brAInfood of the
Knowledge Centre
Data & Society

WAT ARE DEEPFAKES ?

What is real and what is not? **Deepfakes** make the line between reality and fantasy thinner. But what exactly are deepfakes?

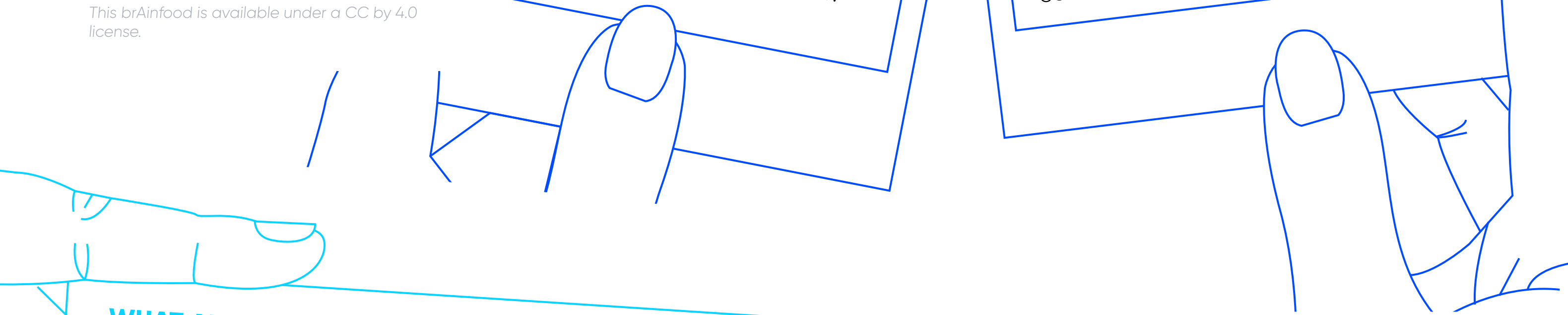
Read more about deepfakes in this brAInfood. We focus on how they **function**, show you some **examples** and point out the **risks** and **benefits**.

Knowledge Centre Data & Society (2021). What are deepfakes? brAinfood of the Knowledge Centre Data & Society. Brussels: Knowledge Centre Data & Society.

WAT ARE DEEPFAKES?

'Deepfakes' are **images, sounds and texts created** or **manipulated** by **artificial intelligence** software. To do this, it uses a set of existing images, audio or texts, from which the software teaches itself certain things (e.g. what a face looks like). With this information the software can, for example, create a new image of a face.

A **General adversarial network** (GAN) is a machine learning model in which **two neural networks compete** to make **more accurate predictions.** One network generates a synthetic example similar to the real data while the other tries to distinguish whether an example is real or synthetic. In this way, the two networks constantly improve, and the software creates new content that is almost indistinguishable from the original content.



WHAT ARE THE RISKS?

Deepfakes are highly realistic and hard to differentiate from real imagery. They distort reality and the truth.

This poses several risks when used for malicious reasons. For example, deepfakes could:

- create and amplify **fake news**;
- **distort** political **campaigns**;
- strengthen **extremism**;
- **falsify** evidence;
- increase general **distrust** in certain people or media.

We are used to take video and images at face value and it is easy to share them in large numbers on social media. This makes the false information in deepfakes able to **influence large groups of people**. In the wrong hands, the technology can have a serious impact on democracy, security and foreign affairs.

WHAT ARE THE OPPORTUNITIES?

Deepfakes and GANs can be of great value for the **entertainment, arts and culture, education and healthcare sectors** in many different ways: from imitating artists or performers who have passed away a long time ago to editing videos without the need for a new recording.

Samsung's AI lab in Moscow was able to transform Da Vinci's Mona Lisa into a video in which the woman moves her eyes, head and mouth. Deepfake technology can bring history to life, and can make education, art and culture more **accessible and interactive** to a wide audience.

With StyleGAN, a specific application of GAN, you can generate an image in a certain visual style that the user teaches to the technology. They are used in **fashion design, web art, animation, music**, etc.

In **healthcare**, deepfakes can reconstruct realistic data that helps researchers to develop **new treatments** without using real patient data. A well-known and promising experiment is the development of artificial MRI scans of the brain in order to train an algorithm that recognises brain tumours.

HOW TO RECOGNISE THEM?

Recognising deepfake videos is not always easy, especially when they are made with professional tools and budgets.

When watching them, you may feel that there is **something 'wrong'** with the video you are seeing. An algorithm has difficulty generating **body language** and **subtle human expressions** and placing them in the right context. Therefore, pay attention to unnatural facial expressions and eye movements (e.g. blinking the eyes too much or too little) and wrong perspectives (of a nose, f.e.) when watching a video.

There are also **(open source) deepfake detection tools** that use various techniques to unmask deepfake videos. For example, there is an algorithm that can analyse digital hashes and metadata of a file for authenticity and originality.

Such tools work properly, but due to advances in technology, making deepfakes and detecting them will remain a **cat-and-mouse game** for a long time.