International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 08 | Aug 2024 www.irjet.net p-ISSN: 2395-0072

How does the rise of deep fake technology undermine democratic processes and erode public trust–what strategies can be implemented to mitigate these threats?

SHRAVIL AGGARWAL

Abstract:

Deep fakes are synthetic media, in which artificial intelligence morphs someone's face on an already existing video of an actor. This research explores the evolution of deep fakes, the risks they present, and their broader meaning in political, psychological, and social backdrops.

Deepfakes have raised serious concerns due to their ability to manipulate information. Developed for entertainment purposes, deepfakes have quickly evolved into a tool for misinformation, harassment, and fraud. The current state of deep fake technology, although still developing, is already capable of causing significant harm to pre-existing democratic models.

Politically, deepfakes pose a serious threat to democracy and societal trust. They can be used to spread false information, manipulate public opinion, and destabilize political systems. The psychological impact of deepfakes is equally concerning, as they can be used to harass individuals, create false narratives, and undermine personal relationships. The potential for deepfakes to damage an individual's reputation, private life, and profession is significant, as they are known to be weaponized and to create damaging content that appears real.

Addressing the threats posed by deepfakes requires practical and feasible approaches discussed in this research, such as developing AI-driven detection tools, legal frameworks, and public awareness campaigns. It is crucial to explore future trends in AI and understand how they might be detrimental to the public.

Introduction:

Artificial Intelligence (AI) has come a long way since its invention in the 1950s by Alan Turing and John McCarthy. Since then, AI has been a topic of many studies. Deep Fakes, a significant advancement in AI and Machine Learning, though a relatively new topic, has sparked considerable interest and debate in the community. By leveraging sophisticated algorithms, particularly deep learning techniques, deep fake softwares can create hyper-realistic but synthetic videos/audio/images that cannot be discerned from reality by an untrained eye. This has raised several security concerns, not only because it contributes to the spread of misinformation, but it also invades personal privacy and can ruin the reputation of public figures. It creates mass hysteria in the public, which is detrimental to national security.

Although Deep fakes are crucial in the cinematography and editing industries, the risks involved in the unrestricted use of deep fakes exceed their benign nature. The potential for misuse and breaking the societal trust is too high. Hence, it is crucial to identify the victims of this rampant technology and to suggest effective solutions (temporary or permanent) for this threat that deep fakes pose.

Many research papers provided solutions to these threats. Some of them even discussed the failure of the judicial system due to the manipulation of truth. However, the victims of this technology remain anonymous and are still suffering the consequences, of something they didn't do because "Legal loopholes don't help victims of deepfakes abuse". This research aims to unveil these problems related to deep fakes and hopes to provide some solutions.

Literature Review:

- **Deep fakes, fake news, and what comes next** by Sean Dack (*The Henry M. Jackson School of International Studies, 2019)*

The article concentrates on how in the 2016 American presidential election as well as in the 2017 French presidential election, Russian syndicates targeted the then-candidates. For example, Emmanuel Macron was a target of such a campaign in the 2017

French election. It has been repeatedly shown that Deep Fakes have been a bane to the political and social world. This campaign uncovered the history-altering information war and marked the underlying terror of Deep Fakes. Sean Dack writes "The Machine Learning format works such that the more audio and video it is fed, the more realistic it seems … and the more people will choose to believe in them … With more time passing, it seems to be less science fiction and more the reality of ourworld." Public opinion can easily be swayed by a single fake audio of a leader indulging in bribery, or a leader committing malpractice. The use of deep fakes can be corresponded to an individual level.

- **Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News** by Cristian Vaccari and Andrew Chadwick *(SagePub, 2020)*

In April 2018, a video went viral on WhatsApp. The footage showed a group of children playing cricket on the street. Two men on a motorbike ride up and grab one of the smallest kids then speed away. This "kidnapping" video created widespread confusion and panic, culminating in 8 weeks of mob violence that killed at least nine innocent people (BBC News, 2018). This footage was a fake—an edit of a video from a public education campaign in Pakistan, which was designed to raise awareness about abductions. The original video opened with one of the hired actors' faces, signaling the parents to look after their children. This part was cleverly edited in the video that went viral on WhatsApp, leaving just the footage of a child being shockingly kidnapped leading to riots and killings for no reason at all. This raises the question of whether any of these communal violence videos are real. The deep fakes not only spread misinformation but also affect public opinion in such a way that propaganda has now become ingrained in our brains.

- **Deepfakes and Their Impact on Society** by Neill Jacobson *(OpenFox, 2024)*

Deepfakes, generated through artificial intelligence and deep learning, have become more advanced, making it challenging to differentiate between genuine and fictitious content. This technology is rapidly evolving, with the number of deepfake videos and audio doubling every six months, projected to reach 8 million by 2025. Social media platforms play a significant role in the spread of deepfakes, often without users being aware that the content is not real, which intensifies problems such as political manipulation, misinformation, and privacy violations. Deepfakes have already been utilized to sway political events, fake news, and exploit individuals' images without their permission, as demonstrated during the 2023 actors' strike. Businesses are also vulnerable, facing threats like CEO impersonations and deceptive content that can harm their reputation and trustworthiness. To address these challenges, it is essential to create and implement advanced deepfake detection technologies, raise awareness among the public and businesses about the associated risks, and establish robust regulations. Governments must revise and enforce laws to tackle the misuse of deepfakes, with international cooperation being vital for global safety. Ensuring transparency in content creation, such as through watermarking and third-party validation, can assist users in verifying authenticity. In the future, the progression of deepfakes may lead to a decline in trust in digital content, which could have serious consequences in fields like law enforcement and justice. Companies like Issured are tackling this issue by developing platforms like MeaConnexus, which offer secure, tamper-proof online environments for interviews, fostering trust and integrity, but this is still insufficient as technology continues to evolve rapidly.

Methodology:

This research has been formulated using data collection from several articles, referencing the deep fake issue. The most likely webpages from where data has been collected are trusted global news sites, which have proved time and time again the importance of authenticity and method.

All the data has been collected through secondary sources, and each site has been given due credit in the References Section.

Results:

Many victims have been recognized these past few years. Most of them are targeted due to their fame (for example, in early 2024, pop star Taylor Swift became the center of attention because of the disturbing and explicit media her face was pasted on). But others, less famous ones are 90% of the time victimized either because of their status or their relations with the deep fake user.

Through my research of the topic, the following results could be inferred from the referenced studies:

- "Using Artificial Neural Networks, which are based on 'real' or biological neural networks, systems can learn how to perform tasks by looking at examples.

  Several technologies can be used for this, but the most popular is based on what is known as Generative Adversarial Networks (GAN) and Variational AutoEncoders." [Bart Van der Sloot, 2022] A GAN is a machine learning model that trains deep fakes on how to generate a more authentic media set. It has a Generator G(x) and a Disciminator D(x). When the real data x is fed to the system, the Generator tries to compete with this data and feeds another data set corresponding to x. The Discriminator then decides whether the Generator pumped out any useful data. If not, it classifies it as fake and backpropagates the data. If it is close to the original data, it classifies it as real and progresses further.

- "A 2019 survey in the United Kingdom, Australia, and New Zealand found that about 14.1% of respondents aged between 16 and 84 have experienced someone creating, distributing, or threatening to distribute their digitally altered media representing them explicitly [without any legislation on it]" [Asher Flynn, 2024]. The laws regarding the creation/production of fake images, videos, or any sort of media are quite ambiguous not only in second and third-world countries but also in first-world countries like the USA, Australia, and the UK. To produce a reliable justice system, governments should aim to strengthen such lax laws and legislate a sterner punishment for those who possess or create deep fakes. The developers should be heavily fined too.

- According to the article published by Neill Jacobson titled "Deep Fakes and their Impact on Society", the number of deep fakes online in 2023 alone was 500,000! And it is doubling every 6 months. By the end of 2025, this number is estimated to skyrocket to an enormous 8 million.

- The development of effective deepfake detection tools is lagging the technology's advancement. While some AI-driven tools can detect certain deepfakes, the technology's constant evolution makes it difficult to create foolproof detection methods. Researchers have highlighted the need for ongoing development in this area to keep pace with emerging threats.

- Deepfakes technology has rapidly advanced, becoming increasingly sophisticated and accessible. Researchers found that the quality of deepfakes has improved to the point where they can convincingly mimic real people in both audio and video formats.

- Deepfakes have significant implications for politics and society. The research showed that deepfakes could be used to spread misinformation, disrupt political processes, and manipulate public opinion. The potential for deepfakes to be used in propaganda, election interference, and social unrest is a major concern.

- The psychological effects of deepfakes on individuals are profound. Victims of deepfake harassment, suffer severe emotional and reputational harm. The research also found that deepfakes could erode trust in media and personal relationships, as people become more skeptical of the authenticity of digital content.

Discussion:

The exploration of deepfake technology reveals an issue with significant adversities across various domains, including politics, psychology, social interactions, and individual privacy. The development of deepfake technology, fueled by advancements in artificial intelligence (AI), has led to both innovative applications and significant risks. This section discusses the research findings, their implications, the limitations of the study, and potential avenues for future research.

Interpretation of Results

The research indicates that deep fake technology, while still in its recent stages, has rapidly evolved, allowing for the creation of highly realistic and convincing fake media. This technology is not only used for benign purposes like entertainment but also poses severe risks, such as the manipulation of public opinion, the undermining of trust in media, and the potential harm to individuals' reputations and professional lives. The study underscores that deepfakes can be weaponized, particularly against public figures and vulnerable populations, leading to significant psychological and social consequences.

The findings align with existing literature that highlights the dual-use nature of AI technologies, where the same tools that drive innovation and creativity can also be exploited for malicious purposes. The study further confirms the growing concern among researchers, policymakers, and the public about the potential for deepfakes to disrupt social order, spread misinformation, and erode trust in institutions.

Implications

The implications of these findings are far-reaching. Theoretical implications include the need for a deeper understanding of the psychological impact of deepfakes on individuals and society.

Practically, the research suggests the urgent need for the development of more robust detection tools and the implementation of legal frameworks to manage and mitigate the risks associated with deepfakes. The study also emphasizes the importance of public awareness and education in recognizing and responding to deepfake content.

In terms of policy, the research suggests that governments should consider implementing regulations that require digital media to carry verifiable markers of authenticity, such as blockchain-based digital watermarks. Furthermore, collaborations between tech companies, academic institutions, and governments are essential to develop and deploy effective countermeasures against deepfakes.

Limitations

Despite the comprehensive nature of the research, there are several limitations to consider. The study primarily focuses on the current state of deepfake technology and its immediate implications. However, as AI continues to evolve, the technology will likely become even more sophisticated, potentially outpacing current detection methods. Additionally, while the research touches on the psychological and social impacts of deepfakes, further studies are needed to quantify these effects and explore the long-term consequences.

The study also has a limited geographic scope, primarily focusing on the impact of deepfakes in Western societies. Future research should explore how deepfakes are perceived and managed in different cultural and political contexts.

Conclusion:

In conclusion, deepfake technology presents both opportunities and significant challenges. While it offers new ways for creativity and innovation to bloom, it also poses substantial risks to individuals and society. Addressing these challenges requires a combination of technological innovation, legal regulation, public education, and ethical reflection. As deepfake technology continues to evolve, it is crucial for researchers, policymakers, and the public to stay vigilant and proactive in managing its impact. So what's an effective solution for these Deep Fake threats that have stagnated our democracy, to the point that hoping for a judiciary that can see through these lies, is considered too much to ask for? Some solutions can be proposed for the government to implement but most solutions come from a technological perspective. Other out-of-box options may arise from a moral and ethical standpoint. Some proposed solutions in currently reviewed articles may include:

1. **Advanced Detection Tools**

    a. **AI-Based Detection**: AI tools that can identify deepfakes by analyzing inconsistencies in videos, such as unnatural facial movements, lighting discrepancies, or audio-visual mismatches should come into play. This alone will greatly improve the situation, but it is not feasible at the moment.

    b. **Watermarking and Verification**: Implement digital watermarking and content verification systems that can trace the origin of media files, making it easier to detect alterations. Though unethical, it is a more viable option, leading to effective legislation.

2. **Legislation and Regulation**

    a. **Comprehensive Laws**: Governments should enact and update laws specifically targeting the creation, distribution, and use of deepfakes, particularly in cases of harassment, defamation, and election interference.

b. **International Cooperation**: Since deepfakes can spread globally, international agreements and cooperation are crucial to create a unified legal approach to combat this issue.

3. **Public Awareness and Education**

a. **Awareness Campaigns**: Launch public awareness campaigns to educate people about the existence of deepfakes, how they can be used to manipulate information, and ways to spot them.

b. **Critical Media Literacy**: Incorporate media literacy programs into educational curriculums to teach individuals, especially young people, how to critically evaluate the authenticity of digital content.

4. **Support for Victims**

a. **Legal Support**: Provide legal assistance and resources for individuals targeted by deepfakes, helping them pursue justice and repair their reputations.

b. **Psychological Support**: Offer psychological counseling and support services to victims of deepfake harassment, addressing the emotional and mental health impacts.

By implementing these solutions, society can better manage the risks associated with deepfakes, ensuring that this powerful technology is used ethically and responsibly while minimizing its potential for harm.

References:

Dack, Sean. "Deep Fakes, Fake News, and What Comes Next." *Jackson School of International Studies*, 20 March 2019,

https://jsis.washington.edu/news/deep-fakes-fake-news-and-what-comes-next/

Vaccari, Cristian, and Andrew Chadwick. "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News." *Social Media +*

Society, 2020, https://doi.org/10.1177/2056305120903408

Jacobson, Neill. "Deepfakes and Their Impact on Society." *CPI OpenFox*, 26 February 2024, https://www.openfox.com/deepfakes-and-their-impact-on-society/

Van der Sloot, Bart, and Yvette Wagensveld. "Deepfakes: regulatory challenges for the synthetic society." *Computer Law & Security Review* 46 (2022): 105716,

https://www.sciencedirect.com/science/article/pii/S0267364922000632

Flynn, Ashler. "Legal loopholes don't help victims of sexualized deepfakes abuse." *Monash Lens*, 18 April 2024,

https://lens.monash.edu/@politics-society/2024/04/18/1386624/legal-loopholes-dont-help-victims-of-sexualised-deepfakes-abuse