Chapter 12
From Copy & Paste to Deep Fakes – Digital Collaging and Image Manipulation

Olivia Hägle

I. Introduction

Binary encoding, the basic principle of digital information processing, delivers access to a variety of new possibilities for the manipulation of visual objects. By dissolving images into their components, reducing them to their essential characteristics and recomposing visual objects, it is not only possible to manipulate existing visual objects. Rather, one may create entirely new visual objects in a deceptively realistic manner. In the context of these deceptions, not only do legal and technical questions arise but, increasingly, ethical ones too.

Of course, the fact that some people try to deceive other people by using deliberately manipulated words or images is not a new phenomenon. However, two remarkable changes can be observed in this context.

Especially in the field of image manipulation, there are new developments in the means used. Whereas for a long time persons manipulating images were constrained to their own manual, still widely primitive forms of manipulation or to extensive and expensive forms of manipulation executed by specialists, modern technologies now allow almost fully automated manipulations of still and moving images. Moreover, the results of these automated operations already exhibit a persuasiveness that can hardly be dispelled by the human eye.[1] Despite these technical advances, the claim of authenticity that images still enjoy goes far beyond what "simple" text-based disinformation can achieve.[2]

Apart from the methods used, the effects of such fakes are also new. Due to the increasing global interconnectivity in the information age, the consequences of such deceptions are both increasing in scale and impact. With a single manipulated item, it is possible to reach an almost infinite

1 Maras/Alexandrou (2019) 257 et seq.
2 Sherwin/Feigenson/Spiesel (2006) 241 et seq.

number of people without significant effort. This fact does not only intensify the consequences for those directly affected, but also paves the way for a variety of other consequences that additionally impact a wider public. One could observe the consequences of such information manipulation, for example, in the US election campaign in 2016. Today it is assumed to have been deliberately manipulated by a Russian campaign of disinformation.[3] And in times of the Covid-19 pandemic, one also observes the effects of such targeted disinformation campaigns.[4]

In the following, the path from other forms of pictorial manipulations to deep fakes will be outlined first, to later identify the particularities of deep fakes in the field of image manipulation (II.). Subsequently, an overview of currently possible application scenarios, the consequences for the affected individuals and the general public, as well as the rights potentially affected by deep fakes will be outlined (III. 1. and 2.). Finally, conceivable approaches to solving the problem of deep fakes will be described (III.3.).

II. From Copy & Paste to Deep Fakes – The Origins of Image Manipulation

Deep fakes are based on new technological developments and are still a relatively new phenomenon. For this reason, there are no rules that are specifically designed to solve the problems arising in this context, but similar phenomena, like other forms of image manipulation, and conflicts of interests are long known and have already been regulated and could therefore help to respond to this new phenomenon.

1. A brief history of image manipulation

a) Starting point: art forgery

Before the invention of photography and the resulting possibility of simple and numerous duplication of images, pictorial representations were

3 DiResta et al. (2019); Timberg/Romm (2018).

4 See for example the refurbishment of a German research network *CORRECTIV*, Coronavirus-Faktenchecks: Diese Behauptungen hat CORRECTIV geprüft, https://correctiv.org/alle-corona-faktenchecks.

basically unique.[5] Since a small number of highly appreciated pieces of art have always been met by a far greater number of wealthy people, the demand on the market could often only be satisfied by means of copies and counterfeits.[6] The origins of image manipulation lead therefore back to the beginnings of art forgery.

Painters have always copied the greatest artists to learn their style of painting. Such copies are not problematic as such. In addition, it is noteworthy that especially artists themselves often even appreciate good fakes.[7] Duplicates have only become problematic when they were used to deceive third parties. In antiquity, this fraudulent intent cannot generally be assumed, since it was common practice at that time for masters to maintain an entire atelier in which work was done in the style of the masters and from which all works were considered to be those of the masters, even if only a fraction of the manufacturing steps had been carried out by themselves.[8] In those days this practice was generally known and therefore not deceptive. In addition, people had a completely different conception of originality back then, but it has changed over the centuries. All too often, originality is now measured by the artist's own execution rather than the independent conception of a work.[9] Especially in the 20th century, the demand on the art market was so high that this century produced a large number of so-called master forgers.[10] Those forgers took advantage from the fact that the market craved for pictures from certain artists and if fakes of highly demanded pictures then appeared, they would subsequently be only superficially examined to see if they were real or faked.[11]

5 See Deussen (2007) 30. This development was reinforced even further by the transition from analogue to digital photography.
6 Ibid.
7 Hebborn (2011) 9. For different forms of copies from forgeries to artifacts created as homages, see Brinkmann (2020).
8 Partsch (2010) 30.
9 See Butin (2020) 38.
10 See Partsch (2010) 115 et seq.
11 For example, Han van Meegeren profited from the fact that the market craved for pictures by Jan Vermeer, see ibid. 122.

b) Making history with fake photographs

With the invention of photography, unexpected possibilities for the editing[12] and duplication of pictorial representations were developed.[13] This is accompanied by the potential to use images as a tool to influence numerous of people.[14] Long before the term "fake news" went viral in social media, the governments of totalitarian regimes already used this circumstance to their own advantage by distributing propaganda in the form of manipulated images to deliberately influence their population.[15] In particular, the former Soviet Union made extensive use of such manipulated photographs. These manipulations started with minor retouching such as the removal of a cigarette butt,[16] but also included the retouching of individual persons,[17] and went up to *damnatio memoriae*, the attempt to erase a particular person from collective memory.[18] The manipulations actually extended to a level where certain scenes were recreated for the photographs,[19] other representations were even completely staged.[20] Recently, manipulated images have also repeatedly appeared in the media (including those of non-totalitarian states).[21] And all these pictorial manipulations for the purpose of deception are enabled by an asymmetry concerning the information about the context in which photographs were being taken.[22]

12 The possibilities of editing range from subtle manipulations such as retouching of minor blemishes to complete photomontages.
13 Cf. Deussen (2007) 30.
14 Ibid.
15 Cf. Schuster (2020) 192.
16 Further information in Jaubert (1989) 110.
17 For some examples from Stalinism, see Stiftung Haus der Geschichte der Bundesrepublik Deutschland (2000) 82 et seq; King (1997).
18 For, e.g., the complete erasure of Trotzki see Jaubert (1989) 32; King (1997) 66 et seq.
19 One famous case of reconstruction was the image of the hoisting of the Soviet flag on the Reichstag in 1945, see Stiftung Haus der Geschichte der Bundesrepublik Deutschland (2000) 44 et seq.
20 See especially for the staged images of the leaders of totalitarian regimes Jaubert (1989) 53 et seq., 63 et seq., 79 et seq., and 99 et seq.
21 See the examples in: Stiftung Haus der Geschichte der Bundesrepublik Deutschland (2000).
22 Usually only the photographer has all the background information on how an image was taken, in contrast, the viewer only gets to see the final image; Boehme-Neßler (2010) 86.

c) "Face swap" as preliminary stage

For a while now, phenomena could be observed that come even closer to the conception of deep fakes. Parts of pictures in the form of faces have been cut out and integrated into other pictures. This was first done manually and later automatically.[23] The progression from these "face swaps" to deep fakes is not so remote. The sole difference is that with the deep fake technology, it is now possible to transfer facial expressions and gestures from one person to another and let the static "face swaps" turn into dynamic deep fakes.

2. The Technology behind: deep fake algorithms

A few years ago, this new phenomenon called deep fakes caused a worldwide wave of attention.[24] These are images[25] created with the help of artificial intelligence, which give the impression of authenticity.

a) Deep learning

The basic technology behind the current developments in artificial intelligence is deep learning.[26] Deep learning algorithms can solve a variety of problems. They work within an artificial neural network, which is a special form of an algorithm that is loosely based on the information processing in the human brain.[27]

---

23 For this approach see, e.g., Mallick (2016).

24 After a Reddit user published a number of pornographic videos, in which various female celebrities were seen, under the pseudonym Deepfakes at the end of 2017, this phenomenon was first reported on Vice in December 2017; see Cole (2017). Subsequently, worldwide reports about this phenomenon appeared; see, e.g., Roose (2018); FAZ Redaktion (2018).

25 Besides images and videos, voices are now also generated with the help of such "deep fake algorithms", see Chesney/Citron (2019) 1753 et seq. and, in particular, 1761 et seq.; Greengard (2020) 18.

26 On the use of deep learning technology in the context of deep fakes see further Nguyen et al. (2019).

27 Further on the information processing in an artificial neural network see Alpaydin (2020) 271 et seq.; especially on learning in an artificial neural network Russell/Norvig (2016) 694 et seq.

The term "deep fake" combines a variety of technologies based on deep learning algorithms. Autoencoders and Generative Adversarial Networks, are two applications that should be emphasised in the following discussion. Both these technologies offer the tools to create deep fakes with only a few images – as a matter of fact, even a single image of the targeted person is sufficient for some algorithms –[28] and a few hours of training with a sufficient processor.[29] Those deep fake algorithms permit the transfer of facial expressions and gestures from one person to another. And these technologies are already freely available to the public on the internet.

b) Autoencoder

One way to create convincing deep fakes is to use a so-called autoencoder. An autoencoder is a special form of an artificial neural network that consists of two nets: an encoder and a decoder. For creating a deep fake, in which person A's face should be exchanged by person B's face, data material of these two different persons is necessary to train the nets. For each of the faces, one encoder-decoder-pair is required. The encoders try to reduce the images to their substantial attributes and the decoders, to reconstruct the respective images from their substantial attributes received from the encoders to create a counterpart of the image.[30] The trick of this technology is that, after the weights have been memorised, decoder A is replaced by decoder B, so that the facial expressions of person A are reconstructed with the face of person B.[31] This process is enabled by the fact that the two encoders share their weights which means that the encoders have learned the common features of these two faces.[32] The decoders can therefore easily be exchanged and yet reconstruct the respective face from the reduced data.

28 See, e.g., the proposal of Siarohin et al. (2019).
29 See, e.g., the attempt to create one's own deep fake as an AI-layman by Schreiner (2019).
30 Nguyen et al. (2019) 2.
31 Ibid.
32 Ibid.

c) Generative adversarial network

Another way to create deep fakes is by way of a so-called generative adversarial network (GAN) which also consists of two different nets: a generator and a discriminator.[33] The learning process here runs in cycles. Imagine an art forger who wants to fool a gallery with his fake paintings.[34] Transferred to an artificial neural network, this art forger is the generator, and therefore receives some original paintings as input and must reproduce them. Now the faked pictures are sent to the gallery with several real pictures. There, a trainee should determine whether it is an original or a fake. After completion of the classification, both the trainee and art forger receive feedback to improve their results in the next learning cycle.

The final result is a picture of a person that was either never taken this way or has no real counterpart at all.

3. The power of images: why images are more than simple information media

These deep fakes are particularly effective when used to deceive other people, as they have the special power of images. Years ago, as the possibilities of digital image manipulation increased, similar problems were faced. Images, especially in the form of photographs, largely claim for authenticity in our society.[35] Therefore, viewers of an image, manipulated in a way that is unrecognisable at first glance, tend to assume that this picture is a representation of reality.[36] This seems surprising, at least on closer inspection, since in pictures in general, and photographs in particular, certain forms of manipulation are already immanent. After all, pictures always only show a part of the whole, a perspective, and are therefore not free of any external influences and thus not purely objective.[37] But why, given this background, do images – especially photographs – continue to have a largely unquestioned power of persuasion?

The outstanding persuasiveness of pictures may be based on the fact that pictures make a certain circumstance generally perceivable and thus

33 Goodfellow et al. (2014) 1.
34 Example based on Sabsch (2018).
35 See Schwarte (2015) 27 et seq.
36 See, e.g., the argument of the German Federal Constitutional Court (Bundesverfassungsgericht, BverfG), 1 BvR 240/04 of 14 February 2005.
37 Similarly Deussen (2007) 32 et seq.; Schürmann (2013) 17 et seq.

accessible to a general evaluation.[38] Conversely, this persuasiveness may also be due to the special way that the human eye and brain perceive an image.[39] The visual perception is the most important source of knowledge for the human mind,[40] therefore we prefer to rely upon the things we can see with our own eyes. However, continuing to refer to images without reflecting seems dangerous, as the meaning of images is already in a process of change which is (also) due to the developments in the field of pictorial illusion. Based on this pictorial power of persuasion, deep fakes can cause a lot of harm.

III. Deceptions Through Image Manipulation in the Information Age

1. State of the art: what AI is already capable of

The origins of deep fakes date back to pornography.[41] And even after several years of using this technology, by far most deep fakes still contain pornographic content.[42] These pornographic videos usually feature female celebrities as actresses and musicians.[43] And in the case of these pornographic deep fakes, the defamation of the person (especially females[44]) is still the main focus. But it is to be expected that the creators and users of such fakes will increasingly pursue additional purposes, especially in the economic and political field. In the political field, currently far more primitive forms of pictorial manipulation are still sufficient to deceive the observers (so-called "cheap fakes" or "shallow fakes").[45] However, the situation is slightly different in the economic field, where there have been

38 Schwarte (2015) 9.
39 This is what already happened in the context of image manipulation; see Deussen (2007).
40 See also, e.g., Anderson (2020) 1 et seq.
41 For the first report on deep fakes see Cole (2017).
42 E.g., for about thousands of deep faked nude pictures of women that appeared on Telegram see Möller (2020). For more information on deep fake sex videos see Citron (2019) 1921 et seq.
43 Ajder/Patrini/Cavalli/Cullen (2019) 2.
44 According to the study by Deeptrace, deep fake pornography even exclusively affects women; see ibid.
45 E.g., on various social media platforms manipulated videos were circulating, showing Nancy Pelosi, the US Democrat and House Speaker, in a slowed down way to give the impression that the politician was drunk; see Harwell (2019); O'Sullivan (2020).

cases of cyber criminals using audio deep fakes to obtain high amounts of money.[46]

Besides all these negative (potential) application scenarios, this technology can also be used in other areas in a meaningful way. Such positive fields of application can, for example, be found in art[47], the economy[48] and entertainment[49], especially in the form of satirical deep fakes[50]. It is even possible to give people who have lost their voices a chance to have them back.[51] Although these fields of application are generally positive and should not be obstructed by excessively strict regulatory measures, ethical principles must also be considered. Here too, the basically desirable end does not justify every means.

2. *Deep learning technology and its consequences*

But every technology has an inherent potential for damage to a certain extent. This also applies to the technology behind deep fakes. And depending on the use of deep fakes, they can have a multitude of direct consequences for the individual as well as further and indirect consequences for our society as a whole. For this reason, not everything that is technically possible using the means available to us is also reasonable from an ethical perspective. A distinction must therefore be made between what is technically possible, what is legally legitimised, and, within these limits, which also appears to be ethically acceptable.[52]

---

46 Stupp (2019).

47 Probably the most famous AI artwork and at the same time a deep fake in a broader sense is the "Portrait of Edmond Belamy", which was auctioned in 2018 at Christie's auction house; see Christie's (2018). In the meantime, however, a whole community of AI artists has already formed, see AIArtists.org, https://aiartists.org.

48 For example, this technology has been used to synthetically create a large number of artificial faces for stock photos, Berger (2019).

49 In the film industry, for example, this technology can be used to bring actors who have already passed away back onto the screen; see Chesney/Citron (2019) 1770 et seq.

50 See, e.g., the satirical video in which Barack Obama seems to give an opinion on his successor Donald Trump to warn of the dangers of deep fakes, BuzzFeedVideo (2018).

51 See for example the Project Revoice of the ALS Association, https://www.projectrevoice.org, in the context of which the technology of Lyrebird is used; see Lyrebird AI, https://www.descript.com/lyrebird.

52 These ethical questions do not only affect creators, but also intermediaries.

a) Consequences for the individual as a social being

In addition to the potential infringements, which will have to be addressed immediately, further consequences for the affected individuals must be feared. Even if the creator of the deep fake did not have fundamentally malicious intentions, such deep fakes can have various consequences for the personal integrity and social behaviour of the affected persons.[53] Those effects do not simply result from the mere creation of the deep fakes, but in particular, from the disclosure and dissemination of such images. For this reason, it is necessary to address these activities in the context of regulation.

b) Potentially affected rights

The most obvious right that could potentially be affected in the context of deep fakes is the right of publicity. This right protects, inter alia, the right to self-determination, including the right to decide if and how to present oneself in public and the right to one's own image. Because one's appearance is a, if not the most, significant aspect of one's personality, special effects on the right of publicity can be observed when images of one's appearance are involved. But does the infringement worsen if manipulations improve? In some contexts, the manipulations even profit from the fact that the final image (the output of the algorithm in form of the deep fake) is blurred or imperfect, because it seems more authentic. Currently, even the most primitive forms of manipulation seem to be sufficient to deceive a large portion of the recipients.[54] Therefore, it is often not necessary to have particularly good manipulations to cause serious harm.

Additionally, deep fakes require at least a few images, so there are also copyright issues to discuss. During the process of creating, the original images or at least parts of them need to be copied and one could argue that also the finished fake is a duplication of the original image or the original images.

53 For the story of Indian journalist Rana Ayyub, who had been a victim of a deep fake, see Citron (2019); see also Chesney/Citron (2019) 1773.

54 See above the examples of the Nancy Pelosi Cheap Fake videos.

c) Indirect consequences: disinformation

From an external point of view, the indirect consequences of this phenomenon, which will be addressed hereafter under the term disinformation, are potentially even more severe than the direct consequences for the individual. The persuasive power of images in combination with deep fake technologies, which enable the automated and uncomplicated creation of a multitude of manipulations, is a dangerous mixture because it can be used for targeted deceptions in sensitive sectors (such as in the political context). Another problem arising from deep fakes is that, at the moment, images have a very high level of credibility, e.g., as evidence in court, and important decisions are based on them. But in times of deep fakes, authenticity can't be guaranteed anymore. If we cannot trust our own eyes and ears anymore, the consequence is a general uncertainty, that can easily be exploited by perpetrators.[55]

With this multitude of potential risks that is intensified by our handling of images, it is necessary to find solutions to address these different aspects.

3. Regulating the consequences: possible solutions for this problem

a) Legal mechanisms

One way of regulating deep fakes is to use legal mechanisms. In law, basically a choice between three courses of action exists. One may completely prohibit, partly prohibit or fully permit.

Full permission is not an option regarding the issues linked to deep fakes, as the technology has an inherent potential for too much damage. A complete ban does not seem appropriate either, as there are certainly positive applications.[56] This points towards a partial ban, to be supplemented by duties of care and supervision for intermediaries.

Legal systems constantly lag behind the technological reality. Currently, there are no[57] provisions specifically designed to address the issue of deep fakes. But to implement some kind of partial ban, there already are regula-

55 Chesney and Citron have described the consequences of this general uncertainty about what is real and what is fake under the term "the liar's dividend"; for further details see Chesney/Citron (2019) 1785 et seq.

56 See above III.1.

57 For a few exceptional cases there already exist regulations, see immediately below.

tions that could be formulated in a way that is sufficiently general to cover new situations and could therefore also be applied to deep fakes. Such general provisions could especially be found in the regulations concerning copyright and the right of publicity (in particular the right to one's own image) and its non-constitutional manifestations.[58] These regulations on the special forms of the right of publicity were basically created with regard to the simple possibility of reproduction of certain personality describing attributes, such as the German Kunsturhebergesetz (KUG) with regard to the invention of photography[59] or the General Data Protection Regulation for the protection of personal data regarding the recent technological developments in connection with globalisation.[60] Since the right of publicity and copyright are open for development and depend on the technological progress,[61] they could provide for appropriate compensation in the context of deep fakes.

Through the mechanisms of the right of personality and copyright, the affected persons are particularly entitled to claim injunctive relief and compensation for damages.[62] In addition, there are also criminal law mechanisms: Defamatory deep fakes may be subject to offences against the personal sphere[63] and against honour, and if further goals are pursued, such as self-enrichment (as in the well-known CEO-fraud cases[64]), offences

58 In Germany, in particular the peronality rights' protection of one's own image provided for by the Kunsturhebergesetz (KUG) may be applied to deep fakes; see, e.g., Hartmann (2019) and (2020). – For constitutional aspects of the protection of one's own image see Eichenhofer (2022). In addition, at European level applying the regulations of the General Data Protection Regulation may also be considered, since pictures are also personal data; see Müller-Tamm (2022).

59 German Reichstag, Reichstagsprotokolle 11. Legislatur-Periode, 1530.

60 Recital 6, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Official Journal of the EU L 119 of 4 May 2016, corrected Official Journal L 127 of 23 May 2018.

61 Regarding the right of publicity see Götting (2019) § 2 note 1. Regarding copyright see Specht (2019) 253 et seq. and Hofmann (2016) 482 et seq.

62 In German law, this is enabled by Section 823 (1) and Section 1004 (1) of the German Civil Code in connection with the general right of personality and by Section 97 et seq. of the German copyright law. Moreover, the General Data Protection Regulation provides further claims for deletion and compensation resulting from the protection of personal data.

63 See, e.g., Section 201a (2) of the German Criminal Code, which deals with the protection of the personal sphere from images.

64 See already above III. 1.

against assets may apply.[65] Moreover, in a few cases, special regulations regarding deep fakes have recently been created.[66]

The affected individuals are therefore already protected by the legal system. However, it might be difficult to enforce their rights against the infringers. This is partly due to the fact that anonymity still prevails on the Internet to a large extent. In addition, information – and in particular, manipulated information[67] – spreads rapidly on social media platforms, reaching a wide range of people within a very short time.[68] Further, the effectiveness of countermeasures is controversial, at any rate it is difficult to stop disinformation once in circulation.[69]

In order to improve law enforcement, additional options are available, for example in intellectual property, one could claim the right to information against intermediaries.[70] In the context of the right of publicity, however, such explicit supplementary options are missing, which creates an imbalance in the enforcement of different rights[71] on the internet. For

65 Besides, there are further special criminal law regulations, for example in Section 33 of the German KUG and in Sections 106, 108 of the German copyright law.

66 In some US states, for example, it is illegal to create and spread deep fakes featuring politicians in temporal connection with elections (California Assembly Bill No. 730, Texas Senate Bill No. 751). In 2019 the state of Virginia extended the existing prohibition of non-consensual pornography to deep fake pornography (House Bill No. 2678). Also, China recently announced new regulations concerning deep fakes, Reuters (2019). In addition, the new Proposal for a Regulation of the European Parliament and of the Council laying down harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) should be mentioned in this context, which contains a labeling requirement for users of AI systems that generate or manipulate media, see Art. 52 (3) of the Proposal of an Artificial Intelligence Act.

67 See, e.g., Vosoughi/Roy/Aral (2018).

68 See Paschke/Halder (2016) 726.

69 See further Del Vicario et al. (2016); Zollo et al. (2015).

70 See Article 8 of Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights (Official Journal of the EU L 157 of 30 April 2004), based upon Article 47 of the Agreement on trade-related aspects of intellectual property rights (TRIPS-Agreement), as well as Article 8 of Directive 2001/29/EC of the European Parliament and the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (Official Journal of the EU L 167 of 22 June 2001), and Articles 15 (2) and 18 of Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Official Journal of the EU L 178 of 17 July 2000); Dreier (2022), § 101 note 1a.

71 For the different layers of legal rights with regard to encoded and materially stored information see Raue (2022).

this reason, there are considerations about creating instruments to improve law enforcement on the internet in the context of violations of personal rights. The German *Netzwerkdurchsetzungsgesetz* (NetzDG), for example, can at least partially be traced back to these considerations.[72]

As already seen in the context of law enforcement, information intermediaries play a key role in the regulation of deep fakes because they significantly contribute to the intensification of infringements by spreading disinformation on the internet. Conversely, they are able to combat infringements more effectively through the measures they take. However, these obligations cannot be unlimited, since, after all, Art. 15 of the eCommerce Directive stipulates a prohibition of a general duty of monitoring. Whereas previous attempts have been made to limit the liability of intermediaries, recently an opposing tendency seems to be observed on a European level. The European Court of Justice seems to be increasingly expanding intermediaries' duties of supervision in recent decisions by extending Art. 3 (1) Information Society Directive to secondary liability.[73] This development, which is also reflected in the introduction of the new Art. 17 of the DSM-Directive,[74] is likewise to be appreciated against the background of deep fakes and disinformation.

However, these legal mechanisms are by no means a complete solution, since they are at least partially powerless in relation to the indirect consequences of disinformation. Once in circulation, the manipulated information and its further implications can hardly be stopped by legal measures alone. Consequently, other mechanisms are needed to additionally address this problem.

---

72 See the explanatory memorandum of the Entwurf eines Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (NetzDG), German parliament (Deutscher Bundestag), BT-Drucksache 18/12356, 2. At the European level, the Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) was recently published (Doc. COM(2020) 825 final of 15 December 2020) that is also intended to address this issue.

73 See, e.g., ECJ Judgements C-682/18 and C-683/18, ECLI:EU:C:2021:503 – YouTube and Cyando; C-160/15, ECLI:EU:C:2016:644 – GS Media; C-527/15, ECLI:EU:C:2017:300 – Stichting Brein (Filmspeler) and C-610/15, ECLI:EU:C:2017:456 – Stichting Brein (Pirate Bay); Spindler (2019), 285.

74 In this context, the Proposal for a Digital Services Act (Doc. COM(2020) 825 final of 15 December 2020) should also be mentioned. According to the Proposal, the privileged position of the intermediaries should not be rejected entirely but replaced by a tiered liability system.

b) Technical solutions

In computer science, two potential ways of addressing the problem can be identified. Firstly, special techniques could be used to detect manipulated images. And why not fight fire with fire?[75] One could use AI technologies – more precisely a binary classifier – to distinguish real and manipulated images. These technologies have drastically improved recently, but so have the manipulations. It is only a matter of time before the manipulations overtake the detection methods again. Additionally, the deceivers benefit from data compression which often occurs when generated data needs to be transferred (for instance when images are shared on the web or audio is transmitted over a phone line). The resulting incompleteness and imperfection of information lead to traces left by the algorithms during generation becoming blurred and thus, detection is hampered which further strengthens the deception.[76] Irrespective of this, such methods are not effective in every respect. Especially for the disabled individual, they only offer a subsidiary benefit since the infringement has already occurred. The infringement can no longer be undone, and detection methods can therefore only help in the context of law enforcement. In contrast, those detection methods could help to reveal disinformation. In general, however, one must ask how effective purely subsequent measures could be, especially in the context of disinformation.

Secondly, for this reason, it is obvious to consider preventive technical protection measurements. There are various ways to protect the original pictures against manipulation. In contrast to those detection methods mentioned above, these protection mechanisms have the advantage that they address the real problem and do not merely intervene afterwards. But are such methods really more effective? The "problem" – or at the same time the advantage – of any technical protection mechanism is that they are constantly being developed. Technology becomes better by breaking or circumventing it, as this is the only way to identify the weak points. It is therefore not surprising that every technical protection measure has been broken so far. In the end, in all likelihood, technical protection measures will only be able to increase the threshold for counterfeiting and manipulation. However, even if such protective mechanisms might not

75 See, e.g., already Clark (1996): "The answer to the machine is in the machine".

76 See Unterrichtung der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale, German parliament (Deutscher Bundestag), BT-Drucksache 19/23700, 464.

completely prevent manipulations, they can at least help to prove existing manipulation.

c) Social measures

In addition to legal and technical measures, society should be considered as a third dimension. At the moment, people still believe what they see. So, it is much easier to deceive people by using images than with simple disinformation. This is the starting point for change. Society needs to be more aware of this issue. All the other possible solutions will remain ineffective or even be taken *ad absurdum* if everyone does not become aware of this problem. For example, legal measures seem meaningless if the enforcement of rights in court is based on "false facts". And classification algorithms become useless if AI is trained to detect fake images with undetected deep fakes.[77] It is therefore necessary to question the authenticity of images.

d) Combination

The effects of this problem will soon also appear in sensitive sectors, as in court where important decisions are made and where images have developed into an important source of knowledge. For this reason, it is necessary that the handling of images is amended, particularly in key positions. Images should no longer be granted unlimited significance in court proceedings. This can be legally implemented by reversing the burden of proof; i.e., it should be assumed that any picture presented as evidence to a trial court is fake unless it is explicitly marked or can be proven beyond reasonable doubt as authentic.[78] Reversing the burden of proof also seems appropriate against the background of the asymmetry of information with regard to photographs.[79] If an image is not verifiably authentic, the technical detection mechanisms described above could help to prove or exclude manipulation.[80] Simultaneously, this provides an incentive for journalists,

77 For the consequences false information can have for the training of AI, see Hurtz (2020).
78 See Deussen (2007) 145 et seq.
79 See above I. 2.
80 To this effect cf. Leone (2022).

photographers and increasingly also for information intermediaries (who all occupy key positions) to work diligently and ethically in order to regain the meaningfulness that images once enjoyed. Such an approach also facilitates the handling of images for everyone.[81]

For this purpose, consistent principles will have to be formulated that regulate the handing of (potentially manipulated) information in general and images in particular, by these key operators. These ethical questions raised by deep learning-technologies are not entirely novel. But, they have become considerably more intense in the context of automated and highly convincing visual deceptions in the form of deep fakes. It is therefore possible to rely on the familiar ethical principles in connection with images, deceptions and disinformation, which will have to be applied and adapted to the particularities of the new phenomenon.

IV. Conclusion

In fact, there already are measures that are – if used properly, combined and adapted selectively – able to adequately, still not extensively, regulate the phenomenon of deep fakes. However, it is neither possible, nor reasonable to regulate deep fakes extensively by using the strict measures of law and technology, since on the one hand, the phenomenon of deep fakes is constantly developing, and on the other hand, a single technology can be used equally for positive and negative purpose.[82] This "gap" can and should be filled by ethical rules, seeking a responsible and transparent handling of this technology.

References

Ajder, Henry/Patrini, Giorgio/Cavalli, Francesco/Cullen, Laurence (2019): *The State of Deepfakes: Landscape, Threats, and Impact* (2019, https://regmedia.co.uk/2019/10/08/deepfake_report.pdf)

Alpaydin, Ethem (2020): *Introduction to machine learning* (4th ed., Cambridge/MA.: MIT Press 2020)

Anderson, John R. (2020): *Cognitive psychology and its implications* (9th ed., New York: Worth Publishers 2020)

81 Similarly, Deussen (2007) 147.

82 See, e.g., the different possible applications of audio deep fakes, above III.1.

Berger, Daniel (2019): 'Stockfoto-Firma veröffentlicht 100.000 KI-Gesichter' (2019, https://www.heise.de/newsticker/meldung/Stockfoto-Firma-veroeffentlicht-100-000-KI-Gesichter-4537889.html)

Boehme-Neßler, Volker (2010): *BilderRecht – Die Macht der Bilder und die Ohnmacht des Rechts* (Berlin/Heidelberg: Springer 2010)

Brinkmann, Franziska (2020): 'Formen der Kopie von der Fälschung bis zur Hommage – Eine Begriffsbestimmung und ihre Grenzen', in: Dreier, Thomas/Jehle, Oliver: *Original – Kopie – Fälschung* (Baden-Baden: Nomos 2020) 57–104

Butin, Hubertus (2020): 'Die Crux mit dem Original – ein Begriff und seine Grauzonen in der Kunstgeschichte und im Rechtswesen', in: Dreier, Thomas/Jehle, Oliver: *Original – Kopie – Fälschung* (Baden-Baden: Nomos 2020) 37–55

BuzzFeedVideo (2018): 'You Won't Believe What Obama Says In This Video!' (2018, https://www.youtube.com/watch?v=cQ54GDm1eL0)

Chesney, Robert/Citron, Danielle (2019): 'Deep Fakes – A Looming Challenge for Privacy, Democracy, and National Security', 107 California Law Review (2019) 1753–1819

Christie's (2018): 'Is artificial intelligence set to become art's next medium?' (2018, https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx)

Citron, Danielle (2019): 'How deep fakes undermine truth and threaten democracy', TEDSummit 2019 (2019, https://www.ted.com/talks/danielle_citron_how_deepfakes_undermine_truth_and_threaten_democracy/transcript)

Citron, Danielle Keats (2019): 'Sexual Privacy', 128 Yale Law Journal (2019) 1870–1960

Clark, Charles (1996): 'The Answer to the Machine is in the Machine', in: P. Bernt Hugenholtz (ed), *The Future of Copyright in a Digital Environment* (The Hague/London/Boston: Kluwer Law International 1996) 139–145

Cole, Samantha (2017): 'AI-Assisted Fake Porn Is Here and We're All Fucked' (2017, https://www.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn)

CORRECTIV (2020): 'Coronavirus-Faktenchecks – Diese Behauptungen hat CORRECTIV geprüft' (2020, https://correctiv.org/alle-corona-faktenchecks)

Del Vicario, Michela/Bessi, Alessandro/Zollo, Fabiana/Petroni, Fabio/Scala, Antonio/Caldarelli, Guido/Stanley, Eugene H./ Quattrociocchi, Walter (2016): 'The spreading of misinformation online', 113 Proceedings of the National Academy of Science of the USA (2016) 554–559

Deussen, Oliver (2007): *Bildmanipulation – Wie Computer unsere Wirklichkeit verzerren* (Berlin/Heidelberg: Springer 2007)

DiResta, Renee/Shaffer, Kris/Ruppel, Becky/Sullivan, David/Matney, Robert/Fox, Ryan/Albright, Jonathan/Johnson, Ben (2019): 'The Tactics & Tropes of the Internet Research Agency' (2019, https://disinformationreport.blob.core.windows.net/disinformation-report/NewKnowledge-Disinformation-Report-Whitepaper.pdf)

Dreier, Thomas (2022): 'Commentary on § 101 of the German Copyright Act', in: Dreier, Thomas/Schulze, Gernot: *Urheberrechtsgesetz* (München: C.H. Beck 7th ed. 2022)

Eichenhofer, Johannes (2022): 'The Constitutional Protection of Images', in: Dreier, Thomas/Andina, Tiziana: *Digital Ethics – The issue of images* (Baden-Baden: Nomos 2022) 357 – 385

Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale (2020): 'Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale' (2020, https://dip21.bundestag.de/dip21/btd/19/237/1923700.pdf)

FAZ Redaktion (2018): 'Jetzt also "Fake Porn"' (2018, https://www.faz.net/aktuell/wirtschaft/kuenstliche-intelligenz/fake-porns-mit-ki-werden-promis-in-pornos-eingebaut-15419920.html)

Götting, Horst-Peter (2019): 'Geschichte des Persönlichkeitsrechts', in: Götting, Horst-Peter/Schertz, Christian/Seitz, Walter: *Handbuch Persönlichkeitsrecht* (München: C.H.Beck 2nd ed. 2019)

Goodfellow, Ian J./Pouget-Abadie, Jean/Mirza, Mehdi/Xu, Bing/Warde-Farley, David/Ozair, Sherjil/Courville, Aaron/Bengio, Yoshua (2014): 'Generative Adversarial Networks' (2014, https://arxiv.org/abs/1406.2661)

Greengard, Samuel (2020): 'Will Deepfakes Do Deep Damage?', 63 Communications of the ACM (2020) 17–19

Hartmann, Frank (2019): 'Der Schutz vor Deepfakes durch das Kunsturhebergesetz', in: Taeger, Jürgen*: Die Macht der Daten und der Algorithmen – Regulierung von IT, IoT und KI* (Edewecht: Oldenburger Verlag für Wirtschaft, Informatik und Recht 2019) 563–579

Hartmann, Frank (2020): 'Der persönlichkeitsrechtliche Schutz vor Deepfakes', Kommunikation und Recht (2020) 350–357

Harwell, Drew (2019): 'Faked Pelosi videos, slowed to make her appear drunk, spread across social media' (2019; https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media)

Hebborn, Eric (1997): *Der Kunstfälscher* (3rd ed., Cologne: DuMont 2011; original ed.: Hebborn, Eric: *The Art Forger's Handbook*, London: Cassell 1997)

Hofmann, Franz (2016): 'Grundsatz der Technikneutralität im Urheberrecht? – Zugleich Gedanken zu einem more technological approach', Zeitschrift für geistiges Eigentum (2016) 482–512

Hurtz, Simon (2020): 'Wikipedia auf Scots – Hälfte der Artikel ist gefälscht' (2020, https://www.sueddeutsche.de/digital/wikipedia-scots-uebersetzung-englisch-1.5030086)

Jaubert, Alain (1989): *Fotos, die lügen – Politik mit gefälschten Bildern* (Frankfurt am Main: Athenäum 1989)

King, David (1997): *Stalins Retuschen – Foto- und Kunstmanipulation in der Sowjetunion* (Hamburg: Hamburger Edition 1997)

Leone, Massimo (2022): 'Semioethics of the Visual Fake', in: Dreier, Thomas/Andina, Tiziana: *Digital Ethics – The issue of images* (Baden-Baden: Nomos 2022) 187–205

Mallick, Satya (2016): 'Face Swap using OpenCV (C++/Python)', (2016, https://www.learnopencv.com/face-swap-using-opencv-c-python)

Maras, Marie-Helen/Alexandrou, Alex (2019): 'Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos', 23 The International Journal of Evidence & Proof (2019) 255–262

Möller, Malin (2020): 'Telegram – Mehr als 100.000 Deepfake-Nacktbilder von Frauen öffentlich geteilt' (2020, https://www.spiegel.de/netzwelt/web/telegram-mehr-als-100-000-deepfake-nacktbilder-von-frauen-oeffentlich-geteilt-a-4b34eeef-43d8-4c90-ba43-8dff569edc09)

Müller-Tamm, Lorenz (2022): 'Portrait or Personal Data – The Rivalry of Image and Data Protection Legislation', in: Dreier, Thomas/Andina, Tiziana: *Digital Ethics – The issue of images* (Baden-Baden: Nomos 2022) 239–252

Nguyen, Thanh Thi/Nguyen, Cuong M./Nguyen, Dung Tien/Nguyen, Duc Thanh/Nahavandi, Saeid (2019): 'Deep Learning for Deepfakes Creation and Detection' (2019, https://arxiv.org/abs/1909.11573)

O'Sullivan, Donie (2020): 'Another fake Pelosi video goes viral on Facebook' (2020, https://edition.cnn.com/2020/08/02/politics/fake-nancy-pelosi-video-facebook/index.html)

Partsch, Susanna (2010): *Tatort Kunst – Über Fälschungen, Betrüger und Betrogene* (München: C.H.Beck 2010)

Paschke, Anne/Halder, Christoph (2016): 'Auskunftsansprüche bei digitalen Persönlichkeitsrechtsverletzungen', Multimedia und Recht (2016) 723–727.

Raue, Benjamin (2022): 'The Multi-Layered Information in a Digital Image', in: Dreier, Thomas/Andina, Tiziana: *Digital Ethics – The issue of images* (Baden-Baden: Nomos 2022) 229–237

Reuters Staff (2019): 'China seeks to root out fake news and deepfakes with new online content rules' (2019, https://www.reuters.com/article/us-china-technology/china-seeks-to-root-out-fake-news-and-deepfakes-with-new-online-content-rules-idUSKBN1Y30VU)

Roose, Kevin (2018): 'Here Come the Fake Videos, Too' (2018, https://www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html)

Russell, Stuart J./Norvig, Peter (2016): *Artificial intelligence – A modern approach* (3rd ed., Boston/Columbus/Indianapolis: Pearson 2016)

Sabsch, Tim (2018): 'Eine kurze Einführung in Generative Adversarial Networks' (2018, https://blog.codecentric.de/2018/11/eine-kurze-einfuehrung-in-generative-adversarial-networks)

Schreiner, Maximilian (2019): 'KI –Deepfake selbst erstellen – so geht es, so lange dauert es' (2019, https://mixed.de/ki-deepfake-selbst-erstellen-so-geht-es-so-lange-dauert-es)

Schürmann, Eva (2013): 'Erscheinen als Ereignis – Zeittheoretische Überlegungen zur Fotografie', in: Alloa, Emmanuel: *Erscheinung und Ereignis – Zur Zeitlichkeit des Bildes* (München/Paderborn: Fink 2013) 17–38

Schuster, Martin (2020): *Fotopsychologie – Fotos sehen, verstehen, gestalten* (Berlin/Heidelberg: Springer 3rd edition 2020)

Schwarte, Ludger (2015): *Pikturale Evidenz – Zur Wahrheitsfähigkeit der Bilder* (Paderborn: Wilhelm Fink 2015)

Sherwin, Richard K./Feingenson, Neal/Spiesel, Christina (2006): 'Law in the Digital Age – How Visual Communication Technologies are Transforming the Practice, Theory, and Teaching of Law', 12 Boston University Journal of Science and Technology (2006) 227–270

Siarohin, Aliaksandr/Lathuilière, Stéphane/Tulyakov, Sergey/Ricci, Elisa/Sebe, Nicu (2019): *First Order Motion Model for Image Animation*, 33rd Conference on Neural Information Processing Systems (NeurIPS) (2019, http://papers.nips.cc/paper/8935-first-order-motion-model-for-image-animation.pdf)

Specht, Louisa (2019): 'Zum Verhältnis von (Urheber-)Recht und Technik', Gewerblicher Rechtsschutz und Urheberrecht (2019) 253–259

Spindler, Gerald (2019): 'Die neue Urheberrechts-Richtlinie der EU, insbesondere "Upload-Filter" – Bittersweet? Analyse der Änderungen beim Text- und Datamining, Leistungsschutz für Presseerzeugnisse und Pflichtenkreis für Hostprovider', Computer und Recht (2019) 277–291

Stiftung Haus der Geschichte der Bundesrepublik Deutschland (2000): *Bilder, die lügen* (Bonn: Bouvier Verlag 2nd edition 2000)

Stupp, Catherine (2019): 'Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case' (2019, https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402)

Timberg, Craig/Romm, Tony (2018): *New report on Russian disinformation, prepared for the Senate, shows the operation's scale and sweep* (2018, https://www.washingtonpost.com/technology/2018/12/16/new-report-russian-disinformation-prepared-senate-shows-operations-scale-sweep)

Vosoughi, Soroush/Roy, Deb/Aral, Sinan (2018: 'The spread of true and false news online', 359 Science (2018) 1146–1151

Zollo, Fabiana/Novak, Petra Kralj/Del Vicario, Michela/Bessi, Alessandro/Mozetič, Igor/Scala, Antonio/Caldarelli, Guido/Quattrociocchi, Walter (2015): 'Emotional Dynamics in the Age of Misinformation', 10 PLoS ONE (2015) 1–22