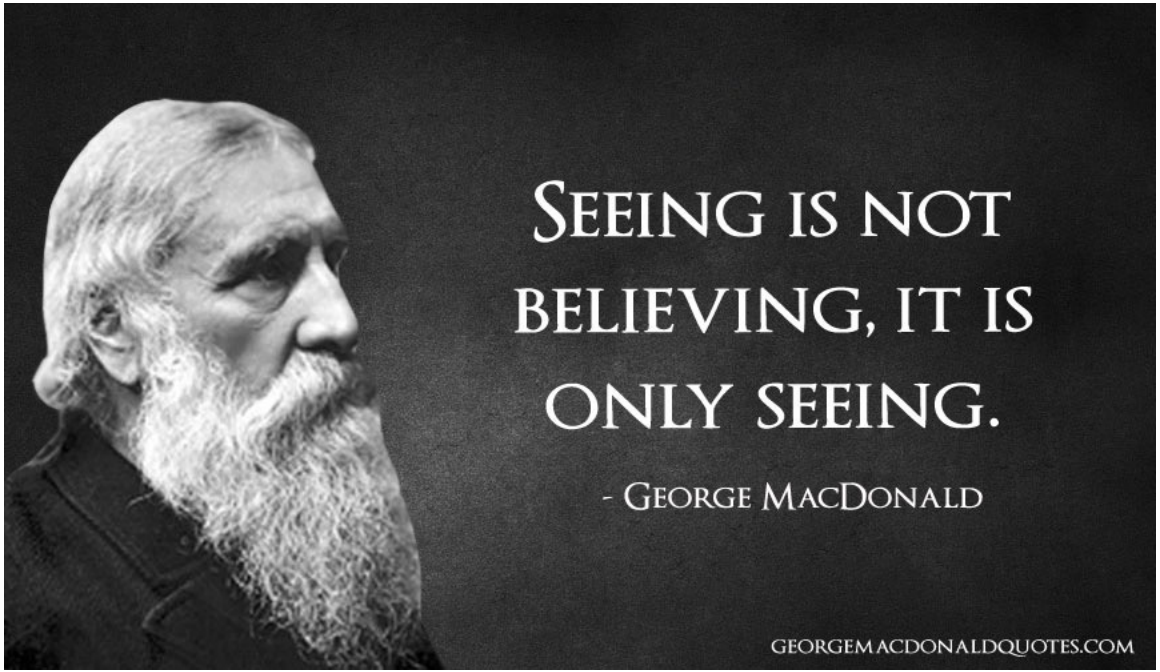
The power of pictures …



Disinformation through picture/video



Disinformation through picture/video



History of media manipulation

Abraham Lincoln

Lincoln's head was added on top of southern politician John Calhoun's portrait.

Joseph Stalin

Leaders remove people (from the images) whom they no longer wanted to associated.

Canadian PM

William Lyon Mackenzie King removes King George VI from a photo with Queen Elizabeth to portray himself more powerful.

Soviet Soldiers

Russian magazine removes the watches from soldiers' wrists to ensure that their readers don't think the soldiers were looting.

Oprah Winfrey

TV guide edited the cover image where they used Oprah's head on the body of Ann-Margaret.

OJ Simpson

Time magazine edited OJ Simpson's image after his arrest and made it darker and more sinister. Actual one was displayed in News Week.

Iranian Missiles

The doctored image was released by the Iranian Government to show successful launch of four missiles when only three were successful.

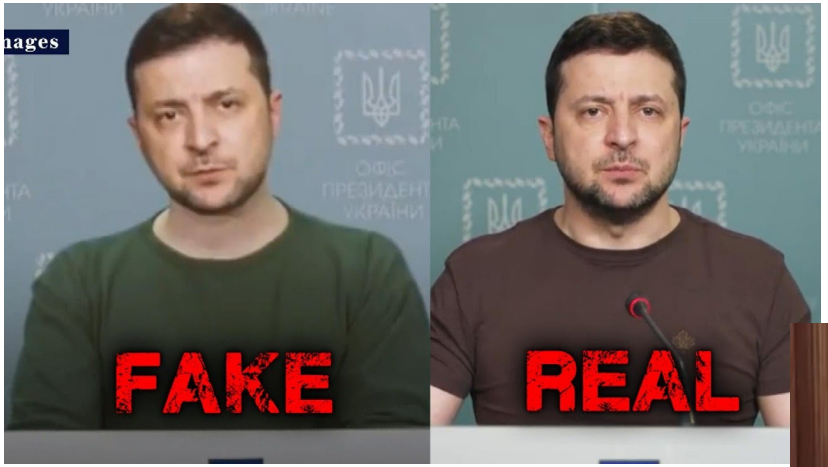
Deepfake Tom Cruise

Near realistic deepfake of Tom Cruise indicates the potential of AI based media manipulation. Image courtesy: Belgium VFX specialist Chris Ume.

AI generated virtual characters

Deepfakes

Deepfake: a new weapon in war time!



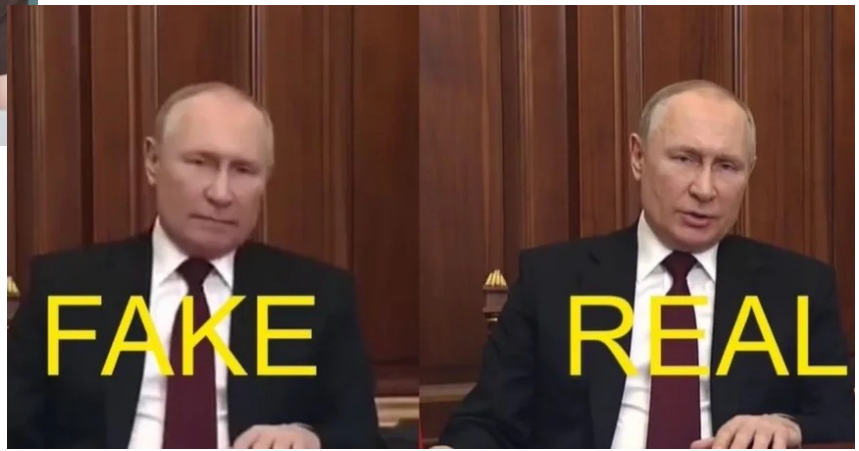
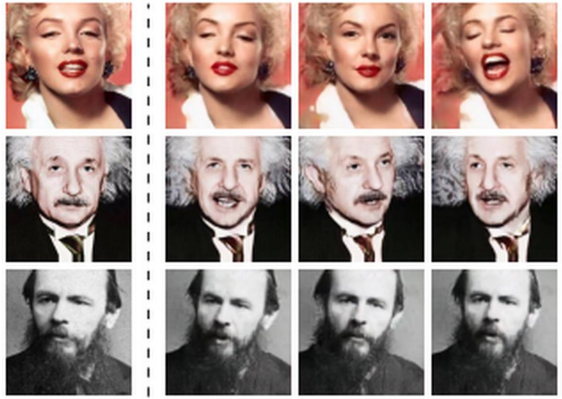


Image Forensics

- Well before the current phenomenon of fake news and deepfakes

Reactive and Proactive approach to image forensics

Reactive approach: detection



Reactive approach: manipulated media detection → Challenges

- Distinguish between **malicious deepfakes** and media manipulated with AI techniques for **creative purposes**.
- Access to **sufficiently large databases** of typical manipulations with **reliable labelling**.
- **Black-box challenge**: identify why and under which circumstances an approach fails and how to improve it.
- **Cat and mouse game**: while detection performance improves, performance of generation methods improves as well.
- **Adoption of technology**: media distribution platforms may not be prompt to integrate the most advanced detection solution.

Proactive approach: provenance annotation

- **Provenance**: a set of information about a media asset including the trail of modifications starting from an actor.

- **Actor**: a human or non-human (software or hardware) that participates in the media ecosystem.
  - The camera, the photographer, the editor, the editing software…

Proactive approach: provenance annotation

Provenance photographer

Origin
Digital camera

Source
Camera RAW

Modifications:
- Cropped
- Contrast ↗
- White balance
- Vignetting
- Compression

Provenance client

Digital Master
High quality JPEG

Modifications:
- Rescaled
- Transcoded

Modified
Reduced size JPEG
Shared on social media

Proactive approach: provenance annotation → Challenges

- In a provenance model, **interoperability** and **security** are essential.
- Model should allow description of information about the **creation** of the asset as well as **modifications**.
- Additional information about **actors** might be required.
- Vision pursued by:
  - Content Authenticity Initiative (CAI)
  - Project Origin
  - JPEG Fake Media

} Coalition for Content Provenance and Authenticity (C2PA)

A multidisciplinary challenge

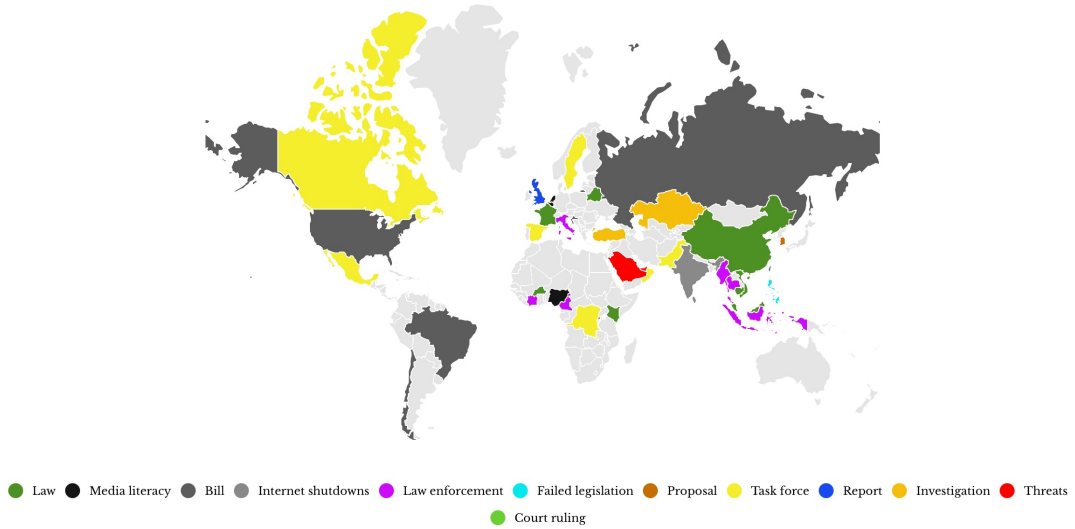
- Fake media is not only a scientific and technical challenge but also:
  - Educational, social challenge
  - Legal, policy challenge

Educational Challenge

- Fake news spread faster than the truth
- High quality synthetic fakes are more trusted than real people!
- Humans, not bots, are primarily responsible for spread of misleading information

Legal Challenge

- There is an urgent need for a General Data Protection Regulation (GDPR) on misinformation, fake news and fake media



Current situation regarding governments actions against misinformation

History in the making

- Moving target

- Not yet well understood nor fully writen yet