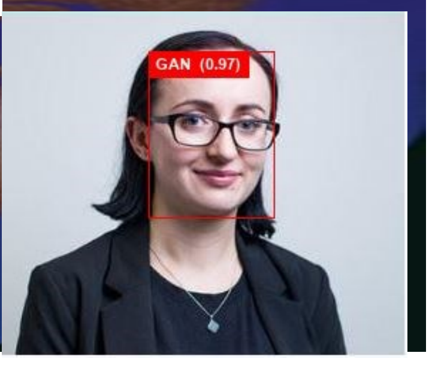
Deepfake Forensics: Exploring the Impact and Implications of Fabricated Media in Digital Forensic Investigations

Dr Áine MacDermott, School of Computer Science and Mathematics,
Liverpool John Moores University, UK
a.m.macdermott@ljmu.ac.uk



Abstract

Deepfake technology is continually evolving, becoming more sophisticated and harder to detect. As a result, law enforcement agencies must constantly keep up with the latest advancements and techniques used in creating deepfakes. There has been a concerning rise in malicious uses of deepfakes, including non-consensual pornography and various criminal activities such as defamation, fraud, and spreading misinformation. Deepfake detection and analysis pose significant challenges for digital forensic practitioners due to time constraints on identification and analysis, and the lack of forensic indicators within said media.

This project explores the impact of deepfakes on criminal investigations and digital forensic practitioners, the growing influence of fabricated media, and the challenges in identifying and combating manipulated digital media.

Background

A tabletop discussion and survey on deepfake forensics was conducted with the participation of 25 skilled digital forensic and cybersecurity practitioners at Liverpool John Moores University (July 2023). This study aimed to explore the evolving landscape of deepfake technology and its impact on the field of forensic analysis. Through their collective insights and expertise, this survey sought to enhance the understanding of deepfake forensics techniques, identify emerging trends, and contribute to the development of effective countermeasures to combat the growing threat of manipulated digital media.

Below are questions from the survey and outputs:

1. *Are you worried about the rise in manipulated digital media/deepfakes?*

90% of participants agreed this is a growing problem and an area to watch.

2. *Is there a certain demographic that you think this affects more?*

Participants identified 'bullying' and 'indecent images' as two main areas in the current scope. Bullying cases involved teenagers and young adults as the main victims, with the created deepfake causing distress and concern due to their perceived realism in the media. Indecent image deepfakes involved revenge porn material being created or indecent media of young adults.

3. *Do you think criminals are likely to use anti-forensics techniques?*

75% of participants agreed with this question. Some suggested that social messaging applications and the growth in 'disappearing messages' features meant that suspects would commit malicious actions with the belief that they could not be traced. There has been a rise in deepfake audio/video in relation to crimes, e.g., 'CCTV' of an incident that has been manipulated or 'threatening recorded audio' suggesting that an individual has threatened someone.

4. *Are there any specific techniques you would use to identify them?*

Respondents conveyed that it often depends on the quality of the deepfake when initially deciding to analyze them. Often, they can be detected by the naked eye based on certain inconsistencies to identify their manipulation.

5. *Are there any challenges to detecting deepfakes?*

Budget and time constraints played a big role in the ability to detect deepfakes promptly. There are issues surrounding limited resources and expertise as deepfake detection requires specialized knowledge, skills, and tools. Training personnel in deepfake forensics and providing access to advanced software can be expensive and challenging.

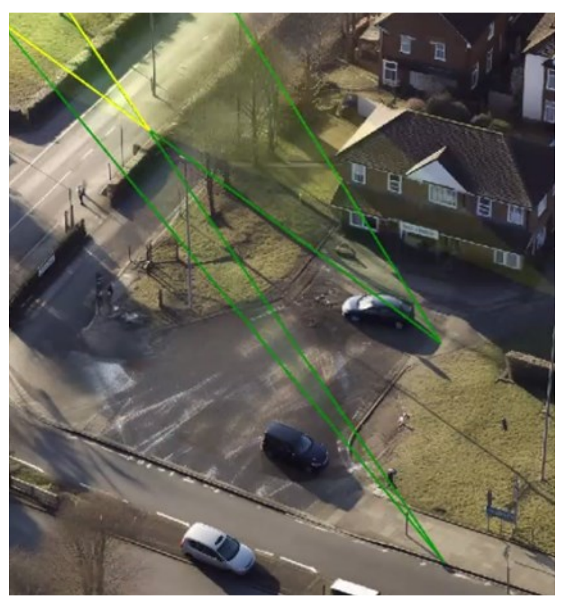
For deepfake detection techniques, the quality and size of an image can affect the accuracy of detection tools. Also, many of the 'suggested' tools work best with images and are not as accurate with videos.

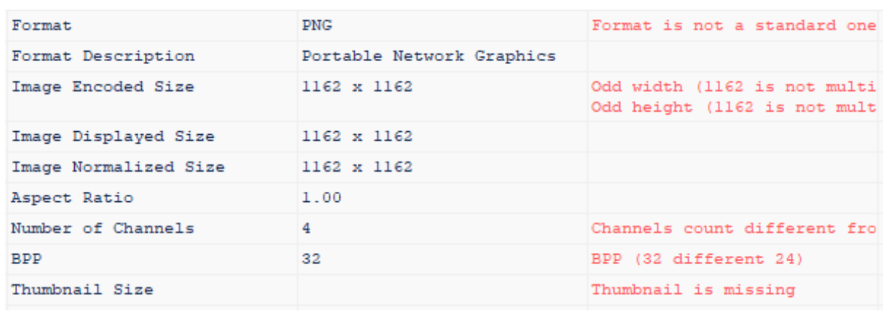
Conclusion: There is a technological divide in the sector where many examiners are not fully aware of the sheer capabilities of deepfake tools. The frequency and use-cases of deepfake media in forensic investigations is rising. We invite forensic and security practitioners/researchers to take part in a global digital forensics needs assessment to gather data on 'deepfake impact in digital forensic units and cyber incidents'.

Research undertaken as part of LJMU QRPSF fund.

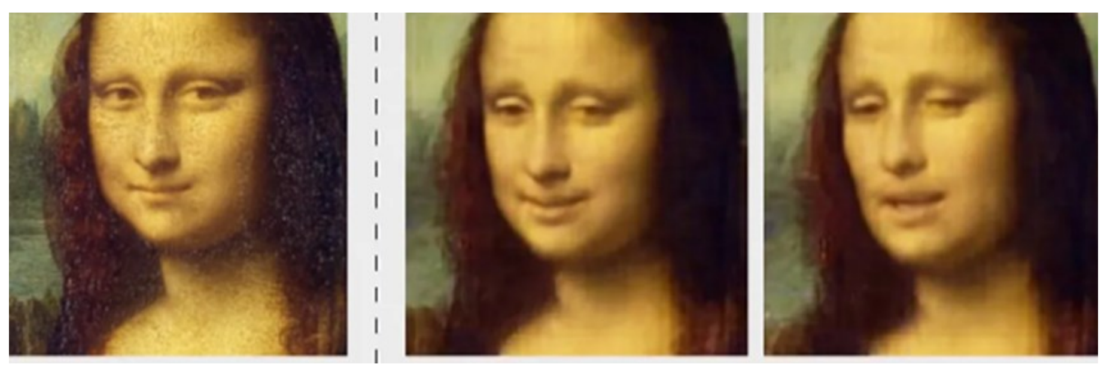
Deepfake Media Indicators

- Unnatural backgrounds
- Inconsistent facial expressions
- Inconsistent lighting/shadows/reflections
- Unnatural movements
- Audio inconsistencies
- Image search





Format	PNG	Format is not a standard one
Format Description	Portable Network Graphics	
Image Encoded Size	1162 x 1162	Odd width (1162 is not multi Odd height (1162 is not mult
Image Displayed Size	1162 x 1162	
Image Normalized Size	1162 x 1162	
Aspect Ratio	1.00	
Number of Channels	4	Channels count different fro
BPP	32	BPP (32 different 24)
Thumbnail Size		Thumbnail is missing



Aim to determine whether the media is 'real' – 'real' means non deepfake but it could be manipulated in other ways not detected by available tools!

Experiment Media

We created 100 deepfake test data images (50 real and 50 deepfake) from *"ThisPersonDoesNotExist"* and *"ThisPersonExists".* We ran the images through a popular online deepfake detection tool 'DFDetect' and then through a premium tool 'Amped Authenticate'. DFDetect generates a percentage score on how likely the image is to be authentic. A sample of findings are below:

Table 1: DFDetect Results

No.	DFDetect (Deepfake)		
1	**100%**	11	100%
2	100%	12	100%
3	**1.23% fake**	13	99.23%
4	100%	14	100%
5	**100%**	15	95.20%
6	100%	16	100%
7	**70% fake**	17	99.55%
8	100%	18	100%
9	100%	19	99.99%
10	100%	**20**	**51.31%**

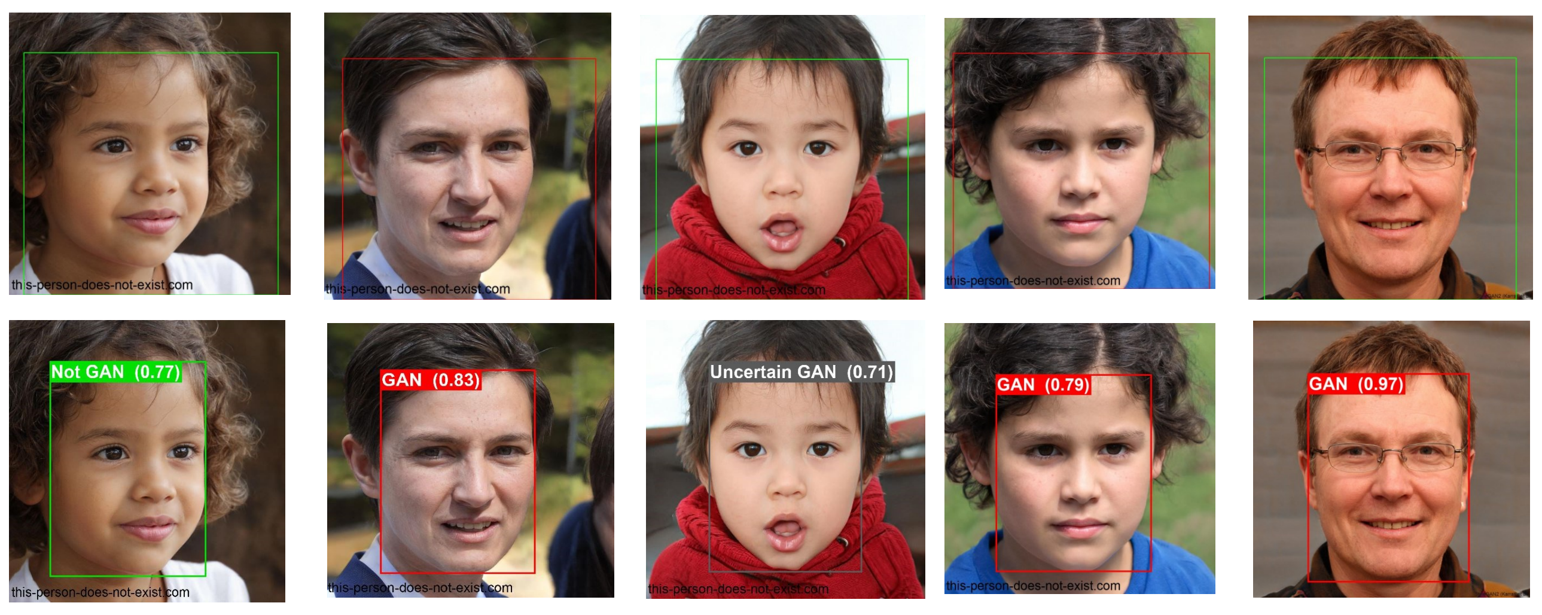
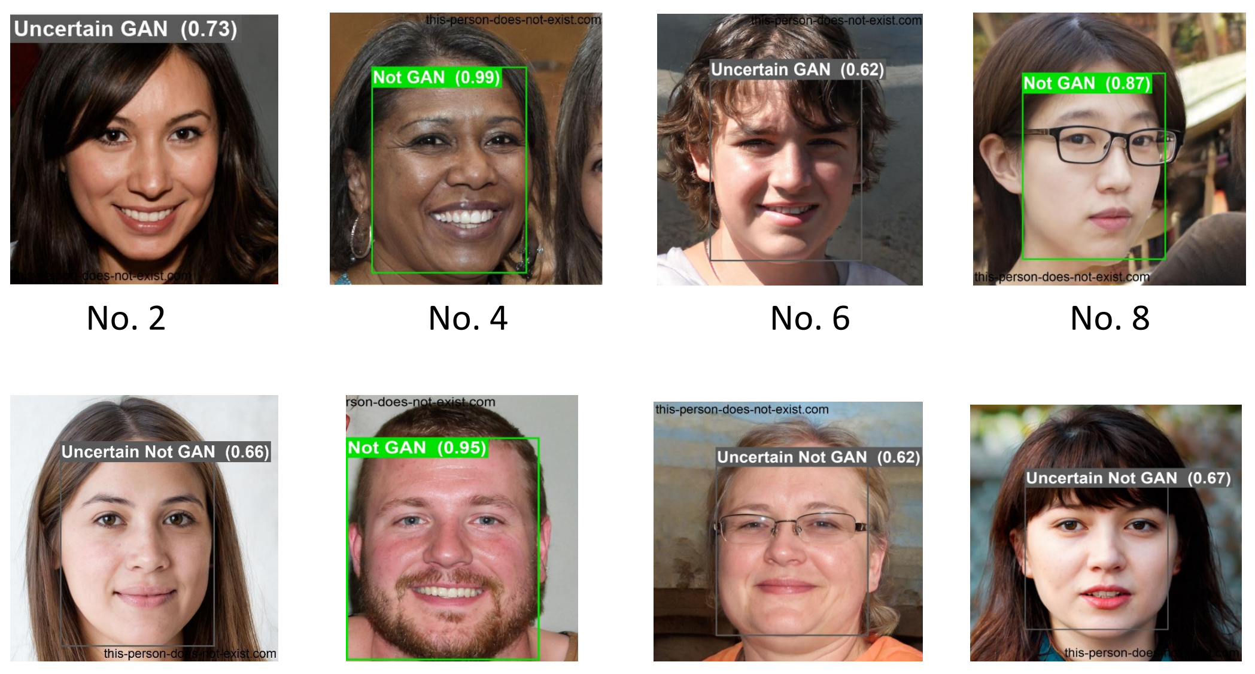


Figure 1: DFDetect vs Amped results - image numbers (left to right) 1, 3, 5, 7, 20.

For DFDetect, accuracy on identifying deepfakes was low but high for real images.

Table 2: Amped Authenticate GAN Deepfake Detection Results

No.	Amped Authenticate (Deepfake)		
1	77% Not GAN	11	92% Not GAN
2	**73% Uncertain GAN**	12	79% GAN
3	83% GAN	13	100% GAN
4	**99% Not GAN**	14	**95% Not GAN**
5	71% Uncertain GAN	15	99% GAN
6	**62% Uncertain GAN**	16	**62% Uncertain GAN**
7	79% GAN	17	81% GAN
8	**87% Not GAN**	18	**67% Uncertain GAN**
9	**66% Uncertain GAN**	19	96% GAN
10	76% GAN	20	97% GAN



No. 2 No. 4 No. 6 No. 8

No. 9 No. 14 No. 16 No. 18

Figure 2: Uncertain GAN Amped results

For Amped Authenticate, there was a good improvement on accuracy but many deepfake samples scored 50-60% or were deemed 'uncertain GAN'. We tested on a range of ages, genders, and angles and are expanding to include more complex photos.

We also explored Face Swap techniques and Amped Authenticate was unable to identify many, such as Figure 3. Face Swap techniques are utilizing GAN models to ensure targets faces are swapped onto bodies more realistically, as GAN models can blend and warp areas which may not be visible to the human eye and therefore should be picked up by a GAN Detector, however, each face is given a certainty result of 100%.

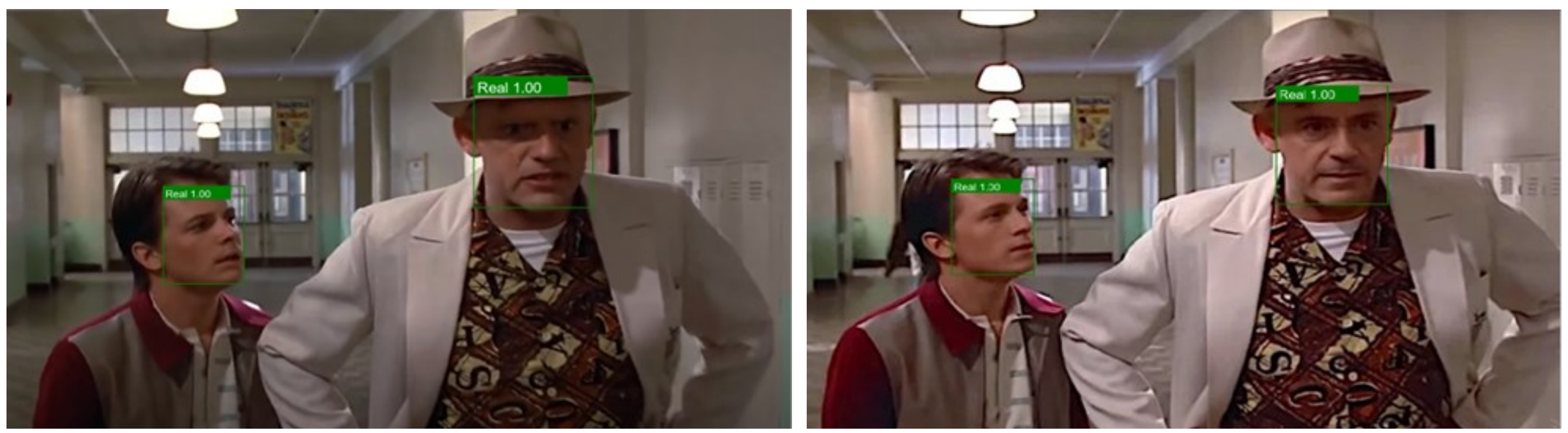


Figure 3 - Back to the Future FaceSwap