AI and Deep Fakes – Safeguarding July

This PDF Poster is now on the Academy website under the safeguarding section with past newsletters.

What are Deepfakes?

A deepfake is an image, video, sound, voice, or GIF which has been manipulated by a computer to superimpose someone's face, body or voice onto something else. This could be done with or without the subject's consent.

Deepfakes can be produced using computer and mobile apps to upload images from a user's camera roll. The quality of deepfakes can vary depending on the sophistication of the tech used and the skills of the creator. However, the standard of the deepfake produced will rarely matter compared to the potential harm it may cause when used to harass, bully, or abuse a victim.

The Harmful Side of Deepfakes

There are already several ways deepfakes are being misused to cause distress and harm. As this technology evolves and becomes more accessible to users, it is important to include in this guide to deepfakes some of the ways it can be misused.

Bullying
Deepfakes have been used in cases of cyberbullying to deliberately mock, taunt, or inflict public embarrassment on victims. The novel appearance of these images may distract from the real issue that they can be used to bully or harass children and young people.

Extortion and Exploitation
Deepfakes can be used to create incriminating, embarrassing, or suggestive material. Some deepfakes are so good that it becomes difficult to distinguish between them and the real thing. Convincing other people that an embarrassing or abusive image is fake

can create additional layers of vulnerability and distress. These images can then be used to extort money or additional 'real' images.

Additionally, deepfakes can be used to create so-called 'revenge porn'. This is a form of image-based sexual abuse as retaliation or vengeance typically associated with the end of a relationship or not agreeing to a sexual relationship with the perpetrator.

Deepfakes can also be used as a form of homophobic abuse, in which a person is depicted in gay pornography. This could then be used to 'out' the person or as an attempt to 'destroy their reputation.' For young people struggling with their sexual orientation, being depicted in any sexualised deepfakes may be particularly distressing.

Image-Based Sexual Abuse

There have been cases where images of children have been harvested and used to generate sexualised deepfakes. The realistic depiction of a victim engaging in a sex act can damage a child's wellbeing and mental health. We know that deepfake software can be used to remove clothing from victims digitally. In some cases, there are commercial services where users can pay to have images professionally manipulated. It is important that parents, carers, and safeguarding professionals are aware of the risks of this form of (non-contact) sexual abuse. In some cases, victims themselves may be unaware that their images have been harvested and misused to create deepfakes. While many children and young people may be aware of and understand how images can be manipulated in this way, others may not. It is important to speak to them about the issue of deepfakes and how they can be misused.

Resources and useful information for parents/carers

[Fake news, hoaxes and misinformation | NSPCC](#)

[Social Media: how to use it safely - NCSC.GOV.UK](#)

Awareness over the Summer holidays:

It has been brought to our attention by Intelligence shared with us by the police and the local authority that criminal gangs are alert to the increased presence of young people in the city centre of Leeds.

This includes the main bus station complex, the area to the rear of the Corn Exchange and the lower part of Briggate near McDonalds. If you son or daughter is meeting with friends in the city centre, we would recommend that you talk to them about those potential dangers.

If you have any concerns around safeguarding,

then please contact:

Mr J Grayston
(Assistant Principal and Designated Safeguard Lead)
jgrayston@rodillianacademy.co.uk