

# Cybersecurity and Deepfake technology\*

## Challenges and educational implications

Chrystala Morfi†  
Computer Science  
University of Central Lancashire  
Cyprus  
CMorfi@uclan.ac.uk

Dr. Andriani Piki  
Computer Science  
University of Central Lancashire  
Cyprus  
APiki@uclan.ac.uk

### ABSTRACT

Deepfake technology is becoming increasingly integrated into daily life, from AI-generated media to its rising use in cyberattacks [1]. The realistic AI-devised content has emerged as a significant cybersecurity threat regardless of its algorithm's sophistication level. This means that even simple, unsophisticated deepfakes launched online can cause damage [2]. A plethora of examples place deepfakes as the perpetrators to a variety of cyberattacks, such as defamation attacks, cyber fraud through impersonation, propaganda during elections and non-consensual sexually explicit deepfake-formed images, videos and audios [3].

Cybersecurity plays a central role in the mitigation of deepfakes in both protecting individuals and the prevention of deepfake-related incidents. Nonetheless, cybersecurity faces several challenges in the mitigation and detection of the variety of deepfake-generated forgeries. By investigating these cybersecurity challenges, this study aims to highlight the need for digital security measures and the importance of education in combating deepfake threats within the online space.

### CCS CONCEPTS

- Security and privacy ~ Software and application security
- Software and its engineering

### KEYWORDS

Deepfakes, Artificial intelligence, Generative AI, Cybersecurity, Education implications

## 1 Research Objectives & Methodology

This study utilises a qualitative research approach, specifically a thematic analysis combining information and knowledge from three key sources. The primary source material will be obtained from a thoroughly devised comprehensive literature review on the main topic of this research the challenges cybersecurity faces due to deepfakes.

Additionally, the study will involve an exploration of the policies and cybersecurity measures against deepfake content employed by key stakeholders such as such as LinkedIn, Google, Instagram, BBC, Microsoft, X, and Facebook.

Then semi-structured interviews will be conducted with computer science professionals and cybersecurity experts in the field that are knowledgeable about deepfake content. In addition, the insights obtained from the interviews will be critically analysed, devised in a critical review of the state-of-the-art on deepfake technology literature review comparison. Finally, the insights gathered from these three sources will be analysed with the resolution to reveal a significant amount of cybersecurity threats associated with deepfake technology.

Moreover, examine the measures currently in place to detect and mitigate these threats, and investigate the challenges highly influential organisations face in implementing effective deepfake detection technologies. Figure 1, illustrates the research process that will be followed for the study to be completed, listing the important milestones.

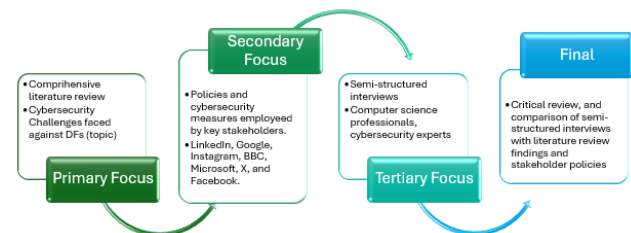


Figure 1: Research process (primary, secondary, tertiary focus)

## 2 Research Questions

The research questions are outlined below:

1. What are the most significant cybersecurity threats associated with deepfake technology?
2. What measures are currently in place to detect and mitigate deepfake threats?
3. What challenges do organisations face in implementing effective deepfake detection technologies?

The **research questions** outlined above will be expanded in the future main research, allowing for further targeted questions.

### 3 Discussion & Expected Results

In the preliminary analysis of the research suggests that this research could potentially yield insights into such topics as the impact of deepfake technology in everyday life, the cybersecurity threats caused by deepfakes, challenges faced by online social media platforms, and how the misuse of deepfake technology can be detected as well as mitigated.

A number of the results heavily lean towards the need for cybersecurity measures against deepfakes, as well as the need for educational programs on deepfakes, and GenAI technologies. Moreover, this study may reveal that preventive measures need to be placed, in a number of stages within the deepfake technology's lifecycle.

Such as the need for monitoring action of the users within the online deepfake development platform. Which in turn will allow for the prevention of the development of explicit deepfakes regardless of gender or age. Another possible preventive measure that may be needed could be the banning of deepfakes in online social services such as X (Twitter), or the flagging of explicit word usage, along with any GenAI content (e.g. images, videos, audio).

The overall implications of such results could potentially showcase the need for creative preventive measures that are placed on both the development tools as well as the social platforms. In addition, the need for educational action would be of paramount importance for the population.

Thus, the implications of such results could be noteworthy for educators, social media users/ platforms, and cybersecurity experts on the development of preventive measures, and understanding of this threat. Also, future research could be developed on how education can assist in the understanding overall navigation of such technologies, all while preventing the unethical use of deepfakes (and GenAI).

### 4 Conclusion

In summary, the study highlights the urgent need for increased awareness and education on the cybersecurity threats posed by deepfake technology and their social, legal, ethical, and organisational implications. Through the understanding of the cybersecurity challenges faced by deepfake technology, it becomes evident that raising awareness and leveraging educational endeavours are crucial steps in equipping both individuals and organisations with the necessary tools to combat cyberthreats, as well as identify deepfake-generated content and preventing it before causing harm. The ultimate goal is to foster a safe, secure, and trustworthy digital environment for everyone, as well as educate people.

### ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my supervisor, Dr. Piki, for her continuous support during my academic journey. Besides my supervisor, I would like to thank

my family for their own patience and overall funding of my academic endeavours throughout the years.

### REFERENCES

- [1] Amerini, I., Barni, M., Battiato, S., Bestagini, P., Boato, G., Bruni, V., Caldelli, R., De Natale, F., De Nicola, R., Guarnera, L. and Mandelli, S., 2025. Deepfake media forensics: Status and future challenges. *Journal of Imaging*, 11(3). DOI: <https://doi.org/10.3390/jimaging11030073>
- [2] Ratnawita, R., 2025. Cybersecurity in the AI Era Measures Deepfake Threats and Artificial Intelligence-Based Attacks. *Journal of the American Institute*, 2(2). DOI: <https://doi.org/10.71364/s3emxx77>
- [3] Kasera, G., Solanki, M., Kaur, H. and Shah, K., 2025, January. A Detailed Exploration to Deepfake: A Cybersecurity Threat. In *2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. IEEE. DOI: <https://doi.org/10.1109/SCEECS64059.2025.10940812>