Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44) 2** : 1–11.
ISSN-0037-8712
doi:10.5269/bspm.80533

Deepfakes - Consequences and Steps to be Taken

B. Nagamani[1], G. Neeraja Rani[2], Nemani Subadra[3], Neha Ratnala[4]

ABSTRACT: Deepfakes is one among many other rapidly advancing technologies in the tech industry. These are fake AI-generated videos or audio resembling a certain subject. The deep learning methods for producing deepfakes involve training generative neural network architectures like autoencoders and Generative Adversarial Networks (GAN). Since 2018, the number of deepfake online videos has doubled. It is a fascinating technology yet mostly misused technology. It was reported that 96% of deepfakes are pornographic videos mainly targeting women, especially female celebrities and even children. The present manuscript is a review paper on consequences of deepfakes in various domains which include digital impersonation, manipulation, blackmail, misinformation, misleading, public shaming, fabricating evidence, financial fraud etc. Deepfakes can be used to damage the reputation of celebrities, politicians and any public figures. (Paper Modi). Initially, deepfakes were mostly created by experts but now many accessible applications are available to create deepfakes thereby drastically increasing the risk of it being misused. Although deep fakes have many drawbacks, they also have advantages such as hyper-personalization, better dubbing, cheaper video campaigns, creating films with lesser utilisation of time and labour, making dynamic images and videos of the deceased, improved video game characters etc. To combat deepfakes, certain laws and safety measures are being implemented.

Keywords: Autoencoders, generative adversarial networks, benefits, drawbacks, measures.

Contents

Introduction

The term "deepfake" is formed after combining the terms "deep learning" and "fake" [2]. Simply put, deepfakes are realistic-looking video or audio recordings of people saying or doing things they never actually said or did which were created using artificial intelligence (AI)[3]. In January 2022, a Deepfake of Barack Obama was released by Jordan Peele spreading awareness regarding the potential of deepfake technology which gained significant attention due to its realism [1]. On June 5, 2023, a video of Vladimir Putin was aired on multiple radio and television networks by an unknown source in which Putin can be seen declaring the invasion of Russia and urging a nationwide army mobilization. This video was later reported as a deepfake which was created and broadcasted by hackers to create a fake emergency.

The common and accurate method of creating deepfakes involves utilising deep learning models, such as autoencoders and Generative Adversarial Networks (GANs)[1]. These models use two distinct AIs

2020 *Mathematics Subject Classification:* 68T07.
Submitted December 01, 2025. Published March 14, 2026

together known as an autoencoder [1]. An autoencoder consists of two components: the encoder and the decoder. First, the encoder scans the face to be deepfaked (such as the face of a celebrity or politician etc) and generates new fabricated images [1]. Later these images are transmitted to the decoder, which scrutinizes both the fake and authentic images [1]. If the decoder detects any differences between them, it communicates with the encoder, prompting necessary adjustments [1]. The encoder then produces new manipulated images. This process continues until the decoder can no longer be able to distinguish between the real and fake images or until an acceptable error rate is achieved [1][4].

Deepfakes are of different forms such as videos, audios, images or even texts [1]. A deepfake video is comparatively difficult to create because it requires specialised software, time, money, and skill for advanced manipulation of both visual and auditory components [1].

As the Deepfake technology advances it can result in potential growth in various domains such as entertainment, education, photography, video games, virtual reality etc. This progression in technology has the capacity to significantly enhance the quality of life. However, its positive applications are majorly used in malicious acts such as spreading false news, fabricating evidence, jeopardising a person's privacy and damaging their reputation, committing financial fraud etc.

Cybercrimes, frauds, manipulation, blackmail and fabrication of evidence using various technological tools have been around for decades but what distinguishes deepfake technology technique is its accuracy and easy accessibility to the public. Earlier it took days or weeks and high-end computing resources to construct a sophisticated fake but now as the availability of easily accessible tools in the market drastically increases which are also inexpensive even people with no technical knowledge or expertise can easily create deepfakes which led to increase in frauds and disinformation[1].

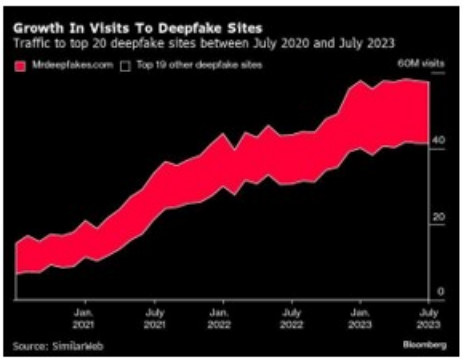
This study provides detailed information regarding the advantages, disadvantages and various detection methods created to detect deepfakes.

1. Disadvantages of Deepfakes in Various Domains

Pornography:

It was reported that 96% of deepfakes are pornographic videos[4] and public figures such as celebrities and politicians are the initial targets since a large number of their videos and images are available online [3]. In fact, the term deepfake was framed and gained attention after a Reddit user created a pornographic deepfake video of Gal Gadot, an Israeli actress whose face was swapped to the face of a porn star back in 2017. Current technological advancements allow the creation of a complete deepfaked video by just using one single photo of the victim enabling people to swap the victim's still images onto pornographic videos. This leads to the immediate threat possessed which is using deepfakes as fake blackmail material. For instance, blackmailers might claim to have access to private information by sending pictures or videos of the victim's own face transposed into pornography.

An application named DeepNude which was released in June 2019 was observed to be more threatening as it uses the generative adversarial network to create nudes of full-clothed images[1][5]. However, this application was removed by its creators but there are many other software available to create pornographic deepfakes. In July 2019 another software was released on Telegram which provided a free deepfake bot service which generated nude images of women within minutes after submitting their photos.



The above graph shows the growth in the number of visits to various deepfake sites within 6 months starting from 2021. Over the years it is observed that the number of pornographic deepfakes produced kept increasing. According to reports so far over 113,000 deepfake videos were uploaded to various websites within nine months of this year whereas in 2022 total of about 73, 000 deepfaked videos were uploaded. This shows that there were about 54% more deepfakes produced this year than there were in 2022.

Financial Frauds

- By Using Audio Deepfakes

High-quality audio deepfakes can replicate a person's voice along with their pitch, intonation and cadence which makes it exceptionally challenging for the human ear to detect abnormalities, leading to the widespread of audio deepfakes in committing financial fraud. A high-profile case reported by the Wall Street Journal in March 2019 is a real-life example of how audio deepfakes can be used to commit financial fraud. In this incident, a British CEO of a UK-based energy company was scammed by cybercriminals who used an audio deepfake to deceive the CEO into believing that he was speaking to the chief executive of a German-based parent company over the phone and deceived him into transferring 243, 000 to a Hungarian supplier [1]. Similarly, in 2020, in Hong Kong, a bank manager was scammed due to a very convincing deepfake call and lost $35 m

- By Influencing Bank Runs

Although bank runs are not primarily caused by social media, the impact of social media along with traditional media and word of mouth can significantly increase the already present rumours about the financial stability of a particular bank, potentially leading to bank runs [6].For example, during the financial instability of the nation, the circulation of a deepfake featuring a government official or a bank executive discussing severe liquidity issues on social media could drastically increase anxiety among depositors[6]. Hence this can lead to a loss of confidence in the affected bank and cause multiple withdrawals hence again leading to a bank run [6].

- By Affecting the Stock Market

Marketplace deception was always present but the introduction of new technology provides new ways of misinformation and one among them is the deepfake technology. A well-convincing deepfake of a CEO being portrayed as behaving unethically or admitting to regulatory fines can amplify the impact of market deception [6].In a similar instance, a deepfake of corporate executives or financial analysts making false statements about the company's performance is circulated can further complicate the issue [6].The trust of investors plays a vital role in determining the stock value of a company hence misinformation and deception can in turn lead to the loss of confidence in the company's operations and products, potentially triggering a drop in the stock value of the company.

Misinformation and Distrust

– **Distrust over the Media** As the deepfake technology is evolving it gets more and more difficult for the general public to distinguish real and fake content. According to one author, "We're not so far from the collapse of reality". Through deepfakes, attackers can make video and audio recordings of public figures such as celebrities, political leaders or government officials saying or doing things they never actually which can damage their reputation and widespread misinformation and manipulate public opinion even after being confirmed as a deepfake can be difficult to recover from. In the upcoming generation, as

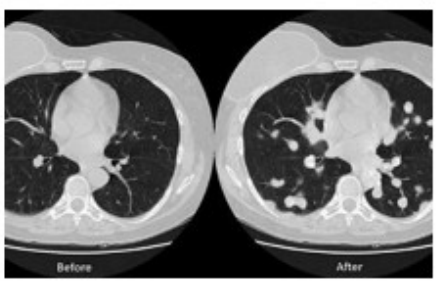
deepfakes get more sophisticated, it will make it more difficult for fact-checkers to correctly recognize unmanipulated content. As a result of all of this, not only does the public lose trust in the media, but journalists and the media themselves may also bear the consequences of not having trustable visual evidence.

– Cause Chaos Nationwide.

In a wider range misinformation caused by deepfakes can cause national threats as it can act as a tool to create chaos in the target country[7].It can be used by rebellious factions and terrorist organisations to provoke incite of opposition among the public[7]. In recent times during the Russia and Ukraine war, a deepfake video of the Ukrainian president Volodymyr Zelenskyy spread around social media in which he ordered his soldiers to lay down their weapons and surrender to Russia which caused great panic and confusion among Ukrainians[4]. The video was not only displayed across social media but also was briefly aired on both news broadcaster's website and Ukrainian televisions.

- Tampering Medical Imagery.

The harm of deepfakes can even spread to the health industry. By adding or removing cancer in a patient's 3D computed tomography scan Israeli researchers showed how attackers can manipulate medical imagery. Three radiologists were assigned to diagnose 100 CT scans of which 70 were tampered and 30 were original out of which not only did radiologists misdiagnose 99% of it but it also fooled a state-of-art AI system that detects cancer.



The above picture shows the medical scan report before and after being tempered using deepfake technology.

2. Influence on Legal Proceedings

Deepfakes have already proven the statement "Seeing is believing" wrong due to its realistic videos and images[2]. According to Gregory "The particular issue is also around the so-called liar's dividend, where it's easy to claim a true video is falsified and place the onus on people to prove it's authentic". In short, the liar's dividend is the phenomenon of misinformation about misinformation with criminals and politicians being the main people gaining benefits from this concept [1]. Politicians can denounce authentic information which damages their reputation as fake and maintain support[7]. Criminals can either deny the truth by generating or spreading altered sounds or images to remove the accusations or claim authentic evidence to be fake, misleading the legal proceedings. In the case of politicians, the liar's dividend can manipulate the public in two ways.

Firstly if there is a political scandal over a politician using deepfakes or spreading false information the public may immediately lower their opinion about that politician but if the politician claims to be a victim of foul play by the opposition to downgrade his reputation, this ultimately causes a sense of uncertainty and confusion among the public in another scenario if the politician denies the accusations made on him of using deepfake technology the public's opinion is anticipated to shift towards the positive side however increasing the uncertainty among the public.

Secondly, when there are multiple accusations against a politician, core supporters or strong co-partisans may seek reasons to continue supporting their preferred politician and avoid feeling conflicted. In that situation, accusing someone or the opposition of spreading fake news using deepfake technology can help supporters justify their loyalty. This strategy can assert the politician's innocence and also criticise the opposition.

3. Advantages of Deepfakes in Various Domains

Immersive Marketing Experiences
The Deepfake technology is an experience that is seductive and memorable with regard to marketing. It has the potential of changing the model of brands advertising their goods or services. Suppose the case, a brand creates a deepfake video in which a celebrity advertises his/her product. It is immediately alluring and leaves a very impressive impression to the customers. The brands can also play with virtual reality (VR) experiences which allow customers to converse with the product in simulated settings and further engage the product even more. In fact, the use of deepfake technology for making ads can save time and energy along with that it can be used to create videos in several languages thereby allowing brands to spread their market and communicate with consumers in their native language. All we require are 20 minutes of audio content and a few video shots of the celebrity or influencer.

Cost-effective video campaigns
One of the major benefits of deepfake technology is its ability to significantly reduce the cost of video campaigns. Traditionally, high-quality videos required expensive equipment to create a video from scratch and hire actors. However, with deepfake technology, brands can cut down on these expenses. They can create videos featuring virtual spokespersons or even digital replicas of celebrities, eliminating the need for costly talent fees and extensive production costs. A deepfake video of Manoj Tiwari, the president of the ruling Bhartiya Janata Party (BJP) in India was the first example of a political party using deepfake technology for video campaigns. In the original video Tiwari was urging the voters to vote the BJP and critiquing his running mate Arvind Kejriwal in the English language which was later deepfaked in the Haryanvi language by matching the movements of the mouth by Tiwari to the voters of Haryana. This fake video by Tiwari has been reported to have reached over 15 million individuals. BJP will apply deepfake technology to reach out to voters in India in over 20 languages.

Enhanced Omnichannel Campaigns
The technology of deepfakes also allows the brands to provide uniform messages on various platforms. The brands can be using deepfake videos with virtual spokespersons to secure their marketing campaign in different channels. As an example, the same virtual spokesperson can be adopted in social media, email campaigns, and other marketing articles. This integrated strategy aids in the formation of an integrated brand experience and the strengthening of the brand identity in the minds of the customers.

Personalisation Experiences
Due to the deepfake technology, it is possible to craft hyper personalized experiences with customers, which puts a spin on their interactions. Deep fakes can be produced by the brands as videos specifically targeted at a particular customer, including their names and personal details. Just picture a video message a brand sends you, calling you by name, it further builds the connection and increases the customization. These intimate details will make clients feel appreciated and empathized with.

Artistic Applications
In addition to marketing, deepfake can be used in art in very powerful ways. Museums, for example, can use deepfake technology to breathe life into historical figures for exhibitions [8]. For example a project named "Dalí Lives" in the Salvador Dalí Museum used deepfakes to create an interactive deepfake of the deceased artist Salvador Dalí in which the visitors can ask questions, hear stories and even take selfies with the artist. Visitors can interact with lifelike simulations, bridging the gap between the past and the present. Additionally, deepfake can be utilized in realistic simulations for training purposes, providing a highly immersive and engaging learning experience in various fields.

Entertainment Purposes
Among the positive uses of deepfakes include the utilisation of visual effects, Snapchat filters, digital avatars, etc[3]. A paper published by researchers at the University of California, Berkeley in August 2018 introduced a phoney dancing app that uses artificial intelligence to replicate skilled dance moves. While much software developed so far focused on the head or other regions of the face, this project extends the use of deepfakes to the complete body.

Currently, many content creators widely use deepfake technology to create memes for entertainment. These can easily be identified as fakes and provide humour through this recognizable imperfection. In 2020 an internet meme surfaced of people singing the chorus of "Baka Mitai" a song from the video game Yakuza 0 in the Yakuza series using deepfake technology. In the series, the player sings the depressing

song in a karaoke minigame. The majority of these memes take their cue from a 2017 video that user Dobbsyrules uploaded, in which he lip-syncs the song.

In 2021, a series of highly authentic deepfake videos of Tom Cruise were released in a TikTok account (@deeptomcruise)[1].These videos had gained over 15.9 million views but at that time training itself lasted two months and involved excessive hours of original Tom Cruise videos to train AI models[1]. Along with that, a skilled actor who can convincingly mimic the movements and mannerisms of Tom Cruise was also an essential element to its success[1]. Currently, deepfakes are getting more advanced with their quality so good that is difficult even for artificial intelligence to detect them.

You can see how precisely Tom Cruise was deepfaked in the image above.

Making of Films

The development of deepfake technology has allowed filmmakers to push the boundaries of creativity and create visually stunning scenes and characters by seamlessly fusing actors with digital elements. This has revolutionized the entertainment industry. The narrative landscape of filmmaking has undergone unprecedented changes as a result of this technological innovation, which has opened up a world of possibilities. These possibilities include the realization of fictional scenarios that were previously thought to be impossible, such as the resurrection of deceased actors on screen[7]. In fact is much cheaper to use deepfake technology than traditional CGI. Using deepfake technology filmmakers can provide more realistic-looking dubbing in various languages[7].

Resurrection

The company My Heritage[1] has introduced a new feature called Deep Nostalgia which animates photographs of deceased loved ones by just uploading old photographs of deceased relatives the system using deepfake technology produces a short animated video of them blinking, smiling and moving their heads. The user can choose from the various combinations of gestures available, and different faces can be animated in the same picture. Through this users could connect with their deceased, loved ones in a unique and emotionally resonant way[7]. Other examples include the resurrection of Joaquin Oliver who was the victim of the Parkland shooting incident. He was brought back to life in 2020 using deepfake technology promoting a gun-safety voting campaign. In this deepfake message, Joaquin is seen pleading with viewers to vote. The ex-Beatles member John Lennon is yet another example of using deepfakes resurrection was murdered in 1980.

Education

Deepfakes will help us to offer an interactive and engaging way of learning experience. The students can be offered to recreate historical events to learn more about events, figures and speeches. Even the deepfakes can be used in terms of military training as they can offer real-life combat situations and soldiers can examine every situation and make their decisions based on the same and it happens to be more in crisis management. The medical professionals can also be assisted in training by realistic medical scenarios as well.

4. Detection Methods to Identify Deepfakes

The identification of deepfake technology is one of the issues that have received massive studies over the past years. Due to the increasing importance of AI-driven misinformation, it is important to create effective methods that will allow detecting and refuting such misleading videos. In the present paper, we are going to examine some of the main approaches that scientists have designed with the purpose of identifying deepfakes.

Subtle Sign Detection

The existence of the subtle signs that can reveal algorithmic manipulation in videos has been revealed through the Detect Fakes project, organized by MIT Media Lab [10]. The project will empower people to recognize deepfakes[10] by on-boarding them on the subtle cues. The appreciation of the significance of these signs can do a great job in addressing the problem of AI-generated misinformation. It is not always easy to identify the slightest indicators of deep fakes: the algorithms used to generate them are complex, and the technology of deep fakes is actively developing. Nonetheless, scholars and professionals are busy at the drawing board trying to come up with methods of examining minute clues that could be used to indicate a deep fake. These are some of the methods and likely nuances that researchers have investigated:

Irregularities in the Facial Features: Deepfakes can be inconsistent in lip syncing and eye movements, as well as facial expressions [2]. In order to discover any strange or unrealistic deviations, sophisticated detection algorithms look at these characteristics in each frame.

Blinking Patterns: Deep fake algorithms might not be able to reproduce a compelling feeling of the features of blinking patterns. Hints at oddities in the frequency of blinking can be retrieved in examining a video to determine the timing of the blink, and how frequent the blink is in addition to the timing, and frequency of the blink[2].

Reflections in the Eyes: Deepfakes may not contain realistic reflections of the eyes. The differences that suggest manipulation are detectable through lighting and reflection of eyes of the subject. Skin Texture and Lighting: The lighting is affected by the skin texture inconsistencies. Deep fakes may not be capable of reproduction the skin texture accurately and reacting to the lighting changes. The investigation of the uniformity of skin texture and lighting frame to frame may show possible indication of manipulation[12].

Temporal Inconsistencies: Deep fakes may not be able to exhibit temporal coherence. It may be useful to consider the motion and feature consistency across time to detect anomalies that may not be apparent in a single frame[12].

Micro-Expressions: Subtle facial expressions that can occur in faces of human beings may not be easy to be imitated by deep fakes. Detection algorithms may focus on these small involuntary motions on the face [10].

Audio-Visual Discrepancies: Audio analysis can be useful with the help of visual content. Deepfakes might have differences between the audio and the audio of the lips, yet another layer of detection is added [10].

Analysis of Metadata: Through the metadata of a video file, it is possible to identify its source and any manipulations that might have occurred. This includes searching file formats, timestamps or compression artifacts discrepancies [10]. **Artifact Analysis:** Deepfake generation processes may add certain artifacts or noise patterns to videos that do not exist in real life. To examine the presence of possible manipulations, it is possible to use image forensics tools to interpret these artifacts.

Consistency with Context: Deepfakes may not be able to accurately imitate the behavior of the subject in a specific scenario. It may be informative to see the general picture of the video and define whether the actions demonstrated are appropriate with the expected behavior of the subject.

5. Machine Learning and Deep Learning

The emergence of advanced technologies such as machine learning, support vector machines, deep learning, and deep neural networks have served as the foundation of the detection of deepfakes. These methods require the training of the techniques to enable them to distinguish real from fake videos. This is conducted based on certain features and clues that can be detected from the features identified in the video processing. By gaining experience through the advancement of artificial intelligence, these methods have already taken over the market with the main goal of improving upon the fast changing technology of the deepfakes that seem to change with every passing day[13]. The use of machine learning also applies in the detection of deep fakes through the use of very sophisticated algorithms meant to identify the subtle characteristics of the manipulation that is often seen in the multimedia content[13]. The performance of the machine learning models to perfect its identification and detection of deep fakes is generally implemented through the use of deep neural networks that are the main methods, for detection

of the hidden implications of patterns and anomalies that may be suggestive of deep fake production[13]. The detection methods would generally use Convolutional Neural Networks (CNN) which are employed in the analysis of images and video pictures [2]. These learn to differentiate the various features and the anomalies in expressions, movement of the body and the movements of the lips no detail may be so obvious to the human observer [2]. There is the use of Capsule Networks which have been tried out as they may have the ability to capture the hierarchical patterns in the character available which means that they are going to be more resilient towards the manipulations. The technology also includes close inspection of audio portions in the totality of the detection methods. Recurrent Neural Networks (RNN) and other methods of deep learning are used to analyze the speech patterns, intonation and other deeper audio indications that may be there. One of the very strong indicators of the presence of the deep fake may be easily identified when there are variations in the movements of the lips as compared and in close inspection observed by the totality that is added in an audio component to the lip movement [2]. Multimodal strategies that consider information from visual and acoustic modalities yield more dependable detection systems, since they take into consideration the merely complementary nature of different media kinds and enable an assessment of the complete content. The detection models are trained on considerable datasets of real and manipulated content. Exposing the models to advanced generative models and adversarial training methods improves their sophistication in recognizing newly emerging deep fake methods [13]. The continuing achievements taking place in machine learning in relation to the detection of deep fakes designate the rapidly changing status of the discipline [13]. The detection procedures pursued by researchers and practitioners are constantly changing since there strive for sophistication with which one can generate deep fakes is changing. In context with deep fakes, as they increase in number, the development of efficient and reliable machine learning based systems will be continuously required for the maintenance of trust and authenticity of digital media [2].

6. Saturation Cues and CNN-LSTM Models

Some detection methods monitor saturation cues to differentiate GAN based generated imagery from that of camera based imagery. Meanwhile, methods using Convolution Neural Networks (CNN) and Long ShortTerm Memory (LSTM) methods have shown some promise correctly identifying video's as being fake or real [11][2]. By analyzing saturation and utilizing these advanced AI models, experts are able to achieve increased accuracy in detecting Deepfakes [9][2].

Saturation Cues: An approach to analyzing color content, more precisely, the saturation of a picture or a video to determine whether a picture has been tampered with, is known as saturation cues. The color gradient and saturation patterns of real content are often difficult to reproduce by deep fakes due to their natural character. In the context of detection:

Colour Anomalies: Saturation cues are used to identify the differences in the color density of the pixel. Deepfakes that have varying saturation rates might have unreliable and unnatural color schemes, which may represent manipulation [2]. Statistical Analysis: Through statistical analysis of the saturation levels in the frames, the algorithms can detect deviations of the common patterns of the real content. Abnormalities in the increases or decreases in saturation may indicate to parts of the picture that have been distorted. Integration with Other Features: The saturation cues of the image may be intertwined with other aspects of the image, e.g. the texture and lighting, to give a holistic analysis. This complex method enhances the strength of deep fake detectors.

CNN-LSTM Models: The CNN-LSTM models take advantage of the advantages that the two construction types offer to sequential data processing by incorporating Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) [11][12]. When identifying deepfakes:

Spatial Analysis and Temporal Analysis: The analysis of the visual components at rest is possible with CNNs as it possesses the ability to detect the spatial details of the images. LSTMs on the other hand are very good at sequential data processing and therefore they are ideal when it comes to recognizing a temporal pattern in a video. With such combinations of architectures, it is possible to do temporal and spatial analysis, which is critical when identifying deepfakes with dynamic content.

Feature Extraction: The CNNs identify patterns and details that are relevant in each frame to feature extract. After these processing attributes have been learned over time, LSTMs query abnormalities or time gaps that may present manipulation.

End-to-End Learning: CNN-LSTM models have the capability to directly learn hierarchical representations on the raw input data and therefore, end-to-end learning can be realized [12]. This proves useful in the event where the subtleties of the deepfakes are intricate and might involve minor temporal and spatial hints.

Multimodal Fusion: Such models are capable of smoothly incorporating information of different modalities e.g. visual and auditory cues [2]. Such a multimodal fusion has the benefit of increasing the overall detection accuracy since it takes into account more features [2].

Hybrid Approaches

A mixture of both has also been proposed as one of the efficient methods of detecting fake images. It is a more powerful approach that incorporates steganalysis feature extractors with traditional deepfake detection models. This method is quite effective in separating tampered and legitimate images involved in the analysis of a range of features and collection of low-level information. The combination of various methods guarantees the use of comprehensive detection capabilities. The most common type of hybrid approach is a combination of the latest machine learning models and traditional computer vision algorithms. Basics of the genuineness of facial traits are given by computer vision algorithms, including the ones that investigate facial scarcities and micro-expressions [11][9]. To learn high-level cues and detect more subtle anomalies that could be indicative of deep fake generation, such cues are subsequently inputted into deep learning models, including Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs)[9].

Multimodal fusion is another relevant element of hybrid approaches. An integrated approach to cross-modal (e.g. visual, aural, even metadata) data can allow detection systems to understand the content more effectively. Inequality of lips and the audio of the same, such as can be strong indicators of manipulation [11]. A more reliable and more complete method of detection is presented through the mixture of audio analysis and visual cues. Hybrid approaches also include forensic techniques that look at the digital evidence generated when the deepfake creation process has taken place. This may involve searching through noise patterns, compression artifacts or other artifacts that the generation algorithm produces[9]. These forensic examinations complement the feature-based techniques by offering an extra degree of examination and verification to identify even expertly crafted deepfakes. Hybrid approaches also include forensic techniques that look at the digital evidence generated when the deepfake creation process has taken place. This may involve searching the noise patterns, compression artifacts or other artifacts that are particular to the generation algorithm. These forensic examinations complement the feature-based techniques by offering an extra degree of examination and verification to identify even expertly crafted deepfakes. In addition, the context-aware models are combined with behavioral analysis, which enhances the sophistication of hybrid approaches. It is possible to tell deepfakes and real content by analyzing the contextual application of actions in a video and understanding the nuances of the behavior of the subject matter [9].

7. Advancements in Deep Learning

Advancements in deep learning methods have helped a lot in the detection of fraudulent photos and videos. However, with the improvement of deepfakes, other issues emerge. Deep learning models need to be enhanced continuously to identify these high quality deepfakes. Researchers will be forced to keep developing innovations to remain abreast with the high rate at which AI generates misinformation [13]. Despite the fact that the fight with deepfake technology is in progress, some progress is achieved due to thorough research and innovative approaches. Researchers are developing trusted mechanisms to identify and avoid deepfakes using machine learning, deep learning and hybrid methods [13]. In order to maintain the integrity of our visual media, detection methods will need to continuously evolve and be modified with the advancement of the deepfake technology.

The breakthroughs in deep learning have resulted in a paradigm shift in the field of many types of technology bringing both opportunities and challenges. In the case of deep fakes, the developments in deep learning have allowed creators of synthetic content to create more extravagant and realistic content. Among these, the ability to reproduce complex patterns and features is one of the main advantages where deep fakes can reproduce gestures, facial expressions, and even details of speech with high accuracy.

Generative models have demonstrated previously unknown capabilities in the generation of realistic visage and sound, and particularly those that are trained on architectures like Transformers and Generative Adversarial Networks (GANs).

Moreover, it is now available to apply the previously trained models to large datasets using transfer learning techniques, accelerating the development of deep fakes and extending their use in other areas of life. Such developments underscore the immense potential of deep learning, but also underscore how urgent it is to come up with deep fake detection methods. Serious countermeasures should be developed to mitigate the dangers of fake news, loss of privacy and trust in digital media that will decrease as deep fakes grow closer to the actual content [13].

8. Conclusion

The advantages of Deepfake technology are also numerous, yet the risks are also severe and should be properly considered. One of the primary problems is the spread of false information. Deepfakes have the potential to create mistrust because people may be misinformed and puzzled by them. Privacy breaches are another issue that may be realized when the personal data is involved in deep fake videos without permission. Moreover, deepfakes could be detrimental to people, particularly when they are directed to them in some malicious or harassing way. The challenge of detecting fake content is cited as a concern over the potential misuse of this technology as deepfakes are invented. This is why it is necessary to be very careful and think of the pros and cons before implementing deepfake technology in other aspects. In summary, the Deepfake technology has various advantages, including immersive marketing, cost-effective video campaigns, enhanced omnichannel strategy, customized interactions, and creative applications. Its potential threats and ethical consequences, nevertheless, should not be neglected, and it requires the sensible application and the elimination of the associated concerns.

9. References

1. T. C. Helmus, “Artificial Intelligence, Deepfakes, and Disinformation: A Primer,” RAND, p. 24, 2022.

2. U. Kosarkar, G. Sarkarkar and S. Gedam, “Revealing and Classification of Deepfakes Video’s Images using a Customized Convolution Neural Network Model,” Procedia Computer Science, vol. 218, pp. 2636-2652, 2023.

3. M. Mustak, J. Salminen, M. Mäntymäki, A. Rahman and Y. K. Dwivedi, “Deepfakes: Deceptions, Mitigations, and Opportunities,” Journal of Business Research, vol. 154, Article 113368, 2023.

4. B. N. Jacobsen and J. Simpson, “The Tensions of Deepfakes,” Information, Communication & Society, pp. 1-15, 2023.

5. T. T. Nguyen, Q. V. H. Nguyen, D. T. Nguyen, D. T. Nguyen, T. Huynh-The, S. Nahavandi, et al., “Deep Learning for Deepfakes Creation and Detection: A Survey,” Computer Vision and Image Understanding, vol. 223, Article 103525, 2022.

6. Z. Akhtar, “Deepfakes Generation and Detection: A Short Survey,” Journal of Imaging, vol. 9, no. 1, Article 18, 2023.

7. J. Bateman, “Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenarios,” Carnegie Endowment for International Peace, 2020.

8. M. Westerlund, “The Emergence of Deepfake Technology: A Review,” Technology Innovation Management Review, vol. 9, no. 11, 2019.

9. N. Caporusso, “Deepfakes for the Good: A Beneficial Application of Contentious Artificial Intelligence Technology,” in Advances in Artificial Intelligence, Software and Systems Engineering, Springer International Publishing, pp. 235-241, 2021.

10. K. Somoray and D. J. Miller, "Providing Detection Strategies to Improve Human Detection of Deepfakes: An Experimental Study," Computers in Human Behavior, vol. 149, Article 107917, 2023.

11. N. Guhagarkar, S. Desai, S. Vaishampayan and A. Save, "Deepfake Detection Techniques: A Review," VIVA-IJRI, vol. 1, no. 4, pp. 1-10.

12. A. Eberl, J. Kühn and T. Wolbring, "Using Deepfakes for Experiments in the Social Sciences: A Pilot Study," Frontiers in Sociology, vol. 7, Article 907199, 2022.

13. T. T. Nguyen, Q. V. H. Nguyen, D. T. Nguyen, D. T. Nguyen, T. Huynh-The, S. Nahavandi, et al., "Deep Learning for Deepfakes Creation and Detection: A Survey," Computer Vision and Image Understanding, vol. 223, Article 103525, 2022.

[1,2,3]*Faculty, Geethanjali College of Engineering and Technology, Cheeryal (V), Keesara (M), Hyderabad, Telangana, India- 501301.*

[4]*B.Tech Student, Geethanjali College of Engineering and Technology, Cheeryal(V), Keesara(M), Hyderabad, Telangana, India- 501301.*

*(*Corresponding author. E-mail:* ***bnsreekar28@gmail.com, nemani.subhadra@gmail.com****)*