Journal of Science Research International (JSRI)
ISSN: 2456-6365 (Online)

OPEN ACCESS

DEEPFAKE TECHNOLOGY AS A CYBER THREAT: RISKS, DETECTION METHODS, AND COUNTER MEASURES

Sabitha Praveen Madamby* and Vansh Singh

Department of Computer Science,

Pillai College of Arts, Commerce and Science (Empowered Autonomous), New Panvel

**Corresponding author E-mail: sabitha.praveen@mes.ac.in*

Received: 11 January 2026	Revised: 20 February 2026	Accepted: 30 March 2026	Published: 17 April 2026

DOI: **https://doi.org/10.5281/zenodo.19622303**

Abstract:

Artificial Intelligence is improving very quickly and it has made it possible to create highly realistic fake videos, images, and audio recordings. This type of manipulated media is known as deepfake technology. Deepfakes are created using machine learning models that can copy a person's face, voice, and expressions. Because of this, it can become very difficult for viewers to understand whether the content they see online is real or fake. Although deepfakes can be used for positive purposes such as movie production, digital entertainment, and educational simulations, they can also be misused. Cybercriminals may use deepfake media to impersonate individuals, spread misinformation, manipulate public opinion, or conduct financial fraud. This research paper studies deepfake technology from a cybersecurity point of view. It explains how deepfake systems work, the risks they create for individuals and organizations, and the methods researchers used to detect fake media. The paper also discusses different strategies that can help reduce the misuse of deepfake technology, including technical tools, legal regulations, and digital awareness among internet users. The study shows that deepfake detection technology is improving, but deepfake generation techniques are also advancing rapidly. Because of this, solving the deepfake problem requires cooperation between governments, technology companies, cybersecurity experts, and the general public.

Keywords: *Deepfake, Cyber Threat, Risks, Detection, Measures, Technology,*

1. Introduction

Technology based on Artificial Intelligence has changed the way digital content is created and shared. One important development in this field is deepfake technology, which allows computers to generate realistic but fake media. These media files can include videos, images, or even voice recordings that appear to be real but are actually created using machine learning algorithms.

Deepfakes are often generated using advanced deep learning techniques such as Generative Adversarial Networks (GANs). These systems study large amounts of video and audio data to learn how a person looks and speaks. After

learning these patterns, the system can produce new content that imitates that person very closely.

While this technology can be useful in areas such as filmmaking, gaming, and digital content creation, it also introduces serious cybersecurity risks. Attackers may create deepfake videos or voice recordings to pretend to be someone else. For example, criminals might imitate a company executive and instruct employees to transfer money to a fraudulent account.

Another major concern is the spread of misinformation. Deepfake videos can be shared quickly on social media platforms and may influence people's opinions before the content is verified. This can create confusion and reduce public trust in digital media.

Because deepfakes are becoming more realistic, researchers and cybersecurity experts are trying to develop ways to detect and prevent their misuse. Understanding how deepfakes work and how they can be identified is important for protecting individuals, organizations, and society from potential cyber threats.

The purpose of this research paper is to examine deepfake technology as a cybersecurity challenge by exploring its risks, reviewing detection techniques, and discussing possible countermeasures.

2. Literature review

Researchers from different fields such as artificial intelligence, cybersecurity, and digital forensics have studied deepfake technology in recent years. Early studies mainly focused on the technical process used to create deepfake content.

One of the most widely used techniques for generating deepfakes is the Generative Adversarial Network (GAN) model. This system consists of two neural networks that work together. One network creates fake images or videos, while the other tries to detect whether the media is real or fake. Through this process, the system gradually improves and produces highly realistic content.

As the technology became more accessible, researchers started studying its potential impact on society. Several studies highlight that deepfakes can be used for spreading fake news, political manipulation, identity theft, and online harassment.

Another important research area focuses on detecting deepfake media. Early detection methods relied on identifying visual errors in manipulated videos, such as unusual blinking patterns, unnatural facial movements, or lighting inconsistencies.

Recent research has introduced machine learning detection systems that analyze patterns in video frames, audio signals, and facial movements. These systems can sometimes detect deepfake content more accurately than traditional methods.

However, many researchers believe that deepfake detection will always be challenging because the technology used to create deepfakes continues to evolve.

3. Methodology

This research uses a secondary research approach. Instead of performing experiments, the study analyzes information from existing research papers, cybersecurity articles, and technical reports.

The collected information was grouped into three main topics:

i. Risks created by deepfake technology
ii. Techniques used to detect deepfake media
iii. Methods used to prevent deepfake misuse

The detection techniques discussed in the research were evaluated based on their accuracy, practicality, and limitations. Prevention strategies such as security policies, awareness programs, and government regulations were also examined. Using this approach helps provide a clear understanding of how deepfake technology affects cybersecurity and what solutions are currently being explored.

4. Results

The research shows that deepfake technology is becoming a growing concern in the field of cybersecurity. One of the most common threats is identity impersonation. Attackers can create fake audio or video recordings that imitate the voice or appearance of a real person.

For example, criminals might create a fake voice message that sounds like a company executive asking an employee to transfer money. Because the voice appears authentic, employees may trust the request and perform the transaction.

Another major risk is the spread of misinformation. Deepfake videos can easily be shared on social media platforms, where they may influence public opinion before the truth is verified.

When studying detection techniques, machine learning based systems showed better performance compared to traditional forensic analysis methods. However, these systems require large datasets and continuous updates to remain effective.

The research also suggests that combining multiple detection approaches, such as analyzing both audio and video signals, can improve the ability to detect manipulated media.

5. Discussion

Deepfake attacks are different from traditional cyberattacks because they mainly target human trust rather than computer systems. Instead of hacking software, attackers attempt to deceive people by presenting fake media that looks real.

Because of this, cybersecurity strategies must include not only technical tools but also awareness programs for employees and internet users. Organizations should implement verification procedures before approving sensitive actions such as financial transfers or confidential information sharing.

Another challenge is that deepfake detection technologies must constantly evolve. As detection systems improve, developers of deepfake tools may also improve their generation techniques.

There is also privacy concerns related to monitoring digital media for manipulation. Detection systems may need access to large datasets, which raises questions about how personal information should be protected.

To effectively deal with deepfake threats, cooperation between governments, technology companies, cybersecurity researchers, and educational institutions is necessary.

Conclusion

Deepfake technology has emerged as a significant challenge in the digital age. While it offers useful applications in entertainment and media production, its misuse can lead to serious cybersecurity problems such as identity theft, misinformation, and financial fraud.

This research paper examined the risks associated with deepfake technology, discussed the techniques used to detect manipulated media, and explored possible countermeasures.

The study suggests that although detection technologies are improving, they cannot completely eliminate the threat. A comprehensive approach that combines technical solutions, cybersecurity policies, legal regulations, and

public awareness is necessary.

Future research should continue developing more accurate detection methods and focus on creating systems that can identify deepfake media in real time.

References

1. Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*(6), 1753–1819. https://doi.org/10.2139/ssrn.3213954
2. Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys, 54*(1), 1–41. https://doi.org/10.1145/3447690
3. Korshunov, P., & Marcel, S. (2019). Deepfakes: Threats and opportunities for multimedia forensics. *IEEE Access, 7*, 152227–152234. https://doi.org/10.1109/ACCESS.2019.2949209
4. Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T., & Nahavandi, S. (2019). Deep learning for deepfakes creation and detection: A survey. *Neural Computing and Applications, 31*, 13145–13164. https://doi.org/10.1007/s00521-019-04428-9
5. Korshunov, P., & Marcel, S. (2018). Vulnerability of face recognition to deepfake videos. In *2018 26th European Signal Processing Conference (EUSIPCO)* (pp. 1–5). IEEE. https://doi.org/10.23919/EUSIPCO.2018.8553102
6. Verdoliva, L. (2020). Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing, 14*(5), 910–932. https://doi.org/10.1109/JSTSP.2020.2994609
7. Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A compact facial video forgery detection network. In *2018 IEEE International Workshop on Information Forensics and Security (WIFS)* (pp. 1–7). IEEE. https://doi.org/10.1109/WIFS.2018.8630765
8. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion, 64*, 131–148. https://doi.org/10.1016/j.inffus.2020.05.011
9. Wang, X., Han, Y., & Li, Y. (2021). Combating deepfake attacks: Detection, analysis, and countermeasures. *Computers & Security, 106*, 102284. https://doi.org/10.1016/j.cose.2021.102284
10. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 1–11). https://doi.org/10.1109/ICCV.2019.00125