Deepfake: An Assessment from the Perspective of Data Protection Rules

Deepfake: An Assessment from the Perspective of Data Protection Rules

Article by Baran Can Yıldırım and Celal Duruhan Aydınlı

Introduction

Fake content is nothing new and has been there in the digital realm particularly after the introduction of photo-editing, image creation and graphic design software such as Adobe Photoshop in 1990. The number of them, however, has been alarmingly increasing within the last months. These contents are now very hard to be identified as fake thanks to the "*Deepfake*". Deepfake has been so far used to create, among others, malicious hoaxes, fake news as well as adult videos[1].

While these contents may interfere with a wide range of laws including the ones related with fraud, crime, and copyright, we will throughout this article try to lay out and focus on its interference with the data protection rules. As such, we do not claim or aim to address every possible aspect regarding the individuals' protected rights.

We will first explain basically what Deepfake is and how it is created, and then discuss how its effects may be relevant with the data protection rules.

Technical Summary

The popular name Deepfake comes from "*deep learning*" and "*fake*". It creates problems when the "*fake*" part is being used for purposes that may violate, among others, data protection rules.

Deepfake is a product of *Deep Learning* (or *machine learning*), which is a subset of artificial intelligence ("**AI**") and based on a specific method called "*Generative Adversarial Networks*" ("**GANs**"). GANs are deep neural net architectures comprised of two nets, pitting one against the other. It can reportedly learn to mimic any distribution of data such as images, music or speech[2]. After the input is scanned, phonemes and visemes are isolated; alignment of the corresponding data begins to track, reconstruct and produce the rendered mix, eventuating as the new content. Therefore, GANs may learn to create fake worlds that are very similar to the real world.

As the visibility and thus popularity of *Deepfake* content grow, demand on the better codes that create and develop such contents have also been increasing. Open-source platforms such as Github or TensorFlow have been used by the developers with a view to increase the intelligence of the algorithm patterns and creating mobile applications. In this regard, the "quality" of the mobile applications creating Deepfake has alarmingly increased.

There are methods that can detect the fake content such as reverse-image search, magnification and Uncanny Valley Hypothesis. Although it is not the very purpose of this article, we will briefly explain these methods before moving on to the legal aspect of the Deepfake.

Reverse-Image Search Method[3]

The method relies on the logic of searching the used images or videos on the internet if the creator has supplied the content from the internet itself as in the *Gonzalez incident*[4]. However, this method is regarded as the most basic detection process and it is relatively easy to be crossed by experienced developers.

Magnification Method[5]

As even behavioral actions can be implemented in the fake content, it has been concluded that detecting the differences of biometric data of the *victim* (such as gait recognitions, gaze checks or heart rate monitoring) could be used to identify whether the content is real or fake.

Uncanny Valley Hypothesis[6]

Masahiro Mori's hypothesis is based on people's emotions about robots and automatons. According to the idea, human likeness of a robot affects people's feelings positively; but beyond a similarity limit, humans tend to develop a strange sense of revulsion or aversion. According to the Uncanny Valley Hypothesis, the incompleteness of the whole mimics and facial expressions might have the chance to develop awareness with a close study.

Could Deepfake Content Itself be regarded as Personal Data?

The fundamental principle of every data protection set of rules, as the name suggests, is to protect the personal data of the data subjects (i.e. the people). Article 4/1 of the General Data Protection Regulation ("**GDPR**") of the European Union, for instance, defines the personal data as follows:

"'*personal data' means* ***any information relating to an identified or identifiable natural person*** *('data subject')*

"

Therefore, in order to speak of a personal data, we first need (**i**) an information and (**ii**) an identifiable natural person, (**iii**) to which such information is related. Then the questions of how we should define the identifiable natural person arises. In this regard, the same article reads read as follows:

"*an identifiable natural person is one who can be identified, directly or indirectly,* ***in particular by reference to*** *an identifier such as a name, an identification number, location data, an online identifier or to* ***one or more factors specific to the physical, physiological, genetic,*** *mental, economic, cultural or social identity* ***of that natural person***"

In Deepfake videos, we generally see the face and/or body of a real person (**i**) talking with a voice that is very similar to their actual voice and (**ii**) acting and using mimics in a way that they usually do in real life. However, the face, body, voice, acts or mimics are actually not their own; they are created by, among others, combining through software a lot of their actual pictures, voices, videos, etc. The contents therefore in a sense are artificial animations or graphics created by the developer. Another type of Deepfake videos is the one when the created face is imposed on another person's actual body. In this case, the body is not created and belongs to someone real.

In the latter scenario, it should be clear that the body is the personal data of the relevant data subject. The former scenario, however, requires an interpretation of the abovementioned definition as to whether such created faces or bodies are the personal data of the *victim*.

Misuse of a created fake content representing someone who is not actually acting or speaking as in the edited video has the potential of affecting the individuals' prestige/reputability even if the content's fake nature can be distinguished from a reasonable person's eye. However, the data protection rules cannot protect the data subject unless the fake content is regarded as personal data. In this context, because of the fake nature of the content, consideration of data protection rules would be irrelevant because the content itself does not belong to a real individual. On the other hand, it might be argued that even the positive use of the person's name or mimics without the data subject's consent might constitute a personal data violation. In light of the above, if the fake content may be considered an information that is related to an identifiable natural person, then such fake content should be protected within the scope of the GDPR. Such protection grants the relevant natural person the right to, among others, request the erasure of the fake content and request compensation, if any damage is suffered as a result of an "*unlawful processing of personal data*" (the concept of which we cover below). It should be noted that the assessment above discusses whether the fake content may be regarded as a personal data (and the protection thereof). The assessment of whether the action of creating such content may constitute a violation of data protection rules will be discussed below.

Unlawful Processing of Personal Data

Regardless of whether the Deepfake content may be regarded as personal data, the concept of "*processing of personal data*" is of importance to evaluate the legality of *creating* the fake content within the scope of the data protection rules.

As defined in Article 4/2 of the GDPR,

*"'**processing**' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction".*

In light of the above, almost every action relating with personal data, including collecting and altering, which are of special importance for creating the Deepfake content, may be considered processing of personal data. As we explained above, the fake content is generally created by, among others, combining through software a lot of the victim's (data subject's) actual pictures, voices, videos, all of which are regarded as personal data. Therefore, there is no doubt that the personal data of the victim is processed at least while the fake content is being created. Then the question of whether such processing is lawful arises.

Processing of personal data is lawful so long as it is based on at least one of the conditions stated in the GDPR. The most common grounds for processing the personal data are (**i**) consent, which requires the explicit consent of the data subject in which they allow processing of their personal data for specific purpose(s) and (**ii**) legitimate interest of the data processor, which means the processing is required for the legitimate interest of the data controller. The full list of grounds, which are six in total, are listed under Article 6 of the GDPR, and may be found here (and their recitals may be found here).

We consider none of these conditions is likely to be met in creating the fake content, especially taking into account that the fake content is generally created to harm the victim.

Conclusion

Apart from the well-established general remedies of law, such as and depending on the jurisdiction (**i**) monetary compensations within the scope of tort law, (**ii**) criminal sanctions, (**iii**) requests as per the data protection rules (including request of erasure and compensation), (**iv**) blocking the access to the harmful or fake content, there is not a specific set of rules that addresses directly to products of Deepfake and the liability of the creator and/or publisher. That being said, potential remedies are being discussed in different fields such as defamation, privacy, copyright infringements, criminal acts, etc. For instance, "*Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019*"[7] was proposed by Representative Yvette Clarke (of NY) in the U.S. House of Representatives "*to combat the spread of disinformation through restrictions on deep-fake video alteration technology*", which aims to criminalize creating the Deepfake content.

As the *quality* of Deepfake grows, we are more likely to observe more infringements in various fields. Although certain authorities have started to take their steps to fight the potential negative effects with legal remedies, such action is very likely to take time to catch up with the issues at this point considering the speed of the disruptive technologies. In this regard, technological remedies are being used in the related field such as the authenticated alibi services or cybersecurity precautions.

Either way, it is certain that both law-makers and white hat developers have to act prompt and precise in order to prevent the Deepfake from affecting vital issues such as taking evidence in courts or even conflicts between sovereign states considering the negative potential effects of these contents.

Whilst the main object of this article is to raise awareness on an ongoing development of a fresh area of potential violations with a focus on data protection rules, it is without doubt that there are a lot to discuss and consider on many different fields that may be affected or related with Deepfake.

Footnotes

1 https://www.cnbc.com/2019/10/14/what-is-deepfake-and-how-it-might-be-dangerous.html

https://www.bbc.com/news/technology-49961089

https://www.technologyreview.com/f/614485/deepfake-porn-deeptrace-legislation-california-election- disinformation/

2 https://skymind.ai/wiki/generative-adversarial-network-gan

3 https://iapp.org/news/a/privacy-law-and-resolving-deepfakes-online/

4 https://www.telegraph.co.uk/news/2018/03/26/fake-images-parkland-shooting-survivor-emma-gonzalez-tearing/

5 https://insights.sei.cmu.edu/sei_blog/2017/08/real-time-extraction-of-biometric-data-from-video-1.html

6 https://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley

7 https://www.congress.gov/bill/116th-congress/house-bill/3230/text