Journal of Intellectual Property and Information Technology Law (JIPIT)
ISSN: 2788-6727 – Volume 2, Issue 1 (2022) p.145-186
https:doi.org/10.52907/jipit.v2i1.208

The Regulation of Deepfakes in Kenya

*Faith Amatika-Omondi**

ABSTRACT

'Truth has become elusive.' 'We are entering into an age of information apocalypse.' 'Seeing is no longer believing unless you saw it live.' These and similar statements characterise most discussions in the present highly digital age. With the borderless nature of the Internet, it is possible to share videos, photos, and information with countless people provided one has a reliable internet source and a smart gadget, for instance, a mobile phone. Technological advancements have also made it possible for tech-savvy individuals to compile computer programs that make it possible to swap faces and replace them with those of celebrities, politicians, et cetera. Yet even more sophisticated technology uses Artificial Intelligence (AI) methods to create videos and photos that are not easily distinguishable from the real ones. 'Deepfakes' has become a buzzword. Along this line, this paper posits that there is widescale misinformation due to deepfakes and assesses the regulation of deepfakes in Kenya to curb the misinformation. It recommends pragmatic ways to train forensic experts and to create awareness among members of the public for detecting deepfakes, hence curbing their negative effects.

Keywords: Deepfakes, Misinformation, Infocalypse, Face Swaps, Artificial Intelligence (AI), Manipulation

* Chief Legal Counsel at the Kenya Copyright Board. She holds an LL.M Degree jointly from Ankara University, Turkey, and the World Intellectual Property Organization (WIPO) Academy. Faith has had various professional engagements for instance serving as Kenya's focal point in the WIPO Committee on Development and Intellectual Property (CDIP) Project on the use of IP in the software sector; consulting for the U.S Civilian Research and Development Foundation (CRDF Global) Arlington, Virginia; and as a Cyber Policy Centre (CPC) Fellow at the Centre for Intellectual Property and Information Technology Law (CIPIT), Strathmore University.

TABLE OF CONTENTS

I. INTRODUCTION

'Can we uninvent the bomb?' asks Donald Mackenzie, a sociologist and science studies' scholar (Paris & Donovan, 2019). The answer is a resounding 'no'. Once a technology is invented, it is impossible to uninvent it. The question that then begs is how one would know that certain technology has a potentially harmful effect *ab initio,* especially if it guarantees open access.

Technology that is potentially harmful or negatively disruptive is usually not granted protection. The same is seen in most patent legislations where there are provisions for the unpatentability of inventions that are contrary to public order, morality, public health and safety, principles of humanity, and environmental conservation (Kenyan Industrial Property Act, No. 3 of 2001, s. 26(b)). The Agreement on Trade-Related Aspects of Intellectual Property (TRIPS Agreement) excludes from patentability inventions the exploitation of which is contrary to *ordre public* or morality and provides that inventions should aim at protecting human, animal, plant life or health or avoid serious prejudice to the environment (a. 27(2)).

Further, section 27 of the Kenyan Industrial Property Act (2001) regulates the patentability of potentially weaponizable technologies by giving the Managing Director authority to restrict the publication of such information. This prevents the public from openly accessing such information and using it negatively (Industrial Property Act, 2001, s. 42). The Managing Director is also required to coordinate with the Cabinet Secretary in charge of defence as well as the National Commission of Science, Technology, and Innovation (where technology involves atomic energy) in determining whether information relating to the patent should be published or not (Industrial Property Act, 2001, s. 27(5)).

While the patentability of technology that is contrary to public order or morality is outlawed in Kenya, how would one address a situation where the technology is not protected by patent or trademarks but by another intellectual property right

(IPR), for instance, copyright, which does not require registration for protection to be granted? In Kenya, computer programs, in which category software and apps fall, are protected by copyright as literary works (Kenya Copyright Act, 2001, s. 2). There are several apps on the Google Play store and App Store that one can easily download and use to manipulate a face. They include Reface: Funny Face swap videos,[1] FakeApp,[2] FaceSwap,[3] and DeepFace Lab,[4] among others, which are available as open-source software.[5] The effect of such apps is that, as most of their names suggest, they swap faces, they 'reface' (alter) a face and overall, they create fake faces and fake content. The fake content often serves as instruments of misinformation among other applicable uses that can be derived from them.

Mobile applications (hereafter referred to as 'apps') and software are composed of different aspects which are protected by various IPRs including the source code, which is protected by copyright (Shemtov, 2021, p. 28; Kenyan Copyright Act, 2001, s. 2). Trademarks protect the name and logo of the app. Industrial designs may be appropriate for the Graphical User Interface (GUI), et cetera (Shemtov, 2021, p. 28). On rare occasions, there are aspects of the software that are protectable by patents in Kenya.

For the apps highlighted above, the component that enables them to perform the function of swapping faces is the source code, which is automatically protected by copyright upon compiling the code. There is no opportunity to test the code for conformity with morality and public order or other requirements for

[1] Reface https://play.google.com/store/apps/details?id=video.reface.app&hl=en_US&gl=US.

[2] FakeApp FakeApp 2.2.0 - Download for PC Free (malavida.com).

[3] Deepfakes, Deepfakes/Faceswap, Python, 2019, https://github.com/deepfakes/faceswap.

[4] Iperov, DeepFaceLab is a tool that utilizes machine learning to replace faces in videos. Includes prebuilt ready to work standalone Windows 7, 8, 10 Binary (Look Readme.Md), Iperov/DeepFaceLab, Python, 2019, https://github.com/iperov/DeepFaceLab.

[5] Open-source software is software that one can freely inspect, modify, and enhance. See Opensource.com https://opensource.com/resources/what-open-source.

patentability. This leads to the question, 'how possible is it to regulate such technologies that receive intellectual property protection without any formal or substantive examination?' The law grapples with these challenges and appears to focus more on the application of such products.

Against this backdrop, this paper examines the implications of deepfakes in Kenya, the legal frameworks established to regulate deepfakes, and how to best alleviate their negative adoption and application. For this examination, this paper is divided into five parts. Part I is the introduction and gives a brief background of the problem. Part II examines the nature of deepfakes and the historical evolution of the same to unearth deepfakes as an advanced form of photo, video, or content manipulation for various reasons. Part III studies the creation of deepfakes and their application to recommend detection tools or skills and government engagement towards capacity building to adequately address the challenges associated with deepfakes. It also addresses both the positive and negative applications of deepfakes. Part IV discusses the legal regulation and some regulatory challenges associated with deepfakes. Part V makes recommendations on the best way to ensure the technology around deepfakes is harnessed for the benefit of humanity. It also concludes the paper.

II. HISTORICAL DEVELOPMENT OF PHOTO AND VIDEO MANIPULATION

The manipulation of photos, videos, or other information did not start with the deployment of Artificial Intelligence (AI) in deepfakes. To understand the nature of deepfakes better, it is necessary to backtrack to when photo or video manipulation began and then examine the technological advancements that have made deepfakes a possibility.

As far behind as eighteenth-century France, images of Maria Antoinette and Louis XVI were depicted in sexually explicit cartoons (Burkell & Goose, 2019). This may not be considered a

form of photo manipulation, but the message was manipulated to give false information with a negative impact on the target audience. The effect was that the populace developed so much bile against the queen and all her close associates that they instituted merciless ostracization against them, which culminated in their date with the guillotine (Encyclopaedia Britannica, 2020).

Photo or video manipulation has been in existence for as long as digital photos and videos have been in existence or even before (Photo Tampering Through History, 2020). An example of this is in the famous Abraham Lincoln presidential portrait (Iconic Photos, 2010). The composite photo where Lincoln's head was placed on Calhoun's body was created by Thomas Hicks in the mid-1860s after Lincoln's assassination (Iconic Photos, 2010). The aim was to create a presidential portrait for the fallen President, which did not exist until then. For some years, Americans believed that the portrait was the image of President Lincoln, and it was only later that Stefan Lorant, the art director for the London Picture Post magazine, noticed that the photo was fake (Iconic Photos, 2010). The irony in the composite photo is the divergent views that the two statesmen supported. While Calhoun supported slavery (History.com Editors, 2019), Lincoln was for the emancipation of slaves (History.com Editors, 2021). Therefore, other than merely being used to improvise for the lack of a presidential portrait of Lincoln, the composite photo might have been used for satirical purposes.

In the year 1975, Steve Sasson, a former Kodak electrical engineer, invented the digital camera (Tutorial Example, 2021). This has made it possible to easily edit a photograph and vary the effects by, for instance, blurring the background. It has given rise to what photograph analysts refer to as the subjectification of photographs. This is as opposed to photo objectification which is easily achieved with analogue (film-based) photography techniques (Biro, 2012). It is argued that the rapid adoption of digital techniques of photography in the 1990s heralded an erosion of trust in the truthfulness of images produced by digital technolo-

gies (Biro, 2012, p. 354). A digital photo has a greater propensity to conceal the truth in comparison to an analogue photo.

In analysing the works of renowned photographers Bernd and Hilla Becher,[6] the truth or objectiveness of the content is evident. First, is the possibility to choose the angle of taking the shots as well as the distance from the object (Biro, 2012 p. 354). In an analogue (film-based) shot, all the photographer's preferences are shelved as the focus is on ensuring that the image is captured as it is as much as possible (Biro, 2012, p. 353). To this end,

> 'the camera is consistently level with the middle of the subject; lighting is even and diffused; contrast is reduced to give all parts of the structure a similar weight and impact; and the background is de-emphasized to direct the spectator's attention to the architecture itself – its iconic form and its indexical connection to a specific construction existing at a particular moment of historical time' (Biro, 2012, p. 354).

Since the photographs are taken to capture the image as it is, this author finds that it leads to the preservation of important historical information that can be used in schools for educational purposes or in courts as evidence especially when the object no longer exists. The reverse is true for digital photographs, which are sometimes edited and modified to conceal or add more information.

In contrast to an objective mindset in photography, a subjective mindset leads to the capturing of a photograph that is more of artwork, and less of a source of information (Biro, 2012, p. 357) as explained in the foregoing paragraph. In a subjective mindset, the photographer aims to communicate what they want the world to see as opposed to conveying information as it is (Biro, 2012, p. 357). In examining Andreas Gursky's[7] works, one notices a wide departure from the Bechers' works (Biro, 2012). Andreas employs photo-editing techniques available for digital cameras

[6] Bernd and Hilla Becher were renowned photographers who specialized in analogue photos (Tate) < https://www.tate.org.uk/art/artists/bernd-becher-and-hilla-becher-718/who-are-bechers > accessed 15 November 2022.

[7] Andreas Gursky was a photographer who employed digital photo-editing skills in photographs hence creating spectacular art forms (Tate) < https://www.tate.org.uk/art/artists/andreas-gursky-2349 > accessed 15 November 2022.

and captures spectacular images of nature and other unnatural objects, but the works lack the objectivity seen in the Bechers' works. His photos are symbolic and subjective, more like works of art (Biro, 2012, p. 357).

Biro (2012) describes Gursky's photograph of the Rhine[8] as something that does not exist in nature (p. 358). He observes that the depth of nature has been compressed through the lens used and that the 'Paintbox' software used removed some buildings from the photo's background. Consequently, a natural geographical feature has been transformed into an abstract artwork due to the interaction of nature and technology (Biro, 2012 p. 358). By employing editing software and other techniques, Andreas reinforces the foregoing argument that with digital cameras, a photographer can be subjective and create art forms that appeal to their tastes. Therefore, the truth is suppressed as photographers manipulate photographs to turn them into mere art that is abstract and no longer documentaries.

The same is the case with videos. The ease with which digital videos can be made today also comes with the ease to edit them. The Internet is replete with several free video-editing tools and software (Seigchrist, 2021). This leaves humanity in the same place with digital photographs. They are no longer objective; rather, they are subjective as the videographer pushes for their likes and preferences hence the video lacks the truthfulness that comes with analogue videos. A video can also be edited such that a caption, which may be true or false, is included to describe certain footage. The false caption is what

[8] 'Andreas Gursky, The Rhine II, 1999' (Tate) < https://www.tate.org.uk/art/artworks/gursky-the-rhine-ii-p78372 > accessed 18 May 2021. Andreas describes how he ended up with the masterpiece of a photograph in the following words: 'there is a particular place with a view over the Rhine which has somehow always fascinated me, but it didn't suffice for a picture as it basically constituted only part of a picture. I carried this idea for a picture around with me for a year and a half and thought about whether I ought perhaps to change my viewpoint ... In the end, I decided to digitalize the pictures and leave out the elements that bothered me.' This explanation by Andreas reinforces the foregoing argument that with digital cameras, a photographer can be subjective and create artforms that they wish to see. The truth is therefore suppressed as photographers manipulate photographs to turn them into artforms that are no longer documentary. They therefore lack probative value.

is commonly known as a 'cheap fake', (Arnold, 2020), where a simple video is edited to include wrong captioning or is slowed down to distort the original message. Depending on how wrong the caption is, or the effect of the slow motion, the video may be interpreted differently.

The possibility to use software to remove some undesirable features in a photo or video distorts their authenticity, as they cease to represent reality. However, the use of digital forensic technologies and expertise makes it possible to detect digital alterations in videos or photographs (Melendez, 2018). Such technologies are used by experts to analyse any inconsistencies in the photo or video for instance the direction of the shadow from the light source, the absence of any realistic expectations in a photo or video, like the absence of reflection of an object in a pool of water, among others.

This digital manipulation of videos or photos has given rise to what has come to be known as dumb or cheap fakes (Dupuy & Ortutay, 2019). Dumb fakes are fake videos or photos that are developed by other means apart from AI deep learning techniques. Such other means include simple video or photo edits through the elimination of some features from the original photo or even a simple mechanical process such as playing the video in slow motion. Such 'dumb' video manipulations have resulted in the passing of wrong information and the subsequent wrong interpretation of the video by the target audience. An example is an infamous video on the US House Speaker Nancy Pelosi where, due to the video being played in slow motion, she appeared to be inebriated and hence not fit to discharge the duties of her office (Dupuy, 2019). The video caused so much negative press due to the interpretation it caused.

In Kenya, social media is replete with funny videos portraying politicians and celebrities in satirical situations. The uptake of the TikTok App,[9] for instance, has made it possible for videos

[9] 'TikTok, Free Video Creation and Sharing App' <https://tik-tok.en.softonic.com/> accessed 20 May 2021.

and jokes to be circulated at astronomical speeds. Most of the videos shared on this platform are outrightly manipulated and not real and this is what makes them popular because they mostly serve entertainment purposes.

Away from the entertainment purpose served by manipulated videos, these videos have the potential to serve other purposes, and Bitange Ndemo (2021) has attempted to propose the enactment of laws to curb the negative uses of this technology. In addition, Peter Kagwanja (2009) reflects on Kenya's 2007-2008 post-election crisis, which was partly sparked and fuelled by misinformation and hate messages sent via SMS. Ndemo (2021) weighs in on the same and foresees a situation where wrong information from manipulated audio-visual snippets can cause unprecedented harm if not checked. According to him, this technology can also be used by politicians to conceal their failure to discharge their official duties hence misleading people concerning their fitness for office.

Other videos on social media may have adverse effects of mudslinging, settling political scores or for campaign purposes. An example is the manipulated image showing Kenya's then Deputy President, William Ruto, grinning while paying his respects to the fallen third president of Kenya, Mwai Kibaki. Some Kenyans must have believed this image to be true as one of the comments on the photo was *'Kutoka nizaliwe, sijawai ona mtu anachekelea mtu amekufa. UDA was happy'* translating to 'since birth, I have never seen someone laugh at a dead person' (PesaCheck, 2022). There were other comments along this line which also included reference to a political party, the United Democratic Alliance Party (UDA). This means that the video served a political purpose of mudslinging the party and painting it in a bad light, probably to swing votes to the opposing political party. However, PesaCheck[10] found the video to have been doctored.

[10] PesaCheck is Africa's largest indigenous fact-checking organisation, debunking misleading claims and deciphering the often-confusing numbers quoted by public figures in 15 African countries. https://pesacheck.org/ accessed July 28 2022.

Political mudslinging has also been witnessed where an image of Kenya's former Prime Minister, Raila Odinga, was manipulated to show that he was asleep during a meeting in the United States in April 2022. Comments on social media about the image were such that Raila would 'wake up when Kenyans have already buried former President Kibaki' (PesaCheck, 2022). PesaCheck also confirmed that the image was doctored as the original one was taken in 2012 while the one in question was purportedly taken in 2022.

The ease of detecting manipulated digital content in comparison with content made using AI deep learning technologies explains why there is much concern and trepidation over photos or videos manipulated using the latter. With digitally manipulated photos and videos, there is a possibility of comparison with still images hence it is possible to see some inconsistencies. This is not the case with AI-derived deepfakes because the same may be entirely new images or videos, which never existed prior. The possibility of comparing with a still or live image is not there. They are deepfakes because they purport to be a representation of something that does not exist; for instance, the creation of a deepfake photo of a bridge over a river where none exists in nature. As such, deepfakes can be completely new images or videos that, *prima facie,* do not have any inconsistencies that are likely to be digitally manipulated but can misinform the public (Chesney & Citron, 2019).

III. THE CREATION AND APPLICATION OF DEEPFAKES

A. The use of Artificial Neural Networks

In creating deepfakes, AI employs the machine learning component known as Artificial Neural Networks (ANNs) which tends to mimic the neurons in a human brain (Kukreja *et al*., 2016, p. 27). ANNs are complex layers of mimicked neurons between the point where data is inputted into the machine and the

outcome (Käde & Von Maltzan, 2019). The first layer, as seen in figure 1 below, is the input layer (red) and it has as many input neurons as the features to be analysed. The features could be the number of legs, the presence or absence of a tail, fur, etc. The last layer (green) is the output layer and gives the result of the analysis of the data that was inputted at the beginning (red). Depending on how the algorithm was trained (what happened in the layer(s) between red and green by applying suitable weights to achieve the expected results) the output may be anything. The machine adapts the weights with the input of the programmer until it can give the expected results (Hedrick, 2019, p. 363 - 367).

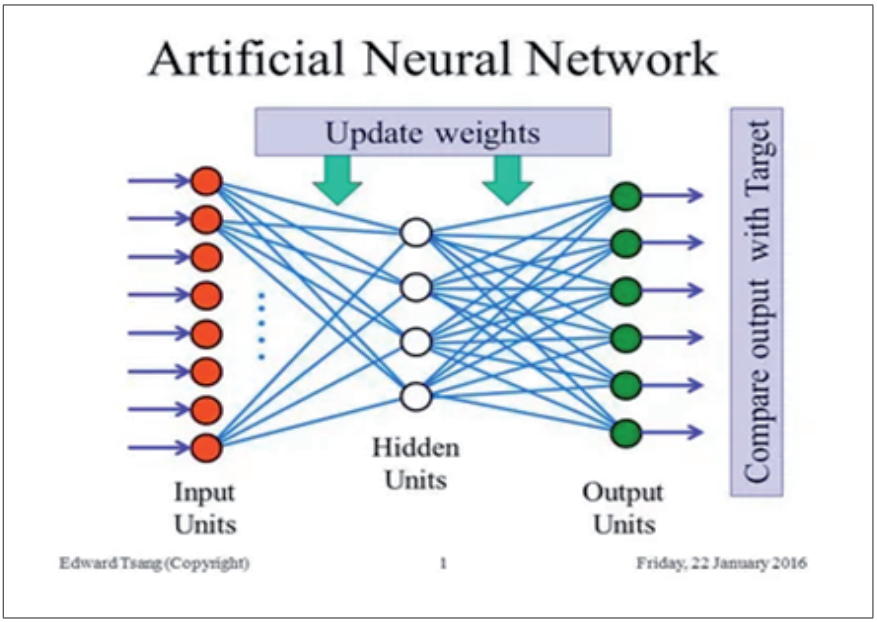

Figure 1 - Input layer, weights, one middle layer, and the output layer.[11]

To obtain accurate results, the machine must be adequately trained. For instance, it must 'see' as many pictures of dogs as possible. Therefore, the input data must be adequate and of

[11] Simplified figure showing how artificial neural networks work. <https://cdn-images-1.medium.com/max/1600/1*22It4FL5aWXX6H9XRbqujg.jpeg > accessed 20 May 2021.

good quality. The algorithm must also interpret external data correctly. For this to happen, the algorithm must be trained on the information to seek. For example, the presence of fur on an animal, the presence of paws as opposed to hooves, et cetera. It must also learn from such data and be able to identify similar pictures correctly. Finally, it must use those learnings to accomplish distinct targets through adjustable transformation. Take for instance the dog, let it be assumed that the algorithm is part of an alarm system to alert the property owner of the trespass to the area by stray dogs. Once it correctly identifies a dog, it should be able to trigger the alarm bells to alert the relevant persons that something is amiss.

AI systems can be used in any industry. However, this paper focuses on the use of deep learning AI technologies in creating non-existent photos or video footage or tweaking original photos or videos in such a way that the difference between the original and the manipulated content is daunting to identify.

B. Creation of Deepfakes

The narrow definition of deepfakes constitutes the merging of the words 'deep learning' and 'fakes' to form 'deepfakes' (Deepfake.com, 2022). They are created by techniques that can superimpose the face of a target person onto the face of another such that the target person appears to be doing the thing that the other person is doing. The resulting image or video is a simple face swap (Nguyen *et al.*, 2021). In a deeper sense, they incorporate AI-synthesized content that may either be in the form of lip-syncing or puppet-mastering (Nguyen *et al.*, 2021, p. 3).

As the name suggests, the lip-sync types of AI involve the manipulation of one's lips to match what someone else says. The puppet-master type involves the use of videos of a target person (puppet) who is animated following the facial expressions, eye, and head movements of another person (master) sitting in front of a camera (Pernalete, 2021). Whether the mode employed is

lip-syncing or puppet mastery, the underlying algorithm uses deep learning methods such as Generative Adversarial Networks (GANs) to synthesize new images or videos based on massive training data (Goodfellow *et al*., 2020, p. 139).

Essentially, the process involves the training of the deepfake program or algorithm using two sets of data, in this case, the faces to be swapped (Albahar & Almalki, 2019, p. 3243-3244). This involves the use of ANNs as illustrated in figure 1. The first set consists of images to be replaced while the second set consists of images to replace the existing ones. The training process begins by passing the images through an encoder and then through a decoder to encode and decode the images (Albahar & Almalki, 2019).

The next step is the swapping step where the decoder is used to reconstruct the image instead of feeding it to the original decoder. This merges the features of one face with another (Albahar & Almalik, 2019). Afterwards, cropping, reshaping or any other mode of editing is done to ensure the swapped face looks as realistic as possible (Albahar & Almalik, 2019). This is when the GANs kick into play. They involve the discrimination of the outputs where the 'fake' outputs are marked as such and the 'real' ones are equally marked as real. In this way, the system checks itself and the output appears authentic.

C. The application of deepfakes

Deepfakes are an erosion of truth and most times, they lead to a situation where news clips cannot be verified. This is commonly referred to as an '*infocalypse*'. Indeed, books and other materials have been published on this and the general message is that in the present age, video footage can no longer be trusted as it is highly probable that it may have been manipulated (Schick, 2020).

In the same vein, on 26 November 2019, Witness,[12] in collaboration with the Centre for Human Rights, Faculty of Law, University of Pretoria, organized a day expert workshop on understanding deepfakes and other forms of Synthetic Media in Sub-Saharan Africa and published a report (Johnson & Faife, 2019, p. 3). The experts who participated in the workshop were from several countries across the continent including Ghana, Kenya, Nigeria, South Africa, Uganda, and Zimbabwe. The countries that represented outside the African continent included Italy and the United Kingdom (Johnson & Faife, 2019, p. 4).

In the report, part of the feedback received was that there was a need to lobby politicians to raise awareness of disinformation as a social problem to be tackled, to which resources must be allocated, and to continue to address existing problems with 'shallowfakes' *(cheapfakes)* - that is, mis-contextualized videos and lightly edited content (Johnson & Faife, 2019, p. 7).

Further, the report indicated that prior to the year 2019, the focus of discussions on deepfakes was centred on the trends in the Global North and America with little focus on the Global South (Johnson & Faife, 2019, p. 11-12). This implies that most of the documented examples of applications of deepfakes are drawn from the Global North. The workshop aimed to blaze a trail by including perspectives of the Global South on this technology (Johnson & Faife, 2019, p. 11-12).

The workshop focused on the negative uses of deepfakes. This can be deduced from the reaction concerning the positive

[12] 'WITNESS: See it, Film it, Change it,' is an organization that helps people use video and technology to protect and defend human rights. < https://www.witness.org/ > accessed 23 June 2021. It also works globally with about fifteen team members in the United States and twenty team members across all continents in Europe, Latin America, Southeast Asia, and Africa. WITNESS works with individuals and organizations in supporting the documentation of human rights violations and abuses. This includes providing video evidence for war crimes, police violence, land rights issues, et cetera. As the world has evolved, WITNESS now works with social media. With this increased volume of evidence comes the problem of an increase in manipulated media and non-accountable social media platforms. The goal of WITNESS is to listen carefully to identify critical challenges and problems in the video-as-evidence field, then advocate for better strategies and approaches to inform interventions to protect human rights and the integrity of trustworthy information.

use of deepfakes in the David Beckham video on the campaign 'malaria must die so that millions can live' where some participants affirmed that deepfakes can be used for good (Johnson & Faife, 2019, p. 17). Similar sentiments were raised regarding the use of deepfakes in film dubbing which makes videos and films more enjoyable. An example was given where the technology was employed such that celebrities from the Global North appeared to be singing songs in South-African local dialects (Johnson & Faife, 2019, p. 17).

Therefore, it follows that there is a silver lining to deepfakes, and they should not be fought with the crudest weapons in the armoury. Rather, a balance of what can be tolerated and what cannot be permitted should be struck. Indeed, it was an observation from some of the participants in the workshop that there is a need to rethink the balance between the positive and adverse effects of this technology (Johnson & Faife, 2019, p. 17).

The impact of such content on the population depends on what message is conveyed. Some content may be satirically curated to entertain, hence a positive use, while some are serious misinformation or disinformation popularly known as fake news, hence a negative application of the technology. Some of the negative impacts include mental or psychological disturbance to the victims. For instance, the digital card depicting Mr Polycarp Igathe, Nairobi's 2022 gubernatorial candidate, as uttering words with sexual undertones was a slur on his reputation. Nation Africa has since denounced the publication of the digital card (PesaCheck, 2022).

1. Positive applications of deepfakes

The positive application of deepfakes today can be seen in various sectors. This includes visual effects to simply modify the appearance of a video or photo for quality enhancement or other purposes; digital avatars commonly used as digital influencers, or digital celebrities (Lu, 2018). Snapchat filters are applied to

photos to create desired effects for instance animal ears or other 'artificial' features (Casey, 2020). Deepfakes also create voices of those who have lost theirs or aid in updating episodes of movies without reshooting them, et cetera (Nguyen *et al*., 2021). A few detailed examples are discussed below.

i). Entertainment and satirical purposes

The creation of deepfakes comes in handy in making satirical content. This is because it is easy to replace what a character said with something else which when analysed in the context of the original snippet, creates a perfect parody. Anyone who watched the 1968 movie, '2001 - A Space Odyssey' (Ebert, 1997) will be thrilled to see the 'trailer' for '2021: A SpaceX Odyssey' (Ctrl Shift Face, 2019). In the former, 'HAL 9000', a robot ejects one of the astronauts from the spaceship 'Discovery' and leaves him floating in space. The movie, though based on science fiction, is about a space mission that was too important to be jeopardised. The robot is heard telling the commander of the spacecraft that it read the *tweets* of Elon Musk. In the original movie, the robot reads the *lips* of the commander.

This is a good example of the use of copyrighted work for parody, caricature, or pastiche. In Kenya, such use falls under the general exceptions and limitations of the use of copyrighted works as provided under part A section 1 of the Second Schedule of the Copyright Act (2001). An interesting Kenyan example is where a photo was circulated on social media showing vehicles on what appeared to be a section of Nairobi's Thika superhighway. The vehicles were branded 'Wajackoyah the 5th'[13] in a bid to brand Wajackoyah as Kenya's fifth president (PesaCheck, 2022).

[13] The photo was satirical because of its association with Mr Wajackoyah. While some dismissed him because they thought he did not stand a chance, some felt that his ideas, expressed through his manifesto, were radical and may just have been what Kenya needed to get out of the political and economic quagmire it has been wallowing in for a long time. Referring to him as 'the 5th' meant he was set to be Kenya's 5th President. Based on his radical manifesto and the number of attendees to his rallies, the photo was quite satirical.

ii). Advertising industry

Using deepfakes, the words in a conversation can be synthetically replaced. It can also be used in redubbing advertisements and films into different languages (Albahar & Almalki, 2019). An example of this is the David Beckham, 'Malaria must die so that millions can live' campaign to end malaria where he appears to speak nine different languages (Global News, 2019). The possibility to apply lip-syncing techniques makes it possible to vary the languages. Such uses make it possible to convey messages without the exorbitant costs involved in getting all the people who speak all the different languages to be recorded. It also gives the video a good aura and as a result, it goes viral, and the message reaches as many people as possible.

iii). Education

One of the other positive uses of deepfakes is in the education sector. Learners at all levels of education tend to internalize the lessons more when they have visual demonstrations of the subject. This is particularly true for history lessons. Deepfakes can be used to recreate historical scenes and make them real and interactive. The costs and logistical expenses involved in developing the videos can be prohibitive. Deepfakes, therefore, are useful in this area.

An example of where this has been done is in the Illinois Holocaust Museum and Education Centre (Braunstein, 2018). At this museum, apart from the traditional display of artefacts, photos, and videos relating to the Holocaust, visitors also get a chance to have live interactions with holograms of survivors. They can ask questions and get responses. This is enabled by the machine and deep learning techniques employed by Apple's Siri[14] where the hologram picks keywords in the questions and

[14] Siri is a virtual assistant that is part of Apple Inc.'s iOS, iPadOS, watchOS, macOS, tvOS, and audioOS operating systems. It uses voice queries, gesture-based control, focus-tracking, and a natural-language user interface to answer questions, make recommendations, and perform actions by delegating requests to a set of In-

gives the relevant responses.

Africa is rich in undocumented history. Although some efforts have been made to recreate the stories through movies and documentaries, such stories are limited in their impact. The use of deepfakes to recreate historical events and figures gives learners a chance to see and hear some long-dead leaders speak and it is a gratifying feeling. Through deepfakes, it has been possible for the speech of former President John F. Kennedy to be recreated hence giving Americans and the entire world a chance to listen to the speech he was to give prior to his assassination, in his voice (Stenbuch, 2018). It would be great to have some little-known information about Africa brought to life using deepfakes. For instance, immortalising leaders such as Patrice Lumumba of the Democratic Republic of Congo (DRC) (Cordell, 2021) through deepfakes, where he is portrayed speaking about his ideologies, may bring to light a side of the DRC that learners have never visualised. It will make them understand the underpinnings of the current political and civil upheavals in the DRC since they are, at most, historical.

iv). Film and movie industry

'2021: A SpaceX Odyssey' is a *'trailer'* for a movie created purely using deepfakes. This means that it is possible to 'direct' and 'produce' full movies using deepfakes only that it may be more time-consuming and more expensive. Deepfakes have also been useful in the wake of the COVID-19 pandemic where large gatherings were discouraged making it impossible for actors and actresses to congregate and shoot episodes. This has left the option of using deepfakes to 'shoot' scenes that have never existed. However, it has been argued that this proposition still belongs in the world of science fiction as Hollywood and other movie industries are yet to embrace the idea of creating a whole movie

ternet services. With continued use, it adapts to users' individual language usages, searches, and preferences, returning individualized results. <https://www.apple.com/siri/ >

with the aid of deepfakes (Arnold, 2020). Presently, deepfakes can be employed in the creation of short snippets of parodies, for instance, '2021: A SpaceX Odyssey'. One of the reasons why the production of full movies using deepfakes will remain in the realm of short parodies like '2021: A SpaceX Odyssey' for some time is that the cost and technology involved in creation are quite prohibitive (Arnold, 2020).

2. Negative uses of deepfakes

On the flip side, deepfakes can be applied to the wrong uses. Since they are fakes, any reliance on the information conveyed by videos or photos is likely to have negative effects on those who rely on them. Unlike when deepfakes are used to recreate undocumented historical events, when used to create scenes that never happened in a way that wrong information is conveyed, the result is misinformation and disinformation which could have adverse effects on an entire population. Some of the negative uses of deepfakes include:

i). Political propaganda, hate speech, and other vices

It is the author's opinion that politicians and celebrities have often borne the brunt of this form of video or photo manipulation. This has been made possible by their easily available photos and videos that provide fodder for training data.[15] Since they are public figures, anything they say or do or appear to say or do is likely to be given attention and acted upon. For instance, if a video of a head of state who appears to be making statements concerning the extermination of their opponent is circulated, it is likely to be watched by a great percentage of the population. It is further likely to be believed by some members of that population. There have been instances where former US President, Barack Obama, was pictured giving a warning about deepfakes and how people

[15] See examples already given through the paper depicting the Kenyan President, Deputy President, and other political aspirants in incorrect terms.

should be wary of them (BuzzFeed Video, 2018). The clip follows from a video showing him speaking about his successor, former President Trump in not-so-kind words (BuzzFeedVideo, 2018).

In Africa, this technology was partly responsible for the attempted coup in Gabon on 7 January 2019 when President Ali Bongo's critics suspected that he was either dead or incapacitated and was therefore unfit to hold office (Washington Post, 2020). It all started with unconfirmed rumours of his ill health coupled with lack of information from any reliable Government sources. Then there was a video on 31 December 2018 where the President appeared to give his new years' address while some features on his body seemed awkward (Washington Post, 2020). His eyes appeared fixated in one direction, there were no wrinkles on his forehead and his right hand appeared puffed up and completely immobile during the whole address. This was the straw that broke the camel's back. His critics took this as a deepfake video only meant to hoodwink the citizenry of the fact that the President was still in control, yet they suspected he was dead, and someone else was in control. The video was subjected to digital media forensic examination through a deepfake algorithm, and the outcome was that it was not a deepfake but had been subjected to extensive video editing to the extent that it appeared unreal (Washington Post, 2020). This fact is in tandem with what was discussed in the first part of this paper that video or photo manipulation of any kind gives rise to subjective as opposed to objective outcomes. The President's supporters wanted the world to believe that he was fine and therefore invested in video editing to have what they wanted the world to see and not what the facts were. The result was a video that was too edited to be authentic hence the attempted coup.

Therefore, it is evident that deepfakes and other forms of manipulation can be abused to cause political tensions (Communications, 2020). They can also mislead the public concerning election campaigns as was seen during the campaigns in Meru County by the UDA party where a small crowd was depicted. The

image used was related to a different campaign event in Karatina, Nyeri County (PesaCheck, 2022). The effect of such manipulation is that it plays on the mental faculties of citizens, and it may result in chaos once election results are announced, and the candidate fails to garner as many or few votes as the citizens were made to anticipate.

ii). Doctoring court evidence

In reaching its finding, a court of law relies on both primary and secondary evidence presented to it by all parties in the case (Kenyan Evidence Act, 1963, s. 65, and 66). Where secondary evidence includes photographs or videos, a court may arrive at a wrong decision if the photo or video is not authentic. Juries have been reported to return a guilty verdict after watching a video in slow motion (Khaleeli, 2016). It is suggested that in such cases, the verdict would have been different had the video been played at its normal playback speed. A manipulated video that is slowed down gives a false impression that the accused person had a guilty mind; that they planned and premeditated the guilty act (Khaleeli, 2016).

The use of manipulated videos or photographs as evidence in court to conceal some facts and avoid incrimination has been witnessed in Africa. For instance, it is reported that the South African Marikana Commission of Inquiry into the massacre of striking miners on 16 August 2012 was subjected to manipulated evidence where they believed, among other things, that the miners staged an attack on the police (Marsden, 2013). The doctoring of the video evidence involved a reversal of the events in the footage probably using common video editing techniques (Marsden, 2013). Other forms of doctoring employed during the hearing included the concealment of information (Marsden, 2013).

Therefore, deepfakes are a threat to the justice system. When used in political mudslinging, or other vices, the ultimate determinant of the truth is the court. Judicial officers, including police, forensic experts, prosecutors, and other stakeholders

must be informed of the existence of the technology and how to detect it to avert the obvious consequence of a miscarriage of justice. It, therefore, behoves governments to investigate the possibility of regulating the use of deepfakes. The next part of this paper examines the implications of this technology vis-à-vis the existing laws.

IV. THE LEGISLATIVE RESPONSE

The challenge with technology regulation is that when laws are passed to regulate technology, sooner than later, the technology advances to a newer, better version—leaving the law behind. The effect is that several laws may be passed and become redundant over time as technology advances. This part, therefore, looks at the practicality and feasibility of enacting specific laws to govern deepfakes and whether it will help to curb the negative effects of the technology. It also looks at the existing laws and discusses whether they are adequate or whether there is a need to enact specific laws to regulate the technology. In Kenya the following laws apply to the application of deepfakes:

A. Constitution of Kenya 2010

1. Freedom of Expression

Article 33(1) of the Constitution of Kenya guarantees the right to freedom of expression. The article provides that the freedom of expression is a fundamental right of every person and includes the right to among others, seek, receive, or impart information or ideas. This implies that the creation of deepfakes to impart information or ideas is a guaranteed right under the constitution. This part particularly applies in situations where deepfakes are used for educational purposes to recreate undocumented information that is relevant to the present circumstances, for instance, the late president John F. Kennedy's recreated deepfake speech (Stenbuch, 2018). It is, however, a challenge

when the information or ideas imparted or received is potentially harmful by causing panic, unrest, or civil strife, or leading to unfavourable judgments by the court.

The other aspect of the guarantee of freedom of expression is artistic creativity. This means that an individual's artistic creativity is encouraged, guaranteed, and safeguarded. The implication is that regardless of the means through which artistic creativity is realised, the resulting artistic creativity is protected. Artistic creativity is part of intellectual property rights which are also safeguarded by the Constitution (a. 11, 40, and 69).

From the foregoing discourse, it has emerged that artistic creativity is one of the positive applications of deepfakes. For instance, deepfakes can be used in the movie industry or in the creation of world-class masterpieces of art which cannot be distinguished from traditional paintings made with paint and brush. An example is the next Rembrandt project where through AI, a team of engineers, artists, curators, and others recreated a painting in the Rembrandt style (The Next Rembrandt, 2016; Microsoft Reporter, 2016).

In Kenya, the use of deepfakes to express oneself freely is not an absolute right. Article 33(2) of Kenya's Constitution provides that the freedom of expression does not extend to propaganda for war; incitement to violence; hate speech; advocacy of hatred that constitutes ethnic incitement, vilification of others or incitement to cause harm. It also does not apply to content based on any ground of discrimination such as race, sex, pregnancy, marital status, health status, ethnic or social origin, colour, age, disability, religion, conscience, belief, culture, dress, language, or birth (Kenyan Constitution, a. 27(4)). Therefore, this implies that the use of deepfakes to create expressive works that fall under any of the foregoing categories is prohibited.

Article 33(3) provides that, 'in the exercise of the right to freedom of expression, every person shall respect the rights and reputation of others.' Under this sub-article, it can be concluded

that the use of deepfakes in videos or photos that are harmful to others' reputations or rights is outlawed. This borders more on defamation as a tortious liability and not a criminal offence. More details on the tortious liability vis-à-vis the criminal offence are stipulated in the section on the Kenyan Defamation Act (1970) and the Kenyan Penal Code (1930) respectively.

2. Freedom of the media

Freedom of the media is another guaranteed right under the Constitution. Article 34(1) provides that the freedom and independence of electronic, print and all other types of media is guaranteed. However, it goes ahead to limit the freedom by excluding any expressions specified in Article 33(2). Deepfake videos, photographs, and other content are often disseminated through social media such as Twitter, Facebook, Instagram, Tik Tok, YouTube, et cetera.[16] This means that the publication and dissemination of information created by deepfakes technologies, which falls under the categories named in Article 33(2), is prohibited.

B. Data Protection Act No. 24 of 2019

Under the Data Protection Act, 'data' is defined under section 2 as information that is processed using equipment operating automatically in response to instructions given for that purpose. From this, it can be concluded that photographs, videos, or

[16] Social media can be regarded as media since Article 34(1) of the Constitution refers to other types of media. Sub-article 3 talks of broadcasting (*mainstream media houses*) and other electronic media. The other electronic media could well mean social media as the same is not expressly excluded. Furthermore, the Kenya Information and Communications Act, No. 2 of 1998 mentions media in a broad sense. For instance, it defines media to mean *broadcast, electronic and other types of media but does not include print and book publishing*. This implies that broadcasts are one type of media and then there are other types of media which the author posits that social media is part of the same since it is not expressly excluded as the print and book publishing. In addition, mainstream broadcasting houses also share their content on social media for instance 'Facebook' and 'TikTok'. The fact that it is shared on social media does not absolve the media house from responsibility over the content just because social media platforms are not governed by Kenyan laws.

sound recordings form part of the data. It goes further to define 'processing' to mean any operation or sets of operations that are performed on personal data or sets of personal data whether or not by automated means, such as storage, adaptation, or alteration. Therefore, the creation of deepfakes is one of the ways in which data is processed. It can further be concluded that deepfakes are governed by the provisions of the Data Protection Act.

Section 25 provides the principles of data protection. One of the principles is that it must be processed in conformity with the privacy of the data subject. However, the term privacy is not defined in the Act. That notwithstanding, Article 31 of the Constitution provides for the right to privacy, which includes information relating to one's family or private affairs not being unnecessarily required or revealed. This leads to the question, when one alters or adapts a photo of a data subject, for instance by placing their head on a body of a different person, does the person adapting contravene the right to privacy of the data subject in question? The obvious response is that the privacy of the data subject is compromised since the resulting photo or video may end up going viral and a wide population gets to see it. This was demonstrated when criticism over the terms and conditions of the Chinese Zao App[17] was voiced by many users. The terms and conditions allowed the app developers to permanently use the images created on the app without any permission from the owners (Coleman, 2019). Deepfakes, therefore, infringe on the privacy of an individual and the effects can be dire as far as the emotional stability of the data subject is concerned (Chesney & Citron, 2019, p. 1775).

Looking at such contravention of the right to privacy from the public's perspective, when such images or videos go viral, the response from the public may be undesired depending on the content. Therefore, ensuring the processing of personal data while considering the privacy principle may deter some negative uses of deepfakes.

[17] Zao, https://zao.en.uptodown.com/android.

C. The Computer Misuse and Cybercrimes Act No. 5 of 2018

Deepfakes are also governed by the Computer Misuse and Cybercrimes Act. Under section 22(1) of the Act, it is a crime to intentionally publish false, misleading, or fictitious data with the intention that the same is relied on as authentic. The crime attracts a fine of not more than five million Kenya shillings or imprisonment for not more than two years or both. Therefore, the publication of deepfake content that is untrue, and which is believed by the public to be true could attract the penalty described.

Further, section 22(2) reiterates the limitations of the freedom of expression as contained in Article 33(2) of the Constitution which limits the sharing of information that is likely to propagate war or incite violence or negatively affect the rights or reputation of others. Therefore, the creation of deepfakes whose effect is to propagate war or incite violence is prohibited.

Additionally, under section 23, a person who knowingly publishes information that is false in print, broadcast, data, or over a computer system, that is calculated or results in panic, chaos, or violence among citizens of the Republic, or which is likely to discredit the reputation of a person commits an offence and shall on conviction, be liable to a fine not exceeding five million Kenya shillings or to imprisonment for a term not exceeding ten years, or to both. This means that the creation and publication of deepfakes that serve the purpose outlined above is a crime punishable as stated.

D. Defamation Act, Chapter 36 (Act No. 10 of 1970)

The Defamation Act is fashioned to provide civil remedies in cases, for instance, in libel. Regarding deepfakes, 'words' under section 2 of the Act include pictures, visual images, gestures, and other methods of signifying meaning. This implies that photos and other visual images created using deepfakes or other forms of manipulation are subject to the Act. Under section 7(1), for an

action of libel to stand, malice must be proved. Under sub-paragraph (3) the section further states that, 'nothing in this section shall be construed as protecting the publication of any matter the publication of which is prohibited by law, or of any matter which is not of public concern and the publication of which is not for the public benefit.'

The implication of this is that whether there is malice or not, any publication that is prohibited by law[18] does not fall under the qualified privilege envisioned by the Act. Hence, the publication of deepfake content that is outlawed, for instance, because it has the potential to damage the reputation of an individual, is not privileged and is actionable in the suit of the subject.

E. Copyright Act No. 12 of 2001

Copyright protects all creative works fixed in a tangible format, for instance, photos and videos, and sound recordings, including derivative works, among others (Copyright Act, No. 12, 2001, s. 22). Copyright is not subject to formalities like other Intellectual property (IP) rights for instance patents or trademarks. The protection is automatic upon fixation (Copyright Act, 2001, s. 22(3) (b)). One of the ways that copyright regulates deepfakes is by invoking the moral rights of copyright holders. Under section 32 of the Kenyan Copyright Act, an author of a copyrighted work has the right to object to the mutilation or adaptation of a copyrighted work in such a way that it is prejudicial to their honour or reputation. For instance, the use of the Chinese-developed app, Zao, which works by allowing users to photoshop themselves into their favourite movies or videos, could be an infringement of the moral right of integrity of the party who owns the copyright in a video or a movie (Copyright Act, 2001, s. 32).

Image rights can also be enforced in Kenyan courts where a person whose image was used in an advert was awarded dam-

[18] For instance, articles 33 and 34 of the Constitution as well as section 25 of the Data Protection Act and sections 22 to 24 of the Computer Misuse and Cybercrimes Act.

ages since their consent was not obtained prior to the use of the image. The case in point is *Jessicar Clarise Wanjiru v Davinci Aesthetics & Reconstruction Centre & 2 others [2017] eKLR* (Constitutional Petition No. 410 of 2016). The court stated that image or personality rights are generally considered to consist of two types of rights: the right of publicity, or to keep one's image and likeness from being commercially exploited without permission or contractual compensation, which is similar to the use of a trademark; and the right to privacy, or the right to be left alone and not have one's personality represented publicly without one's permission.

Kenyan courts take the position adopted by the US courts where the image rights jurisprudence was set in the 1953 landmark case of *Haelan Laboratories, Inc v Topps Chewing Gum Inc. ((1953) 202F).* In this case, the second circuit found that the privacy right was not adequate to protect celebrities against the use of their images in a way that is commercially beneficial to a third party without the consent of the celebrity in question. The right to publicity, therefore, protects against the commercial use of one's persona. It aims at protecting and safeguarding the 'publicity value and commercial magnetism inherent in the name, likeness or identity of a person' (Perot & Mostert, 2020, p. 5). Therefore, where deepfakes are used to create an individual's image and the image is subsequently used in an advert, such an individual is entitled to compensation if their consent was not sought.

F. Penal Code, Chapter 63 of 1930

Among the relevant provisions in the Penal Code is section 66(1), which criminalizes the dissemination of alarming publications and any false statement or other information that is likely to cause fear or disturb the peace. However, if the person who publishes the information can prove that they took the necessary steps to ascertain the accuracy of the information, they may be absolved from liability (Penal Code, s. 66(2)). This means that

the publication of deepfake content that is likely to cause panic or disturb the peace is prohibited.

The defence thereof is important as it puts some level of accountability on individuals to ascertain the truthfulness of a photo or video before circulating it. The result is that there will be less deepfake content being circulated as individuals are likely to question the accuracy of such before pressing the send button. However, there is a need to educate the public on such provisions of the law and the importance of adhering to them to avoid the negative consequences that may arise since ignorance of the law is not a defence.

G. Evidence Act, Chapter 80 (Act No. 46 of 1963)

Under the Evidence Act, section 78 provides that photographic evidence is admissible based on a certificate (Evidence Act, First Schedule) originating from the office of the Director of Public Prosecutions (DPP). However, it is not clear whether the DPP has the means of ascertaining that the photograph is genuine and not a deepfake. Subsection 3 provides that the maker of the certificate may be summoned for examination concerning the signature and for purposes of providing any further information on how the photograph was developed from the film. The certificate as well as the possibility to summon the maker of the photograph helps to authenticate the photographic evidence.

Digital and electronic evidence is also admissible under section 78A. However, there is a lacuna on the admissibility of digital or electronic evidence as the section pegs the probative value of such evidence on factors such as the reliability of the way the electronic and digital evidence was generated (Evidence Act, 1963, s. 78(A)(3)(a)). This means that where, for instance, the process of generating electronic evidence is not clear, such evidence will not be relied on. Where it is not possible to tell when a photograph was taken, the device used to take the photograph and other basic pieces of information, the law provides that such

evidence ought not to be relied on. In examining deepfakes, it is usually not possible to tell when the deepfake 'photo' was taken. As such, it should be disregarded. A deeper explanation of how to detect a deepfake photo or other content is discussed in the next part of this paper.

Without proper forensic expertise and tools, it may be difficult to ascertain, for instance, the way any electronic or digital evidence was generated. As was seen in the case of Gabon's President's video, due to the uncertainty surrounding the veracity of the video, it had to be subjected to digital media forensic examination through a deepfake algorithm to confirm that it was a genuine video. It is therefore important to invest in the proper training of forensic experts as well as requisite tools to help in ensuring that digital evidence that is tendered in court is of high probative value.

From the extensive magnification of the legal documents above, it is hence evident that Kenya has some regulatory frameworks around deepfakes. Ergo, there is no compelling reason to enact specific laws for the same seeing as technological advancements will occur and may render the legislated laws redundant. The next part makes recommendations on the pragmatic ways for detecting deepfakes to curb misinformation since this is where the main challenge with deepfakes lies.

V. RECOMMENDATIONS AND CONCLUSION

A. Recommendations: Detection of deepfakes

Over-regulating AI applications may stifle creativity and it is, therefore, better to give it time as other ways of adhering to the proper uptake of this technology are investigated. The Organization for Economic Cooperation and Development (OECD) Network of Experts is discussing the possibility of regulating this technology. This paper, therefore, recommends the adoption of the OECD AI principles in regulating the technology. This is

indeed a good place to start as the trajectory of the technology is observed and since Kenya already has considerable laws on the application of deepfakes. Specifically, the aim is to boost expertise in the detection of deepfakes and sensitize the public on the same.

One of the OECD AI Principles is that there should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them. This means that training and awareness creation on the deepfakes technology can also help in curbing its negative effects.

Witness Media Lab made some recommendations on how to curb the negative effects of deepfakes and proposed the training of journalists, forensic experts, citizens, and other stakeholders (Witness Media Lab, 2019, p. 12). In line with Witness Media Lab's recommendation and proposal, this author proposes that citizens should be trained, through social media campaigns on the basic tell-tale signs of deepfakes.

The Twitter policy of doctored media is one such example of social media campaigns. During the workshop held in South Africa organized by Witness in collaboration with the Centre for Human Rights, Faculty of Law, the University of Pretoria, on understanding deepfakes and other forms of Synthetic Media in Sub-Saharan Africa, (Johnson & Faife, 2019, p. 40) participants were taken through the proposed Twitter policies on deepfakes (Harvey, 2019). Some of the input required from participants was how they would like Twitter to treat manipulated videos or photos. The answers included: -

i). Placing a notice next to Tweets that share synthetic or manipulated media.

ii). Warning people before they share or like Tweets with synthetic or manipulated media.

iii). Adding a link – for example, to a news article or Twitter Moment – so that people can read more about why

various sources believe the media is synthetic or manipulated.

This is a great initiative by Twitter, and it is recommended as one of the ways to help identify manipulated media. Placing a notice next to the Tweets warning the public that the content is manipulated, may help further the principle of responsible AI. It may also serve to warn the public not to rely on it and act in any way that may be negative. This goes hand in hand with the first point because once there is a notice, it serves as a warning to any further action by anyone else who consumes the manipulated content. Therefore, it helps to curb instances of chaotic outcomes and reduces the chances of the manipulated content going viral. Other social media platforms could employ similar policies.

The third point of adding a link so that a viewer of the tweet can read what others say about it helps in stirring critical thinking in people. It helps to keep people alert and on the lookout for other tell-tale signs. This reduces the negative impact of jumping to conclusions by relying on fake content to cause chaos or to judge individuals whose image has been tainted wrongly. It also helps judicial officers to note that not all electronic or digital evidence is authentic. Generally, exposing people to others' thoughts about whether content is fake or genuine makes them understand that not all content is genuine.

Furthermore, there have been global discussions that deepfakes can be detected and probably regulated through blockchain technologies (Video, 2018). Blockchain operates on what is commonly known as 'immutable life logs' (Video, 2018). This involves the cryptographic signing of photographs and videos at the source and the creation of smart contracts (Video, 2018). The cryptography assigns a hash to the photo or video which cannot be altered or modified. Every instance of modification will require an agreement from all the parties who cryptographically signed it. This creates an audit trail, and it is therefore possible to detect any tampering with any photo or video (Martinez, 2018).

The other proposed method of detecting deepfakes is the use of a technique known as Photo Response Non-Uniformity (PRNU) analysis (Nguyen et al, 2021, p.10). The PRNU pattern of a digital image is a noise pattern created by small factory imperfections in the light-sensitive sensors known as silicon wafers in a digital camera (Nguyen et al, 2021, p.10). The noise pattern is caused by factory imperfections and the wafers' inconsistent sensitivity to light pixels. When a photo is taken, the sensor imperfection is reflected in the high-frequency bands of the content as invisible noise (Nguyen et al, 2021, p.10). Because the imperfections in the silicon wafers are not uniform, even the sensors reflected in the high-frequency bands are not uniform (Nguyen et al, 2021). This implies that each photo has a unique noise. This noise is considered the digital fingerprint left in the images by cameras. In analysing deepfakes using this method, a deepfake will have a different 'noise' from a genuine photo or video. There will be a significant statistical difference in terms of mean normalized cross-correlation scores between deepfakes and genuine content (Nguyen et al, 2021). It is, therefore, possible to use this method in detecting deepfakes, but the downside is that large sets of data will be required.

Further, it is proposed that deepfakes can be detected by analysing the upper body language of victims in comparison to the deepfake (Yasrab *et al*., 2021, p. 304). The proponents of this method of detection have put forward a hypothesis that 'the body language is distinct for different individuals and can be used to expose deepfakes by Recurrent Neural Networks (RNNs)' (Yasrab, *et al*., 2021, p. 311). It involves the data pre-processing method which extracts twelve key body points that represent the upper body pose. Subsequently, a Long-Short Term Memory (LSTM) model is designed and trained to analyse the upper body language. After several experiments, the LSTM model has proved the proposed hypothesis to be highly accurate (Yasrab, *et al*., 2021, p. 311). It is, therefore, one of the methods by which forensic experts can be trained to detect deepfakes.

More recently, Convolutional Neural Networks (CNN)-based approaches that decompose videos into frames have been deployed to extract salient and discriminative visual features associated with deepfakes automatically (Hussain, *et al*., 2021, p. 3348). Other efforts include the segmentation of the entire input image to detect facial tampering resulting from face swapping, face morphing, and other forms of manipulation (Hussain *et al*., 2021, p. 3349). It has also been reported that eye blinking is usually not well reproduced in fake videos and there is, therefore, a proposal of using a combination of CNN and RNN to detect the lack of eye blinking when attempting to flash out deepfakes (Hussain, *et al*., 2021, p. 3349). Any inconsistencies in head poses can also be detected using this method (Hussain *et al*., 2021 p. 3349).

Apart from Twitter's Policies on Deepfakes, the other methods of detecting deepfakes are more technical and may be difficult for the common citizens to understand. The equipment and tools needed are also expensive and not easily available to everyone. As such, it is recommended that the Government of Kenya considers investing in forensic labs and further training of forensic experts to equip them with the necessary skills and expertise. Such will come in handy in instances where content must be examined to make further decisions as was in the case of the Gabon President's video. It would also be important to have such experts in courts of law where the evidence tendered (photo or video evidence) is questionable as to its authenticity. Forensic experts should also be proactive enough to be on the lookout for deepfakes as opposed to waiting until an incident is reported. Where potentially inciteful content is published, it should be immediately pulled down and investigations commenced as to its source.

B. Conclusion

Photo, video, text, or other manipulation of information has been used by mankind since time immemorial. Such manipulation has evolved from simple cartoon depictions with wrong or misleading captions to photo or video edits to deepfake content that seems to be authentic. Such doctored content has both positive and negative applications. The positive applications include use in the entertainment industry, advertising industry, and education among others. The negative applications include use in hate speech, incitement to violence, attacking one's reputation and political mudslinging. Such hate speech has the potential of further negative effects for instance causing chaos and civil unrest. Other negative applications include the doctoring of court evidence leading to wrong verdicts by judges and judicial officers.

As such, deepfake technology has introduced several challenges. Some of the challenges brought about by the deepfake technology include the difficulty in regulation since it advances quite fast making it almost impossible to pin it down. One aspect of the technology could be regulated today by the enactment of laws. However, soon after, the technology may advance to a newer version not covered by the law thus leaving a redundant law behind.

The other challenge is the difficulty involved in detecting deepfake content so as not to rely on it for instance as a judicial officer while adjudicating a case. The dearth of forensic experts to offer guidance and direction in this area is a major technical challenge. It has emerged that most of the technical modes of detecting deepfake content are advanced and require well-equipped laboratories and highly trained personnel to use them. This problem can be solved by the Government of Kenya making such investments and training forensic experts.

From the foregoing, the author posits that this technology should not be disregarded in totality. The existing laws are sufficient to offer some regulation of the technology. There is no need

to enact a specific law as the challenges posed by that approach have already been elucidated. The Data Protection Act No. 24 of 2019, the Penal Code (Chapter 63), the Evidence Act (Chapter 80), and the Copyright Act (No. 12 of 2001), among others offer a regulatory platform for the technology. The positive applications should be encouraged while the negative ones should be managed through existing laws and the employment of other techniques to curb their negative effects and impacts. Social media campaigns should be adopted by all social media platforms to reach out to all citizens about the existence of the technology. Watermarking of content that is doctored could also be adopted by social media.

REFERENCES

Agreement on Trade-Related Aspects of Intellectual Property Rights (Marrakesh) April 15, 1994 (as amended on January 23, 2017) https://www.wto.org/english/docs_e/legal_e/31bis_trips_e.pdf (entered in force January 1, 1995) (hereinafter TRIPS Agreement).

Albahar, M., & Almalki, J. (2019). Deepfakes: Threats and Countermeasures Systematic Review, *Journal of Theoretical and Applied Information Technology, 97:22 3242* www.jatit.org

Arnold, (2020, November 16). What are Cheapfakes? are they different from Deepfakes? *Deepfake Now,* https://deepfakenow.com/what-are-cheapfakes-meaning/

Arnold, (2020, October 1). Why Doesn't Hollywood use Deepfakes in their movies? *Deepfake Now,* https://deepfakenow.com/hollywood-deepfake-movies/

Biro, M, (2012). From Analogue to Digital Photography: Bernd and Hilla Becher and Andreas Gursky, History of Photography. https://doi.org/10.1080/03087298.2012.686242

Braunstein, E, (2018, January 23). At this Holocaust Museum, you can speak with holograms of survivors. *Holocaust Testimony*. https://www.timesofisrael.com/at-this-holocaust-museum-you-can-speak-with-holograms-of-survivors/

Britannica, T. Editors of Encyclopaedia (2022, August 17). *Marie-Antoinette. Encyclopedia Britannica.* https://www.britannica.com/biography/Marie-Antoinette-queen-of-France

Burkell J & Gosse C. (2019). Nothing New Here: Emphasizing the social and cultural context of Deepfakes. *Peer Reviewed Journal on the Internet* https://journals.uic.edu/ojs/index.php/fm/article/download/10287/8297?inline=1#author

BuzzFeedVideo (2018). 'You won't believe what Obama says in this video', YouTube https://www.youtube.com/watch?v=cQ54GDm1eL0

Chesney, R., & Citron, D. K. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *107 California Law Review 1753.* http://dx.doi.org/10.2139/ssrn.3213954

Coleman, A. (2019, September 4) 'Deepfake' app causes fraud and privacy fears in China,' *BBC Monitoring, 4 September 2019* https://www.bbc.com/news/technology-49570418

Communications. (2020, January 6). Elections in Africa: AI generated deepfakes could be the greatest digital threat in 2020, *Paradigm Initiative*, 6 January 2020 https://paradigmhq.org/deepfakes/

Computer Misuse and Cybercrimes Act, No. 5 of 2018

Constitution of Kenya, 2010

Copyright Act, No. 12 of 2001

Cordell, D.D. (2021, June 28). 'Patrice Lumumba' *Britannica, 28 June 2021* https://www.britannica.com/biography/Patrice-Lumumba

Ctrl Shift Face (2019) '2021: A SpaceX Odyssey', YouTube https://www.youtube.com/watch?v=sTksmoTdT4Q

Data Protection Act No. 24 of 2019

Deepfake.com, (2022). https://deepfake.com/knowledge-center/what-is-a-deepfake/

Deepfakes/Faceswap, Python, (2019). https://github.com/deepfakes/faceswap

Defamation Act, Chapter 36

Dupuy, B. (2019, May 24), Not Real News: Altered Video makes Pelosi Seem to Slur Words. *AP News, May 24, 2019,* https://apnews.com/article/social-media-donald-trump-nancy-pelosi-ap-top-news-not-real-news-4841d0ebcc704524a38b1c8e213764d0

Dupuy, B., & Ortutay, B. (2019, July 19). Deepfakes Pose a Threat but 'Dumbfakes' may be worse *AP News, July 19, 2019*, , https://apnews.com/article/media-nancy-pelosi-ap-top-news-politics-technology-e810e38894bf4686ad9d0839b6cef93d

Ebert, R (1997). 2001 - A Space Odyssey https://www.rogerebert.com/reviews/great-movie-2001-a-space-odyssey-1968

Evidence Act, Chapter 80

FakeApp FakeApp 2.2.0 - Download for PC Free (malavida.com)

Fakers. app https://play.google.com/store/apps/details?id=fakers.app&hl=en_US&gl=US

Global News (2019) 'David Beckham 'speaks' nine languages in call to end malaria' YouTube https://www.youtube.com/watch?v=U-mg7a1vwkw

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020), Generative Adversarial Networks, *Communications of the ACM, Volume 63 Issue 11 November 2020 pp 139–144* https://doi.org/10.1145/3422622

Haelan Laboratories, Inc v Topps Chewing Gum Inc. ((1953) 202F) (1953) 202 F. 2D 866 cert. denied 346 US 816, 98L. Ed. 343, 74 S. Ct. 26 (2nd Cir).

Harvey, D. (2019). Help us shape our approach to synthetic and manipulated media. https://blog.twitter.com/en_us/topics/company/2019/synthetic_manipulated_media_policy_feedback

Hedrick, S. F, (2019). I 'Think,' Therefore I Create: Claiming Copyright in the Outputs of Algorithms', *N.Y.U. Journal of Intell. Prop. & Ent. Law 8:2 324-374.* https://jipel.law.nyu.edu/vol-8-no-2-1-hedrick/

History.com Editors (2019). John C. Calhoun https://www.history.com/topics/us-politics/john-c-calhoun

History.com Editors (2021). 'Emancipation Proclamation' https://www.history.com/topics/american-civil-war/emancipation-proclamation

Hussain, S., Neekhara, P., & Koushanfar, F. (2021). Adversarial Deepfakes: Evaluating Vulnerability of Deepfake Detectors to Adversarial Examples. *IEEE Xplore, 3348*

Iconic Photos (2010) Lincoln-Calhoun Composite. https://iconicphotos.wordpress.com/2010/04/24/lincoln-calhoun-composite/

Industrial Property Act, No. 3 of 2001.

Johnson, A., & Faife, C. (2019). Report of a One-Day Expert Workshop on Understanding Deepfakes and Other Forms of Synthetic Media in Sub-Saharan Africa. https://blog.witness.org/wp-content/uploads/2020/02/WITNESS-South-Africa-Deepfakes-Workshop-Report.pdf

Käde, L., & von Maltzan, S. (2019). Towards a Demystification of the Black Box—Explainable AI and Legal Ramifications. *Journal of Internet Law, 4-6.*

Kagwanja, P. (2009). Courting Genocide: Populism, Ethno-Nationalism, and the Informalisation of Violence in Kenya's 2008 Post-Election Crisis. *27:3 Journal of Contemporary African Studies, 27:3 365-387.* http://dx.doi.org/10.1080/02589000903187024

Kenya Information and Communications Act, No. 2 of 1998

Khaleeli, H. (2016, August 2). How slow-motion video footage misleads juries. *The Guardian,* https://www.theguardian.com/law/shortcuts/2016/aug/02/how-slow-motion-video-footage-misleads-juries

Kukreja, H., Bharath, N., Siddesh, C., & and Kuldeep, S. (2016), An Introduction to Artificial Neural Network *KuVol-1 Issue-5 IJARIIE-ISSN (O)-2395-4396 C-1399 www.ijariie.com 27*

Lu, J. (2018, July 20). From Sci Fi to Commercialization: The Rise of Digital Avatars. *LinkedIn,* https://www.linkedin.com/pulse/from-sci-fi-commercialization-rise-digitalavatars-jerry-lu/

Marsden, C. (2013, September 28). South Africa's police concealed evidence, lied to cover up Marikana massacre, *World Socialist Web Site,* https://www.wsws.org/en/articles/2013/09/28/mari-s28.html

Martinez, A.G. (2018, March 26). The Blockchain Solution to Our Deepfake Problems. *Wired.,* https://www.wired.com/story/the-blockchain-solution-to-our-deepfake-problems/

Melendez, S. (2018, April 4). 'How DARPA's Fighting Deepfakes', *Fast Company,* https://www.fastcompany.com/40551971/how-darpa-is-fighting-deepfakes

Microsoft Reporter, (2016, April 13). The Next Rembrandt, Blurring the Line between Art, Technology and Emotion. *Microsoft,* https://news.microsoft.com/europe/features/next-rembrandt/

Ndemo, B (2021). [blogpost] https://bitangendemo.io.ke/

Ndemo, B. (2021, March 25). Kenya Needs Artificial Intelligence Law. *Business Daily,* https://www.businessdailyafrica.com/bd/opinion-analysis/colum-

nists/kenya-needs-artificial-intelligence-law-3335026

Nguyen, T. T., Nahavandi, S., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Thanh Tam Nguyen, T. T., Pham, Q-V., & Nguyen, C. M. (2021). Deep Learning for Deepfakes Creation and Detection: A Survey. *IEEE,* http://export.arxiv.org/pdf/1909.11573

OECD AI Principles, https://www.oecd.org/going-digital/ai/principles/#:~:text=The%20OECD%20Principles%20on%20Artificial%20Intelligence%20promote%20artificial,approved%20the%20OECD%20Council%20Recommendation%20on%20Artificial%20Intelligence

Omdena (2021). [blogpost] 'AI meets Art: Can Creativity be Replicated?' https://omdena.com/blog/ai-art/

Paris Convention for the Protection of Industrial Property (Paris) March 20, 1883 (as amended on 28 September 1979) https://wipolex.wipo.int/en/treaties/textdetails/12633 (entered into force June 3, 1984) (hereinafter Paris Convention).

Paris, B & Donovan, J (2019, September 18). Deepfakes and Cheapfakes, the Manipulation of Audio and Visual Evidence. *Data & Society's Media Manipulation Research Initiative,* https://datasociety.net/library/deepfakes-and-cheap-fakes/

Penal Code, Chapter 63

Pernalete, J., (2021, May 31) 'Lip Sync and Deepfakes: A novelty or a great tool for the film industry?' *Budapest Reporter,* https://www.budapestreporter.com/lip-sync-and-deepfakes-a-novelty-or-a-great-tool-for-the-film-industry/

Perot, E., & Mostert, F. (2020). Fake it Till You Make it: An Examination of the US and English Approaches to Persona Protection as Applied to Deepfakes on Social Media. *Journal of Intellectual Property Law & Practice, Vol. 15, No. 1 32.*

PesaCheck https://pesacheck.org/

PesaCheck, (2022) https://pesacheck.org/altered-this-image-claiming-to-show-former-prime-minister-raila-odinga-napping-during-a-us-tour-is-bc-c9aedb87b0

PesaCheck, (2022) https://pesacheck.org/altered-this-image-of-vehicles-on-a-highway-branded-wajackoyah-the-5th-is-altered-cfe96da63331.

PesaCheck, (2022) https://pesacheck.org/altered-this-image-showing-kenyas-dp-ruto-grinning-while-paying-last-respects-to-the-late-mwai-e35a2396fffc

PesaCheck, (2022) https://pesacheck.org/fake-this-digital-card-attributed-to-nation-africa-quoting-nairobi-gubernatorial-candidate-db-c67c53e33

Reface https://play.google.com/store/apps/details?id=video.reface.app&hl=en_US&gl=US

Schick, N. (2020). Deep Fakes and the Infocalypse: What You Urgently Need to Know*, Monoray, Kindle Edition, 2020.*

Seigchrist, G. (2021, May 7). 6 Best Free Video Editing Software Programs for 2021. *Livewire,* https://www.lifewire.com/best-free-video-editing-software-programs-4128924

Shemtov, N. (2018). Intellectual Property and Mobile Applications. *WIPO,* https://www.wipo.int/export/sites/www/ip-development/en/agenda/pdf/scoping_study_mobile_apps.pdf

Simplified figure showing how artificial neural networks work. https://cdn-images-1.medium.com/max/1600/1*22It4FL5aWXX6H9XRbqujg.jpeg

Stenbuch, Y. (2018). Listen to JFK speak from beyond the grave. https://nypost.com/2018/03/16/jfks-voice-delivers-speech-he-never-gave-day-of-assassination/

Tate (2021) 'Andreas Gursky', https://www.tate.org.uk/art/artists/andreas-gursky-2349

Tate (2021) 'Who are Bernd and Hillda Becher', https://www.tate.org.uk/art/artists/bernd-becher-and-hilla-becher-718/who-are-bechers

Tate (2021). 'Andreas Gursky, The Rhine II, 1999' https://www.tate.org.uk/art/artworks/gursky-the-rhine-ii-p78372

The Next Rembrandt, a website, https://www.nextrembrandt.com/

TikTok, Free Video Creation and Sharing App https://tik-tok.en.softonic.com/

Tutorial Example, (2021, July 4). Who invented the first digital Camera? (2021). *Tutorial Example* https://www.tutorialandexample.com/who-invented-first-digital-camera

Video, A. (2018, September 19). 'How blockchains can be used in authenticating video and countering deepfakes,' *Medium,* https://medium.com/amber-video/how-blockchains-can-be-used-in-authenticating-video-and-countering-deepfakes-25d596ad7a5

Washington Post, (2020). The suspicious video that helped spark an attempted coup in Gabon | The Fact Checker. YouTube https://www.youtube.com/watch?v=F5vzKs4z1dc

Witness Media Lab, (2019). Deepfakes: Prepare Now (Perspectives from Brazil), https://lab.witness.org/wp-content/uploads/sites/29/2019/10/WITNESS-Deepfakes-Brazil-Prepare-Now-Updated.pdf

Witness, (2019). Deepfakes: Prepare Now Workshop Report (English) Brazil. https://lab.witness.org/wp-content/uploads/sites/29/2019/10/WITNESS-Deepfakes-Brazil-Prepare-Now-Updated.pdf

Yasrab, R., Jiang, W., & Yasrab, A. R., (2021) Fighting Deepfakes Using Body Language Analysis. *Forecasting 3, 303–321* https://doi.org/10.3390/forecast3020020

Zao, https://zao.en.uptodown.com/android