Research Workx

Journal of Computational Innovation

https://researchworkx.com/index.php/jci

Vo1 4 (1) 2024

Deep Fake Detection and Mitigation: Securing Against AI-Generated Manipulation

Anwar Mohammed[1,2]

[1] Rakbank, National Bank of U.A.E

[2] Singhania University, India

Corresponding Email: anwarmohammed567@outlook.com

Abstract:

Deep fake technology has been advanced in the past decade and this advancement in deep fake technology has brought major problems for many organizations and individuals. Machine learning and neural network has been the main source through which deep fakes are being generated and Generative Adversarial Network (GAN) has been the most prominent neural network for generation of deep fake imaged. This study has analyzed and discussed the role of machine learning in creation of deep fakes including a thorough discussion about different types of deep fakes that are being used. This paper also discusses the history of deep fakes and when it gained attention globally. Different literatures by researchers have also been reviewed about deep fake that are included in this study. As deep fakes have imposed many risks globally, these risks along with social and political risks are included in the study. The development of deep fake detection and mitigation technology and the ways it has helped individuals in securing against deep fakes are included in this paper. Furthermore, this paper looks into the challenges that are still present from deep fakes, their impacts, and the ways these challenges can be overcome. In conclusion this paper suggests advancement in deep fake detection technologies, introduction of policies, and awareness programs for the use of deep fakes and its detection.

Keywords: *Deep fake, Artificial intelligence (AI), Machine learning, Neural networks, mitigation, manipulation.*

Introduction:

Deep fake also known as synthetic media is typically a video or audio generated and manipulated through Artificial intelligence (AI) through learning algorithms. Deep fake videos are much harder to identify as being fake as there has been great focus given to detailing these videos and audio [1]. The AI algorithms are very advanced as they can manipulate the facial expression in the videos with changes in the environment and the audio tones also change manipulating speech and making the audio more realistic [2]. These deep fake videos and audio are entirely fabricated and nowhere near reality, the main purpose of these are to spread misinformation and deceive viewers into believing that what they are watching and listening to is authentic [3].

Reputable organisations and researchers are working tirelessly on an enhancement in deep fake detection and mitigation as it has been a great concern for cybersecurity globally by targeting the reputation of the government and private organisations and also respected individuals [4]. The misuse of AI and its easy availability to people has initiated many cyberwarfare too as deep fake videos and content get viral in some seconds causing difficulty for individuals to prevent such situations. At this time the availability of deep fake detection has been a relief for people who are constantly in fear of losing their reputation through misinformation [5]. Although deep fake detection has been introduced there are still people unaware of its existence and there are people who are unaware of deep fake information and video generation so they often get fooled by the data they see on the internet without checking its reliability [6].

This research aims to write about the use of deep fake detection and mitigation which can help in securing against AI-generated manipulation. AI has been advanced all around the world and it has proved to be helpful for many people in many sectors but as the advantages have increased the increased misuse of AI has also started which is creating a lot of problems [7]. From solving daily problems and simplifying everyday life AI has done all but there has also been a rise in increased deep fake videos and pictures which are often used for fake news generation and manipulation. Deep fake detection and mitigation are of high importance today as these tools help in detecting AI-generated deep fake videos which are most often used for malicious purposes [8]. The detection of fake videos helps organisations and individuals to take proper action against these videos which can eventually help in stopping the spread of fake information. The increased fraud, blackmailing, and cyberbullying are most often done through deep fake AI-generated videos to harass individuals and organizations and damage their reputations in society [9].

Deep fake detection and mitigation can be used in many sectors for data validation including healthcare sectors, financial sectors, and other legal sectors where data accuracy is of great importance deep fake detection tools can help in detecting authentic data and it also helps in identifying the authenticity of videos, audio recordings to ensure that any decision or action can be taken on reliable and verified information [10]. Many individuals have tirelessly worked on various deep fake detection tools to minimize the risks imposed by AI-generated deep fake content. Deep fake detecting technology is not only limited for the detection of deep fake videos but it can also help in detecting deep fake pictures, audios, and documents [11]. There are many deep fake detecting tools being developed everyday some of which include intel’s real-time deep fake detector, sentinel, Microsoft’s authenticator tool for videos, and We Verify [12].

The main purpose of this study is to highlight the need of bringing advancements in deep fake detection tools to be much better in detecting and mitigating deep fakes. This research will also deeply analyze the technologies of deep fake detection and their major importance in securing data maintaining the reputation of organizations and individuals. The study also highlights the issue of fake videos and news article which are usually used for defamation purposes [13]. Deep fake technologies and tools have been of popular use since they have been developed but deep

fake detection tools are also getting popularity now and their access to users allows them in identifying deep fake content and protect their data [14]. Deep fake content is utilized for many reasons but apart from entertainment the negative intentions being employed by people include data theft, cyberbullying, fake pornographic videos, monetary frauds, and blackmailing [15].

Aims and objectives:

Aim:

The main research question of this study is to unveil the problematic aspect of deep fakes' application and to describe how this technology causes social and political instabilities. This study also concerns advances that are being made in the field of deep fake detecting technologies and tools in data prevention.

Objectives:

- This study's primary goal is to examine how crucial deep fake detection and mitigation are to having a safe platform against manipulation by artificial intelligence
- The research aims to understand the importance of deep fake detection and mitigation and the ways these tools can help protect organizations and individuals from privacy breaches and misinformation.

Research questions:

- How is deep fake detection and mitigation important against AI-generated manipulation?
- How can deep fake detection and mitigation help protect organizations and individuals from privacy breaches and misinformation?

Background and Related Work:

Historical context of deep fake technology and its evolution:

The term 'deep fake' was first coined in 2017 when a Reddit moderator used face swap technologies on the faces of celebrities by swapping those faces in pornographic videos and he created a 'Subreddit' for users to share such videos [16]. The subreddit forum has long been deleted but the deep fake trend went through great evolution and AI is still being used for the generation of deep fake. Besides the term 'deep fake' being coined in 2017 the use of deep fake can be traced back to the 1990s when many researchers contributed to creating realistic human images through CGI [17]. This technology gained major attraction in 2010 as largely available datasets and major developments in machine learning led to more advancement in the field. The year 2014 became a no-return year for deep fakes as a new development in machine learning by Lan Goodfellow and his team introduced Generative Adversarial Network (GAN), this technology in machine learning brought major advancement in enabling people to create detailed and sophisticated deep fake images and videos [18]. Deep fake advancement was not only the contribution of great researchers but the average internet users also contributed to it, the openly available deep fake tools were used by people to generate images and videos for everyday entertainment purposes.

Current landscape:

Deep fake is mostly created through face swap tools and voice synthesis technology, these tools and technologies are much advanced as they enable users to give precise attention to detailing like motion in videos, change in face expression, background, and detailing to voice tones [19]. Since the use of deep fake emerged it has become a threat for cybersecurity as it targets people and organizations by creating fake news and spreading misinformation. Deep fakes mostly target social media platforms as people majorly tend to believe anything that they see on the internet [20]. In social media platforms any news takes seconds to spread globally as these platforms are used all over the world and the ongoing 'infopocalypse' has created a belief in people that the news they hear from friends or family cannot be true till they see evidence supporting the news on social media [21]. People majorly used deep fakes in the past just for entertainment purposes like creating memes, making funny videos, and swapping people's images in funny videos. However, the major advancement of AI in generating highly sophisticated and detailed deep fake videos has created a lot of complications [22]. People use deep fake technology globally to damage the reputation of individuals, creating cyberwarfare, and causing political unrest by spreading misinformation.

Literature review:

Deep fake has been the most recent development in the AI industry which allows users to create deep fake videos or photos by swapping the faces of individuals with other faces and this has created major problems globally. Although deep fake is mostly used in video generation it is not only limited to that but is also used in audio recordings and other important documentation [23]. This misuse of deep fake has ended up resulting in damaging the reputation of reputable organizations and individuals. Many people indulge in such activities to harm the reputation of rivals by spreading misinformation about them and such controversial misinformation gets viral quickly which causes a huge problem. Research has been done to overcome this issue and to secure people from AI-generated manipulation deep fake detection and mitigation tools have been developed most recently [24]. These tools have been proven to be a blessing for organizations and people who always have the urge to prove their nobility in society by disregarding any misinformation spread against them. The purpose of these tools is to analyze the deep fake content thoroughly and to detect the fake features included in the content, this helps in identifying the reliability of the content being generated and shared [25].

Studies have shown that the rapid progress in AI tools and machine learning has been the main reason for the use of new tools and various techniques that are being used for generating misinformation and deep fake content for manipulating multimedia. The main purpose of advancement in AI was not for it to be misused but it was most importantly introduced and enhanced for entertainment and educational purposes and it has proven to outdo itself in both these fields [26]. The misuse of AI has been used in generating fake videos, pictures, and recordings to particularly damage the reputation of politicians, organizations, and individuals which has caused political unrest in states, blackmailing of reputable organizations and

individuals. As the need to address such threats gained attention more people got their attention on studying and generating tools that can help in avoiding such problems and checking the reliability of content [27]. Researchers and technicians started developing tools for deep fake detection and mitigation, these tools have proved to be a blessing as they detect deep fake content and check the reliability and validity of the information being spread [28]. Deep fake detection and mitigation have been a source of identifying misinformed data and deep fake content as they work as a security measure against AI-generated manipulation.

As the issue of deep fakes was on the rise a study was conducted by [29], in this study a new technique for detecting deep fakes was introduced which was named The Deep Fake Detection Challenge (DFDC) Dataset, this technique was introduced to counter the most basic threat from AI manipulation which is of face swapping of individuals in videos. The technique was used on a large number of face swap video datasets to test the detection models and enable their training this was accompanied by the deep fake detection challenge dataset (DFDC) Kaggle competition. The main purpose of this model was to create much better models for the detection of manipulated data [30]. As this data is available for use outside too it enables individuals to work and bring more advancement in developing deep fake detection and mitigation tools. More techniques are being introduced day by day to bring more development in these tools to make life easier for individuals and to stop the spread of misinformation and the negative use of deep fake [31]. Securing against AI-generated manipulation is of great concern and undoubtedly the Detection challenge dataset (DFDC) has provided a standardized benchmark for researchers and developers to bring more advancement in bringing new innovative techniques and tools for deep fake detection and mitigation.

Deep fake detectors have resulted in being the most important countermeasure against deep fake content. There are several deep fake datasets available online to test the reliability of deep fake detectors allowing individuals to bring enhancement in deep fake detecting techniques and tools. A study by [32] indicates the existence of deep fake as a worldwide issue in producing fake information and manipulation of masses by means of producing deep fake content to the targeted individuals and organizations. Another dataset namely Wilddeep fake has been exclusively developed to address the deep fake videos type [33]. This data especially contains a massive number of videos of actual situations with mimic voices and faces in these settings, the main goal is to classify deep fake videos in the dataset by observing changes in faces and other qualities of a video. As a result, this dataset has contributed to progress in deep fake studies and related creation of techniques and tools for that [34]. Deep fake data sets have been introduced and their availability has been easy for researchers which is positive as there are many more data sets online for testing of deep fake contents. These techniques and tools have minimized the risks associated with deep fake content as people are getting more into checking the reliability of available information.

Deep fake and synthetic media have been making great advancements recently bringing many challenges for deep fake detection tools and the techniques being used for it. Deep fake is having

a negative impact on society as it is misleading people by spreading misinformation [35]. The most important way of dealing with such problems is bringing advancement in deep fake detection and mitigation. There are many problems in differentiating between real and fake data as the manipulated data is very much advanced but there are various techniques and methodologies to differentiate [36]. Forensic analysis technique which includes pixel-level analysis and metadata examination is a very useful technique in identifying real and fake data [37]. Pattern recognition and anomaly detection are commonly used AI-based detection techniques which can identify deep fake data. There are algorithms and computational methods also being introduced majorly for the purpose of identifying deep fake content that is being developed day by day.

Technical Foundation of Deep Fakes:

AI and Machine Learning Techniques:

The roots and technical foundation of deep fakes can be seen in the continuous technological advancements in AI and machine learning techniques. The technological advancements in AI and machine learning have impacted many industries, platforms, and majorly human existence. The introduction of a Generative Adversarial Network (GAN) in machine learning is the main source from which deep fakes started being generated [38]. The two neural networks in GANs known to be the generator and discriminator work simultaneously, the generator creates fake images and videos whereas the discriminator helps in distinguishing between real and fake data. In the training process of GAN, the generator works on creating highly realistic images and videos so that the discriminator does not detect the fake elements whereas the discriminator works on getting better and detects the fake elements, this process in GAN continues till the generator ends up creating realistic images content. There are two popular GAN variants that are widely being used which include StyleGAN and CycleGAN. StyleGAN is used for creating highly realistic images while in CycleGAN the user can do image-to-image translation for example converting the same style images like a horse into a zebra [39]. Other AI and machine learning techniques that are commonly used other than GAN are Face recognition and manipulation techniques Recurrent neural networks (RNNs) and Transformers. Face recognition and manipulation techniques are used for face swapping, alignment, and creating a 3D model of the face for manipulation whereas RNNs and transformers are used for lip-syncing and voice synthesis [40].

Common Tools and Platforms:

Deep fake creation has become easy for people because of the easily available data sets, deep fake creation tools, and techniques [41]. The easy access to these tools allows people to use these tools without difficulty to create deep fake content for negative or positive purposes [28]. The most popular tools and techniques used for deep fake creation include:

Deep Face Lab:

It is the most common and easily accessed tool extensively used for deep fakes since it has tools for face swaps and other deep fake making activities [42]. Its features include advanced face

swapping and tools used for mask editing, face alignment, and face extraction [43]. People majorly use applications allowing face swap for swapping the faces of people in pictures or videos to mislead people by providing fake information [44].

Face swap:
Face swap tool uses a technique of swapping faces in videos like including the face of a person on another person doing certain things in a video [45]. This tool has a vast configuration system allowing it to be used with different supporting systems (Windows, Linux, MacOS).

Zao:
A mobile application named Zao has recently been launched and has gained great attention among many users as it provides with accurate results and is easily accessible by individuals [46]. The most common use of Zao is for users majorly use it for swap images of public figures and celebrities in images and videos in which they are not originally involved [45]. It provides users with high-quality content and it has a huge library of already present videos which are easily available for user access [47].

Types of Deep fakes:
Essentially, while deep fake can manifest in several ways, it is useful to know that these different configurations often involve different methods and instruments that are used differently [48]. The subcategories of deep fakes are, visuals, audio, and videos and these are the most common kinds of fake content [20]. The following is a discussion of the unique traits and characteristics of each sort of deep fake:

Audio deep fakes:
It is noted that deep fakes of this nature are mainly used to generate phone audio samples which imitate other people's voice to have the right voice intonation. It is primarily used for manipulative operations on audio data with the intention of influencing other people [49]. It is primarily used for manipulative operations on audio data with the intention of influencing other people [50]. The features of this deep fake are TTS, voice cloning, S2S, and the uses in voice dubbing, pranking, call recordings or spreading of fake news [51].

Video deep fakes:
Video deep fakes mostly are used to synthesize and alter realistic human-like (Ciftci et al. , 2020) videos [52]. The individuals in the videos are placed by face swapping, and the wages give the impression that the persons are engaged in something they have not been involved in. The elements of this deep fake are substitution of lips, face and body [28]. It is mostly utilized for recreation, spreading fake news, watching movies, and making political content [53].

Image deep fakes:
Image deep fakes generate novel synthetic images as it swaps the backgrounds and manipulates the facial structure of the persons in a picture [54]. The main traits of image deep fakes include

face swap, style transfer, face blend, and synthetic images [55]. These types of deep fakes are for identity theft, beauty filters, photography and also art works [56].

Threats Posed by Deep Fakes:

Security Risks:

Through deep fakes, the security of the people and organizations is being threatened because deep fakes are within people's reach and most importantly individuals with bad intentions are already using them for personal regain and commit heinous crimes [57]. The following are the most prevalent crimes committed by deep fake activities; phishing, misinformation and impersonated attacks [58]. The ways in which deep fake compromises security are discussed below:

Phishing:

Targeted victims are harassed by face swapping apps and voice impersonation applications [59]. This is mostly achieved for extortive intentions and for tarnishing the image of a noble individual [60]. It is also used to make victims release the cash, or give out significant information to the con artists [61].

Misinformation:

It has been seen that deep fakes are employed for disinformation campaigns against organizations and famous persons and personalities [62]. Misinformation against politicians or government officials in most cases leads to political instabilities and a threat to security [54]. Manipulated videos or audio of politicians engaging in immoral activities are employed to fuel anarchy, aggression, and the erode people's confidence on the politicians [63].

Impersonation:

Deep fake has also been used for impersonation as people often on social media platforms impersonate by face swapping in images which leads to a breach in privacy of the individual and causing reputational damage [64]. This can also invoke personal harm, blackmailing, and exploitation through harassment [65].

Impact on Privacy:

Deep fake advancement has significantly impacted individual privacy due to its ability to create realistic videos and images. These techniques are often used to create videos and images without the permission of individuals by face-swapping their images in videos and pictures in which they were not originally present. This non-consensual act leads to personal and professional harm as they are presented in controversial and defamatory scenarios [66]. One of the most damaging uses of deep fake is its use in pornographic videos where face-swapping technology is used to impose the faces of other individuals in the videos which ends up destroying the image of the individual in front of others [67]. Identity theft is also a risk to individual privacy as deep fake audio and images can be used for impersonation enabling unauthorized access to services and sensitive information leading to financial fraud. Social status and personal relationships are majorly affected by deep fake as it compromises the reputation of individuals by creating fake

data about them [65]. Social media is the main source through which deep fake content gets viral as a larger audience is available on social media to believe in such content without verification causing psychological harm to individuals whose reputation gets affected.

Social and Political Risks:

AI has brought change globally and AI manipulation through deep fakes is posing risks in social and political situations. Politicians and rivals opposing each other often use AI to generate deep fakes without understanding and analyzing the great risks it can pose for society [68]. Deep fakes are used to create social unrest, manipulate the opinion of the public, and invoke elections. This misuse of deep fakes has been witnessed in many countries and has ended up creating great distress in the social and political situations of the country. Social unrest happens when deep fakes are used to generate audio or video recordings of influencing politicians indulging in violent behaviors and provoking violent behavior. Moreover, deep fake videos of police or mobs indulging in violence can also end up creating distress in society and provoke the emotions of the general public [41]. Deep fakes are not only limited to videos or audio but creating deep fake images regarding tweets and social media posts of public figures and politicians is also used to manipulate public opinions. Elections are often invoked as opponents indulge in activities including creating deep fake videos of politicians indulging in unacceptable behaviors to sway voter opinions and make the public go against the political leaders [69]. Once deep fake content is created regarding an influencing public figure or politician it creates social unrest and debunking is majorly not possible at a high level as the fake content spreads much more quickly and reaches a larger audience than the original content which is why it becomes merely impossible to reverse the impact of deep fakes.

Detection Techniques:

Machine learning-based detection:

Deep fake mitigation and detection techniques are majorly connected to machine learning as they provide the major basis of deep fake detection tools. There is major advancement being made in this field through neural networks and machine learning enabling users to get better results in mitigating deep fake content [70]. Neural networks use the technique of pre-processing to analyze image properties and detect deep fake images. Many of the neural network tools about image recognition have also aided in identifying deeply faked image that are produced though GAN. A convolutional neural network will adopt the process of two image processing when it comes to the identification of deep fake images [71]. CNNs are also deemed the forensic model because pre-processing steps assist in recognizing the fraudulent picture created by GANs. And as it focuses on the properties of the depicted image and assists in the task of discriminating between a real and fake image more efficiently. CNNs are able to detect artefacts that deep fake generating tools leave which include the blurry and fuzzy areas of the images as well as the spatial erratic movements. To overcome the problems that exist in deep fakes, Recurrent Neural Network (RNNs) is used as a type of a neural network that targets at the abnormal and irregular face expression changes in the video [72]. The field of deep fake mitigation and detection as new

techniques are being introduced for instance the new hybrid approach to deep fake detection this involve the use of several neural networks like the CNNs, RNNs and the LTMs for deep fake detection.

Feature-Based Detection:

Feature-based detection for deep fakes is applied by considering the artefacts and overlooked discrepancies with the help of machine learning and hand-designed features [73]. The feature-based detection techniques can be highly effective as they show the abnormalities in deep fakes which are most probably intended to remain there during the production of deep fake media [74]. A few of the widely applied feature based approaches are as follows: Hand crafting based techniques and Machine learning generated feature based techniques. [75]. The main characteristics of these techniques through which deep fakes are detected are discussed below:

The handcrafted feature-based technique:

This feature-based technique is produced using ample crafting, relates to facial artefacts, color artefacts, and geometric characteristics [76]. Facial parings include eye blinks, mouth movements, and overall facial symmetry in deep fakes natural eye movement is missing, speech and mouth do not movements are not in sync, and there are asymmetrical facial features revealed [13]. Lighting color is in artefacts such as the differences in lightning, skin color of the body, and the shadowing as these cannot, for the most part, be correct in deep fakes. It also recognizes geometric characteristics as it takes into account the motion of the body and the synchronization of eyes, these characteristics are usually conspicuous in deep fakes [77].

Machine-learned feature-based technique:

The major area of deep fake detection through Machine-learned feature-based technique is temporal artefacts and frequency-domain analysis using deep neural networks [78]. Temporal artifacts pay much attention to frame consistency while motion analysis, as in deep fake, flickering and temporal discrepancies can be easily spotted through detailed analyses [79]. It is also possible to learn basic models from facial landmarks and motions that deviate from what a normal human body should be [28]. In this technique frequency domains are used and artefacts which are invisible to the spatial domain are identified. These are accomplished by deep neural networks including CNNs, RNNs and LTMs for detecting these inconsistencies.

Block chain and Digital Watermarking:

Out of all the tools, block chain and Digital watermarking have turned out to be the most effective in combating deep fakes. The most distinctive characteristic of these tools is to ascertain the validity of data using and safeguarding the original data [80]. Blocks are a secure method of doing transactions and securing data as it comes with, an unalterable record and decentralized technology, which is favorite for ascertaining the legitimacy of the data. Based on block chain technology the verification of new blocks decentralized, real time authentication of each field of data and developing records of origin, to enhance data security [81]. Digital watermarking adds a code or an identifier into the file and it cannot be seen by naked eyes but it

can be deciphered by another software. Such watermarks consist of clear, concealed, and graphic watermarks which are applied to protect data against changes [82]. Both the block chain technology and digital watermarking when integrated and used also have the influence of increasing the effectivity and reliability of data by improving the features of provenance and security in distribution.

Emerging Technologies:

Generally, there is always improvement being made in deep fake detection, with improvement, there arises more emerging technologies. There is a newer form of neural network to detect deep fakes known as Vision Transformers (VTs) more effective than CNN is as this is a transformer model employed for image analytics and natural language processing [83]. Besides, the current developments include the works made on GANs and CycleGANs are introduced as a model for detecting deep fake images through a process of converting the images forward and backward and the differences between photos and fakes are identified [84]. The audio-visual cross-model analysis is an emerging model introduced for analyzing deep fake data and detecting the inconsistencies found in the movement of lips and production of sound as deep fake videos majorly fail to have perfect lip-syncing [85].

Mitigation Strategies:

Policy and Legislation:

Deep fakes have become a great concern globally as besides bringing threat to privacy invasion and individual threats it has also contributed to creating political and social unrest [86]. As many countries have faced concerns regarding these issues they have worked on creating laws and legislations to address these concerns [87]. The current legislations imposed by some countries are discussed below:

United States (US):

In US National Defense Authorization Act (NDDA) was introduced in 2020 this act requires the homeland security of the country to create an annual report about the impact of deep fake technology to combat the harm caused by deep fakes [88].

European Union (EU):

The European Union (EU) introduced a strict law of data privacy known as be General Data Protection Regulation (GDPR), this law does not majorly target deep fakes but has strict policies regarding privacy invasion and it also addresses the issue of unauthorized personal data use [89]. Digital Services Act (DSA) is also introduced which includes strict laws regarding the use of illegal content on deep fakes and social media platforms [87].

United Kingdom (UK):

In the UK an Online safety bill has been proposed which aims to impose a strict duty on individuals, social media platforms, and companies to protect users against harmful content including deep fakes [90].

Proposals for new policies:
Laws should be created for deep fake data creation and distribution making it a criminal offence when used for harassment, spreading misinformation, and engagement in fraudulent behavior [91]. Consent has to be obtained from people concerning media that may infringe on their privacy and image [92]. It can be argued that deep fake content should be compelled to indicate that it is fake content so that it shouldn't influence the public [93].

Technical Mitigation:
A lot of progress has been made recently regarding the use of technology in tackling prevention of the creation and supply of deep fake information. Such technical solutions are a vast array of tools and methods that strengthen data protection and identify and counter deep fakes [82]. The technical mitigation solutions include some of the measures that have to be taken to protect data these can be done using the methods such as block chain and the watermark technology [94]. TEEs are part of Content Authentication; securing the original data and the process of changing the data are also another part of the same framework; is used to counter deep fakes [95]. Real-time content verification technique is also employed in the prevention of the generation and spreading of deep fakes for the simple reason that this technology can be easily verified in real-time on the device of users.

Public Awareness and Education:
To deal with the problem of deep fake data the frequent usage of different educational and increasing public awareness programs needed. Such awareness programs can assist in enhancing the provision of relevant information to the members of the public and will contribute to the creation and implementation of deep fake detection tools [96]. For public awareness programs governments required to take many steps, step one should be National awareness campaigns pertaining to media and public seminars to aware the public about risks related to deep fake sharing and how a normal person can detect deep fake data [97]. School and University curricula should also incorporate the topic of deep fakes as a part of their syllabi so the learners can learn about the various challenges as well as the harms of deep fakes and this can assist in the learners learn different approaches to identify deep fake content [67]. However, social media, printed media, and television media should also begin to raise awareness by posting information on deep fake cases and the repercussions of disseminating similar data.

Industry Practices:
Many industries that arose with the deep fake trend had no tools and methods to prevent deep fake impact on them but over time many sectors introduced the best practices and methods creating the risks of deep fake. The financial industries have recently implemented identity verification systems that consists of biometric verification and multi-factor authentication such strategies aid in ensuring financial security by protecting the individuals' identity as well as their information [98]. The financial industries have also incorporated the use of AI-based monitoring and training of the employees in identifying factors/features of deep fake data [99]. The media and entertainment services are also employing the strategies and applying instruments to identify

deep fake data along with the validation of data. This is usually done by using content verification tools like metadata standards and digital watermarking, metadata standards are used to track down the original source of data while, digital watermarking are invisible watermarks in the videos which can be used for verification of data authenticity [100].

Case studies (Notable incidents):

Deep fake has been affecting different domains of life of an individual including reputable organizations and popular personalities [86]. Some major incidents of deep fakes have had serious impact on society by damaging the image of individuals and manipulating public opinions regarding politicians and government [101]. Some of the notable incidents that were witnessed regarding deep fakes are discussed below:

The incident of deep fake video regarding Nancy Pelosi (2019):

A popular incident of a deep fake video was witnessed by US in 2019 regarding Nancy Pelosi as she was seen slurring in a public speech. This manipulated video gained attention by viewers on social media and it went viral and damaged the public image of Nancy Pelosi [102]. This video caused misinformation among public and a popular opinion of unstable mental status of Pelosi was made [102]. The impact of this incident was that it manipulated the opinion of general public regarding a political figure. Social media platforms were criticized for their inability to quickly react to the spread of misinformation on these platforms [103].

Deep fake audio incident of a CEO:

In 2019 another prominent deep fake incident occurred in UK as an individual used a deep fake audio of a CEO to communicate with the senior executive of the company for a fraudulent transaction of a €220,000 in an account [104]. This impacted the company by giving it a major financial loss and it also ended up highlighting the lack of stability and verification detection in securing financial transactions [105].

Deep fake revenge porn:

Many people use deep fake as a source of taking revenge or for blackmailing purposes, they use it to create deep fake porn videos of people from whom they have to take revenge by damaging their reputation or for blackmailing them to take financial favors [106]. This use of deep fake has effected the image of individuals by damaging their reputation and it rises an important question regarding privacy invasion and lack of data security [107].

Lessons learned:

There have been many lessons learned by the past deep fake incidents in the past and some of these lessons are discussed below:

Subject to Accountability Act of 2019:

Technology advancements have enabled the creation of deep fakes, which can cause harm to individuals or the public [108]. Currently, the US Congress drafted the Defending Each Person

from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019, which says that every creator and splitter of such fakes will face civil and criminal sanctions [109].

Fostering international cooperation, policymakers:

The case of CEO draws attention to how AI presents threats to regulation and law enforcement because of the kind of specific legal loopholes that AI-based hackers use to their advantage, and the necessity for strengthening cooperation and implementation of legislation at the international level to address global threats [110]. It is high time that the policymakers formulated and complied accurate legal mechanisms and promote international cooperation considering the fact that AI is advancing significantly to perpetrate cybercrimes [111].

Legal regulations and protection methods:

Non-consensual sharing of fake nude photos continues to be an emerging problem, revenge porn as well as deep fake porn reveals the failure of legal frameworks [112]. Comparing the current laws of nine EU Member States, three have separate incriminations for revenge porn, but the conceptual scope of its definition differs [113]. Legal regulations and protection methods for victims can be updated to benefit victims by assessing legal and technological solutions to address this problem [114].

Incident	*Reference*	*Citation*
Nancy Pelosi Deepfake Video (2019)	*Deep Fakes Accountability Act: Overbroad and Ineffective*	*Schapiro, Z. (2020). Boston College Intellectual Property and Technology Forum.*
	Deep fakes: The algorithms that create and detect them and the national security risks they pose	*Dunard, N. (2020). James Madison Undergraduate Research Journal (JMURJ), 8(1), 5.*
Deepfake Audio Incident of a CEO (2019)	*Cybercrime threats and responsibilities: the utilization of artificial intelligence in online crime*	*Rasyid, M. F. F., et al. (2024). Jurnal Ilmiah Mizani: Wacana Hukum, Ekonomi Dan Keagamaan, 11(1, April), 49-63.*
	A survey on the detection and impacts of deepfakes in visual, audio, and textual formats	*Mubarak, R., et al. (2023). IEEE Access.*
Deepfake Revenge Porn	*Legal protection of revenge and deepfake porn victims in the European Union: findings from a comparative legal study*	*Mania, K. (2024). Trauma, Violence, & Abuse, 25(1), 117-129.*
	'A Deepfake Porn Plot Intended to Silence Me': exploring continuities	*Maddocks, S. (2020). Porn Studies,*

	between pornographic and 'political' deep fakes	*7(4), 415-423.*
Legislation and Legal Frameworks	*Deepfakes: a new content category for a digital age*	*Pesetski, A. (2020). Wm. & Mary Bill Rts. J., 29, 503.*
	Audio deepfakes: A survey	*Khanjani, Z., Watson, G., & Janeja, V. P. (2023). Frontiers in Big Data, 5, 1001063.*
	Virtual revenge pornography as a new online threat to sexual integrity	*Šepec, M., & Lango, M. (2020). Balkan Social Science Review, 15, 117-135.*
	Pornographic deepfakes: the case for federal criminalization of revenge porn's next tragic act	*Delfino, R. A. (2020). Actual Probs. Econ. & L., 105.*
	Reporting revenge porn: a preliminary expert analysis	*De Angeli, A., et al. (2021). Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter.*

Future Directions:

Advancements in detection:

The continuous advancements in deep fake technology is going side by side with advancements being made in deep fake detection tools and technologies [58]. Some of the potential advancements that can majorly be made in deep fake detection technologies are discussed:

Advanced AI algorithms and machine learning:

The main advancements in deep fake detection are majorly done by advancements in machine and AI algorithms [115]. The reverse-GAN technology can be used to detect the generation deep fakes by identifying the inconsistencies present in the media left by GAN [116]. Moreover, detection algorithm should continuously be updated with update in the latest technologies of deep fake creation for deep fake detection technologies to be able to detect deep fakes [117].

Forensics based on deep learning:

Deep learning can be used to detect micro expression by making advancements in it and introducing micro-facial expression analysis [72]. Minor expression detailing is often not possible through the use of deep fakes so deep learning technologies analyzing detailed features can help in identifying deep fakes [118].

Advancements in Real-time detection:

Advancements in real-time detection tools can help in securing data by providing access to authorized users only [119]. These tools can also detect deep fake content in real time allowing users to take accurate actions regarding the spread of deep fakes on platforms [26].

Challenges Ahead:

As technological advancement is being made in every sector there are also advancements being made in the deep fake creation industries [120]. The AI manipulated data are being so advanced that the detection of deep fakes is becoming difficult and the deep fake mitigation and detection field is getting great concern [118]. This can pose a serious risk to society as difficulties in detecting deep fake data can lead to spread of misinformation and fraudulent behavior [121]. The easy accessibility of deep fakes is also of concern as these tools are in the use of average internet users they can use it for any purpose they want including blackmailing, fraudulent behavior, and spread of misinformation.

Interdisciplinary Approaches:

Interdisciplinary approaches can prove to be a successful way of dealing with challenges imposed by deep fakes. This approach majorly requires collaboration between different domains for successful implication regarding the detection and mitigation of deep fakes [122]. The collaboration between cyber security technologies and AI algorithms and machine learning can lead in bringing advancements in development of technologies and tools used for deep fake detection and mitigation [123]. Legal frameworks allow in making law and policies both at international and national level these policies highlight the risks associated with deep fakes and legal actions that can be made against the distribution of deep fake content [124]. Educational institutes and government awareness programs are of great importance as these can be used to create awareness regarding deep fakes and allowing people to get knowledge about deep fake detection and mitigation.

Conclusion:

This paper looks at the use of deep fake instances in the global arena and its effects on social and political circumstances. It draws the attention to dissemination of misinformation and creation of higher-level deep fake data using neural networks. In the same fight as with the regular fake news, efforts are being made to curb the deep fake proliferation and protect people's information. But, detection continues to be a worry internationally. There is a need for the legislation of tangible severe laws, and polices regarding the penalization of anyone who is share fake data. Similar to DeepFake detection, social media platforms should also incorporate deep fake technology. To remove malicious content from spreading, people and organizations need to protect their information against AI-pumped posts.

References:

[1] D. Gragnaniello, F. Marra, and L. Verdoliva, "Detection of AI-generated synthetic faces," in *Handbook of digital face manipulation and detection: From deepfakes to morphing attacks*: Springer International Publishing Cham, 2022, pp. 191-212.

[2] J. Frank, T. Eisenhofer, L. Schönherr, A. Fischer, D. Kolossa, and T. Holz, "Leveraging frequency analysis for deep fake image recognition," in *International conference on machine learning*, 2020: PMLR, pp. 3247-3258.

[3] A. Mukherjee, "Safeguarding Marketing Research: The Generation, Identification, and Mitigation of AI-Fabricated Disinformation," *arXiv preprint arXiv:2403.14706*, 2024.

[4] J. Yamagishi *et al.*, "ASVspoof 2021: accelerating progress in spoofed and deepfake speech detection," in *ASVspoof 2021 Workshop-Automatic Speaker Verification and Spoofing Coutermeasures Challenge*, 2021.

[5] S. Suganthi, M. U. A. Ayoobkhan, N. Bacanin, K. Venkatachalam, H. Štěpán, and T. Pavel, "Deep learning model for deep fake face recognition and detection," *PeerJ Computer Science*, vol. 8, p. e881, 2022.

[6] H. S. Shad *et al.*, "[Retracted] Comparative Analysis of Deepfake Image Detection Method Using Convolutional Neural Network," *Computational intelligence and neuroscience*, vol. 2021, no. 1, p. 3111676, 2021.

[7] Y. Liu and Y.-F. B. Wu, "Fned: a deep network for fake news early detection on social media," *ACM Transactions on Information Systems (TOIS)*, vol. 38, no. 3, pp. 1-33, 2020.

[8] S. R. Ahmed, E. Sonuç, M. R. Ahmed, and A. D. Duru, "Analysis survey on deepfake detection and recognition with convolutional neural networks," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2022: IEEE, pp. 1-7.

[9] R. Chowdhury, *Technology Facilitated Gender-Based Violence in an era of Generative AI*. UNESCO Publishing, 2023.

[10] H. Zhao, W. Zhou, D. Chen, T. Wei, W. Zhang, and N. Yu, "Multi-attentional deepfake detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2185-2194.

[11] Y. Wang, "Synthetic realities in the digital age: Navigating the opportunities and challenges of ai-generated content," *Authorea Preprints*, 2023.

[12] T. C. Helmus, "Artificial intelligence, deepfakes, and disinformation," *RAND Corporation*, pp. 1-24, 2022.

[13] S. Agarwal, H. Farid, T. El-Gaaly, and S.-N. Lim, "Detecting deep-fake videos from appearance and behavior," in *2020 IEEE international workshop on information forensics and security (WIFS)*, 2020: IEEE, pp. 1-6.

[14] E. Temir, "Deepfake: new era in the age of disinformation & end of reliable journalism," *Selçuk İletişim,* vol. 13, no. 2, pp. 1009-1024, 2020.

[15] Y. Hua, S. Niu, J. Cai, L. B. Chilton, H. Heuer, and D. Y. Wohn, "Generative AI in User-Generated Content," 2024.

[16] R. Krohn and T. Weninger, "Subreddit links drive community creation and user engagement on Reddit," in *Proceedings of the International AAAI Conference on Web and Social Media*, 2022, vol. 16, pp. 536-547.

[17] M. Westerlund, "The emergence of deepfake technology: A review," *Technology innovation management review*, vol. 9, no. 11, 2019.

[18] A. O. Kwok and S. G. Koh, "Deepfake: a social construction of technology perspective," *Current Issues in Tourism*, vol. 24, no. 13, pp. 1798-1802, 2021.

[19] S. Pashine, S. Mandiya, P. Gupta, and R. Sheikh, "Deep fake detection: Survey of facial manipulation detection solutions," *arXiv preprint arXiv:2106.12605*, 2021.

[20] H. Farid, "Creating, using, misusing, and detecting deep fakes," *Journal of Online Trust and Safety*, vol. 1, no. 4, 2022.

[21] B. Chesney and D. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," *Calif. L. Rev.*, vol. 107, p. 1753, 2019.

[22] N. Kaloudi and J. Li, "The ai-based cyber threat landscape: A survey," *ACM Computing Surveys (CSUR)*, vol. 53, no. 1, pp. 1-34, 2020.

[23] A. Malik, M. Kuribayashi, S. M. Abdullahi, and A. N. Khan, "DeepFake detection for human face images and videos: A survey," *Ieee Access*, vol. 10, pp. 18757-18775, 2022.

[24] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake detection: A systematic literature review," *IEEE access*, vol. 10, pp. 25494-25513, 2022.

[25] F. Sahran, H. H. Altarturi, and N. B. Anuar, "Exploring the Landscape of AI-SDN: A Comprehensive Bibliometric Analysis and Future Perspectives," *Electronics*, vol. 13, no. 1, p. 26, 2023.

[26] S. Chaudhary, R. Saifi, N. Chauhan, and R. Agarwal, "A comparative analysis of deep fake techniques," in *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 2021: IEEE, pp. 300-303.

[27] C. Tarsney, "Deception and Manipulation in Generative AI," *arXiv preprint arXiv:2401.11335*, 2024.

[28] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," *Information Fusion*, vol. 64, pp. 131-148, 2020.

[29] B. Dolhansky *et al.*, "The deepfake detection challenge (dfdc) dataset," *arXiv preprint arXiv:2006.07397*, 2020.

[30] S. A. Khan, A. Artusi, and H. Dai, "Adversarially robust deepfake media detection using fused convolutional neural network predictions," *arXiv preprint arXiv:2102.05950*, 2021.

[31] B. Lamichhane, K. Thapa, and S.-H. Yang, "Detection of image level forgery with various constraints using DFDC full and sample datasets," *Sensors*, vol. 22, no. 23, p. 9121, 2022.

[32] B. Zi, M. Chang, J. Chen, X. Ma, and Y.-G. Jiang, "Wilddeepfake: A challenging real-world dataset for deepfake detection," in *Proceedings of the 28th ACM international conference on multimedia*, 2020, pp. 2382-2390.

[33] S. Das, S. Seferbekov, A. Datta, M. S. Islam, and M. R. Amin, "Towards solving the deepfake problem: An analysis on improving deepfake detection using dynamic face augmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3776-3785.

[34] N. Waqas, S. I. Safie, K. A. Kadir, S. Khan, and M. H. K. Khel, "DEEPFAKE image synthesis for data augmentation," *IEEE Access*, vol. 10, pp. 80847-80857, 2022.

[35] S. Zobaed *et al.*, "Deepfakes: Detecting forged and synthetic media content using machine learning," *Artificial Intelligence in Cyber Security: Impact and Implications: Security Challenges, Technical and Ethical Issues, Forensic Investigative Challenges*, pp. 177-201, 2021.

[36] M. C. El Rai, H. Al Ahmad, O. Gouda, D. Jamal, M. A. Talib, and Q. Nasir, "Fighting deepfake by residual noise using convolutional neural networks," in *2020 3rd International Conference on Signal Processing and Information Security (ICSPIS)*, 2020: IEEE, pp. 1-4.

[37] B. Khoo, R. C. W. Phan, and C. H. Lim, "Deepfake attribution: On the source identification of artificially generated images," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 12, no. 3, p. e1438, 2022.

[38] A. Aggarwal, M. Mittal, and G. Battineni, "Generative adversarial network: An overview of theory and applications," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100004, 2021.

[39] G. Wang, H. Shi, and Y. Chen, "Self-augmentation with dual-cycle constraint for unsupervised image-to-image generation," in *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, 2021: IEEE, pp. 886-890.

[40] S. Reza, M. C. Ferreira, J. J. Machado, and J. M. R. Tavares, "A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks," *Expert Systems with Applications*, vol. 202, p. 117275, 2022.

[41] P. Fraga-Lamas and T. M. Fernandez-Carames, "Fake news, disinformation, and deepfakes: Leveraging distributed ledger technologies and blockchain to combat digital deception and counterfeit reality," *IT professional*, vol. 22, no. 2, pp. 53-59, 2020.

[42] L. Verdoliva, "Media forensics and deepfakes: an overview," *IEEE journal of selected topics in signal processing*, vol. 14, no. 5, pp. 910-932, 2020.

[43] I. Perov *et al.*, "DeepFaceLab: Integrated, flexible and extensible face-swapping framework," *arXiv preprint arXiv:2005.05535*, 2020.

[44] J. Botha and H. Pieterse, "Fake news and deepfakes: A dangerous threat for 21st century information security," in *ICCWS 2020 15th International Conference on Cyber Warfare and Security. Academic Conferences and publishing limited*, 2020, p. 57.

[45] Z. Yi and N. H. Romainoor, "A Systematic Literature Review for Interface Design of Pelvic Floor Muscle Training Mobile App base on mHealth 2017-2022," *Journal of Advanced Computing Technology and Application (JACTA)*, vol. 5, no. 1, pp. 28-42, 2023.

[46] M. Masood, M. Nawaz, K. M. Malik, A. Javed, A. Irtaza, and H. Malik, "Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward," *Applied intelligence*, vol. 53, no. 4, pp. 3974-4026, 2023.

[47] M. Albahar and J. Almalki, "Deepfakes: Threats and countermeasures systematic review," *Journal of Theoretical and Applied Information Technology*, vol. 97, no. 22, pp. 3242-3250, 2019.

[48] J. T. Hancock and J. N. Bailenson, "The social impact of deepfakes," vol. 24, ed: Mary Ann Liebert, Inc., publishers 140 Huguenot Street, 3rd Floor New …, 2021, pp. 149-152.

[49] F. Matern, C. Riess, and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations," in *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, 2019: IEEE, pp. 83-92.

[50] S. Agarwal, H. Farid, O. Fried, and M. Agrawala, "Detecting deep-fake videos from phoneme-viseme mismatches," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 660-661.

[51] A. M. Almars, "Deepfakes detection techniques using deep learning: a survey," *Journal of Computer and Communications*, vol. 9, no. 05, pp. 20-35, 2021.

[52] U. A. Ciftci, I. Demir, and L. Yin, "How do the hearts of deep fakes beat? deep fake source detection via interpreting residuals with biological signals," in *2020 IEEE international joint conference on biometrics (IJCB)*, 2020: IEEE, pp. 1-10.

[53] K. Shiohara and T. Yamasaki, "Detecting deepfakes with self-blended images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18720-18729.

[54] M. Mustak, J. Salminen, M. Mäntymäki, A. Rahman, and Y. K. Dwivedi, "Deepfakes: Deceptions, mitigations, and opportunities," *Journal of Business Research,* vol. 154, p. 113368, 2023.

[55] N. M. Müller, K. Pizzi, and J. Williams, "Human perception of audio deepfakes," in *Proceedings of the 1st International Workshop on Deepfake Detection for Audio Multimedia*, 2022, pp. 85-91.

[56] Á. Vizoso, M. Vaz-Álvarez, and X. López-García, "Fighting deepfakes: Media and internet giants' converging and diverging strategies against hi-tech misinformation," *Media and Communication,* vol. 9, no. 1, pp. 291-300, 2021.

[57] J. Langa, "Deepfakes, real consequences: Crafting legislation to combat threats posed by deepfakes," *BUL Rev.,* vol. 101, p. 761, 2021.

[58] M. Sharma and M. Kaur, "A review of Deepfake technology: an emerging AI threat," *Soft Computing for Security Applications: Proceedings of ICSCS 2021,* pp. 605-619, 2022.

[59] P. Neekhara, B. Dolhansky, J. Bitton, and C. C. Ferrer, "Adversarial threats to deepfake detection: A practical perspective," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,* 2021, pp. 923-932.

[60] A. Ali, K. F. K. Ghouri, H. Naseem, T. R. Soomro, W. Mansoor, and A. M. Momani, "Battle of deep fakes: Artificial intelligence set to become a major threat to the individual and national security," in *2022 International Conference on Cyber Resilience (ICCR)*, 2022: IEEE, pp. 1-5.

[61] R. Katarya and A. Lal, "A study on combating emerging threat of deepfake weaponization," in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 2020: IEEE, pp. 485-490.

[62] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, and H. Li, "Protecting World Leaders Against Deep Fakes," in *CVPR workshops*, 2019, vol. 1, p. 38.

[63] A. Firc, K. Malinka, and P. Hanáček, "Deepfakes as a threat to a speaker and facial recognition: An overview of tools and attack vectors," *Heliyon,* vol. 9, no. 4, 2023.

[64] K. R. Harris, "Video on demand: What deepfakes do and how they harm," *Synthese,* vol. 199, no. 5, pp. 13373-13391, 2021.

[65] D. Fallis, "The epistemic threat of deepfakes," *Philosophy & Technology,* vol. 34, no. 4, pp. 623-643, 2021.

[66] T. C. Helmus, "Artificial Intelligence, Deepfakes, and Disinformation," 2022.

[67] E. Meskys, J. Kalpokiene, P. Jurcys, and A. Liaudanskas, "Regulating deep fakes: legal and ethical considerations," *Journal of Intellectual Property Law & Practice,* vol. 15, no. 1, pp. 24-31, 2020.

[68] I. Kalpokas and J. Kalpokiene, *Deepfakes: a realistic assessment of potentials, risks, and policy regulation*. Springer, 2022.

[69] S. Ahmed, "Who inadvertently shares deepfakes? Analyzing the role of political interest, cognitive ability, and social network size," *Telematics and Informatics,* vol. 57, p. 101508, 2021.

[70] A. Ghai, P. Kumar, and S. Gupta, "A deep-learning-based image forgery detection framework for controlling the spread of misinformation," *Information Technology & People,* vol. 37, no. 2, pp. 966-997, 2024.

[71] W. Shahid *et al.*, "Detecting and mitigating the dissemination of fake news: Challenges and future research opportunities," *IEEE Transactions on Computational Social Systems,* 2022.

[72] R. Rafique, R. Gantassi, R. Amin, J. Frnda, A. Mustapha, and A. H. Alshehri, "Deep fake detection and classification using error-level analysis and deep learning," *Scientific Reports,* vol. 13, no. 1, p. 7422, 2023.

[73] B. Kaddar, S. A. Fezza, W. Hamidouche, Z. Akhtar, and A. Hadid, "On the effectiveness of handcrafted features for deepfake video detection," *Journal of Electronic Imaging,* vol. 32, no. 5, pp. 053033-053033, 2023.

[74] S. R. Sahoo and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," *Applied Soft Computing,* vol. 100, p. 106983, 2021.

[75] I. Demir and U. A. Ciftci, "Where do deep fakes look? synthetic face detection via gaze tracking," in *ACM symposium on eye tracking research and applications*, 2021, pp. 1-11.

[76] A. Hamadene, A. Ouahabi, and A. Hadid, "Deepfakes Signatures Detection in the Handcrafted Features Space," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 460-466.

[77] R. Tolosana, S. Romero-Tapiador, J. Fierrez, and R. Vera-Rodriguez, "Deepfakes evolution: Analysis of facial regions and fake detection performance," in *international conference on pattern recognition*, 2021: Springer, pp. 442-456.

[78] A. Choudhary and A. Arora, "Linguistic feature based learning model for fake news detection and classification," *Expert Systems with Applications,* vol. 169, p. 114171, 2021.

[79] N. u. Huda, A. Javed, K. Maswadi, A. Alhazmi, and R. Ashraf, "Fake-checker: A fusion of texture features and deep learning for deepfakes detection," *Multimedia Tools and Applications,* pp. 1-25, 2023.

[80] B. Wang, S. Jiawei, W. Wang, and P. Zhao, "Image copyright protection based on blockchain and zero-watermark," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 4, pp. 2188-2199, 2022.

[81] A. Garba *et al.*, "A digital rights management system based on a scalable blockchain," *Peer-to-Peer Networking and Applications*, vol. 14, pp. 2665-2680, 2021.

[82] A. Qureshi and D. Megias Jimenez, "Blockchain-based multimedia content protection: Review and open challenges," *Applied Sciences,* vol. 11, no. 1, p. 1, 2020.

[83] C. Zhang *et al.*, "3d talking face with personalized pose dynamics," *IEEE Transactions on Visualization and Computer Graphics,* vol. 29, no. 2, pp. 1438-1449, 2021.

[84] L. Liu, L. Gao, W. Lei, F. Ma, X. Lin, and J. Wang, "A survey on deep multi-modal learning for body language recognition and generation," *arXiv preprint arXiv:2308.08849,* 2023.

[85] C. Zhang, "Talking Human Synthesis: Learning Photorealistic Co-Speech Motions and Visual Appearances From Videos," The University of Texas at Dallas, 2023.

[86] A. Rahman *et al.*, "A qualitative survey on deep learning based deep fake video creation and detection method," *Aust. J. Eng. Innov. Technol,* vol. 4, no. 1, pp. 13-26, 2022.

[87] C. Whyte, "Deepfake news: AI-enabled disinformation as a multi-level public policy challenge," *Journal of cyber policy,* vol. 5, no. 2, pp. 199-217, 2020.

[88] H. Farid and H.-J. Schindler, "Deep Fakes," *On the Threat of Deep Fakes to Democracy and Society. Berlin: Konrad Adenauer Stiftung,* 2020.

[89] A. Fernandez, "“Deep fakes”: disentangling terms in the proposed EU Artificial Intelligence Act," *UFITA Archiv für Medienrecht und Medienwissenschaft,* vol. 85, no. 2, pp. 392-433, 2022.

[90] A. Ray, "Disinformation, deepfakes and democracies: The need for legislative reform," *TheUNIVERSITY OF NEW SOUTH WALES LAW JOURNAL,* vol. 44, no. 3, pp. 983-1013, 2021.

[91] T. T. Nguyen *et al.*, "Deep learning for deepfakes creation and detection: A survey," *Computer Vision and Image Understanding,* vol. 223, p. 103525, 2022.

[92] N. Aslam, I. Ullah Khan, F. S. Alotaibi, L. A. Aldaej, and A. K. Aldubaikil, "Fake detect: A deep learning ensemble model for fake news detection," *complexity,* vol. 2021, no. 1, p. 5557784, 2021.

[93] S. Tariq, S. Jeon, and S. S. Woo, "Am I a real or fake celebrity? Evaluating face recognition and verification APIs under deepfake impersonation attack," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 512-523.

[94] T. Geppert, S. Deml, D. Sturzenegger, and N. Ebert, "Trusted execution environments: Applications and organizational challenges," *Frontiers in Computer Science,* vol. 4, p. 930741, 2022.

[95] A. Muñoz, R. Rios, R. Román, and J. López, "A survey on the (in) security of trusted execution environments," *Computers & Security,* vol. 129, p. 103180, 2023.

[96] Y. Hwang, J. Y. Ryu, and S.-H. Jeong, "Effects of disinformation using deepfake: The protective effect of media literacy education," *Cyberpsychology, Behavior, and Social Networking,* vol. 24, no. 3, pp. 188-193, 2021.

[97] K. Shu *et al.*, "Combating disinformation in a social media age," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 10, no. 6, p. e1385, 2020.

[98] W. Yang, S. Wang, M. Shahzad, and W. Zhou, "A cancelable biometric authentication system based on feature-adaptive random projection," *Journal of Information Security and Applications,* vol. 58, p. 102704, 2021.

[99] M. Tahraoui, C. Krätzer, and J. Dittmann, "Defending Informational Sovereignty by Detecting Deepfakes: Risks and Opportunities of an AI-Based Detector for Deepfake-Based Disinformation and Illegal Activities," in *Weizenbaum Conference" Practicing Sovereignty: Interventions for Open Digital Futures"*, 2023: DEU, pp. 142-161.

[100] J. C. Simmons and J. M. Winograd, "Interoperable Provenance Authentication of Broadcast Media using Open Standards-based Metadata, Watermarking and Cryptography," *arXiv preprint arXiv:2405.12336*, 2024.

[101] A. Chadha, V. Kumar, S. Kashyap, and M. Gupta, "Deepfake: an overview," in *Proceedings of second international conference on computing, communications, and cyber-security: IC4S 2020*, 2021: Springer, pp. 557-566.

[102] Z. Schapiro, "Deep Fakes Accountability Act: Overbroad and Ineffective," in *Boston College Intellectual Property and Technology Forum*, 2020, vol. 2020, pp. 1-16.

[103] N. Dunard, "Deep fakes: The algorithms that create and detect them and the national security risks they pose," *James Madison Undergraduate Research Journal (JMURJ)*, vol. 8, no. 1, p. 5, 2020.

[104] M. F. F. Rasyid, M. A. SJ, K. Z. Mamu, S. R. Paminto, W. A. Hidaya, and A. Hamadi, "CYBERCRIME THREATS AND RESPONSIBILITIES: THE UTILIZATION OF ARTIFICIAL INTELLIGENCE IN ONLINE CRIME," *Jurnal Ilmiah Mizani: Wacana Hukum, Ekonomi Dan Keagamaan*, vol. 11, no. 1, April, pp. 49-63, 2024.

[105] R. Mubarak, T. Alsboui, O. Alshaikh, I. Inuwa-Dute, S. Khan, and S. Parkinson, "A survey on the detection and impacts of deepfakes in visual, audio, and textual formats," *IEEE Access*, 2023.

[106] K. Mania, "Legal protection of revenge and deepfake porn victims in the European Union: findings from a comparative legal study," *Trauma, Violence, & Abuse*, vol. 25, no. 1, pp. 117-129, 2024.

[107] S. Maddocks, "'A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political' deep fakes," *Porn Studies*, vol. 7, no. 4, pp. 415-423, 2020.

[108] N. Nnamdi, O. Oniyinde, and B. Abegunde, "An Appraisal of the Implications of Deep Fakes: The Need for Urgent International Legislations," *American Journal of Leadership and Governance*, vol. 8, no. 1, pp. 43-70, 2023.

[109] A. Pesetski, "Deepfakes: a new content category for a digital age," *Wm. & Mary Bill Rts. J.*, vol. 29, p. 503, 2020.

[110] Z. Khanjani, G. Watson, and V. P. Janeja, "Audio deepfakes: A survey," *Frontiers in Big Data*, vol. 5, p. 1001063, 2023.

[111] V. A. Jones, "Artificial intelligence enabled deepfake technology: The emergence of a new threat," Utica College, 2020.

[112] A. De Angeli, M. Falduti, M. Menendez Blanco, and S. Tessaris, "Reporting revenge porn: a preliminary expert analysis," in *Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter*, 2021, pp. 1-5.

[113] R. A. Delfino, "Pornographic deepfakes: the case for federal criminalization of revenge porn's next tragic act," *Actual Probs. Econ. & L.*, p. 105, 2020.
[114] M. Šepec and M. Lango, "Virtual revenge pornography as a new online threat to sexual integrity," *Balkan Social Science Review*, vol. 15, pp. 117-135, 2020.
[115] A. Heidari, N. Jafari Navimipour, H. Dag, and M. Unal, "Deepfake detection using deep learning methods: A systematic and comprehensive review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 14, no. 2, p. e1520, 2024.
[116] S. Negi, M. Jayachandran, and S. Upadhyay, "Deep fake: an understanding of fake images and videos," *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, vol. 7, no. 3, pp. 183-189, 2021.
[117] S. Kingra, N. Aggarwal, and N. Kaur, "Emergence of deepfakes and video tampering detection approaches: A survey," *Multimedia Tools and Applications*, vol. 82, no. 7, pp. 10165-10209, 2023.
[118] Z. Almutairi and H. Elgibreen, "A review of modern audio deepfake detection methods: challenges and future directions," *Algorithms*, vol. 15, no. 5, p. 155, 2022.
[119] L. A. Passos *et al.*, "A review of deep learning-based approaches for deepfake content detection," *Expert Systems*, p. e13570, 2022.
[120] Y. Patel *et al.*, "Deepfake generation and detection: Case study and challenges," *IEEE Access*, 2023.
[121] T. Zhang, "Deepfake generation and detection, a survey," *Multimedia Tools and Applications*, vol. 81, no. 5, pp. 6259-6276, 2022.
[122] A. Godulla, C. P. Hoffmann, and D. Seibert, "Dealing with deepfakes–an interdisciplinary examination of the state of research and implications for communication studies," *SCM Studies in Communication and Media*, vol. 10, no. 1, pp. 72-96, 2021.
[123] C. Sample *et al.*, "Interdisciplinary lessons learned while researching fake news," *Frontiers in psychology*, vol. 11, p. 537612, 2020.
[124] C. F. Brooks, "Popular discourse around deepfakes and the interdisciplinary challenge of fake video distribution," *Cyberpsychology, Behavior, and Social Networking*, vol. 24, no. 3, pp. 159-163, 2021.