JournalsKart ISSN: ISSN: 2231-2498

International Journal of Communication Development
Vol17 Issue 1 January-June 2026

Rise of deepfakes in the ai-generated content era: a review of trends, challenges, and future perspectives

Sakshi Aggarwal[1], Prof. (Dr.) Durgesh Tripathi[2]

Research scholar, Guru Gobind Singh Indraprastha University[1]
Dean, USMC, Guru Gobind Singh Indraprastha University[2]

Abstract

Generative Adversarial Networks (GANs) have transformed the way of content creation. It has introduced a way where two neural networks work against each other to generate highly realistic content like images, videos, and audio on various social media platforms. This paper focuses on the trends, ethical challenges, and prospects of deepfakes, which are a result of GAN technology. In the last few years, deepfakes have been further developed, driven by generative models and constant data advancements. It is a digital media forgery technique that can not only manipulate fake face content, but also generate it. This is making it difficult for users to differentiate between real and fake content. Although multiple studies have been conducted on deepfake detection, it is currently lagging behind deepfake generation, underscoring the need for comprehensive surveys on deepfake generation. The objective of this empirical study is to explore the growing use of deepfakes and associated risks and promote ways of deepfake detection. In this research, Clifford G. Christians' Media Ethics Theory of Global Justice and Uses and Gratification Theory will be used to examine the issue of deepfakes. The researcher analyzes the types of deepfakes and the factors driving their proliferation, understanding deepfake detection methods, understanding the current challenges of deepfakes, and the developing trends through a mixed-method approach integrating a youth survey and literature review, among the youth in Delhi. The sample includes 384 participants aged between 18 and 24 years.

Keywords: deepfakes | AIGC | deepfake detection

Introduction:

Fake news is any fabricated content or fictitious news style used to deceive the public. In the past few years, fake news has become a threat to society and democracy as it spreads quickly and has the power to impact the lives of millions of users who consume information through social media platforms. (Westerlund, 2019). It has become easier than ever to spread misinformation or disinformation through these platforms, thereby underlining the importance of media literacy so that the audience knows what to trust and what not.

What are deepfakes?

Deepfakes are defined as AI-generated digital content, including audio, video, and images, which are fabricated using deep learning algorithms to change, replace, or superimpose the original content with new content that looks real. (Abbas & Taeihagh, 2024). The word 'deepfake' is a combination of 'deep learning' and 'fake.' (Arora & Soni, 2021).

The word 'deepfake' is made from two words- 'deep learning' and 'fake'. Deep learning has its roots back to 2017 when an unknown Reddit user used it to explain the entry of this technology in the world of digital media. Since then, various upgradations in technology have been witnessed for deepfakes in machine learning (ML), generative adversarial networks (GANs), which allow users to create highly realistic videos that are hard to identify as synthetic.

However, an interesting fact about deepfakes is that anyone with a computer can create such fabricated videos, which are not practically distinguishable from the real media. (Fletcher, 2018). There are multiple purposes of using deep fake media, the most common of which are modifying human faces or synthesising them with the help of Artificial Intelligence (AI) for e-commerce, advertisement, etc. Some other applications are dubbing in movies or the reanimation of historic figures for educational purposes.

Earlier, deepfakes were used to defame actresses, political leaders, entertainers, comedians, etc., with their faces changed into porn videos, but deepfakes in the future will be used for bullying, revenge porn videos, false evidence videos in courts, blackmailing, propaganda for anti-national activities, etc. Also, this technological advancement invites the risk of cyber crimes such as identity theft. Other harmful uses of deepfakes are seen in fake news, financial fraud, and hoaxes.

As a result, the efforts to detect synthetic images and videos have been increasing in contrast with the traditional research areas of general media forensics. The growing importance of understanding fake face detection is also witnessed in the growing number of discussions in top conferences.

Types of deepfakes

Deepfakes are mainly classified into 3 categories: video-based, audio-based, and text-based. Many software tools allow the creation of such content. Two of the most commonly used tools are Generative Adversarial Networks (GANs)
(Malik, Kuribayashi, Abdullahi & Khan, 2022) and autoencoders (Thanh Thi Nguyen et al., 2022). With these tools, various deepfake content can be created, like face manipulation, voice cloning, political deepfakes, celebrity impersonation, synthetic influencers, cross-modal deepfakes, etc.

Benefits of deepfakes

There are many benefits of deepfake technology in multiple sectors, such as the education sector, healthcare, fashion, gaming, digital communication, e-commerce, and various other business fields. It can be used to update film footage instead of reshooting it, to create voices of actors who may not be able to dub due to any illness, etc. (*The Best (And Scariest) Examples Of AI-Enabled Deepfakes | Bernard Marr*, n.d.). The technology also allows for realistic dubbing in any other language to increase the movie's audience. Apart from movies, it is also used for educational purposes.

Challenges of deepfakes

Deepfakes are a major threat to democracy. They create pressure on the journalists to distinguish between fake and real news, especially during elections. With deepfakes on the internet, citizens are unable to trust the authorities with the information they share online. The technology also puts individuals and organisations at risk of cybersecurity matters. (Westerlund, 2019) Deepfakes are even worse than traditional fake news, as these are harder to identify, and the audience believes that it is true, thereby putting the reputation of the journalists at risk. For instance, during India-Pakistan tensions in 2019, Reuters found around 30 fake videos being circulated on popular platforms, most of which were old, but with new captions. Mika Westerlud describes the phenomenon as 'information apocalypse' or 'reality apathy', where the audience, when served with so much misinformation, simply dismisses genuine footage as fake. Hence, multiple challenges with deepfakes include consent and identity theft, trust erosion, youth vulnerability, reputation damage, misinformation, political manipulation, and non-consensual pornography.

Ways to combat deepfakes

Studies show that there are 4 ways to fight the issue of deepfakes. Firstly, legislation and regulation. Then, prepare corporate policies and voluntary action. Then, education and training, and lastly, anti-deepfake technology that allows detection of such content, its authentication, and deepfake prevention.

Review of literature:

A report by DeepStrike (2025) gives a comprehensive view of the growth of deepfakes and the risks linked with this technology. It tells how the technology can change from a technological development to a main tool in cybercrime. The study also declares that the ability of humans to detect deepfakes is significantly low, which makes them highly vulnerable to being deceived by misinformation or disinformation. The report states that the dominant attack vector is choice cloning as it is low-cost, accessible, and has huge financial implications. All in all, the study focuses on how the technology is growing day by day to even encompass the detection abilities of the users, thereby creating a trust deficit amongst them.

Another research paper titled 'Comparative Study of Deep Learning Techniques for Deepfake Video Detection' emphasises the strengths, limitations, performance, and architectures of deepfakes. The paper discusses various deep learning techniques used for the detection of deepfakes, but also states that there is no single technique that is universally available to detect deepfakes. Therefore, it suggests that continuous advancement should be made to detect systems, to keep up with the pace of growing technology.

Another study titled '*Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence*' gives in-depth insights into deepfakes, their evolution, implications, mechanisms, etc., in the digital world. As stated in *Expert Systems with Applications (2024)*, various methods of deepfake detection and generation have been discussed. The study highlights how deepfakes have penetrated the digital media after becoming realistic. Multiple state-of-the-art detection tools and techniques have been discussed along with their broader societal impact. Policy changes and the need for integrated solutions are suggested in the study.

As per the overview shared by ScienceDirect, deepfakes are considered synthetic media created with generative media and neural networks to produce manipulated content such

as audio, video, and images. Literature shows that innovations like GANs, diffusion models, and autoencoders have improved the quality of deepfakes, making it more difficult to detect.

While deepfakes share important applications in education, entertainment, and other fields, their widespread use in digital content creation also fuels privacy violations, misinformation, identity theft, etc.

Overall, the literature emphasises the need for strong detection systems, regulatory mechanisms, and ethical guidelines to address the complex implications of deepfakes on society.

Methodology:

This study incorporates a quantitative method using a survey with a questionnaire as a tool to analyse the trends, challenges, and future aspects of deepfakes, particularly with youth aged between 18 and 24 years, who are the most active users of digital media. Non-probability convenience sampling was used, considering the time constraints, and a total of 384 respondents were surveyed, considering Cochran's formula. The respondents were selected from the Delhi region, and the questionnaire was circulated online through Google Forms, which ensured reach and participation. The questionnaire consisted of closed-ended multiple-choice questions divided into two parts: part A with demographic details and part B with questions related to awareness and exposure to deepfakes, ethical challenges, and the future of deepfakes.

Additionally, the study includes a review of the already existing literature on deepfakes to provide theoretical grounding and contextual understanding. Multiple research articles were accessed through platforms like MyLoft, IEEE Xplore, Google Scholar, ScienceDirect, and ResearchGate using keywords deepfakes, AIGC, and deepfake detection. The data analysis is done using descriptive statistical techniques like frequency distribution and percentages, to understand patterns, behaviour, and awareness related to deepfakes. Ethical guidelines have been met in maintaining voluntary participation, informed consent, and keeping the identities of the respondents anonymous.

Results:

Through the questionnaire, it was found that most of the respondents interacted with deepfakes casually through social networking sites, but some of them responded actively as well, because of a lack of awareness. This highlights the role of social media in spreading AI-generated content. The results also tell about the societal impact of AI techniques, especially in terms of content creation. The results also show that deepfakes are used the most during elections to spread disinformation and misinformation. Earlier, most of the deepfakes were targeted towards obscene imagery of women, leading to defamation and raising other ethical concerns. Deepfakes blur the gap between real and fake and, hence, not only put pressure on journalists but also diminish the trust of the audience in the media. Detection of deepfakes is still reactive and not preventive, considering that there is a need for AI literacy among youth. As the understanding of deepfakes is largely superficial, it depicts confusion and scepticism among users. The results also showed that many users, even after knowing the ill effects of deepfakes, still do not verify the content before sharing it. Also, the users strongly think that there should be regulatory guidelines by the government for engagement with deepfakes.

On the other hand, some users also identified the positive applications of deepfakes, like in the field of education, entertainment, media production, etc. All in all, the findings indicate that the users are looking at deepfakes as an

opportunity as well as a critical challenge, suggesting the need for media literacy and ethical and regulatory frameworks to ensure responsible use.

Discussion:

These results are important as they indicate a critical gap between technological development and the preparedness of the users. While deepfakes are all across social networking sites, the users are still not literate enough to understand their proper consumption and engagement. This vulnerability can have far-reaching consequences. The study shows a credibility crisis in the growing digital world, as people are unable to distinguish between real and fake content.

The results also matter from a societal and policy aspect. The users are concerned regarding misinformation spread through deepfakes, highlighting a need for proper governance and prompt action by policymakers, digital platforms, and media organisations. They need to develop guidelines to prevent the misuse of deepfakes.

Next is the need for media literacy, as just knowing about deepfakes is not enough. Digital and AI literacy should be integrated into the education system so that users can critically evaluate the content before consuming it or engaging with it through like, share, or comment.

Also, the results are relevant for solving issues like climate change, owing to sustainable communication, as deepfakes, if used to spread fake information, can hinder public opinion and their understanding, and the potential in achieving sustainable goals. Hence, it is important to ensure that authentic information reaches the users for collective progress.

Lastly, deepfakes should not be seen only as causing harm to society, but treated as a two-edged sword that, if used with proper regulation, can be useful in multiple fields like education, media innovation, entertainment, etc.

Conclusion:

The study shows the importance of deepfakes on digital media platforms. The findings show that though the social media users are aware of the concept of deepfakes, they lack knowledge for identifying them, thereby increasing their deeper penetration on the media platforms. The study also highlights that social media users' trust is being reduced due to the presence of synthetic content online. To conclude, deepfakes are a technological development as well as a challenge for society. There is a need for a multi-dimensional approach to address the implications of deepfakes, such as ethical AI practices, AI and digital media literacy, increased public awareness, and strict policy measures. The study provides insights into youth behaviour and perceptions, thereby focusing on the need for better communication practices in the growing media landscape.

Acknowledgement:

I would like to express my gratitude towards all those who contributed to the successful completion of this research paper. I am thankful to my Ph.D. supervisor and my co-author, Prof. (Dr.) Durgesh Tripathi for his constant support and guidance throughout my research journey. His valuable insights and encouragement have been instrumental in shaping this study.

I am also thankful to the respondents for taking out their valuable time to participate in the survey, providing valuable data for the study. Also, I would like to acknowledge the support of academic platforms like IEEE Xplore,

ScienceDirect, ResearchGate, and Google Scholar for providing me with relevant literature related to the topic.

I am also thankful to my university and colleagues for supporting and encouraging me throughout, without which the study would not have been possible.

References

1. Abbas, F., & Taeihagh, A. (2024). Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence. Expert Systems with Applications, 252(8),
2. 124260. https://doi.org/10.1016/j.eswa.2024.124260
3. Abdolahnejad, M., & Liu, P. X. (2020). *Deep learning for face image synthesis and semantic manipulations: A review and future perspectives*. **Artificial Intelligence Review**, 53(8), 5847–5880. https://doi.org/10.1007/s10462-020-09835-4
4. Almutairi, S., Alharbi, A., & Alotaibi, N. (2024). Deepfake video detection using deep learning techniques with optimized hyperparameters. Applied Sciences, 14(5), Article
5. 1218. https://doi.org/10.1016/j.array.2024.100121
6. Arora, T., & Soni, R. (2021). A review of techniques to detect the GAN-generated fake images. *Generative Adversarial Networks for Image-to-Image Translation*, 125–159. https://doi.org/10.1016/B978-0-12-823519-5.00004-X
7. Cybenko, A. K., & Cybenko, G. (2018). *AI and fake news*. **IEEE Intelligent Systems**, 33(5), 3–7. https://doi.org/10.1109/MIS.2018.2877280
8. De Keersmaecker, J., & Roets, A. (2017). *'Fake news': Incorrect, but hard to correct: The role of cognitive ability on the impact of false information on social impressions*. **Intelligence**, 65, 107–110. https://doi.org/10.1016/j.intell.2017.10.005
9. DeepStrike. (2025). *Deepfake statistics 2025: AI fraud data & trends*. Retrieved from Deepfake Statistics 2025
10. Fletcher, J.
11. 2018. Deepfakes, Artificial Intelligence, andSome Kind of Dystopia: The New Faces of OnlinePost-Fact Performance. Theatre Journal, 70(4):455–471. Project MUSE,https://doi.org/10.1353/tj.2018.0097
12. Hasan, H. R., & Salah, K. (2019). *The emergence of deepfake technology: A review*. Retrieved from https://www.researchgate.net /publication/ 337644519_The_Emergence_of_Deepfake_Technol _A_Review
13. Khan, M. A., Hussain, M., Rehman, A., et al. (2022). *A survey on deepfake detection using machine learning and deep learning techniques*. **Information Sciences**, 608, 1089–1111. https://doi.org/10.1016/j.ins.2022.06.077
14. Malik, H., Satoshi Kuribayashi, Abdullahi, S. M., & Khan, S. (2022). *Deepfake detection for safeguarding media authenticity: A survey*. **IEEE Access**, 10, 123–145

15. Maras, M. H., & Alexandrou, A. (2019). *Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos*. **International Journal of Evidence & Proof**, 23(3), 255–262. https://doi.org/10.1177/1365712718807226

16. Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T., & Nahavandi, S. (2022). *Deep learning for deepfakes creation and detection: A survey*. **Computer Vision and Image Understanding**, 220, 103525.

17. ScienceDirect. (n.d.). *Deepfakes*. In ScienceDirect Topics. Retrieved from https://www.sciencedirect.com/topics/computer-science/deepfakes

18. *The Best (And Scariest) Examples Of AI-Enabled Deepfakes | Bernard Marr*. (n.d.). Retrieved April 1, 2026, from https://bernardmarr.com/the-best-and-scariest-examples-of-ai-enabled-deepfakes/

19. Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, *9*(11), 39–52. https://doi.org/10.22215/TIMREVIEW/1282