iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61625

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Deepfake, Disinformation and Social Media

Sushant Bhatta[1], Prabesh Chhetri[2]
St. Xavier's College, Maitighar, Nepal

Abstract: ***The advancement of Artificial Intelligence (AI) has uplifted the era of new possibilities. Along with the limelight positive aspects, there are some neglected negative aspects as well. With growing stardom, popularity and the need for power, AI and deep fakes have seen tremendous growth. Challenging the notions of privacy, identity and consent, AI clones and deepfake technology overshadow the reality and authenticity, producing fakely-realistic audiovisual contents. This paper deals with the impacts of AI clones and deepfaces, and lists out the possible solutions to reduce their impact.***

I. INTRODUCTION

In this rapidly growing technological world, data and information are as delicate as a researcher working on a nuclear power plant. A single mistake and a total havoc. Likewise, a "hot" "controversial" topic, usually from an influential person, requires a few seconds to spread its wrath. It may be true or maybe falsely created, but chaos is guaranteed. Social media: a primary source of information is also a source for disinformation. Presently, one out of five Internet users get their news via YouTube, second only to Facebook. Given the ease in obtaining and spreading misinformation through social media platforms, it is increasingly hard to know what to trust, which results in harmful consequences for informed decision making, among other things.[4]

This rise in popularity of video highlights the need for tools to confirm media and news content authenticity, as novel technologies allow convincing manipulation of video (Anderson, 2018)..[4] Deepfakes only surfaced on the Internet in 2017, scholarly literature on the topic is sparse.[4]

The technology relies on deep learning techniques, particularly generative adversarial networks (GANs), first introduced by Ian Goodfellow and his team in 2014. GANs consist of two neural networks, a generator, and a discriminator, that work together to create synthetic data that closely resembles accurate data. This breakthrough laid the foundation for the development of deepfake technology.[8]

A. Deepfake

The term "deepfake" is derived from a combination of "deep learning" and "fake."[8] The term came to be used for synthetic media in 2017 when a Reddit moderator created a subreddit called "deepfakes" and began posting videos that used face-swapping technology to insert celebrities' likenesses into existing pornographic videos.[7] Deepfake videos are synthetically altered footage in which the depicted face or body has been digitally modified to appear as someone or something else. Such videos are becoming increasingly lifelike, and many fear that the technology will dramatically increase the threat of both foreign and domestic disinformation. These synthetic videos' images are developed through generative adversarial networks (GANs). The GAN system consists of a generator that generates images from random noises and a discriminator that judges whether an input image is authentic or produced by the generator. [1]

B. Disinformation

Disinformation is getting an upgrade. A primary tool of disinformation war-fare has been the simple meme: an image, a video, or text shared on social media that conveys a particular thought or feeling (Sprout Social, undated). Russia used memes to target the 2016 U.S. election (DiResta et al., 2019); China used memes to target protesters in Hong Kong (Wong, Shepherd, and Liu, 2019); and those seeking to question the efficacy of vaccines for coronavirus disease 2019 used memes as a favorite tool (Wasike, 2022; Helmus et al., 2020).[1]

C. Social Media

Social media refers to the means of interactions among people in which they create, share, and/or exchange information and ideas in virtual communities and networks.[2] Social Media has become a necessity for the general public. Platforms on social media are increasingly starting to provide online commerce, jobs, and access to education etc. Social media allows us to communicate and build social relationships with the public.

Adults and even children are among the most active users of social media. There are more than 3.8 billion social media users in this world. In Indonesia, from 274.9 million people, 170 million of them are active users of social media.[3] The number of social media users has increased especially during the covid-19 pandemic. With increasing social media uses, deep fakes and disinformation follows a growing trend. .

II. CREATION OF DEEP FAKES

Deepfake creation is now easier than ever before, thanks to the increased sophistication possible through deep neural networks and realistic content offered through GANs (Whittaker et al., 2021). These video manipulations have been witnessing an alarming rise in numbers, with over 85 thousand harmful deep fake videos detected up to December 2020, with the numbers doubling every six months (Petkauskas, 2021; Sensity, 2020).[5]

1) The growing sophistication of the GAN approach has meant that it is now possible to create increasingly convincing deep fakes which could go undetected by the untrained eye (Maras & Alexandrou, 2019).
2) Digital forensics has largely focused on detecting low-level alterations in images, while research on the detection of face manipulation is growing but still sparse (Maras & Alexandrou, 2019). This, in turn, would mean that it would take years before deepfakes are detected reliably by the systems (Porter, 2020).
3) The widespread availability of deepfake creation technologies has meant that it is easy to produce deep fakes without the need for expert intervention (Gosse & Burkell, 2020)
4) A fourth factor in the form of network effects may also have a role in the creation of deepakes, while it is unclear whether this factor is exclusive to a specific region.[5]

III. IMPACT

A. *Positive Impact*

1) *Accessibility*

Artificial intelligence can create tools that can hear, see, and, soon, reason with increasing accuracy.[9] It gives them independence by making accessibility tools smarter, affordable, and personalizable. Moreover, AI-based tools can make solutions more accessible to everyone in a simple and generalized way.

2) *Education*

Deepfakes can assist a teacher in delivering engaging lessons. Also, these lessons would go beyond traditional visual and media formats. A synthetic video of reenactments or a voice and video of a historical figure will have a greater impact. It might increase engagement and be a more effective learning tool. In 1963, President John F. Kennedy was on his way to deliver a speech in Dallas when he was assassinated. The beloved politician did not get to bring those words to the world that day, but thanks to modern technologies and innovative techniques, we can hear them now.[10] With deep fake technology, the same can be done on a bigger scale. Historical figures can be brought back to life, and more interactive historical classes can be created for schools. This practice already exists, and deepfakes can take it to the next level. [10]

3) *ART*

Deepfake has the potential to democratize expensive VFX technology. It can also become a powerful tool for independent storytellers at a fraction of the cost.[9] The emergence of a network of film producers, researchers, and AI technologists in the UK signals a proactive approach to harnessing the positive potential of deepfakes in creative screen production.[10]

4) *Digital Reconstruction & Public Safety*

Reconstruction of crime scenes requires reasoning as well as evidence. Artificial intelligence-generated synthetic media can aid in the reconstruction of a crime scene. Also, a team of civil investigators created a virtual crime scene using cell phone videos. It used autopsy reports and surveillance footage.[9]

5) *Innovations*

Nowadays, synthetic voices are in rising trends in instagram reels and youtube shorts which is possible through AI and deep fakes. Customers' faces, bodies, and even micro mannerisms can also make an exciting app. This will generate a deep fake and allow them to try on the latest fashion trends. [9] Graphics designing has become easier than before. AI-generated graphics and imagery can speed up game development in the video gaming industry.[9]

B. Negative Impacts

1) Social Impact

Only a few studies have examined the social impact of deep fakes,4 despite the popularity of face swap platforms (e.g., the Zao app5). Although there have been dozens of studies looking at false memory acquisition and social influence from altered still images (i.e., Garry and Wade6), the psychological processes and consequences of viewing artificial intelligence (AI)-modified video remain largely unstudied. [19] As the technology underpinning deepfakes continues to improve, we are obligated to confront the repercussions it will have on society. The introduction of educational initiatives, regulatory frameworks, technical solutions, and ethical concerns are all potential avenues for addressing this matter. [21]

In a study looking at the effect of deep fakes on trust in the news, Vaccari and Chadwick19 found that although people were unlikely to be completely misled by a deepfake (at least with the technology they were using), exposure to the deep fake increased their uncertainty about media in general. [19] This, in response, prompts users of social media platforms to exercise caution and verify information from multiple sources as a means of ensuring the credibility of the news. [22]

Besides, the rapid utilization of the technology led to the creation of adult content and materials with the potential to be exploited for blackmail purposes.[22] There are also many interpersonal effects of deep fake videos such as their capacity to modify our memories and even implant false memories in people's minds. This can alter a person's attitude towards another without any actual reason. [23]

During specific periods of heightened sensitivity or significant events, the likelihood of the dissemination of deepfake content increases substantially. For instance, during elections, it is highly likely that the opponents will share deepface videos to bring each other down.[24] Therefore, if deepfakes are continuously shared on the internet, the next generation of users will find it difficult to rely on internet-based information. Therefore, Society needs to recognize that sharing deep fakes can help deep fake creators fulfil their desires to spread misinformation.

2) Political Impact- Election

The threat became serious when the Defense Advanced Research Project Agency (DARPA), an agency of the U.S. Department of Defense, realized that even an unskilled person can tamper any visual media [12]. Siekierski [12] states that when a fake video of Barack Obama, former U. S President, was released by researchers at the University of Washington in July 2017, the general public was warned about the potential disruptive interference of deep fake technology. Afterward, in May 2018, a low quality deep fake video of President Donald Trump was uploaded to social media, telling Belgians to withdraw from the Paris Climate Change agreement. This showed that this technology is continually evolving and has the ability to mislead a large segment of the public.

Figure 1: How Deepfake Videos are Produced

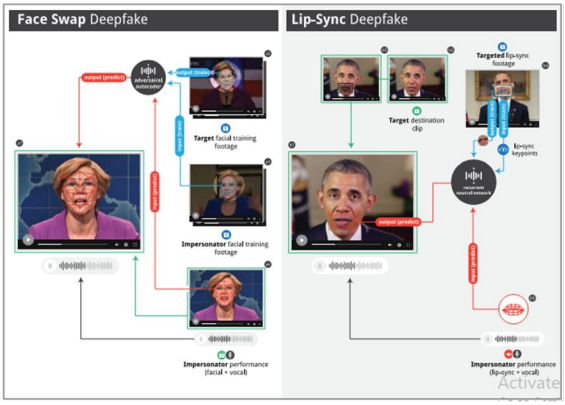


Fig. [15]

Moreover, many deepfakes are produced by essentially swapping the faces of political elites with their impersonators' in comedic sketches, including the example depicted in Figure [15] with Senator Elizabeth Warren inserted into Saturday Night Live actor Kate McKinnon's performance.Deep Fakes that swap the face of a target (e.g., President Barack Obama) with an actor (e.g., Hollywood actor Jordan Peele) – dubbed face-swaps in Figure [15] – are synthesized via a particular class of artificial neural networks called Adversarial Autoencoders.[15]

Figure 2: Voters were unable to discriminate between a real video and a deepfake (Study 1)

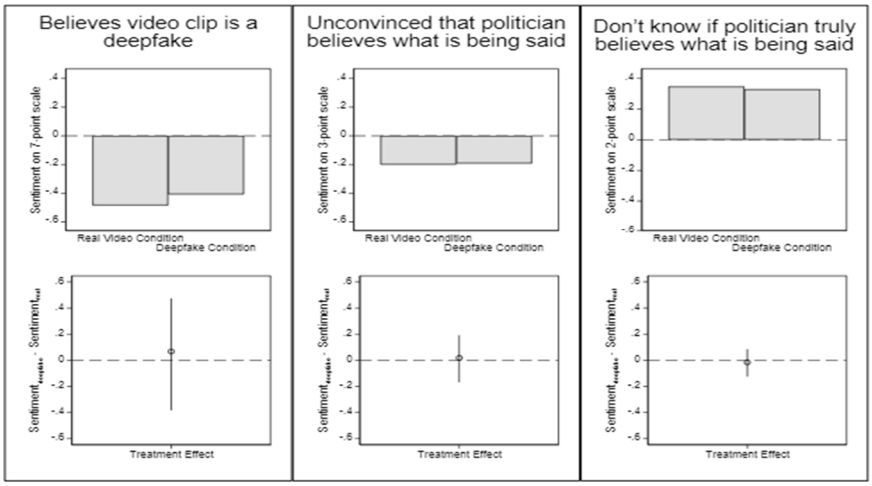


Fig [18]

Our first question is whether subjects can discern between deep fakes and real videos. In Figure [18], we show that participants were unable to do so. The leftmost column in Figure [18] uses an outcome measure that asked participants how much they agreed with the statement "This video was doctored, manipulated and/or faked by a computer (i.e. it is a 'Deep Fake')" on a 7-point agree/disagree scale. Participants in the real video condition had a mean sentiment of -0.48, which is nearly equidistant between "somewhat disagree" and "neither agree nor disagree." The mean sentiment of participants in the deepfake condition was not significantly different from that value (only 0.07 points higher). As shown in the remaining two columns, treatment effects were even smaller across alternative measures of disbelief. In other words, a well-made deepfake video is unlikely to be detected by the naked eye of a typical American voter. In fact, it appears that Americans are somewhat confident that a given political video is not manipulated by a computer—even when it is.[18]

Figure 3: Warnings induced disbelief in accompanying video regardless of whether the video is real or fake

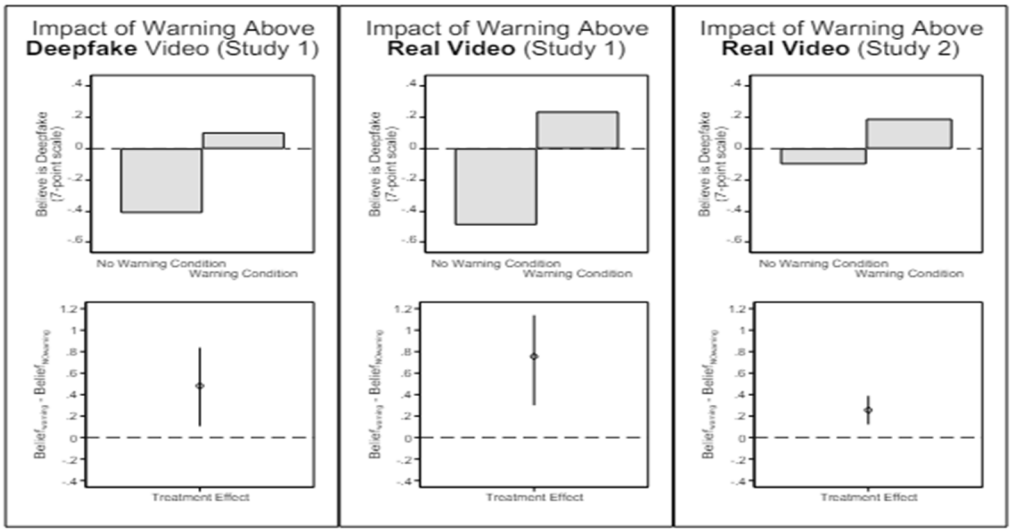


Fig. [18]

From Figure 18, we observed that when participants watched a deepfaked video without a warning, they were fairly confident that what they were watching was not a deep fake (-0.4 points on our 7-point scale); when the warning was added, they became more uncertain about whether the video was a deepfake (0.1 points).[18]

3) Global Economy

A study conducted by the University of Baltimore and Cybersecurity firm CHEQ, found that in 2020, fake news will cost the global economy $78 billion.[13] Synthetic identities could be created, whereby they take elements of a real or fake identity and attach them to a non-existent individual. Deepfakes also represent a threat to economic actors. In a study by Euler Hermes, published at the end of 2021, two thirds of the companies surveyed stated that they had been victims of fraud attempts in the last twelve months.[14]

Table 1. Impact of false content on the economy

FAKE NEWS	CONSEQUENCE	ECONOMIC OUTCOME
COVID 19 vaccination	Deaths	Decline of the labour force
5G network	Local regulations and aggravated circumstances	Slower development of local areas
Elections	Radical political options	Radical trade restrictions

Fig. [20]

Firstly, promoting fake news about COVID 19 and conspiracy theories impact people's minds about quarantine and measures regarding behavior of people and raise the percentage of deaths in this pandemic. More deaths mean a change of structure in the labour force, which could impact wages.

Secondly, risks of promoting a false content about infrastructure development of the 5G network could make circumstances in the local area aggravated, which could lead to slower development of these areas.

Thirdly, placing fake news about election candidates and creating fake content like deep fake videos could impact people's minds to vote for another candidate. This could help radical political options to win the elections. Later it could lead to radical policies like radical trade restrictions or derivation of bad relationships between the countries. [20]

4) Bullying

Deepfakes can be used to create fake videos or images that are sexually explicit, violent, or otherwise harmful. These videos can be used to harass, intimidate, and abuse others, especially women and marginalized groups. [11] A 2023 Internet Watch Foundation (IWF) report warned of increasing AI-generated child sexual abuse material (CSAM). They identified over 20,000 of these images posted to one dark web CSAM forum over a one-month period. They judged more than half of these as "most likely to be criminal."[16]

"AI-generated child sexual abuse material is already filling the queues of our partner hotlines, NGOs and law enforcement agencies. The inability to distinguish between children who need to be rescued and synthetic versions of this horrific material could complicate child abuse investigations by making it impossible for victim identification experts to distinguish real from fake," Ms Inman Grant said.[17] Hard to distinguish real and fake video, especially for bullying children and vulnerable people cause immense harm.

IV. SOLUTIONS

Knowingly or unknowingly, deep fakes and disinformation have become a part of our life. And the fact is that everyone is vulnerable to it. Though cannot completely be stopped, some solutions help minimize it:

A. Detection of Deepfakes

1) Across all videos, substantial percentages of respondents did not correctly identify a video's authenticity when receiving deep fake *or* authentic videos. [6] The graph below presents the percent respondents identified the original video and the deep fake video:

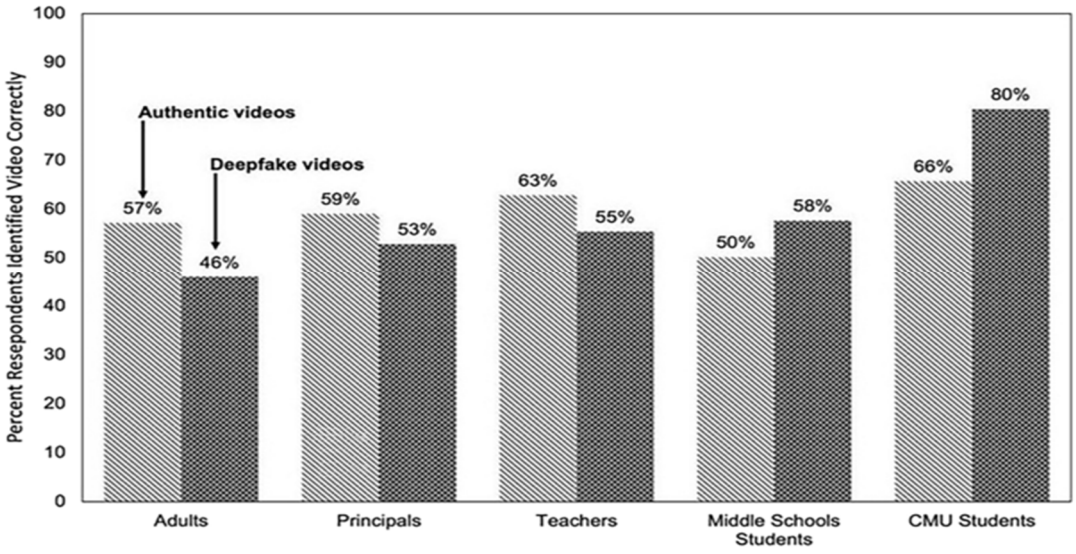


Fig. https://www.nature.com/articles/s41598-023-39944-3

Several public and private organizations have started to realize the importance of this technological advancement by launching bespoke initiatives aimed at deepfake detection.
Several platform players have taken notice of this trend and are devising strategies to combat it. Facebook is funding initiatives aimed at creating a corpus of videos that can aid researchers in combating deepfakes through precise detection mechanisms. For eg. Twitter is taking a similar approach through an intricate set of rules which go about identifying tweets carrying manipulated content and warning users about them alongside carrying the authentic source wherever possible and eliminating the doctored content. [5]

2) Another approach toward mitigating deepfakes is content provenance: Through the Content Authenticity Initiative (CAI), Adobe, Qualcomm, Trupic, the New York Times,and other collaborators have developed a way to digitally capture and present the provenance of photo images (CAI,undated-a). Specifically, CAI developed a way for photographers to use a secure mode on their smartphones,which embeds critical information into the metadata of the digital image. This secure mode uses what is described as "cryptographic asset hashing to provide verifiable,tamper-evident signatures that the image and metadata hasn't been unknowingly altered" (CAI, undated-b). When photos taken with this technology are subsequently shared on either a news site or a social media platform, they will come embedded with a visible icon: a small, encircled i (see Figure 6). When clicked, the icon will reveal the original photo image and identify any edits made to the photo.[1]

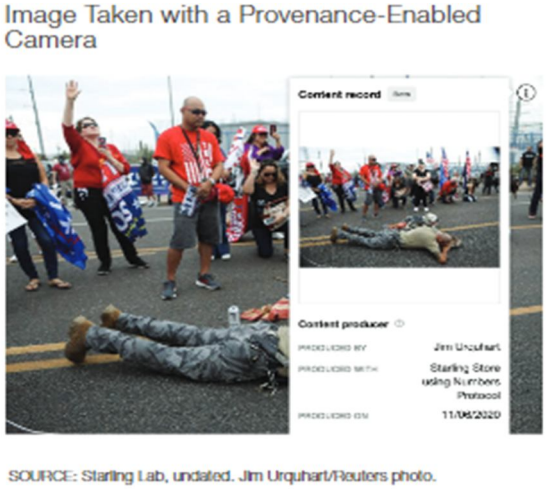


Figure [6]

3) Another approach to countering the risks associated with deep fakes is through regulation and the creation of criminal statutes. Several such initiatives have been either proposed or adopted. Several bills have been adopted at the state level in the United States. In 2019, Texas passed a law that would make it illegal to distribute deep fake videos that are intended "to injure a candidate or influence the result of an election" within 30 days of an election (Texas State Legislature SB-751, 2019). [1]

4) Governments, companies, academics, journalists, and all parties of the issue should strive to raise awareness of the individuals about artificial intelligence, including deep fakes, in terms of news security. It would be beneficial to prevent the use of these technologies for 'malicious purposes' legally.[25]

5) In spite of the precautions followed, if anyone encounters deep fakes, swift action is crucial. Begin by collecting the URL of the website hosting the content and register your case with StopNCII.org. (https://stopncii.org/).[26]

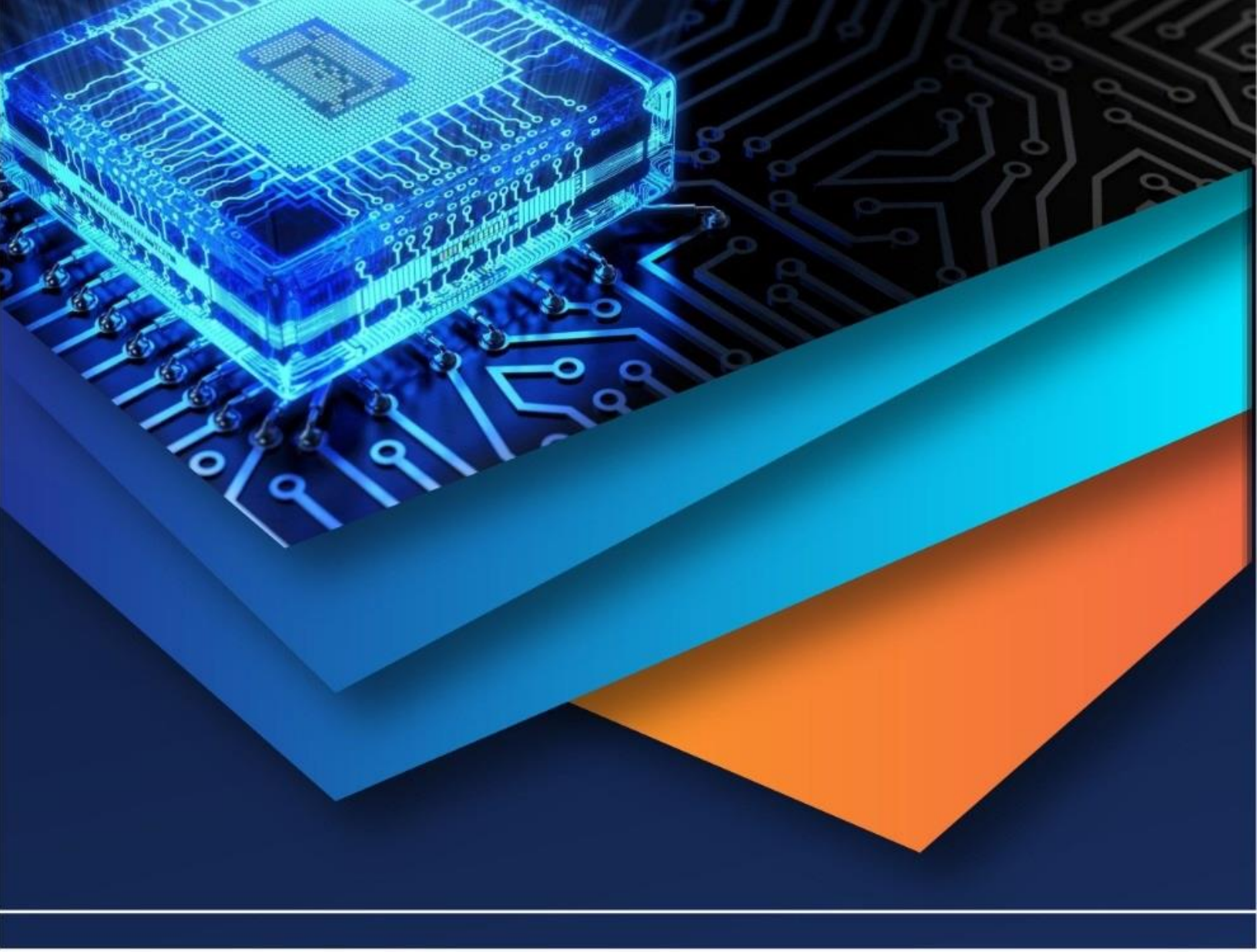
V. CONCLUSION

In this growing age of technological advancement, there is no guarantee on how good or bad effects some technologies may leave behind. The emergence of AI and social media have significantly boosted the entertainment sector. Along with the advancement of AI, deepfake has also seen significant growth allowing people to recreate memories from the past, create funny engaging video lessons for students and help investigators recreate the crime scenes. Despite several such positive aspects, people can still go beyond limits for their own good. The worst part is an attack on the general public's right to valid information. A simple misleading information is more than sufficient to create havoc between two individuals, groups of individuals or even between two countries.Therefore, the best solution that we can think of is being well prepared from our side and creating awareness. Besides, we must not jump to any sorts of conclusion just by looking at a random video seen on social media whosoever or whatever it might be. We must verify the source of information, especially controversial ones before sharing it in any social media platform, preventing the spread of disinformation. Addressing such challenges and working on its prevention will benefit its future prospects as well. A technology can never be labeled as either good or bad, but its user can definitely label it as construction or a destruction.

REFERENCES

[1] ("Artificial Intelligence, Deepfakes, and Disinformation: A Primer a Primer on JSTOR," n.d.) Artificial Intelligence, deepfakes, and Disinformation: A primer A primer on JSTOR. (n.d.). www.jstor.org. https://www.jstor.org/stable/resrep42027?mag=artificial-intelligence-and-education-a-reading-list&typeAccessWorkflow=login&seq=1
[2] (Social Media Overview - Communications, 2024)
[3] https://repo.undiksha.ac.id/7745/3/1802041037-BAB%201%20PENDAHULUAN.pdf
[4] Westerlund, Mika. "The Emergence of Deepfake Technology: A Review." Technology Innovation Management Review, vol. 9, no. 11, 1 Jan. 2019, pp. 39–52, timreview.ca/sites/default/files/article_PDF/TIMReview_November2019%20-%20D%20-%20Final.pdf, https://doi.org/10.22215/timreview/1282.
[5] Vasist, P, and S Krishnan. "Deepfakes: An Integrative Review of the Literature and an Agenda for Future Research." Communications of the Association for Information Systems, vol. 51, 2022, p. pp-pp, web.archive.org/web/20220814175942id_/aisel.aisnet.org/cgi/viewcontent.cgi?article=4403&context=cais.
[6] Doss, Christopher, et al. "Deepfakes and Scientific Knowledge Dissemination." Scientific Reports, vol. 13, no. 1, 18 Aug. 2023, p. 13429, www.nature.com/articles/s41598-023-39944-3, https://doi.org/10.1038/s41598-023-39944-3.
[7] Payne, L. (2024, April 28). Deepfake | History & Facts. Encyclopedia Britannica. https://www.britannica.com/technology/deepfake
[8] Kumar, N. (2023, June 25). What is Deepfake Technology? Origin and Impact. Analytics Insight. https://www.analyticsinsight.net/what-is-deepfake-technology-origin-and-impact/
[9] Applications of Deepfake Technology: Its benefits and Threats. (2023, November 3). https://www.knowledgenile.com/blogs/applications-of-deepfake-technology-positives-and-dangers#:~:text=Deepfake%20can%20also%20be%20used,using%20a%20personal%20digital%20avatar.
[10] Deepfake technology in video industry. (n.d.). https://www.dataart.com/blog/positive-applications-for-deepfake-technology-by-max-kalmykov
[11] What are Deepfakes and Why are They Dangerous? | Enterprise Tech News EM360. (n.d.). https://em360tech.com/tech-article/what-are-deepfakes#:~:text=The%20real%20danger%20of%20deepfakes,and%20ultimately%20ruin%20people's%20lives.
[12] Albahar, Marwan, and Jameel Almalki. "DEEPFAKES: THREATS and COUNTERMEASURES SYSTEMATIC REVIEW." Journal of Theoretical and Applied Information Technology, vol. 97, 2019, p. 22, www.jatit.org/volumes/Vol97No22/7Vol97No22.pdf.
[13] Jacobson, N. (2024, February 26). Deepfakes and their impact on society. CPI OpenFox. https://www.openfox.com/deepfakes-and-their-impact-on-society/#:~:text=Misinformation%20and%20Fake%20News,the%20global%20economy%20%2478%20billion.
[14] https://www.buster.ai/post/deepfake-a-social-and-economic-threat
[15] Files.osf.io, 2021, files.osf.io/v1/resources/cdfh3/providers/osfstorage/5fff6b75e80d370500a564c9?action=download&direct&version=1.
[16] Internet Matters Ltd. (2024, April 19). What is a deepfake? | Internet Matters. Internet Matters. https://www.internetmatters.org/resources/what-is-a-deepfake/#:~:text=Perpetrators%20might%20also%20use%20deepfakes,child%20abuse%20might%20look%20li
[17] AI-generated deepfake images create bullying danger - Stacks Law Firm. (2023, December 11). Stacks Law Firm. https://stacklaw.com.au/news/personal/criminal-law/ai-generated-deepfake-images-create-bullying-danger
[18] Ternovski, J., Kalla, J. and Aronow, P.M., 2021. Deepfake warnings for political videos increase disbelief but do not improve discernment: Evidence from two experiments. OSF Preprints, 10.
[19] Hancock, J.T. and Bailenson, J.N., 2021. The social impact of deepfakes. Cyberpsychology, behavior, and social networking, 24(3), pp.149-152.

[20] https://www.efzg.unizg.hr/UserDocsImages/TRG/Proceedings%20Trade%20Perspectives%202020.pdf#page=82
[21] Wazid, M., Mishra, A.K., Mohd, N. and Das, A.K., 2024. A Secure Deepfake Mitigation Framework: Architecture, Issues, Challenges, and Societal Impact. Cyber Security and Applications, p.100040.
[22] Alanazi, S., Asif, S. and Moulitsas, I., 2024. Examining the Societal Impact and Legislative Requirements of Deepfake Technology: A Comprehensive Study. International Journal of Social Science and Humanity, 14(2).
[23] J. T. Hancock and J. N. Bailenson, "The social impact of deepfakes," Cyberpsychol. Behav. Soc. Netw., vol. 24, no. 3, pp. 149–152, 2021.
[24] A. Jaiman, "Debating the ethics of deepfakes," Tackling Insurgent Ideologies in a Pandemic World, ORF and Global Policy Journal, New Delhi, pp. 75–79, 2020.
[25] Temir, E. (2020). Deepfake: New Era in The Age of Disinformation & End of Reliable Journalism. Selçuk İletişim, 13(2), 1009-1024. https://doi.org/10.18094/josc.685338
[26] Justice, J. (n.d.). What to do if someone shares your deep fake private picture? JUNIOR JUSTICE. https://juniorjustice.in/f/what-to-do-if-someone-shares-your-deep-fake-private-picture https://timreview.ca/sites/default/files/article_PDF/TIMReview_November2019%20-%20D%20-%20Final.pdf https://journals.sagepub.com/doi/full/10.1177/2056305120903408



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY