

Deepfakes

Deepfake *A video or sound recording that replaces someone's face or voice with that of someone else, in a way that appears real.*
Cambridge Dictionary

Deepfakes are on the rise and on the cusp of changing our relationship with recorded audio or video content forever.

In this lesson, students learn how deepfakes are made and what technological advances enabled their development. After looking at some deepfakes examples and their implications, students will discuss how societies should deal with deepfakes in the future.

LESSON GOALS

- Learning about deepfakes and how they are created

Theme **Technology**

Lesson duration
45 minutes

Student level
Introductory

Main method of instruction
Theory

Materials
Smartboard/projector, Pen and paper

Prep time for teacher

- Understanding the real-life implications of deepfakes and possible solutions

ACTIVITIES

THEORY (15 MINUTES) - TEACHER-CENTERED

Give the students the introduction to deepfakes and show the accompanying videos.

Aim: Students learn how deepfakes are made and used.

EXERCISE (30 MINUTES) - CLASS

Students debate the motion “deepfakes should be prohibited” and explore arguments in favor and against, implications of such legislation, and alternative solutions to the issues raised by deepfakes.

Aim: Students think critically about how to governments should cope with deepfakes.

DISCUSSION (OPTIONAL) - CLASS

Discuss some of the discussion questions with the students.

Aim: Reflect on the topic.

KEYWORDS

Fake news

Deepfakes

Technology

THEORY (15 MINUTES)

History

Image manipulation has been around for a while. Historically it has been used by dictators, but also by artists, to change people's perception of reality. From removing people who fell out of favor with Stalin, to adding a photo of a deceased family member to a family portrait. Scissors, glue, and patience were the main ingredients in "photoshopping" avant la lettre.

Of course, all of this changed when computers got involved and manipulating pictures digitally turned into a possibility. As tools progressed and became easier to use, the number of fake images—from blackmail material to memes—in

circulation grew. The internet played a big role in the process of their creation and distribution.

With fake news on the rise and Photoshop in the back of our minds, judging the content you see online can be difficult. Video seemed to be the last medium that could sort of speak for itself. That is, before technology caught up and introduced the era of deepfakes.

A deepfake is a video or sound recording that replaces someone's face or voice with that of someone else, in a way that appears real.

(Cambridge Dictionary)

Deepfakes have been around since the 90s, but for most of their existence, they looked terrible. Computer scientists laid the groundwork for future development, but it would take advances in processing power, artificial intelligence, and other technologies for more convincing deepfakes to emerge.

While it is difficult to say if a perfect deepfake exists yet, more recent attempts have fooled plenty of people. Societies and governments struggle to come to terms with ethical, political, and journalistic implications of deepfake development.

How deepfakes work

At its core, deepfakes use similar technology as Snapchat or Instagram filters. Software scans a person's face, and figures out which part of the image represents which bodypart. In other words, the computer understands what a face looks like. Once the app has found your eyes and nose, for example, it can figure out when you blink and how your lips move when you speak.

Deepfakes use machine learning to interpret how an actor moves and behaves, and overlay it with the image of another person, like a celebrity. To do this, the machine learning algorithm must be trained with as much source video material as possible from the person that will be "overlaid" on the actor. If the actor blinks, the AI draws from the source material to make the deepfake image blink too.

- Watch a couple of minutes of a deepfake portraying Dutch prime minister Mark Rutte, apologizing for not showing leadership in the climate crisis (turn on English subtitles in the settings).

Beste Mark Rutte, zo klink je als je
#klimaatleiderschap toont

- Video of what would've happened if the Apollo 11 moon landing had failed with an interactive quiz after asking what parts were

real or fake.

<https://moondisaster.org/film>

Even if the person in the video looks convincing enough, getting their voice right is usually an issue. In the past, voice actors would do imitations of the person they wanted to portray. But now, AI is also being used to create deepfake voices. By scanning voice recordings, AI can be trained to use someone's voice to say new things that this person never said—much like using source video materials are used to make the image of someone do new things.

Jay-Z raps the "To Be, Or Not To Be" soliloquy from Hamlet (Speech Synthesis)

Some examples of vocal synthesis

As deepfake videos and deepfake voices are combined, it becomes increasingly difficult to distinguish between real and fake. As we saw earlier, deepfakes can be used to impersonate people in power, like prime minister Mark Rutte. The deepfake was made by a newspaper, stating that the video is a deepfake and how it was created. But imagine a deepfake of a political leader declaring a nuclear attack without a statement mentioning its lack of authenticity.

Ethical implications of deepfakes

Deepfakes have been used to blackmail people, commit fraud, and spread fake news. Also, the technology has been used to create pornographic material of celebrities, and so-called “revenge porn”.

In the country of Gabon, Central Africa, rumors about a video of the president being a deepfake led many people to believe the president had actually died. The president had had a stroke, which explained why he looked slightly “off” in the video that led to the deepfake rumors. The speculation, however, had real life implications, fueling a flame that eventually led to an attempted coup.

But the biggest risk of deepfakes perhaps not even the spread of fake news. Rather, if every video could be fake, it gives people who were caught on video doing something wrong a way out. This is called “plausible deniability”, which means the ability to deny responsibility for damnable actions. Deepfakes raise a philosophical question: how can we determine what is true when everything can be faked?

The first way to tackle adverse effects of this new technology is by spreading awareness and knowledge about it. When people know how

convincing deepfakes can be, they are less likely to take every online video at face value. Knowing about deepfakes is becoming a part of media literacy.

As deepfake technology develops, so does technology to detect whether a video is generated or not. Forever a cat and mouse game, creators of deepfakes are always the next ones to make a move while those trying to detect it try to keep up.

Finally, the blockchain might offer a solution to the issue of verifying video content. By automatically stamping video with a “watermark” that is registered in the blockchain while filming, “watermarks” can be used as evidence that a piece of video is real later. All people or media have to do to determine whether a video is fake is check if it has a watermark.

EXERCISE (30 MINUTES)

Students debate the motion “deepfakes should be prohibited” and explore arguments in favor and against, implications of such legislation, and

alternative solutions to the issues raised by deepfakes.

DISCUSSION (OPTIONAL)

- What are the ethical implications of deepfakes?
- Should creating deepfakes be prohibited?
- Who is responsible for regulating the spread of deepfakes on social media?
- Who should ensure that technology is developed in a way that minimizes harm?
- How should judicial systems deal with “plausible deniability” caused by deepfakes?

FURTHER READING

[Samsung's AI: Megapixel DeepFakes! 📷](#)

[Teaser -- Synthesizing Obama: Learning Lip Sync from Audio](#)

[This is not Morgan Freeman - A Deepfake Singularity](#)

SOURCES

[Dear Prime Minister Mark Rutte, this is what #climateleadership looks like](#)

[Deepfakes: Why you can't believe everything you see and hear | Joseph Palermo | TEDxToronto](#)

[Six U.S. Presidents Introduce the Vocal Synthesis YouTube Channel \(Speech Synthesis\)](#)

[In Event of Moon Disaster](#)

[How Photos Became a Weapon in Stalin's Great Purge](#)

[Deepfakes, Revenge Porn, And The Impact On Women](#)

[How misinformation helped spark an attempted coup in Gabon](#)

[The Blockchain Solution to Our Deepfake Problems](#)

[Back to the list](#)



Co-funded by the
Erasmus+ Programme
of the European Union



The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.