



Deepfake Technology Policy (Surrey and Sussex) (1241/2024)

Abstract

This policy outlines Surrey Police and Sussex Police's approach to verifying the authenticity in relation to AI generation or manipulation of crime-related media, including close circuit television footage, video recordings, audio files, and still images.

Policy

1. Introduction

1.1 The policy aims to provide a consistent and thorough procedure to determine whether such media is genuine or has been manipulated using deepfake technology or generative artificial intelligence (AI).

While media may never be definitively proven to be a deepfake, particularly as technological advancements continue to enhance realism, this policy and procedure provide the best opportunity to identify potential manipulation and maintain the integrity of our investigations.

1.2 This policy has been developed to address the anticipated increase in the likelihood of deepfakes being used for criminal activities. Such emerging threat areas includes, but are not limited to:-

- Child sexual abuse material (including Pseudonymised imagery) – already manifesting at scale.
- Fraud – already manifesting but not yet at scale.
- Non-consensual intimate image abuse, and related violence against females – offences such as coercive control, harassment, stalking and extortion – already manifesting, likely at scale.
- Defending our democracy, including from the threats of foreign interference – manifested already in attempts to utilise Artificial Intelligence to influence democratic debate and electoral outcomes.
- Terrorist content online – already manifesting but not yet at scale.
- Perverting the course of justice and other offences against the administration of justice (police evidential / Criminal Justice System (CJS) risk) – isolated incidents, not yet manifesting significantly but not being monitored.
- Overwhelm or diversion of crime reporting systems – isolated incidents, not yet manifesting significantly but it is not being monitored.

- Grooming for online child sexual abuse, county lines, modern slavery, money laundering and other crimes where victims are recruited for exploitation online – isolated incidents, not yet at scale.
- Harassment, blackmail, extortion – already manifesting in relation to sextortion and arguably happening in relation to harassment of public figures too. Scale not established.
- Public order, inciting violence / hate, reducing public confidence in the police and other institutions – already manifesting in some aspects.
- Digital identity and misuse of Government systems, including immigration, passport and visa offences – not currently manifesting but a future concern.

(above data taken from a letter to all Chief Constables from the Home Office, Minister of State for Crime and the Security Minister dated 17th May 2024 and headed 'Current and emerging deepfakes risks')

2. Scope

2.1 This policy applies to all Surrey Police and Sussex Police (hereafter referred to as the Forces) personnel involved in the collection, analysis, and investigation of crime-related media. It includes guidelines for initial assessment, advanced verification, and collaboration with external experts.

3. Policy Statement

3.1 This policy outlines the Force's approach to identifying evidence of AI manipulation in crime-related media, including close circuit television (CCTV) footage, video recordings, audio files, and still images. The policy aims to provide a consistent and thorough procedure to determine whether such media is genuine or has been manipulated using deepfake technology or generative artificial intelligence (AI) and the accompanying procedure provides information on how this may be achieved.

As the technological advancements improve the quality of deepfakes, the simple reality is that on occasion it will not be possible to definitively conclude whether material is genuine or fake. Following this procedure allows for the best opportunity for all parties involved to make a sound assessment and come to the best available conclusion based on the information presented.

Procedure

1. Crime-Related Media

1.1 Crime-Related Media is any form of media (CCTV footage, video recordings, audio files, or still images) collected or used in the course of a criminal investigation. This will include some pseudo-photographs / images.

1.2 A "pseudo-photograph" under UK law, particularly in the context of the Sexual Offences Act 2003, refers to an image that appears to be a photograph but has been created by other means, such as computer-generated imagery (CGI) or other forms of digital manipulation. It includes images that are not actual photographs but are made to

look like real photographs, often depicting illegal or explicit content involving minors.
<https://www.legislation.gov.uk/ukpga/2003/42/part/1/crossheading/indecent-photographs-of-children>

1.3 The relevant legal definition can be found in the Criminal Justice Act 1988 (as amended), particularly in Section 7(7), which defines a "pseudo-photograph" as an image, whether made by computer-graphics or otherwise howsoever, which appears to be a photograph.

<https://www.legislation.gov.uk/ukpga/1988/33/part/XI/crossheading/possession-of-indecent-photograph-of-child>.

1.4 Images that are realistic should be considered under Section 1 of the [Protection of Children Act 1978 \(legislation.gov.uk\)](#)

1.5 Images that are not realistic should be considered under section 62 of the [Coroners and Justice Act 2009 \(legislation.gov.uk\)](#)

1.6 In the context of the Sexual Offences Act 2003, pseudo-photographs are treated similarly to real photographs when it comes to illegal content, particularly indecent images of children, meaning they are subject to the same legal restrictions and penalties.

2. Deepfake Media

2.1 Deepfake media is media that has been digitally manipulated, often using AI, to replace or alter the appearance or sound of a person in a way that can deceive viewers or listeners.

Merriam-Webster defines deepfake to be 'an image or recording that has been convincingly altered and manipulated to misrepresent someone as doing or saying something that was not actually done or said'.

3. Generative Artificial Intelligence (AI)

3.1 AI technology that can create new content, including images, videos, and audio, which may appear to be real but is entirely or partially generated by AI.

4. Source Media

4.1 Source media is the 'base' media from which a deepfake has been generated. This term encompasses the original images, video, or audio recordings that are manipulated or altered using deepfake technology to create the final fabricated content.

5. Deepfake Technology

5.1 Examples of deepfakes (not definitive) are –

- 'Face Swap' – one face projected on to another.
- 'Lip-sync' – produced using AI and genuine footage to adjust the mouth to make it appear as something else is being said, usually accompanied by audio generated by AI or an impersonator.
- 'Puppet master' – the use of AI to animate the entire head and face.

5.2 A deepfake may consist solely of an AI generated audio track.

5.3 There are readily available apps that allow creation of all the above that can be made even more convincing, especially if 'tidied up' by the additional use of conventional video editing software.

6. Roles and Responsibilities

6.1 Investigating Officer (IO): Responsible for the initial assessment of crime-related media and determining whether further analysis is required.

Digital Forensics Team (DFT) / Forensic Digital Media Unit (FDMU): Responsible for coordinating advanced analysis of suspected manipulated media using specialised tools, techniques, and outsourcing expertise where necessary.

Digital Investigation Support Unit (DISU): Responsible for the capture of crime related media that has been displayed, hosted, or otherwise published online including within social media apps.

7. Current Legislation

7.1 Draft legislation proposed making a sexually explicit deepfake image to be a new offence (announced April 2024). Information can be found via this link: [Government cracks down on 'deepfakes' creation](#).

8. Initial Assessment

8.1 Collection of Media

Upon receiving crime-related media, the Investigating Officer (IO) must document the source, chain of custody, and any relevant details about the media. Such documentation will include recording allegations from members of the public about material being faked. This will include victims self-reporting that they have been made aware of material that purports to be them or a person known to them as part of a wider crime (such as blackmail or fraud) or as a substantive offence in itself (such as pseudo photographs). It is possible that suspects will also be the first to report deepfakes where they are effectively raising this as a line of defence when confronted with visual or audio material such as CCTV. Where the reporting party is able to provide source media this should be preserved by the IO as it may be used for later comparison with the crime related media to show how the deepfake was likely created.

8.2 Preliminary Examination

The IO should conduct a basic visual and auditory inspection to identify any obvious signs of manipulation, such as unnatural movements, inconsistencies in lighting, or distorted audio.

8.3 Metadata Analysis

Examine the metadata associated with the media to check for inconsistencies, such as anomalies in time stamps or editing history. Note metadata may not be immediately available to the IO and may require forensic services. The IO should in all instances ask for the most original version of the source file.

8.4 Decision to Escalate

If the media is deemed suspicious or if the IO cannot confidently verify its authenticity, it should be escalated to the FDMU for further analysis.

9. IO Awareness

9.1 A 'Deepfakes' interactive learning module is currently available from the College of Policing. This gives frontline officers and staff the skills and knowledge to detect and investigate deepfakes, explaining how to deal with deepfakes as part of an investigation and when to seek specialist advice.

10. Preparedness and Capability

10.1 Currently there is no capacity or expert capability within the Forces to definitively analyse media suspected to be a deepfake or of being generated by Artificial Intelligence. However, within Forensic Investigations an outsourcing partner has been identified who have experience in establishing, and reporting upon, the provenance of both still and moving imagery, demonstrating where there has been clear editing or media manipulation.

11. Submission of media

11.1 When the preliminary examination or disclosure is suggestive that deepfake material is present then the IO should follow the steps detailed in the sections 11.2 to 11.4 below.

All media must be recovered in compliance with the relevant section/s of the Forensic Science Regulators Code of Practice (see Document References).

11.2 Portable Storage

If the suspected deepfake media is on a portable storage device such as DVD, USB drive, portable hard drive, SD card etc. the submitting IO should contact the FDMU. The submitting officer must ensure that the media file is, as far as practicable, an exact digital copy of the original along with any associated meta and Exif data present. It should be noted that Video footage of the media filmed on a mobile device; Mobile Data Terminals (MDT) or Body Worn Video (BWV) will not be accepted for analysis.

Following a discussion with the FDMU, if it is agreed that the material can be submitted for outsourcing, the Officer will be provided with appropriate contact details for the contracted outsourcing company and should liaise with the company directly.

11.3 Digital Device

If a digital device is suspected of containing a deepfake or evidence relating to the creation, publication, and dissemination of a deepfake, then the device will need to be submitted to the Digital Forensics Team (DFT). The standard DFT submission process will need to be followed, including the submission of the appropriate Digital Processing Notice.

11.4 Online (including social media)

If the suspected deepfake media is on-line and being reported as a criminal offence, existing processes, and protocols to capture on-line information and to evidence and attribute account holders will be followed. Typically, DISU III (Internet Intelligence and

Investigation) practitioners would be utilised for capturing on-line information where the reporting party has been unable to supply this information themselves. For more information refer to Internet Intelligence and Investigation Policy (Surrey and Sussex) (1155).

12. Document References

12.1

College of Policing / College Learn - Deepfake module.

Forensic Science Activities: Statutory Code of Practice.

National Police Chiefs' Council (NPCC) Framework for Video Based Evidence.

Defence Science and Technology Laboratory (DSTL) Digital Imaging and Multimedia Procedure.

DSTL Recovery and Acquisition of Video Evidence.

Home Office (Minister of State for Crime and the security minister) letter to all Chief Constables dated 17th May 2024 – 'Current and emerging deepfakes risks' Deepfakes Chief Constables Letter.

Snapshot Paper - Deepfakes and Audiovisual Disinformation.

Team: Sponsor: Specialist Crime Command. Forensic Investigations