IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AN EMPIRICAL STUDY AMONG PUBLIC AWARENESS ON DEEPFAKES

Benjamin Jasper. P, Dr. Remya Mariam Raju
Student, Faculty
II MSc Criminology & Forensic Science
Dr. MGR Educational & Research Institute, Chennai

Abstract

Deepfakes, a new breed of manipulated media, have emerged as a significant concern due to their potential to spread disinformation and cause reputational harm. This study investigates public awareness of deepfakes in Chennai, India. Utilizing a quantitative survey methodology, researchers surveyed 103 residents using a 42-question Google Forms questionnaire. The survey assessed participants' knowledge of deepfakes, their perceptions of the technology's potential risks and benefits, and their ability to identify deepfakes. Data analysis was conducted using SPSS software. This research aims to contribute to a more comprehensive understanding of public awareness of deepfakes in a specific Indian context. The findings will inform the development of educational initiatives to equip citizens with the skills necessary to critically evaluate media and mitigate the potential harms associated with deepfakes.

Keywords: Deepfakes, Media Manipulation, Societal Impacts, Public Awareness

Introduction

The term “deepfake” is derived from a combination of “deep learning” and “fake.” 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. **(Deep Learning for Deepfakes Creation and Detection: A Survey [arxiv.org] by Xiaoxiao Zheng et al. (2019)).**

Initially, deepfakes were primarily used for harmless and entertaining purposes, such as creating celebrity face swaps or inserting faces into movies. The origins of deepfake technology date back to advancements in computer graphics and machine learning in the early 2000s. Techniques like face morphing and video texture synthesis paved the way for more advanced deepfake algorithms.

Deepfakes emerging over the years

- In 2016, researchers from the University of Erlangen-Nuremberg, Max Planck Institute for Informatics, and Stanford University developed "**Face2Face**" a method for real-time facial reenactment of a target video sequence. This innovation highlighted the potential for real-time deepfake applications and fueled further interest and development in the field. Deepfake algorithms became more sophisticated, allowing for more seamless and convincing manipulations.

- In late 2017, a **Reddit user** going by the handle "deepfakes" posted pornographic movies in which adult film stars' bodies were replaced with the faces of celebrities via deep learning techniques. This was the first significant instance of deepfake technology receiving global notice. This raised concerns about the potential for malicious use, such as spreading disinformation, defamation, or even political manipulation.

This raised serious concerns about how this technology could be abused to make maliciously fabricated videos.

- By 2018, deepfake technology had progressed to creating realistic video and audio impersonations of public figures. Researchers at the University of Washington created a deepfake video of former **U.S. President Barack Obama,** showcasing how the technology could be used to create convincing fake speeches. This example underscored the potential dangers of deepfakes in spreading political misinformation and propaganda.

Prevalence

In a number of countries, including the US, UK, France, India, and Brazil, deepfakes have become effective means of disseminating false information and undermining public confidence in the media and institutions. They raise doubts about the authenticity of digital content, worsen social tensions, and damage the credibility of information sources by focusing on public figures, celebrities, and regular people. Beyond their influence on politics, deepfakes have been used in the Philippines, Mexico, Indonesia, and Nigeria to influence elections, change public opinion on controversial topics, and to spread misinformation. Studies such as those by Li et al. (2020) have extensively documented the proliferation of deepfakes in countries like the US, UK, France, India, and Brazil, where they have been utilized to disseminate misinformation and erode public confidence in media and institutions**. (Li, Yuezun, et al. "Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.)**

Deepfakes also have consequences in the banking sector, as they take advantage of confidence to make illegal money. Deepfakes are a threat to business security and financial stability, as evidenced by cases of them posing as CEOs and carrying out financial crime in Germany, South Korea, and Japan. Cases documented by **Matern et al. (2019)** illustrate how deepfakes have been exploited for financial crimes, including impersonation of CEOs and identity theft, in these countries. Similar financial scams and identity theft have affected people and companies in Canada and Australia, highlighting the necessity for improved cybersecurity safeguards and legislative frameworks to counter such risks.

By taking advantage of weaknesses, especially in rural and vulnerable groups, these dishonest strategies aggravate already-existing societal divides and threaten democratic processes.Whether it's targeting celebrities in China, journalists and activists in Russia, or ordinary individuals in Italy, Spain, and South Africa, deepfakes have been employed to create non-consensual pornography, spread defamatory content, and orchestrate personal attacks. **Suwajanakorn et al. (2017)**

Despite the diverse manifestations of deepfake-related challenges across different countries, there is a common recognition of the urgent need for collaborative efforts to address this global phenomenon. Developing robust technological solutions, implementing effective regulatory frameworks, enhancing media literacy, and promoting digital ethics are essential components of a comprehensive strategy to mitigate the harmful impact of deepfakes on trust, information integrity, financial security, and individual rights. By fostering international cooperation and collective action, stakeholders can work towards safeguarding the integrity of digital communications and preserving the foundations of a trustworthy and informed society.

The growing popularity of deepfakes in India can be attributed in large part to the easier availability of free and user-friendly deepfake applications. These programs let even those with little technological knowledge to produce realistically altered audio and video content. This lowers the entrance barrier for creating deepfakes, which in turn causes a spike in their production and distribution. Furthermore, the influence of deepfakes is increased by the size and activity of India's social media population. Millions of people are active on social media sites like Facebook, Twitter, and WhatsApp, creating a sizable audience ready to consume digital material. (**Rajmohan, Ganga. (2020). DEEPFAKES AND THE INDIAN LAW. 10.13140/RG.2.2.16798.77128.)** Deepfakes can propagate quickly and extensively over a variety of social media platforms, reaching a wide range of users.

The impact of deepfake technology on society is significant and multi-faceted. On the one hand, deepfakes raise concerns regarding privacy and consent, as individuals' identities can be easily manipulated without their knowledge or permission. (**van der Nagel, Emily. (2020). Verifying images: deepfakes, control, and consent. Porn Studies. 7. 1-6. 10.1080/23268743.2020.1741434.**). This poses a threat to personal and professional reputations, as well as the trust we place in visual and audio evidence. Deepfakes also have the potential to exacerbate misinformation and disinformation campaigns, as they can be used to create realistic

fake news or propaganda.Morphs represent a form of biometric attack where the facial features of multiple individuals are amalgamated into a single, unique face. This technique poses a significant threat to facial recognition systems as it can potentially deceive them into granting unauthorized access by blending elements of both an authorized user and an unauthorized user's face. Moreover, morphing can be exploited to produce counterfeit identity documents, such as passports, enabling individuals who are ineligible for lawful documentation to circumvent restrictions. In such cases, a morph is created by combining the features of the individual lacking proper documentation with those of an individual possessing legitimate identification.

SIGNIFICANCE OF THE STUDY

Deepfakes—highly realistic manipulated videos or images—are a growing concern, especially in places like Chennai, India, known for its rich cultural and linguistic diversity. A survey among adults in the area showed that many people have seen these manipulated media but are not aware of the term "deepfake," which makes them vulnerable to potential harms such as misinformation and emotional distress. This gap underscores the urgent need for educational programs to help people recognize and handle these manipulated images and videos. Increasing awareness and understanding through education is crucial for helping individuals discern reality from manipulation in digital content and supports broader efforts to uphold societal norms and digital integrity. The importance of this study is in its ability to influence policy makers and educational authorities, emphasizing the need to develop strategies to address and lessen the impact of manipulated media. By fostering an informed and watchful public, the research contributes to the wider discussion on media manipulation, offering valuable insights for both national and international efforts to limit the spread and impact of these manipulations.

Statement of the Problem

To study the level of understanding and awareness the public has regarding the concept, creation, and detection of deepfakes and also examine the amount of deepfake content is on various social media platforms and evaluate how it affects public trust.

Objective Of The Study

The Objective Of The Current Study As Follows:

1. To Study the level of understanding and awareness the public has regarding the concept, creation, and detection of deepfakes.
2. To Examine the amount of deepfake content is on various social media platforms and evaluate how it affects public trust.

To identify the major concerns surrounding deepfake technology, it is crucial to understand how it threatens public trust by making it hard to distinguish real media from fake. It also enables fake explicit content, causing emotional distress and damaging reputations without consent. Additionally, deepfakes spread misinformation, deceive the public, and can lead to financial scams. In politics, deepfakes raise concerns about manipulating democratic processes by casting doubt on the authenticity of events and statements.

Research Questions

In India, a nation characterized by its vibrant cultural heritage and linguistic diversity, deepfakes pose a unique challenge. Surprisingly, there's a dearth of research investigating public awareness of deepfakes within the Indian context. This pioneering study aims to bridge this gap by examining the current level of knowledge about deepfakes among the Indian public. Specifically, the research will explore the ability of Indians to identify deepfakes and the potential societal ramifications of this technology. The following questions will guide our investigation:

1. Socio-demographic Factors
2. Level of understanding the Concept and Creation of Deepfakes
3. Level of Trust in Media Platforms
4. Reporting Behaviour and Legal Issues
5. Public Recommendations for Enhanced Security

Research Methodology

Recognizing the dearth of research on public awareness of deepfakes in India, this study adopted a rigorous exploratory survey design. This method is very beneficial for initial studies into new phenomena. This study developed a well-structured survey to shed light on an area that had not been explored before: the public's current understanding and awareness of deepfakes in India. This approach fosters the development of fundamental knowledge in a vital and little-studied area. This Study Focuses on the Adults [Aged 18 Above] residing in Chennai. The Research Tool used for this study is Structured questionnaire that was used as a Google form comprising of 43 Questions. A random sampling technique was employed in this study to achieve a representative sample and strengthen the generalizability of the findings. This approach, involving the selection of 103 adult participants from Chennai, India, aligns more effectively with the established research questions and objectives. The study's data was obtained from two sources: primary and secondary. Primary data was gathered online from residents of Chennai city using Google Forms.

We conducted a pilot test with a small group in Chennai (n=20) using Google Forms for the online survey to refine our questionnaire. This ensured it was clear, comprehensive, and well-structured. Initial analysis using SPSS showed that the questions lacked sufficient consistency (Cronbach's alpha = 0.6). Based on this, we revised the questions and retested them. This iterative process significantly improved reliability (Cronbach's alpha = 0.7).

With the pilot study complete and the questionnaire's reliability improved (Cronbach's alpha = 0.727), we proceeded to the main study. This phase involved a larger sample of 103 adults residing in Chennai, India. The comprehensive questionnaire focused on public awareness of deepfakes, considering sociodemographic factors and their potential negative impact on media trust. By using a reliable survey and a representative sample, this research aimed to gain comprehensive insights into the public's understanding and experiences with deepfakes in Chennai.

The data collected from the survey was carefully analysed using the Statistical Package for Social Sciences (SPSS) software. This software helped generate clear and informative visualizations, such as pie charts, to effectively showcase the results. Simple calculations like frequencies and percentages were used to present the findings in a straightforward and easy-to-understand manner. Open-ended questions received a different approach, with researchers manually reviewing the responses to identify and summarize key points in a descriptive format. By utilizing SPSS, the research ensured a thorough and systematic analysis, ultimately strengthening the reliability and depth of the study's conclusions.

Major findings

A. Socio-Demographic Factors

- Age of Respondents

The majority of the respondents of this study are in the age group of 22 – 30 years. As shown in the Figure 1, the majority respondents are in the age group of 22 – 30 years, or about 65.05% (67 respondents). Apart from that, there were 22.30% (24 respondents) in the age group of 18 – 21 years and 6.80% (7 respondents) were in the age group of 41 above and finally, the minority respondents are 4.85% (5 respondents) were in the age of 31 – 40 years.

- Gender Identity of the Respondents

The Majority respondents of this study are Male (59.22%) 61 respondents and the other respondents of this study are Female (40.78%) 42 respondents. As shown in the Figure 2, the Male respondents are more used in this study then the Female respondents.

- Educational Level of the Respondents

The major respondents are degree graduates. As shown in the figure 3, The 37 respondents are Masters graduate and 58 Respondents are Bachelor’s graduate and the other 8 respondents are high school and diploma graduates.

B. Level of Understanding the Concept and Creation of Deepfakes

- Deepfake Encounter of Respondents

Among the 103 respondents, the majority respondents (58.25%) 60 respondents have encountered deepfakes before whereas, the other respondents (41.75%) 43 respondents have not encountered the deepfakes before.

- Need Education and Awareness in Deepfakes

Majority (94) of the respondents believe that there should be more education and awareness about deepfakes. This shows that people are very new to that term deepfakes.

- Measures to Mitigate Negative Effects of Deepfakes

The majority of respondents (54) suggested technologies to detect and prevent deepfakes followed by the second majority (33) voting to have legal regulations followed by 13 respondents choosing increased media literacy efforts to mitigate deepfakes and 3 respondents resorted to other measures.

C. Level of Trust in Media Platform

- Respondents most trusted media platform

The two major sources that respondents trust the most are social media platforms (42) and mainstream news outlets (41), followed by 17 respondents trusting independent news websites as their source of information, with other sources mentioned by 3 respondents.

D. Reporting Behaviour and Legal Issues

- Attempted to Remove and Report

The majority of respondents who encountered deepfakes personally did not make any reports to remove the deepfake (44), while the rest of the respondents (10) attempted to report and remove the deepfake.

- Specific Laws for Regulation of Deepfakes

The majority of the people agree (76) that there should be specific laws regulating the creation and distribution of deepfakes while the rest disagree (21) and the remaining 6 respondents were unsure.

E. Public Recommendation for Security Enhancement

- Responsibility for Security Measures

The respondents mainly believed that individuals (44) should be held responsible for the creation and dissemination of deepfakes, followed by government (36) and tech companies (20), whereas the remaining respondents (3) chose other options.

Discussion

The study participants clearly highlighted the need for increased public education and awareness about deepfakes, which is similar to (Hancock & Bailenson,). This empowers individuals to identify and defend against them. A crucial aspect of this education is fostering a healthy scepticism towards online content, emphasizing critical thinking and questioning information encountered online (Vaccari & Chadwick, 2020). Furthermore, the study suggests the importance of staying informed about advancements in fake media technology. By keeping ourselves updated on the latest methods used to create deepfakes (similar to Qureshi & Khan, 2024), we can better detect and avoid falling victim to them. In essence, education and a critical approach to online information are vital tools in the fight against deepfakes.

When it comes to social media, the study suggests a multi-pronged approach to tackling deepfakes. First, participants emphasized the importance of privacy settings and limited sharing **(Diakopoulos & Johnson, 2021; Westerlund, 2019).** By controlling the Information, you share publicly and tightening your privacy settings, you make it significantly harder for others to collect the source material needed to create deepfakes of you. Imagine it like locking down your online presence – the fewer valuables lying around, the less likely someone is to be able to steal them. The study also highlighted the importance of scrutiny of sources **(Vaccari & Chadwick, 2020)**. In today's digital age, critical thinking is essential.

While some participants suggested using detection tools to identify deepfakes, these tools might not be as effective as we hope. Studies **(Akhtar, 2023)** show they have limitations. This ongoing challenge highlights the constant battle between creating deepfakes and detecting them. Just like a cat-and-mouse game, as deepfake creators get better, so too must detection methods. This suggests we need to explore other solutions alongside detection tools.

The study revealed some gaps in public understanding of deepfakes. A key area is reporting – there was no mention of people telling social media platforms or the authorities if they suspect something is a fake video or picture **(Diakopoulos & Johnson, 2021).** Just like reporting false information online, flagging deepfakes is essential to stop them from being shared widely. This highlights the need for public education campaigns to emphasize the importance of reporting suspicious content.

Additionally, the discussion mainly focused on social media. It's important to also explore how deepfakes can be addressed in other online spaces like discussion boards and messaging apps (these could be areas for future research). By tackling deepfakes across various online environments, we can create a stronger defense against this growing threat. The study offered real-world examples, like image morphing and financial scams, that showcase the potential negative impacts of cleverly altered videos and pictures (deepfakes), aligning with concerns raised in research by **Qureshi & Khan (2024).** These stories illustrate the emotional and financial harm deepfakes can inflict. In the case of image

morphing, the altered picture caused the victim to isolate themselves due to mental distress, highlighting the potential damage deepfakes can cause to a person's reputation and social standing. This aligns with broader research on the social and psychological impacts of these manipulated media **(Hancock & Bailenson),** which suggests they can erode trust and social connections.

Similarly, the financial scam highlights the ability of deepfakes to exploit trust and cause financial loss. This finding is concerning, as it echoes warnings from other studies about the potential for deepfakes to disrupt financial markets and exploit vulnerable individuals **(Qureshi & Khan, 2024).** These real-world examples underscore the urgency of raising public awareness about deepfakes and the importance of developing effective strategies to combat them, not just to protect individuals from emotional harm but also from financial exploitation.

Recommendations

If we truly want to understand the awful effects deepfakes have on people's lives, future research should focus on talking directly to the people who have been hurt by them (Kocsis, 2022). Hearing their stories firsthand would be incredibly valuable in understanding the emotional and social damage these fake videos cause. Along with this, there's a critical need to figure out the best laws and how well they can be enforced to stop people from making or sharing harmful deepfakes (Kocsis, 2022).

Furthermore, making it easier for people to report videos that seem suspicious and developing tools to catch these fakes before they spread widely are essential steps. Social media platforms have a big responsibility here. They can create reporting features that are simple to use and launch educational campaigns to teach people about deepfakes (Gosse & Burkell, 2020). Additionally, tools like Microsoft's Video Authenticator and contests like Google's Deepfake Detection Challenge show great promise (Josephs, Fosco, & Oliva, 2023). These technologies can check videos to see if they're real, identify deepfakes before everyone starts seeing them, and ultimately protect users from false information and harmful content (Maras & Alexandrou, 2019). An even better idea would be features that highlight the parts of videos that have been changed, making them easier for everyone to spot as fakes (Korshunov & Marcel, 2020). By focusing on these areas of research and development, we can learn more about how to protect people from the dangers of deepfakes and create a safer online space for everyone (Vaccari & Chadwick, 2020).

Conclusion

This study provides valuable insights into awareness and understanding of deepfakes among adults in Chennai, India, despite a significant knowledge gap about the term "deepfake." Many participants were unfamiliar with it, highlighting their vulnerability to harms like misinformation and emotional distress. The study underscores the need for educational programs to empower individuals to recognize and respond to manipulated images and videos effectively, upholding societal norms and digital integrity. By fostering a more informed public, this research contributes to broader efforts to mitigate the impact of manipulated media. Future research should engage directly with individuals who have experienced harm from deepfakes to understand the emotional and social damage caused. Exploring legal frameworks, enhancing reporting behaviours, and improving detection tools are crucial steps forward. These efforts can protect individuals from the dangers of deepfakes and enhance public trust in digital content.

Governments can play a critical role by enacting and enforcing laws to regulate deepfake creation and dissemination. Media literacy campaigns should expand to raise awareness and equip the public with tools to identify and report deepfakes. Collaboration between social media platforms and tech companies is essential to develop and deploy effective detection tools. In conclusion, this study serves as a foundational reference for future research and policy development aimed at mitigating the risks posed by deepfakes and safeguarding digital information integrity. Implementing these measures will contribute to a safer digital environment and protect individuals from the damaging effects of manipulated media.

References

1) Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. *IEEE access*, *10*, 25494-25513.
2) Mahmud, B. U., & Sharmin, A. (2021). Deep insights of deepfake technology: A review. *arXiv preprint arXiv:2105.00192*.
3) Doss, C., Mondschein, J., Shu, D., Wolfson, T., Kopecky, D., Fitton-Kane, V. A., ... & Tucker, C. (2023). Deepfakes and scientific knowledge dissemination. *Scientific Reports*, *13*(1), 13429.
4) Lyu, S. (2020, July). Deepfake detection: Current challenges and next steps. In *2020 IEEE international conference on multimedia & expo workshops (ICMEW)* (pp. 1-6). IEEE.

5) Naitali, A., Ridouani, M., Salahdine, F., & Kaabouch, N. (2023). Deepfake attacks: Generation, detection, datasets, challenges, and research directions. *Computers*, *12*(10), 216.
6) Guarnera, L., Giudice, O., Guarnera, F., Ortis, A., Puglisi, G., Paratore, A., ... & Battiato, S. (2022). The face deepfake detection challenge. *Journal of Imaging*, *8*(10), 263.
7) Verdoliva, L. (2020). Media forensics and deepfakes: an overview. *IEEE Journal of Selected Topics in Signal Processing*, *14*(5), 910-932.
8) Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, behavior, and social networking*, *24*(3), 149-152.
9) Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. *IEEE access*, *10*, 25494-25513.
10) Doss, C., Mondschein, J., Shu, D., Wolfson, T., Kopecky, D., Fitton-Kane, V. A., ... & Tucker, C. (2023). Deepfakes and scientific knowledge dissemination. *Scientific Reports*, *13*(1), 13429.
11) Fallis, D. (2021). The epistemic threat of deepfakes. *Philosophy & Technology*, *34*(4), 623-643.
12) Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Aff.*, *98*, 147.
13) Rini, R. (2020). Deepfakes and the epistemic backstop.
14) Akhtar, Z. (2023). Deepfakes generation and detection: A short survey. *Journal of Imaging*, *9*(1), 18.
15) Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New media & society*, *23*(7), 2072-2098.
16) Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, *107*, 1753.
17) Barber, A. (2023). Freedom of expression meets deepfakes. *Synthese*, *202*(2), 40.
18) Okolie, C. (2023). Artificial intelligence-altered videos (deepfakes), image-based sexual abuse, and data privacy concerns. *Journal of International Women's Studies*, *25*(2), 11
19) Qureshi, J., & Khan, S. (2024). Deciphering Deception–the Impact of AI Deepfakes on Human Cognition and Emotion.
20) Josephs, E., Fosco, C., & Oliva, A. (2023). Artifact magnification on deepfake videos increases human detection and subjective confidence. *arXiv preprint arXiv:2304.04733*.
21) Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology innovation management review*, *9*(11).
22) Figueira, Á., & Oliveira, L. (2017). The current state of fake news: challenges and opportunities. *Procedia computer science*, *121*, 817-825.