A trick of the eye? Video verification in the age of deepfakes.

Hazel Baker  Global Head of UGC Newsgathering, Reuters

#mtm19
Deepfake: “a portmanteau of ‘deep learning’ & ‘fake’”

Realistic synthetic video that has been created using computer-generated imagery powered by artificial intelligence (AI)
Is synthetic media really a serious threat for the news industry?
"Deepfake propaganda is not a real problem. We've spent the last year wringing our hands about a crisis that doesn't exist."

Russell Brandom, The Verge, March 5 2019
“Journalists and media (also) need to have robust processes in place for identifying deepfake video footage, in order to ensure high quality news reporting.”

(Automated tackling of disinformation, A. Alaphilippe, A. Gizikis, C. Hanot & K. Bontcheva - March 2019)
What if we created our own ‘deepfake’ to test how our teams react?
What did we learn?
IDENTIFYING THE ‘DEEPFAKE’

1. Audio to video synchronisation
IDENTIFYING THE ‘DEEPFAKE’

1. Audio to video synchronisation
2. Mouth shape & sibilant sounds
IDENTIFYING THE ‘DEEPFAKE’

1. Audio to video synchronisation
2. Mouth shape & sibilant sounds
3. Static subject
IDENTIFYING THE ‘DEEPFAKE’

1. Audio to video synchronisation
2. Mouth shape & sibilant sounds
3. Static subject

NOTE: ‘red flags’ will become less evident as technology progresses
IDENTIFYING THE ‘DEEPFAKE’

Awareness of threat altered viewer’s assessment
IDENTIFYING THE ‘DEEPFAKE’

“Listening to your senses tells you – there’s something strange here!”

“Hmmm, it seems like it still needs production... in the first few seconds the sound doesn’t match.”

“Is this not her voice? She seems to breathe in a bit while talking?”

“Waaaah! You can’t tell it at the first watch.”
A SLIDING SCALE OF DECEIT

LEVEL OF MANIPULATION

NONE
A SLIDING SCALE OF DECEIT

LEVEL OF MANIPULATION

NONE

EDITED
A SLIDING SCALE OF DECEIT

- NONE
- STAGED
- EDITED

LEVEL OF MANIPULATION
A SLIDING SCALE OF DECEIT

LEVEL OF MANIPULATION

NONE  STAGED  EDITED  CGI
A SLIDING SCALE OF DECEIT

NONE  STAGED  ‘DEEPFAKE’

EDITED  CGI

LEVEL OF MANIPULATION
UGC VERIFICATION TECHNIQUES

- Reverse Image search
- Geo-location
- Meta data examination
- Corroborating imagery
- Directly questioning source
- Subject expert consultation
REVERSE IMAGE SEARCH THROUGH INVID-PROJECT.EU

Video keyframes

With the contextual menu (right click) you can query multiple reverse image search engines or use the magnifier and forensic tools on each keyframe.

Show more keyframes

Zuckerberg: 'We can make it much harder' | CBC News - CBC.ca
https://www.cbc.ca > ...

620 x 349 - Nov 5, 2018 - Facebook CEO outlines coming changes after alleged U.S. election interference by Russia.
UGC VERIFICATION TECHNIQUES

- Reverse Image search
- Geo-location
- Meta data examination
- Corroborating imagery
- Directly questioning source
- Subject expert consultation
LOOKING AHEAD: CAN TECH SOLVE THE PROBLEM?

CREDIT: FaceForensics project (via MIT Technology Review)
KEY CONCLUSION – DETECTING A ‘DEEPFAKE’

HUMAN JUDGEMENT
+
SUBJECT EXPERTISE
+
VERIFICATION FRAMEWORK

> (CURRENT)
TECHNOLOGY
SOLUTIONS