LIVERPOOL JOHN MOORES UNIVERSITY

How robust is the United Kingdom justice system against the advance of deepfake audio and video?

Dr Karl O. Jones
Bethan S. Jones

36th International Conference on Information Technologies (InfoTech-2022)

Introduction

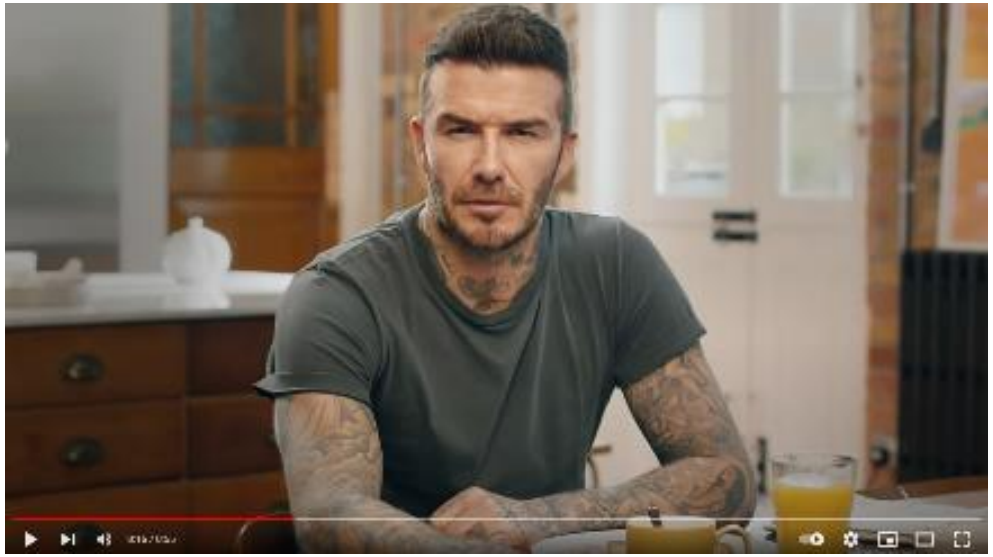
- Deepfakes are defined as ‘artificial intelligence or machine-learning applications that merge, combine, replace and superimpose images and video clips onto a video, creating a fake video that appears authentic’
- Arguably, knowledge of deepfakes is incredibly uncommon in the justice system
- Many individuals, government agencies and policy makers misunderstand their importance and possible impact, especially the risks they pose to legal systems
- Remarkably, a lack of findings in legislation, establishes that legal professionals are unable to protect the public from deepfake technology.
- Further, it is worthy to note there is a lack of standards and processes governing deepfakes and their presence within the justice system

Audio and Video Deepfakes

- There are numerous examples of deepfakes, many encompassing superimposed images and videos of celebrities
- The history of deepfakes can be split into two categories; fakes and deepfakes
  - fakes are created by humans undertaking the work themselves
  - deepfakes require deep learning systems to create something that has never been real
- Creating a deepfake video requires the developer to train a neural network with many hours of video footage of the person being 'faked' so that an understanding of what they look like and how they move is gained. Following this, the trained neural network works with computer generated graphics to superimpose the 'faked' person onto a different actor
- Audio deepfakes are created in a similar manner but then use text-to-speech software to make the deepfake "say" the chosen words.
- Deepfakes are now becoming "commercial", such as Alexa
  - https://www.youtube.com/watch?v=22cb24-sGhg&t=3753s

Audio and Video Deepfakes

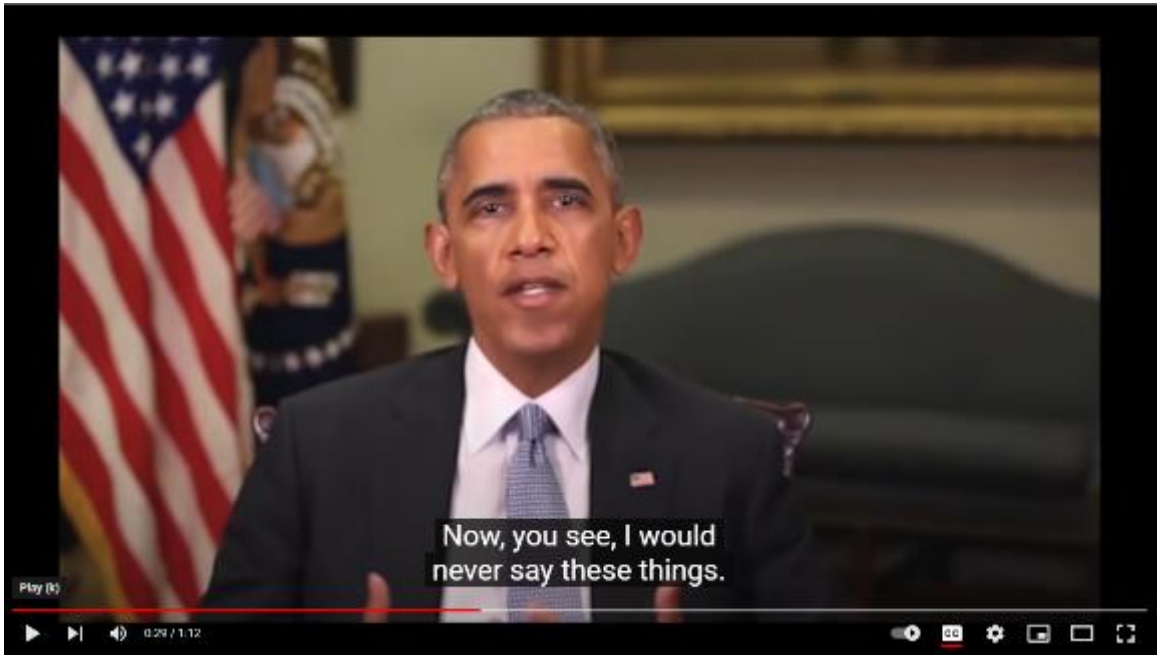
- One particularly ‘good’ audio deepfake is David Beckham apparently speaking 7 languages in a campaign for Malaria



www.youtube.com/watch?v=QiiSAvKJIHo

Audio and Video Deepfakes

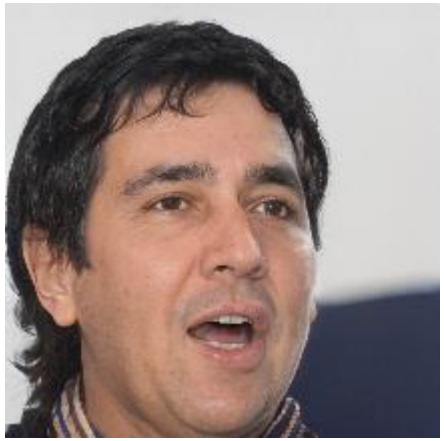
- University of Washington researchers created a realistic version of President Barack Obama, including a precise model of how his mouth moves allowing them to make their deepfake Obama 'say' anything they wished



www.youtube.com/watch?v=cQ54GDm1eL0

Audio and Video Deepfakes

- The complexity of deepfake technology allows it to create faces of people that do not exist
- Without suitable procedures in place only obvious flaws in facial generation might be noticed giving a hint at a deepfake image



Audio Deepfakes – in 'Court'

- In 2020, a Hong Kong bank manager received a telephone call from a man whose voice he recognized as company a director with whom he had spoken before.
- This company was about to make an acquisition, and hence required the bank to authorise a transfer of $35 million.
- Later, it was discovered that the bank manager had been deceived - fraudsters had utilised deepfake technology to clone the director's voice.

- Deepfake audio was presented, by a mother, to the court in a UK child custody dispute in an attempt to portray the father as threatening.
- The deepfake audio was created using freely available systems on the internet to create an excellent copy of the father's voice.

Knowledge of UK Justice System Professionals

- Video and audio recordings are now an inherent part of everyday life and are key technologies for the police service. Evidence showing police awareness of deepfakes is sparse, arguably suggesting they are still unaware of deepfakes and associated malicious capabilities.
- Forensic technicians are also generally unaware of deepfake technology and its far-reaching capabilities: they need to be fully aware of deepfake capabilities since audio/video evidence might have been manipulated or faked.
- Nooner was convicted in the 1990s of murder from enhanced footage from a surveillance tape. The case shows how barristers are totally unaware of potential doctored evidence. It was stated that '*relevant computer-enhanced still prints made from videotape recordings are admissible in evidence when they are verified as reliable representations of images recorded on master videotapes*'. No attempt was made to verify the reliability of the evidence. Similar difficulties are still present, namely that technology is speeding ahead of the justice system, especially in relation to the knowledge of those employed.
- The recent Kyle Rittenhouse Trial in the USA exemplifies the lack of knowledge Judges have surrounding video technology. During the trial, a video was zoomed into. Rittenhouse's lawyer argued that '*using an iPad to zoom in on a video should not be allowed because Apple's AI creates "what it thinks is there, not what necessarily is there"*'. While not about deepfake technology, arguably, most people are familiar with zooming on photographs taken on their mobile phones, thus having a judge not fully aware of what happens when a zoom is used is of some concern.

Evidence in the UK Legal System

- '*Evidence is the information with which the matters requiring proof in a trial are proved*'. Munday states, '*the evidence of a fact is that which tends to prove it… something that may satisfy an inquirer of the fact's existence*'.
- In a court of law the principle of evidence is used to determine a belief in something, whether it be through physical or verbal evidence, such as blood evidence or witness testimony. Thus, the notion of evidence is an extremely important factor when discussing deepfakes, since it establishes the court's ability to not only detect but handle the possibility of both perverting the course of justice, and miscarriages of justice through doctored evidence.
- Video evidence within the justice system is an ever-growing phenomenon, encapsulating different types of recordings such as from, CCTV, police body cameras, mobile phones, dash cameras and Ring™ doorbells, which might include audio.

Evidence in the UK Legal System

- All evidence, whether it be audio, video, blood, fingerprints etc must be handled correctly to avoid corruption, as Horsman and Sunde state '*evidence must be reliable if it is to be used as part of any legal decision making*'. Camacho *et al.* state '*an audio recording can be used as evidence in a legal process only if the integrity of the recording is demonstrated… the file has not been manipulated either by the victim, the suspect or by a third part*'.
- This demonstrates the lack of knowledge and understanding within the justice system relating to possible deepfake evidence since currently, there are no identifiable practices defined either by custom or statute to handle this type of evidence. For example, if the evidence introduced into court was already manipulated prior to seizure by the police, this creates serious concerns regarding the fairness of the law.
- The British Standards Institute assert '*an organisation should adopt policies and plans to assure the preservation of digital evidence and… the organisation should maintain processes that assure the integrity of investigations, the independence of experts, and the evidential value of binary information*'. It is quite worrying to note that police forces have no processes or procedures in place to establish, maintain or preserve the integrity of digital evidence.
- The case of Victoria Breeden demonstrates how law authorities are blind to the ever-growing phenomenon of potential deepfake digital evidence. This case involved a recording of Breeden stating '*how easy would it be to make someone disappear*', regarding hiring a hitman to kill her ex-husband. The police took the recording at face value, carrying out no work to determine the authenticity of the recording since it was made by a third party.

UK Legislation

- Although legal professionals are aware of fabricated evidence, such as creating fake wills for financial gain, the same individuals have little knowledge of the endless possibilities of deepfake evidence and their impact.
- Pavis states, '*the UK is a jurisdiction ripe for reform on the issue of deepfakes as the government is undertaking a series of reviews in connected areas of law*'. Furthermore, '*surprisingly little has been written on deepfakes in relation to UK law*'. Pavis continues, arguing, '*there are significant differences in the legal provisions applicable to Deepfakes between national laws*', indicating that the UK justice system is ill-equipped to handle the advance of deepfake technology.
- While there is no current legislation governing deepfakes, existing laws should be kept up to date and fit-for-purpose. The Protection of Children Act 1978, is '*an Act to prevent the exploitation of children by making indecent photographs of them; and to penalise the distribution, showing and advertisement of such indecent photographs*'. Section 7, sub-section 7 states that a '*pseudo-photograph means an image, whether made by computer-graphics or otherwise howsoever, which appears to be a photograph*'. Furthermore, sub-section 6 identifies that a "child" is '*a person under the age of 18*'. While the term 'pseudo-photograph' accepts an image can be computer generated, debatably, the definition of a 'child' under the Act is a largely contentious issue. If an image has been created through deepfake technology, the individual in the photograph, arguably, does not exist. It is therefore necessary to put forward an argument of whether the image truly depicts a real 'person'.

Improvement Recommendations

Lack of Professional Knowledge

- It is imperative that legal professionals become educated about the ever-growing presence deepfakes in the courtroom - provided by specialists in audio/video technology, and by specialists in artificial intelligence.
- Forensic technicians must also be trained in correct processing of audio/video evidence in general, as well as in methods for attempting to identify deepfake material.
- UK police require training in their approach to seizing audio/video material for evidential purposes, for example currently the technical specifications of video cameras or audio recording devices are not required to be documented, thus making appropriate forensic processing of the material problematic.

Standards, Processes and Procedures

- Although the UK justice system has standards and processes regarding the reliability of evidence, debatably there is an apparent absence of standards and processes addressing deepfake technology. In contrast, standards and processes surrounding deepfakes exist within different legal systems.

Legislation

- If reforms are not taken seriously by legal professionals and policy makers, then there will be severe ramifications from the existence of evidence created/modified through deepfake.
- The challenge of tackling the reliability of digital evidence within the courtroom is an epidemic the UK justice system is ill-equipped to handle, something that will only get worse if reforms are not made promptly throughout the judicial system.

Concluding Comments

- The UK justice system is wholly unaware and oblivious to the ever-growing presence of audio/video deepfake technology
- We have identified that there is a significant absence of legal professional knowledge relating to deepfake technology and its capabilities, creating concern regarding the operational procedures of the courtroom
- The deficiency of law around deepfakes attests to the argument that the UK justice system is ill-equipped and unable to cope
- Logically then, the justice system cannot be shown to be robust against the advance of deepfake technology….. Perhaps deepfakes are such an exclusive and unknown marvel that the law will never be able to catch up
- It is reasonable to suggest that the UK justice system could be identified as robust against deepfake audio and video technology if professional knowledge is improved, new law is brought into force and evidential processes, standards and procedures are developed
- There is an urgent need for the UK Ministry of Justice, as the lead organisation within the justice system, to begin a process of informing people across the justice system about the existence of deepfakes
- If the above is true of the UK, then one can only assume that other countries will be affected by the same problem