JOURNAL OF THE AMERICAN INSTITUTE
ISSN: 3066-3776
United States

Cybersecurity in the AI Era Measures Deepfake Threats and Artificial Intelligence-Based Attacks

Ratnawita[1]

Email Correspondent: witadosen@gmail.com

Keywords:
AI-powered cybersecurity, Deepfake detection techniques, Zero Trust Architecture in cybersecurity, AI-driven cyber threats.

Abstract
The development of artificial intelligence (AI) has had a significant impact on cybersecurity, both as a defense tool and as a threat. One of the biggest emerging risks is deepfake attacks and AI-based cyberattacks, which are increasingly difficult to detect and mitigate. Deepfake technology, which uses Generative Adversarial Networks (GANs), enables video and audio manipulation with a high level of realism, which can be used for disinformation, fraud, and threats to digital security systems. This research aims to analyze cybersecurity threats caused by deepfakes and AI-based attacks, evaluate the effectiveness of conventional security systems in dealing with them, and propose more adaptive AI-based mitigation strategies. The method used in this study is a literature study with a qualitative approach, collecting data from various academic sources, industry reports, and regulations related to cybersecurity. The analysis was carried out using thematic analysis and data triangulation techniques, which allowed mapping of the latest threat trends and solutions that had been implemented. The results show that signature-based security is increasingly ineffective in the face of evolving AI attacks. The implementation of AI in cyber defense systems, such as machine learning-based detection, Zero Trust Architecture (ZTA), and incident response automation systems, is the main solution in dealing with increasingly complex threats. Therefore, an adaptive security approach that combines technology, regulatory policies, and public education is needed to reduce the risk of deepfake attacks and other AI threats.

INTRODUCTION

The development of artificial intelligence (AI) has brought revolutions in various fields, including in cyber security (Nguyen et al., 2021). AI's ability to quickly analyze large amounts of data and detect attack patterns has increased the effectiveness of digital security systems. However, on the other hand, AI is also being used by malicious actors to create increasingly sophisticated threats, such as AI-based attacks and deepfakes that can deceive security systems as well as manipulate public opinion (Citron & Chesney, 2019). As the reliance on digital systems increases, the need to understand and anticipate these threats is becoming increasingly urgent (Goodfellow et al., 2016).

[1] Universitas Mitra Bangsa, Indonesia, witadosen@gmail.com

Cyber Security is an effort to protect computer systems, networks, and data from digital attacks that can lead to information theft, service disruption, or system damage (Von Solms & Van Niekerk, 2013). Along with increasing digitalization in various sectors, cybersecurity threats such as ransomware attacks, phishing, and Advanced Persistent Threats (APTs) are increasingly complex and difficult to detect (Wagner et al., 2019). According to a Cybersecurity Ventures report (2022), global losses due to cybercrime are estimated to reach $10.5 trillion per year by 2025, demonstrating the urgency of increasing cyber protection for companies and government institutions. Technologies such as artificial intelligence (AI) and machine learning have begun to be applied to detect suspicious attack patterns, but major challenges remain in data management and response to increasingly sophisticated cyberattacks (Giraldo et al., 2017).

To address cyber threats, organizations and individuals need to implement proactive security strategies, such as multi-factor authentication (MFA), data encryption, and Zero Trust Architecture (ZTA) that restricts access to only authorized users (Rose et al., 2020). In addition, cybersecurity awareness among employees is also a crucial aspect, considering that more than 85% of security breaches are caused by human error (Verizon Data Breach Report, 2021). Therefore, a holistic approach that includes a combination of advanced technologies, strict security policies, and regular cybersecurity training is needed to mitigate the growing risk of cyberattacks. In addition, regulations such as the General Data Protection Regulation (GDPR) in the European Union and the Personal Data Protection Regulation (PDP) in Indonesia provide guidelines for organizations in managing and protecting user data from the threat of cybercrime (Fink et al., 2017).

One of the biggest threats in the AI era is deepfakes, which are technologies based on Generative Adversarial Networks (GANs) that allow the creation of highly realistic but completely fake videos or audio (Tolosana et al., 2020). This technology has been used to spread disinformation, extort, and damage the reputation of individuals and institutions (Verdoliva, 2020). According to a Deeptrace report (2020), the number of deepfake videos on the internet has increased by more than 900% since 2018, indicating that the challenges in detecting and addressing these threats are getting greater. AI-based cybercrime not only creates risk at the individual level but can also be used for the purposes of espionage, political manipulation, and hacking of high-level security systems (Yu et al., 2021).

Deepfakes are artificial intelligence (AI) technologies that use machine learning and deep learning to manipulate images or videos to make it look as if someone said or did something that actually never happened (Citron & Chesney, 2019). This technology generally uses Generative Adversarial Networks (GANs) to create increasingly realistic results by improving the details of facial expressions, movements, and synchronizing voices with images (Tolosana et al., 2020). Advances in deepfakes have created positive opportunities in the creative industry, such as in film and entertainment, but they also present significant risks, such as the spread of false information, defamation, and threats to cybersecurity and democracy (Verdoliva, 2020).

The impact of deepfake technology is increasingly felt in the political, legal, and social worlds. In some cases, deepfakes have been used to spread political disinformation, with the aim of manipulating public opinion and disrupting elections (Agarwal, 2019; Rana et al., 2022). In addition, cybercrimes such as deepfake-based extortion are also on the rise, where individual faces are faked in explicit content used for criminal purposes (Kietzmann et al., 2020). Although deepfake detection technologies continue to be developed by researchers and social media platforms, such as AI forensics and metadata analysis, the challenge of distinguishing real and fake videos is still a major concern. Therefore, stricter regulations and public education about the dangers of deepfakes are needed to reduce the negative impact they cause (Maras & Alexandrou, 2019).

In addition to deepfakes, AI is also used in cyber attacks that are increasingly difficult to detect. An example is AI-driven phishing attacks, where AI systems are able to adjust attack techniques based on the victim's behavior patterns in real-time (Kreps et al., 2022). These attacks rely on machine learning to generate messages that are more convincing and difficult to distinguish from the original communication (Zhang, 2022). In the industrial world, AI-based automated attacks have resulted in significant economic losses, with a global estimate of $10.5 trillion per year by 2025 due to cybercrime (Weforum, 2025). Therefore, traditional security systems need to be upgraded with AI-based solutions that are able to effectively counter these threats.

The rise of AI-based attacks and the gap between conventional security systems and advanced attack capabilities are constantly evolving (Maras & Alexandrou, 2019). Security systems that still rely on signature-based detection methods are increasingly ineffective in dealing with AI attacks that can automatically adjust (Dash et al., 2022). In addition, regulations and policies related to deepfakes and AI-based cyber threats are still lagging behind compared to the speed of technological development (Bertino et al., 2022). Therefore, studies that not only identify existing threats but also offer adaptive and AI-based strategies to counter modern cyberattacks are needed.

Several previous studies have addressed cybersecurity threats in the context of AI and deepfakes. Chesney & Citron (2019) highlights how deepfakes are a major threat in the spread of disinformation and political manipulation (Citron & Chesney, 2019), while research by Tolosana et al. (2020) focuses on deepfake detection techniques using AI-based digital forensics (Tolosana et al., 2020). On the other hand, Kreps et al. (2022) examined AI-based phishing attacks that are able to tailor attack messages with a very high level of precision (Nobles, 2024). However, previous research was still limited to threat detection, while research on AI-based security strategies to deal with deepfake attacks and other AI threats was minimal. Therefore, this research seeks to fill this gap by exploring adaptive and AI-based security approaches to measure the threat of deepfakes and modern cyberattacks.

This research aims to analyze the threat of deepfakes and AI-based attacks in cybersecurity, as well as develop more adaptive AI-based mitigation strategies. The main focus of this research is to identify the types of threats, assess the effectiveness of conventional security systems, propose AI-based solutions, and evaluate policies related to cybersecurity to deal with artificial intelligence-based attacks more effectively.

METHOD

This study uses a qualitative approach with a literature study method to analyze the threat of deepfakes and AI-based attacks in the context of cybersecurity. The literature study was chosen because it allows for an in-depth exploration of previous research, existing cybersecurity policies, as well as technological solutions that have been developed in the face of AI-based attacks (Snyder, 2019). This approach provides a comprehensive understanding of the latest cyber threat trends and mitigation strategies that have been implemented across various sectors.

The data sources in this study were obtained from scientific journal articles, industry reports, policy documents, and international conference proceedings relevant to AI-based cybersecurity topics. Key references were collected from academic databases such as IEEE Xplore, ScienceDirect, Springer, and Google Scholar, over the last five years (2019–2024) to ensure that the data used is current and relevant to the latest technological developments (Boell & Cecez-Kecmanovic, 2015). In addition, the study also considers cybersecurity reports from global organizations such as Cybersecurity Ventures, the National Institute of Standards and Technology (NIST), and Europol, which provide insights into threats and policy responses in the face of deepfakes and AI-driven cyberattacks.

Data collection was carried out through a systematic review of the literature relevant to this study. This process involves identifying, selecting, evaluating, and synthesizing from various sources related to AI-based cybersecurity. In the selection of data, this study used inclusion and exclusion criteria, where only research that discussed deepfakes, AI-based attacks, as well as cybersecurity policies was included in the analysis (Xiao & Watson, 2019). Each selected document is analyzed based on its relevance, methodological validity, and contribution to the understanding of cybersecurity threats and solutions in the AI era.

Data analysis was carried out using a thematic analysis approach, where findings from various literature sources were grouped into several main categories, such as the types of deepfake threats and AI-based cyberattacks, the impact on cybersecurity, AI-based mitigation strategies, and challenges in regulation and policy (Braun & Clarke, 2021). The data triangulation technique is also applied by comparing results from various sources to increase the validity of research findings (Patton, 2002). In addition, this study uses cybersecurity analysis frameworks that have been developed in previous studies, such as Zero Trust Architecture (ZTA) and AI-driven cyber defense mechanisms, to evaluate the effectiveness of mitigation strategies that have been implemented in various sectors.

RESULT AND DISCUSSION

In this study, a literature study was conducted to explore various academic findings related to cybersecurity in the era of artificial intelligence, focusing on the threat of deepfakes and AI-based attacks. From the various articles found, ten articles have been selected based on their relevance, scientific contributions, and novelty in discussing this topic. The articles cover a wide range of perspectives, from deepfake detection, AI-based cyberattacks, to ethical challenges in the application of this technology. The following table presents a summary of the results of the literature selection which is the main basis for the analysis of this study.

Table 1. Literature Review

No	Author	Title	Findings
1	Poli Reddy Reddem	The Rise of AI-powered Cybercrime: a Data-driven Analysis of Emerging Threats	Deepfake-based fraud is rising in AI-driven cybercrime, with a 238% surge in attacks and a 67% higher success rate, highlighting the need for advanced defense mechanisms.
2	Chirag Gajiwala	Artificial Intelligence in Cybersecurity : Advancing Threat Modeling and Vulnerability Assessment	The paper explores AI's role in cybersecurity for threat modeling and defense but does not cover deepfake or AI-based threats.
3	Nicolae Sfetcu	Threats of Artificial Intelligence for Cybersecurity	The paper highlights AI-related cybersecurity risks, including system manipulation and deepfakes, stressing the need for effective security measures.
4	Godwill Chimamiwa	Managing cyber risks in the face of AI- and ML - Driven Adversarial Attacks	The paper discusses how AI and ML enhance deepfake and social engineering threats, stressing the need for stronger cyber risk management.
5	Zarif Bin Akhtar, Ahmed Tajbiul Rawol	Harnessing artificial intelligence (AI) for cybersecurity: Challenges,	The research highlights AI's dual role in cybersecurity, stressing the need for

	opportunities, risks, future directions	strategic solutions to counter deepfake risks and ethical challenges.

In the increasingly advanced digital era, cybersecurity faces new challenges due to the emergence of artificial intelligence (AI)-based attacks, including deepfakes and other sophisticated cyberattacks. Research conducted by Poli Reddy Reddem highlights that the increase in AI-based cybercrime has reached alarming levels. The study shows that deepfake-based attacks have increased by 238%, with the success rate of AI-based methods reaching 67% higher than conventional methods. These findings show that cyberattacks no longer rely on traditional methods that rely solely on social engineering or software exploits but have now transformed by leveraging AI to trick security systems. Deepfakes, which were originally only used for entertainment and media manipulation, are now a major threat in cybercrime schemes, including identity-based fraud and communication forgery. Therefore, this study emphasizes the importance of more adaptive defense mechanisms to counter this threat, especially by developing security systems that are able to detect and respond to AI-based attacks more effectively.

Meanwhile, research conducted by Chirag Gajiwala highlights how AI can be leveraged as a tool to improve cybersecurity through threat modeling and vulnerability assessment. The study focuses on how AI can be used proactively in identifying threats before an attack occurs. However, unlike previous research, this article does not specifically discuss the threat of deepfakes or other AI-based cyberattacks, but rather emphasizes AI's ability to develop more sophisticated cyber defense models. These findings provide important insights into the potential of AI in building adaptive security systems, which can analyze threat patterns more quickly and accurately than conventional methods. As such, this research contributes to understanding how AI can be used as a tool to strengthen cyber defenses, although it has not specifically discussed mitigation strategies against deepfakes and other AI attacks.

In the article, Nicolae Sfetcu highlights various threats that have emerged due to the development of AI in the context of cybersecurity. The study emphasizes that AI-based system manipulation can have unforeseen consequences, potentially damaging existing security systems and digital infrastructure. One of the biggest challenges discussed in this study is how AI can be used to outwit threat detection systems, such as by creating attacks that are able to circumvent traditional security mechanisms. The research also outlines the importance of a deep understanding of AI threats, so that organizations and individuals can develop more resilient security measures in the face of AI-based cyberattacks. These findings further confirm that AI is not only a tool used to improve security, but can also be a weapon for cybercriminals to create more complex and hard-to-detect threats (Sfetcu, 2024).

Another study conducted by Godwill Chimamiwa reinforces findings on how AI and machine learning (ML) are being used by cybercriminals to create more sophisticated attacks. The study highlights that AI-based attacks are not only limited to data manipulation, but also include increasingly convincing social engineering techniques, including deepfakes and attack automation methods. One of the key aspects discussed in the study is how AI and ML can be used to create deepfakes that are almost indistinguishable from real videos or voice recordings, which are then utilized for criminal purposes such as identity fraud, extortion, or information manipulation on a large scale. This research emphasizes the urgency of the need for better cyber risk management strategies, especially in the face of the growing threat posed by the use of AI in cybercrime. With the increasing skills of hackers in exploiting AI, this article suggests that companies and organizations adopt more innovative defense methods, including AI-based security systems capable of detecting anomalies and unusual attack patterns in real-time (Chimamiwa, 2024)sssssssss.

Finally, research conducted by Zarif Bin Akhtar and Ahmed Tajbiul Rawol provides a broader overview of how AI plays a role in cybersecurity. This article discusses both the opportunities and challenges presented by AI in the world of digital security, including how AI can improve the effectiveness of security systems, but also introduces new risks, especially in the form of deepfakes and other AI-based exploits. One of the key findings in the study is that while AI has great potential to strengthen cybersecurity, it can also be misused to develop new attack methods that are more difficult to detect. This article highlights the need for more innovative mitigation strategies, including the use of AI to detect threats that are also created with AI, as well as the development of stricter regulations to control the use of deepfake technology. In addition, the study also discusses the ethical and regulatory challenges that arise due to the increasing use of AI in cybersecurity, especially related to how companies and governments can control the spread of deepfakes without infringing on the freedom of expression and privacy rights of individuals (Akhtar & Rawol, 2024).

Overall, the results of the research from these five articles show that AI-based cybersecurity threats are on the rise and require a more adaptive security approach. Deepfake, as a form of AI abuse, has become a powerful tool for cybercriminals, both in identity-based fraud, information manipulation, and social engineering attacks. Meanwhile, while AI can be used as a defense tool, research also reveals that AI itself can be a weakness in security systems if not properly controlled. Therefore, further research on mitigation strategies, AI-based threat detection, and strengthening digital security policies are urgently needed to reduce the risks presented by AI technology in cyberspace.

Discussion

Deepfake Threats and AI-Based Attacks in Cybersecurity

In the era of artificial intelligence (AI), cybersecurity faces new challenges that are increasingly complex. The threat posed by deepfakes and AI-based attacks is becoming increasingly serious, with potential impacts extending to various sectors, including finance, government, and national security. The existence of this technology not only brings benefits in various aspects of life, but also provides opportunities for cybercriminals to exploit weaknesses in existing security systems.

One of the biggest threats that has emerged is deepfake manipulation, where AI technology is used to create highly convincing fake videos, audios, or images. This ability can be used to spread disinformation, commit identity fraud, or defame the reputation of individuals and institutions. In the context of cyberattacks, deepfakes can be used in video or audio-based phishing attacks, where criminals fake a person's voice or face to trick the victim into providing sensitive information or performing certain actions in favor of the attacker.

In addition, AI is also increasingly being used in automated cyberattacks that are able to speed up and increase the effectiveness of hacking methods. For example, in brute force attacks, AI allows hackers to guess passwords at a much higher speed compared to conventional methods. Not only that, AI-generated phishing-based attacks are now becoming more difficult to distinguish from genuine communication because AI algorithms are able to mimic language patterns and communication styles very convincingly, making them more effective in deceiving victims.

Another threat is AI-based malware and ransomware, where malicious programs are enriched with AI technology that allows them to learn from user behavior and evade detection by traditional security software. AI can also be used to customize encryption methods in ransomware, making it more difficult for existing security systems to stop. With its ability to adapt, AI-powered malware can infiltrate networks in more sophisticated and hard-to-detect ways.

In addition to attacks that directly target security systems, the threat of adversarial attacks is also a major concern. In these attacks, criminals manipulate AI algorithms by entering misleading

data, so that AI models can make wrong decisions. These attacks have the potential to threaten a wide range of systems, including biometric security, autonomous vehicles, and other data processing systems that rely on AI in decision-making.

Another threat that is no less dangerous is data poisoning attacks, where perpetrators infiltrate malicious data into machine learning models with the aim of changing the model's behavior undetected. These attacks can cause AI security systems to fail to recognize the actual threat or even misidentify legitimate entities as threats, opening up loopholes for hackers to exploit system weaknesses more easily.

Effectiveness of Conventional Security Systems

Conventional security systems today face major challenges in dealing with cyberattacks powered by artificial intelligence (AI). While it has long been used to protect digital infrastructure, traditional approaches to cybersecurity have a number of limitations that make them less effective at warding off evolving threats.

One of the main drawbacks of conventional security systems is their overly static nature. These systems are generally rule-based and work by detecting previously known attack patterns. However, with AI being able to dynamically create new attacks and adapt to the target environment, this static approach becomes less effective. AI-based attacks can adapt in real-time, change attack patterns, and evade detection, so traditional security systems often lag behind in anticipating new threats.

In addition, conventional security systems rely heavily on signature-based detection methods, which detect threats based on digital signatures from attacks that have already occurred. This approach is useful for dealing with familiar attacks, but it becomes less effective when dealing with AI attacks that are capable of creating new variations that have never been seen before. With its ability to generate unique attack patterns, AI allows hackers to avoid detection by systems that only recognize threats that have been documented.

Another disadvantage of conventional security systems is the lack of real-time processing. Many traditional systems require time to update threat databases or apply security updates after detecting new attack patterns. On the other hand, AI is able to perform attacks faster and more flexibly, adjusting its techniques directly without giving security systems time to react. This leads to significant security vulnerabilities, where attacks can occur before traditional systems have had a chance to identify and mitigate them.

With all these limitations, it's clear that conventional security systems aren't robust enough to deal with increasingly sophisticated AI-based threats. A new, more adaptive, dynamic, and artificial intelligence-based approach is needed to ensure that security systems can continue to evolve as the complexity of cyberattacks increases in today's digital era.

AI-Based Solutions in Cybersecurity

To deal with increasingly complex threats due to artificial intelligence, cybersecurity systems also need to adopt more adaptive AI-based technologies. One approach that can be applied is AI-Based Threat Detection and Response, where machine learning (ML) and deep learning are used to recognize never-before-known attack patterns. In addition, behavioral analysis can be used to monitor user and network behavior patterns to detect anomalous activity that indicates a cyberattack.

In detecting deepfake threats, Generative Adversarial Networks (GANs) technology can be used to compare artifact patterns in AI-generated video or audio, as well as identify inconsistencies in facial expressions, voices, or other unusual patterns. Meanwhile, endpoint security can be

improved through AI-Driven Endpoint Security, where AI-based security software is able to automatically adapt to changes in attack patterns without relying on signature database updates like conventional antiviruses.

The security approach can also be strengthened by implementing the Zero Trust Architecture (ZTA), which adopts the principle of "never trust, always verify". In this model, each entity must be rigorously verified before being granted access to the system, including through AI-based multi-factor authentication (MFA) that is able to detect changes in user access patterns in real-time.

As a preventive measure and quick response to attacks, security systems can be equipped with Automated Incident Response, which allows AI to respond automatically to attacks in real-time. AI can detect and isolate infected devices or restrict access to sensitive data when suspicious anomalies are found. By implementing this AI-based solution, cybersecurity can become more resilient in the face of increasingly sophisticated attacks that are difficult to detect by conventional methods.

Cybersecurity strategy recommendations

In the face of artificial intelligence-based cyber threats, stricter regulations and policies are needed to ensure the safe and responsible use of AI. Standardization of AI regulations in cybersecurity is the main step that must be taken by governments and international institutions to prevent the misuse of AI technology for criminal purposes. In addition, law enforcement against the misuse of AI needs to be strengthened by increasing cooperation between countries in identifying and dealing with cybercrimes, including deepfake attacks and AI exploits in hacking.

In addition to legal aspects, education and increasing public awareness are important factors in building resilience to AI-based threats. Digital literacy must be improved so that the public is more aware of the potential for information manipulation and cyber threats generated by AI. On the other hand, the development of AI ethics and safety guidelines must also be carried out, by establishing strict policies in the development and application of AI technology so that it is not misused by irresponsible parties.

Collaboration with the industrial sector is also key in strengthening cybersecurity. Technology companies need to play an active role in developing deepfake detection systems and AI-based security solutions that are able to anticipate cyberattacks more effectively. With cooperation between the government, law enforcement, academia, and the technology industry, cybersecurity systems can be more adaptive in facing increasingly sophisticated threats in the AI era.

CONCLUSION

This research reveals that the threat of deepfakes and AI-based cyberattacks continues to grow with an increasing level of sophistication. AI technology, which was originally developed for innovation and efficiency, has now become a powerful tool for cybercriminals in launching manipulative attacks that are difficult to detect. Conventional signature-based security systems are increasingly unable to anticipate attacks created dynamically by AI, so a more adaptive and AI-based mitigation strategy is needed. The application of machine learning-based detection technology, Zero Trust Architecture (ZTA), and incident response automation systems are the main solutions in countering this threat. In addition, stricter regulations and increased public awareness of the risks of deepfakes and AI attacks are also important factors in strengthening cyber resilience globally.

As a suggestion, more research is needed to develop more accurate and efficient deepfake detection methods, considering that the level of realism of deepfakes continues to increase. In addition, cooperation between the government, the industrial sector, and academia is urgently needed in developing cybersecurity policies that are more responsive to AI-based threats. The

implementation of AI in cybersecurity must also be accompanied by strengthening ethical and regulatory systems that control its use so that it is not abused. With a combination of advanced technology, strict regulations, and extensive education, it is hoped that the threat of deepfakes and AI-based attacks can be effectively minimized.

REFERENCE

Agarwal, S. (2019). *A study on creativity: Detection and network structures*. University of Illinois at Urbana-Champaign.

Akhtar, Z. Bin, & Rawol, A. T. (2024). Harnessing artificial intelligence (AI) for cybersecurity: Challenges, opportunities, risks, future directions. *Computing and Artificial Intelligence*, *2*(2), 1485.

Boell, S. K., & Cecez-Kecmanovic, D. (2015). On being 'systematic'in literature reviews in IS. *Journal of Information Technology*, *30*(2), 161–173.

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*.

Chimamiwa, G. (2024). *Managing cyber risks in the face of AI-and ML-Driven Adversarial Attacks*. SBS Swiss Business School.

Citron, D. K., & Chesney, R. (2019). Deepfakes and the new disinformation war. *Foreign Affairs*.

Dash, B., Ansari, M. F., Sharma, P., & Ali, A. (2022). Threats and opportunities with AI-based cyber security intrusion detection: a review. *International Journal of Software Engineering & Applications (IJSEA)*, *13*(5).

Fink, G. A., Edgar, T. W., Rice, T. R., MacDonald, D. G., & Crawford, C. E. (2017). Security and privacy in cyber-physical systems. In *Cyber-physical systems* (pp. 129–141). Elsevier.

Giraldo, J., Sarkar, E., Cardenas, A. A., Maniatakos, M., & Kantarcioglu, M. (2017). Security and privacy in cyber-physical systems: A survey of surveys. *IEEE Design & Test*, *34*(4), 7–17.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, Issue 2). MIT press Cambridge.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, *63*(2), 135–146.

Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, *9*(1), 104–117.

Maras, M.-H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, *23*(3), 255–262.

Nobles, C. (2024). The weaponization of artificial intelligence in cybersecurity: A systematic review. *Procedia Computer Science*, *239*, 547–555.

Patton, M. Q. (2002). *Qualitative research & evaluation methods*. sage.

Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. *IEEE Access*, *10*, 25494–25513.

Rose, S., Borchert, O., Mitchell, A., & Connelly, S. (2020). Zero trust architecture NIST special publication 888-207. *NIST, Aug/2020.[Online]. Available: Https://Nvlpubs. Nist. Gov/Nistpubs/SpecialPublications/NIST. SP*, *800207*.

Sfetcu, N. (2024). *Threats of Artificial Intelligence for Cybersecurity*.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339.

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.

Verdoliva, L. (2020). Media forensics and deepfakes: an overview. *IEEE Journal of Selected Topics in Signal Processing*, *14*(5), 910–932.

Von Solms, R., & Van Niekerk, J. (2013). From information security to cyber security. *Computers & Security*, *38*, 97–102.

Wagner, T. D., Mahbub, K., Palomar, E., & Abdallah, A. E. (2019). Cyber threat intelligence sharing: Survey and research directions. *Computers & Security*, *87*, 101589.

Weforum. (2025). *Global Cybersecurity Outlook 2025*. https://www.weforum.org/publications/global-cybersecurity-outlook-2025/

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, *39*(1), 93–112.

Yu, P., Xia, Z., Fei, J., & Lu, Y. (2021). A survey on deepfake video detection. *Iet Biometrics*, *10*(6), 607–624.

Zhang, T. (2022). Deepfake generation and detection, a survey. *Multimedia Tools and Applications*, *81*(5), 6259–6276.