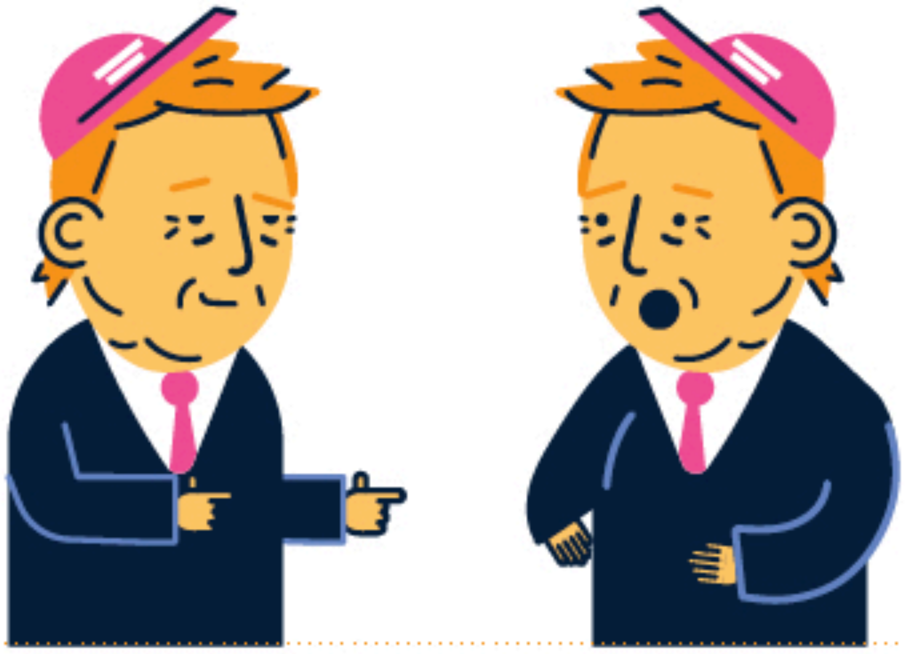
Worksheet 2.3:

Deepfakes Explained



What are Deepfakes?

Deepfakes are fake videos created using digital software, machine learning and face swapping. Deepfakes are computer-created artificial videos in which images are combined to create new footage that depicts events, statements or action that never actually happened. The results can be quite convincing. Deep fakes differ from other forms of false information by being very difficult to identify as false.

How does it work?

The basic concept behind the technology is facial recognition, users of Snapchat will be familiar with the face swap or filters functions which apply transformations or augment your facial features. Deep Fakes are similar but much more realistic.

Fake videos can be created using a machine learning technique called a "generative adversarial network" or GAN. For example a GAN can look at thousands of photos of Beyonce and produce a new image that approximates those photos without being an exact copy of any one of the photos. GAN can be used to generate new audio from existing audio, or new text from existing text – it is a multi-use technology. The technology used to create deepfakes is programmed to map faces according to "landmark" points. These are features like the corners of your eyes and mouth, your nostrils, and the contour of your jawline.

How to spot deepfakes

Like all types of information we encounter online the most important thing we can do when deciding if videos or images online are authentic and real is to be critical.

We need to use critical thinking and ask ourselves key questions such as:

- Who and why is someone sharing this video?
- Who or what is the original source?
- Is the person in the video saying something you'd never expect them to say?
- Does the video advance someone else's agenda? Who benefits from this video?

When seeing is no longer believing

While the technology used to create deep fakes is relatively new technology, it is advancing quickly and it is becoming more and more difficult to check if a video is real or not. Developments in these kinds of technologies have obvious social, moral and political implications. There are already issues around news sources and credibility of stories online, deep fakes have the potential to exacerbate the problem of false information online or disrupt and undermine the credibility of and trust in news, and information in general.

The real potential danger of false information and deepfake technology is creating mistrust or apathy in people about what we see or hear online. If everything could be fake does that mean that nothing is real anymore? For as long as we have had photographs and video and audio footage they have helped learn about our past, and shaped how we see and know things. Some people already question the facts around events that unquestionably happened, like the Holocaust, the moon landing and 9/11, despite video proof. If deepfakes make people believe they can't trust video, the problems of false information and conspiracy theories could get worse.