Do Now

# Lesson 4: : malicious use, the law and staying safe

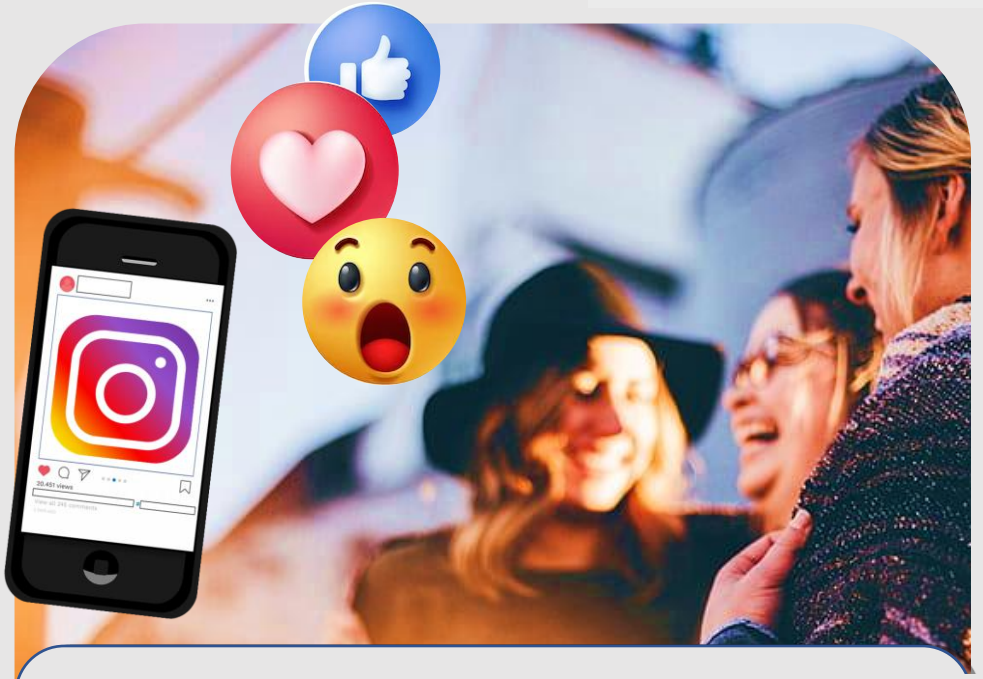
HEALTH AND WELLBEING

BELONG CARE ASPIRE SUCCEED

Three girls from Amy's class were in the school corridor, laughing at a video which was sent to them in a group message. When Amy walked over, the three looked terrified and quickly put their phones away.

Amy's heart sank. She heard a bit of the video, and it sounded like her voice. But she didn't remember saying those words.

**Discuss: What might the three girls have been watching? What should Amy do about this?**



Define the term AI and what we might mean when we talk about 'deepfakes.'

***AI is…..*** ***When we talk about deepfakes we mean…***

# Deepfakes: malicious use, the law and staying safe

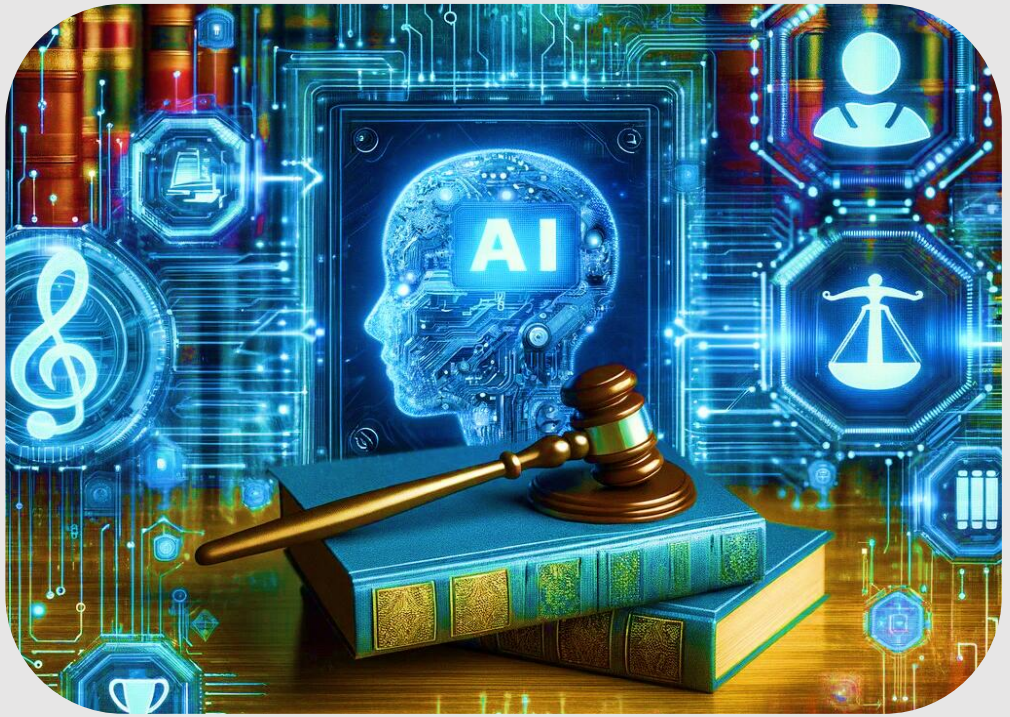
## Learning outcomes:

Explain how new technology produced by artificial intelligence, such as deepfakes, can be used to malicious ends, as well as for entertainment purposes.

Describe the legal penalties for malicious deepfakes and where to go for help and support.

**AI:** Artificial intelligence is the science of making machines that can think like humans. This could be used to create images, text, video footage or music.

**Deepfakes :** 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.




HEALTH AND WELLBEING

BELONG CARE ASPIRE SUCCEED

## *Did you know?*

It was only April 2024 that the UK Government outlawed the creation and sharing of explicit deepfakes. **Amy's predicament could have been much worse…**

https://www.youtube.com/watch?v=eZon6XQoYv8

Amy was right to be worried. She reported what she heard to her Head of Year, and the girls were called into a meeting.

All were unaware of the new UK law against the creation of malicious deepfakes, which are often **'immoral, often misogynistic, and a crime.'** (UK Department of Justice).

Fortunately, after seeing the video contents, the HoY found it was not an explicit video, but it was poking fun at Amy and using her voice and image without consent.

The girls were punished through the school's anti-bullying policy.

# Deepfakes: malicious use, the law and staying safe

HEALTH AND WELLBEING

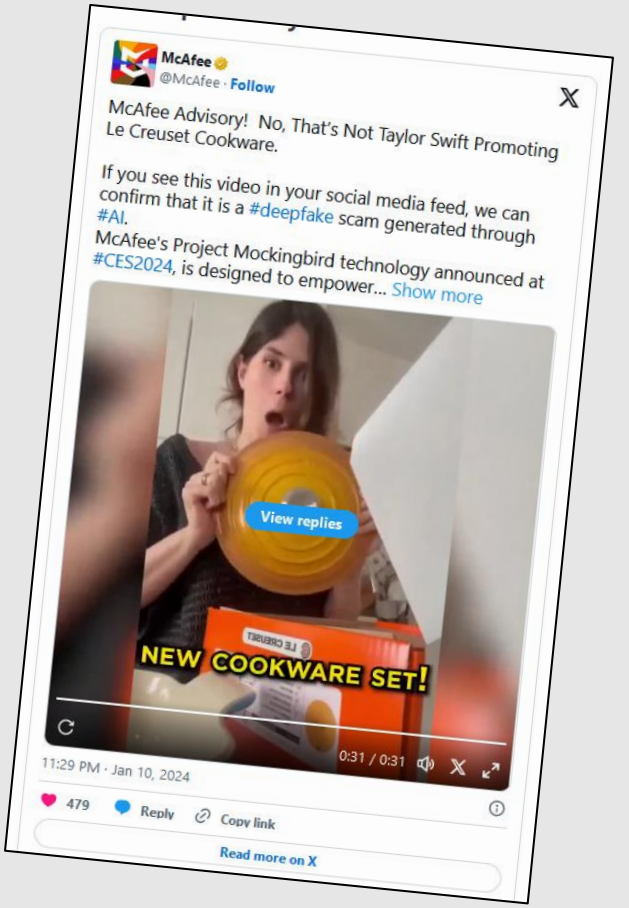
BELONG CARE ASPIRE SUCCEED

## How convincing are deep fakes?

**Watch the second video on this page, of Taylor Swift supposedly advertising cookware here:**

**https://www.creativebloq.com/features/deepfake-examples**

**DISCUSS: As you watch, be ready to feedback:**
**How might this have misled people?**
**Could you tell it was fake? How?**



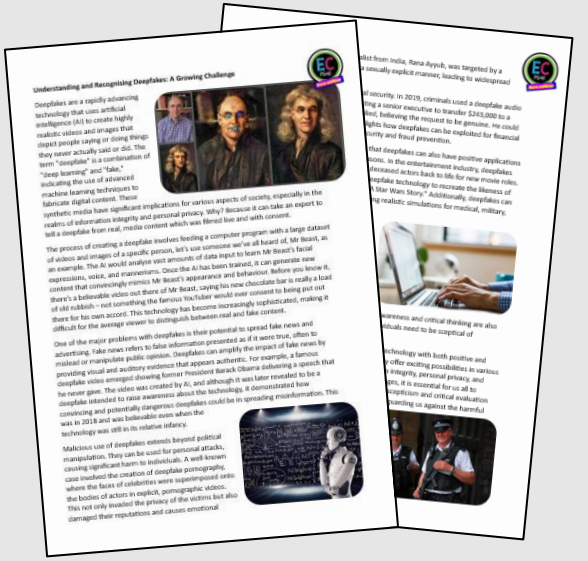
# Deepfakes: malicious use, the law and staying safe

THE THORNDEN THREE

BELONG CARE ASPIRE SUCCEED

We will now find out more about the significant issues and legal consequences surrounding the creation of deepfakes, as well as some examples of when they have been used maliciously.

**THORNDEN THREE GUIDED READING**
**Follow the text in your booklet. This is an important, new and developing topic.**



We will now find out more about the significant issues and legal consequences surrounding the creation of deepfakes, as well as some examples of when they have been used maliciously.

**THORNDEN THREE GUIDED READING**
**Follow the text in your booklet. This is an important, new and developing topic.**

THE THORNDEN THREE
BELONG CARE ASPIRE SUCCEED

**Understanding and Recognising Deepfakes: A Growing Challenge**

Deepfakes are a rapidly advancing technology that uses artificial intelligence (AI) to create highly realistic videos and images that depict people saying or doing things they never actually said or did. The term "deepfake" is a combination of "deep learning" and "fake," indicating the use of advanced machine learning techniques to fabricate digital content. These synthetic media have significant implications for various aspects of society, especially in the realms of information integrity and personal privacy. Why? Because it can take an expert to tell a deepfake from real, media content which was filmed live and with consent.

The process of creating a deepfake involves feeding a computer program with a large dataset of videos and images of a specific person, let's use someone we've all heard of, Mr Beast, as an example. The AI would analyse vast amounts of data input to learn Mr Beast's facial expressions, voice, and mannerisms. Once the AI has been trained, it can generate new content that convincingly mimics Mr Beast's appearance and behaviour. Before you know it, there's a believable video out there of Mr Beast, saying his new chocolate bar is really a load of old rubbish – not something the famous YouTuber would ever consent to being put out there for his own accord. This technology has become increasingly sophisticated, making it difficult for the average viewer to distinguish between real and fake content.

One of the major problems with deepfakes is their potential to spread fake news and advertising. Fake news refers to false information presented as if it were true, often to mislead or manipulate public opinion.

We will now find out more about the significant issues and legal consequences surrounding the creation of deepfakes, as well as some examples of when they have been used maliciously.

**THORNDEN THREE GUIDED READING**
**Follow the text in your booklet. This is an important, new and developing topic.**

Deepfakes can amplify the impact of fake news by providing visual and auditory evidence that appears authentic. For example, a famous deepfake video emerged showing former President Barack Obama delivering a speech that he never gave. The video was created by AI, and although it was later revealed to be a deepfake intended to raise awareness about the technology, it demonstrated how convincing and potentially dangerous deepfakes could be in spreading misinformation. This was in 2018 and was believable even when the technology was still in its relative infancy.

Malicious use of deepfakes extends beyond political manipulation. They can be used for personal attacks, causing significant harm to individuals. A well-known case involved the creation of deepfake pornography, where the faces of celebrities were superimposed onto the bodies of actors in explicit, pornographic videos. This not only invaded the privacy of the victims but also damaged their reputations and causes emotional distress. In another instance, a journalist from India, Rana Ayyub, was targeted by a deepfake video that portrayed her in a sexually explicit manner, leading to widespread online harassment and threats.

Deepfakes also pose a threat to financial security. In 2019, criminals used a deepfake audio clip to mimic the voice of a CEO, instructing a senior executive to transfer $243,000 to a fraudulent account. The executive complied, believing the request to be genuine. He could not tell the difference. This incident highlights how deepfakes can be exploited for financial gain, creating new challenges for cybersecurity and fraud prevention.

We will now find out more about the significant issues and legal consequences surrounding the creation of deepfakes, as well as some examples of when they have been used maliciously.

**THORNDEN THREE GUIDED READING**
**Follow the text in your booklet. This is an important, new and developing topic.**

Despite these risks, it is important to note that deepfakes can also have positive applications if the technology is used for the correct reasons. In the entertainment industry, deepfakes are used to create special effects and bring deceased actors back to life for new movie roles. For instance, the Star Wars franchise used deepfake technology to recreate the likeness of the late actor Peter Cushing in "Rogue One: A Star Wars Story." Additionally, deepfakes can be employed in education and training, offering realistic simulations for medical, military, and other professional training scenarios.

To combat the malicious use of deepfakes, researchers and technologists are developing tools to detect them. These tools analyse videos for subtle inconsistencies, such as unnatural facial movements or lighting anomalies, that may indicate a deepfake. However, as deepfake technology continues to improve, so too must the methods for detecting them. Basically – if they can fool a human, they can fool a fellow machine too. Public awareness and critical thinking are also crucial in mitigating the impact of deepfakes, as individuals need to be sceptical of suspicious or sensationalist content.

Deepfakes represent a powerful and rapidly evolving technology with both positive and unfortunately, significant negative potential. While they offer exciting possibilities in various fields, their misuse poses significant risks to information integrity, personal privacy, and financial security. As society grapples with these challenges, it is essential for us all to develop effective detection tools and foster a culture of scepticism and critical evaluation towards all media. This attitude will go some way to safeguarding us against the harmful effects of deepfakes.

We will now find out more about the significant issues and legal consequences surrounding the creation of deepfakes, as well as some examples of when they have been used maliciously.

**THORNDEN THREE GUIDED READING**
**Follow the text in your booklet. This is an important, new and developing topic.**

If you or someone you know has been a victim of a deepfake, you can call the police as they must take this attach seriously. On 16th May 2024, the UK government revealed that creating deepfakes, especially for malicious purposes relating to targeting women in a sexually explicit sense, will be illegal. Under the new offence, those who create these horrific images without consent face a criminal record and an unlimited fine. If the image is then shared more widely offenders could be sent to jail for two years. The new law will mean that if someone creates a sexually explicit deepfake, even if they have no intent to share it but purely want to cause alarm, humiliation or distress to the victim, they will be committing a criminal offence.

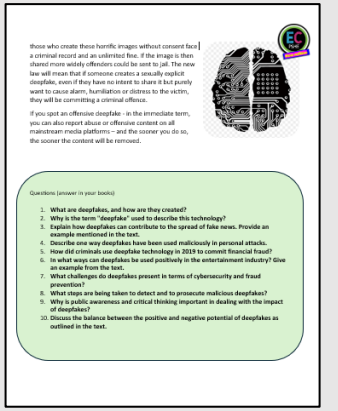
If you spot an offensive deepfake - in the immediate term, you can also report abuse or offensive content on all mainstream media platforms – and the sooner you do so, the sooner the content will be removed.

**Now answer the questions in your booklets.**

HEALTH AND WELLBEING

CARE ASPIRE SUCCEED

1. What are deepfakes, and how are they created?
2. Why is the term "deepfake" used to describe this technology?
3. Explain how deepfakes can contribute to the spread of fake news. Provide an example mentioned in the text.
4. Describe one way deepfakes have been used maliciously in personal attacks.
5. How did criminals use deepfake technology in 2019 to commit financial fraud?
6. In what ways can deepfakes be used positively in the entertainment industry? Give an example from the text.
7. What challenges do deepfakes present in terms of cybersecurity and fraud prevention?
8. What steps are being taken to detect and to prosecute malicious deepfakes?
9. Why is public awareness and critical thinking important in dealing with the impact of deepfakes?
10. Discuss the balance between the positive and negative potential of deepfakes as outlined in the text.


those who create these horrific images without consent face a criminal record and an unlimited fine. If the image is then shared more widely offenders could be sent to jail. The new law will mean that if someone creates a sexually explicit deepfake, even if they have no intent to share it but purely want to cause alarm, humiliation or distress to the victim, they will be committing a criminal offence.

If you spot an offensive deepfake - in the immediate term, you can also report abuse or offensive content on all mainstream media platforms – and the sooner you do so, the sooner the content will be removed.

Questions (answer in your books)

1. What are deepfakes, and how are they created?
2. Why is the term "deepfake" used to describe this technology?
3. Explain how deepfakes can contribute to the spread of fake news. Provide an example mentioned in the text.
4. Describe one way deepfakes have been used maliciously in personal attacks.
5. How did criminals use deepfake technology in 2019 to commit financial fraud?
6. In what ways can deepfakes be used positively in the entertainment industry? Give an example from the text.
7. What challenges do deepfakes present in terms of cybersecurity and fraud prevention?
8. What steps are being taken to detect and to prosecute malicious deepfakes?
9. Why is public awareness and critical thinking important in dealing with the impact of deepfakes?
10. Discuss the balance between the positive and negative potential of deepfakes as outlined in the text.

# Deepfakes: malicious use and staying safe

HEALTH AND WELLBEING

BELONG CARE ASPIRE SUCCEED

**Useful helplines and charities:**

Young Minds. Child and adolescent mental health charity for teens struggling with any subject.

Call: 0808 802 5544

Teen Line | Teens Support hotline - Connect, talk, get help! Teen Line's highly trained teen listeners provide support, resources and hope to any teen who is struggling.

SANE. National out-of hours mental health charity offering emotional support, guidance and information. Call: 0300 304 7000 (4.30pm to 10.30pm)

Mental Health Foundation. Information and support for anyone with

mental health problems or learning disabilities.

Rights online (coe.int) Your rights online as a young person using

social media sites

# Need Support? You're Not Alone

If anything in today's lesson has affected you, or you want to talk to someone, there is help available.

HEALTH AND WELLBEING

BELONG CARE ASPIRE SUCCEED

## 🧠 Mental Health & Low Mood

YoungMinds – https://www.youngminds.org.uk
Kooth – https://www.kooth.com
Mind – https://www.mind.org.uk

## 💬 Talk to Someone

Your Tutor or Head of Year – We're here to help.
Wellbeing Team and School Nurse
Report a Concern on Satchel
Safeguarding Team with the Purple lanyards

## 🍎 Healthy Lifestyle

NHS Every Mind Matters – https://www.nhs.uk/every-mind-matters
Change4Life – https://www.nhs.uk/change4life

## 📱 Apps That Can Help

Calm – For mindfulness and sleep.
Headspace – Meditation and stress relief.
Clear Fear – Manage anxiety (designed for young people).
MeeTwo – Anonymously talk to other teens, moderated by experts.

## 🍽️ Eating Concerns

Beat Eating Disorders – https://www.beateatingdisorders.org.uk
NHS Live Well – Eating Disorders – https://www.nhs.uk/mental-health/conditions/eating-disorders/


HEALTH AND WELLBEING
BELONG CARE ASPIRE SUCCEED