Volume XI Issue 3 Year 2026 | Page 1000-1011 | ISSN: 2527-9866
Received: ***15-06-2026*** | *Revised:* ***30-07--2026*** | *Accepted:* ***16-08-2026***

Navigating Deceptive Realities: Public Perceptions and Cybersecurity Threats of Deepfake Technology

Nur Anis Shafiqah Mazlan[1], Mohamad Fadli Zolkipli[2], Hapini Awang[3], Nur Suhaili Mansor[4], Bingxin Jin[5]
[1,2,3,4,5]Utara Malaysia University, Sintok, Kedah, Malaysia, 06010
[5]Northwest A&F University, Shaanxi, China
e-mail: nur_anis_shafiqah@soc.uum.edu.my, hapini.awang@uum.edu.my, nursuhaili@uum.edu.my, m.fadli.zolkipli@uum.edu.my, jin_gxin@gsgsg.uum.edu.my
*Correspondence: nur_anis_shafiqah@soc.uum.edu.my

Abstract: *The rapid advancement of deepfake technology presents a profound cybersecurity threat by seamlessly fabricating synthetic media, which severely erodes digital trust. This study aims to evaluate the community's awareness of deepfake threats and assess the necessity of multifaceted mitigation strategies. Employing a quantitative methodology, an online survey was conducted with 67 respondents, predominantly young adults, to measure their exposure, psychological vulnerability, and perspectives on cybersecurity countermeasures. The findings reveal that while the public is generally aware of deepfakes, 75 percent struggle to visually differentiate manipulated content from authentic media. Consequently, the proliferation of deepfakes has diminished perceived societal trust in online information. Notably, there is a unanimous consensus among respondents demanding strict legal frameworks and comprehensive public education to combat this menace. The study concludes that relying exclusively on technical detection algorithms is insufficient. Instead, preserving information integrity requires a multidisciplinary approach combining robust technological defenses, proactive policymaking, and widespread digital literacy trainings.*

Keywords: *Deepfake technology, Cybersecurity, Synthetic media, Digital trust, Public awareness.*

I. INTRODUCTION

The advent of deepfake technology, which involves sophisticated synthetic media generated through deep learning systems such as Generative Adversarial Networks (GANs) and advanced artificial intelligence encoders, poses a profound and rapidly evolving threat to modern cybersecurity [1], [2]. By seamlessly fabricating audio, images, and video to mimic real human characteristics, deepfakes blur the boundary between authentic and manipulated content [3]. This technological capability facilitates malicious activities, including disinformation campaigns, identity theft, and critical vulnerabilities within biometric authentication systems [4], [5]. Consequently, deepfakes possess the unprecedented power to manipulate public perception, deceive individuals and organizations, and severely erode societal trust in digital media[1], [6].

As the sophistication of deepfake generation outpaced detection efforts, the resulting deceptive realities become increasingly difficult to distinguish from authentic media [7]. The continuous evolution of these artificial intelligence algorithms makes detection highly challenging, requiring constant updates to security protocols [8]. Mitigating this pervasive menace necessitates a multidisciplinary approach. Collaboration among experts in machine learning, cybersecurity, psychology, and policymaking is imperative to formulate robust technical defenses, implement strict legal frameworks, and promote comprehensive digital literacy. However, while existing literature heavily focuses on the technological underpinnings of

deepfake generation, detection methodologies, and technical cybersecurity threats, there remains a distinct research gap regarding the "human element" of cybersecurity specifically, the public's actual awareness, visual detection capabilities, and psychological vulnerability in daily online interactions.

To address this critical research gap and align the study's focus with its empirical survey methodology, this research shifts the focus from a purely technical evaluation of detection algorithms to an empirical assessment of public perceptions. While previous literature has extensively evaluated the technical adaptability of detection technologies [3], [7] and aimed to contribute to the advancement of cybersecurity protocols [1], [9]. The primary objective of this study is to evaluate public awareness, exposure levels, perceived cybersecurity threats, and community perspectives on multi-layered mitigation strategies regarding deepfake technology. To address this critical research gap and align the study's focus with its empirical survey methodology, this research shifts the focus from a purely technical evaluation of detection algorithms to an empirical assessment of public perceptions. While previous literature has extensively evaluated the technical adaptability of detection technologies, and aimed to contribute to the advancement of cybersecurity protocols, the primary objective of this study is to evaluate public awareness, exposure levels, perceived cybersecurity threats, and community perspectives on multi-layered mitigation strategies regarding deepfake technology. To guide this empirical inquiry, the study explicitly investigates the baseline level of public deepfake awareness and visual detection capability, the resulting impacts of synthetic media on personal and organizational digital trust, and the public consensus regarding the necessity of legal, educational, and software-based countermeasures.

The significance of researching this human dimension is paramount in the contemporary digital age. Investigating public perception yields critical insights regarding the evolving nature of synthetic threats, their consequences on information integrity, and their broader societal impacts[10]. This study provides two explicit research contributions. First, it offers empirical data exposing a critical vulnerability: active digital natives struggle with visual detection despite having high baseline awareness. Second, it highlights a unanimous public consensus demanding structural policy and educational interventions. Ultimately, these insights can empower individuals and organizations to safeguard themselves, contributing to the development of more stringent cybersecurity measures and collaborative efforts to fortify the digital landscape and sustain public trust in an increasingly manipulated digital era [11].

II. LITERATURE REVIEW

A. Theoretical Foundation of Deepfake Technology

Deepfakes are created using advanced AI and deep learning, particularly Generative Adversarial Networks (GANs) and encoder-decoder frameworks. As previous studies point out, these systems learn by analyzing huge datasets of faces and voices to replicate real human features [12], [13]. Typically, an encoder maps out a target's unique facial features, and a decoder overlays them onto another person to produce a seamless visual or audio edit [6], [14]. The theoretical foundation of this technology highlights its capacity to independently learn and improve, resulting in highly convincing fabricated content that perfectly mimics human behavior, thereby blurring the lines between reality and manipulation [15].

B. Cybersecurity Threats and Malicious Applications

Empirical research consistently demonstrates that the rise of deepfakes introduces massive vulnerabilities to cybersecurity Researchers have documented various malicious uses, ranging from personal identity theft and financial scams to coordinated political disinformation campaigns. Hackers are also using highly realistic voice and video clones to bypass biometric

logins and access sensitive data [5], [16]. More recently, enterprise threats like deepfake-driven social engineering and impersonation of company executives have become growing concerns [17], [18]. The literature indicates a massive surge in harmful deepfake content in recent years, proving that this technology has transitioned from a sophisticated novelty into a weaponized tool capable of deceiving both individuals and large organizations.

C. The Evolution of Detection Mechanisms and Limitations

A critical review of existing deepfake detection strategies reveals an ongoing technological arms race between deepfake creators and cybersecurity experts [3], [19] . Early empirical studies focused on identifying visual artifacts and rendering flaws, such as unnatural blinking patterns, patchy skin tones, flickering edges, and poorly synchronized lip movements [20], [21]. However, a major issue is that once a detector starts flagging these errors, deepfake algorithms are immediately updated to correct them, which means single-feature detectors quickly lose their effectiveness over time [6], [15]. Even though governments and tech giants have invested heavily in building better detection tools, deepfakes continue to evolve faster than our automated defenses can keep up [2], [3], [22]. This unresolved issue leaves existing security protocols vulnerable to newer, more polished iterations of deepfakes.

D. Mitigation Strategies and Research Gaps

To address the escalating threat, researchers have proposed various mitigation strategies, including the use of encrypted communication channels, digital watermarking to ensure document integrity, and routine cybersecurity training for professionals. Additional proposals include content provenance systems, blockchain-based verification, and multimodal biometric defenses [1], [2], [10]. While these technical countermeasures are valuable, there remains a notable research gap regarding the human element of cybersecurity specifically, the public's awareness and psychological vulnerability to deceptive realities [23], [24]. Many prior studies focus purely on algorithmic detection, overlooking the necessity of community education and strict legal frameworks.

E. Contribution of the Current Study

Our study aims to address this gap by looking at both technical concerns and the often-ignored human element of digital security. Rather than trying to test or evaluate detection software ourselves, we examine how the public actually perceives and interacts with these threats. By evaluating both the adaptability of current detection technologies and the community's awareness of synthetic threats, this research bridges the gap between technical cybersecurity and human digital literacy. The findings of this study intend to contribute to the development of more holistic, proactive defense strategies that combine technological innovation with comprehensive public education to preserve trust in digital media [1], [2], [24].

III. METHODS

A. Target Population and Sampling Design

To explore how everyday people perceive deepfake risks, this research focused on Malaysian digital natives, particularly university-aged young adults. Since this group is highly active online and frequently consumes social media, they are the most likely to encounter synthetic media. We used a non-probability convenience sampling technique to recruit participants, distributing our questionnaire digitally to peer groups and student networks [25]. This process yielded a sample of 67 respondents. While we acknowledge that this sample size is small and limits our ability to generalize the findings to the entire Malaysian population, it provides an important, targeted look at the demographic most exposed to these digital threats. The demographic characteristics of our collected data revealed that the vast majority of the

respondents fell within the 18 to 23 age demographics, with female participants comprising 64.2 percent of the total sample size.

B. Research Location

Given that the core subject matter revolved around digital awareness and cyber-centric threats, the research was conducted entirely in a virtual, online environment. The survey was strategically distributed through digital communication channels, which significantly maximized the accessibility, efficiency, and reach of the data collection process. While the digital nature of the survey meant there were no strict physical geographical boundaries, the contextual environment of the target demographic was heavily rooted in their daily interactions with social media platforms and online information ecosystems. Geographically, the participant pool was centered around Universiti Utara Malaysia (UUM) in Sintok, Kedah, Malaysia. Conducting the research in this online setting was fundamentally crucial and highly appropriate, as it allowed the researcher to capture authentic responses and assess contextual factors directly within the very environment where deepfake exposure and cybersecurity threats typically manifest.

C. Questionnaire Design and Measurement Indicators

The study adopted a quantitative research methodology to systematically explore, identify, and evaluate numeric patterns related to the societal understanding of deepfake technologies [26]. The primary instrument selected for data collection was an online survey platform (Google Forms), chosen for its accessible user interface and its robust built-in security features designed to ensure respondent privacy.

The implementation of the survey involved a meticulously structured questionnaire divided into three core evaluation segments. The first segment captured foundational demographic profiles to establish the sample's contextual background. The second segment utilized a strict numerical structure, employing a predefined 1-to-7 Likert scale (where 1 represented "Not Familiar/Concerned at All" and 7 represented "Extremely Familiar/Concerned"). This scale scientifically quantified the respondents' levels across three key indicators [27] which are Familiarity with synthetic media concepts, Exposure Frequency to manipulated online content, and the intensity of their Security Concerns. The final segment utilized direct binary questions to evaluate public consensus on proposed mitigation strategies, such as the enforcement of strict legal frameworks, the necessity of specialized educational training, and the development of preventive software applications.

D. Instrument Validity and Reliability

To ensure the validity and reliability of the survey before full deployment, the instrument underwent a systematic review process [26]. For content and face validity, the draft questionnaire was reviewed and vetted by senior academic experts in cybersecurity and informatics. Their feedback helped refine the phrasing of the questions to ensure they were clear, unambiguous, and directly aligned with our research objectives. For reliability, a preliminary pilot test was conducted with a small group of university students. The scale items demonstrated strong internal consistency, with a Cronbach's alpha coefficient exceeding the standard 0.70 threshold [28], confirming that the questionnaire could reliably and consistently capture the respondents' attitudes and perceptions.

E. Ethical Approval and Informed Consent

This study was conducted with strict adherence to ethical research standards [26]. Before participants could access or answer any survey questions, they were presented with an electronic informed consent statement on the first page of the Google Form. This section clearly explained the purpose of the study, guaranteed complete anonymity, and assured participants

that their responses would be kept secure and used solely for academic research. Respondents had to actively check a "Consent" box to proceed, indicating they understood that their participation was completely voluntary and that they could withdraw at any time [26]. No personal identifying information such as names, phone numbers, or email addresses was collected, ensuring the absolute privacy and integrity of all 67 respondents.

F. Analytical Approach

To ensure the reliability and reproducibility of the results, all collected responses were automatically compiled and securely stored in integrated digital spreadsheets for efficient analytical processing. Our evaluation phase employed descriptive statistical methods to critically analyze the response distributions. In line with our survey-based methodology, we did not perform complex algorithmic modeling or program-based experiments. Instead, we systematically analyzed frequencies, counts (n), percentages, and median values to compare the empirical outcomes against our research questions, allowing us to identify clear trends in how everyday digital natives experience and respond to deepfake threats.

IV. RESULTS AND DISCUSSION

The quantitative data gathered from our 67 participants offers crucial insights into how everyday users perceive deepfake threats, their psychological vulnerabilities, and their consensus on how to deal with these risks. Our sample represents active digital natives, predominantly young adults aged 18 to 23, with female participants representing 64.2% (*n*=43) and male participants representing 35.8% (*n*=24) of the total pool. This demographic profile provides a focused look at the exact population group that interacts most heavily with social media and online ecosystems on a daily basis.

To test the reliability of the scale items before analyzing these results, a reliability analysis was done. The scale items measuring user familiarity, exposure, and concern yielded a Cronbach's alpha of range more than 0.82, well above the standard 0.70 threshold, confirming that the survey is highly reliable for measuring public perceptions.

A. Public Awareness and Exposure to Deepfake Content

The initial phase of the analysis measured the public's baseline familiarity and direct exposure to manipulated media. while deepfakes are frequently discussed in the media, a notable minority of an estimated 15 percent of the respondents had never heard of deepfakes prior to participating in the study. For the remaining 85 percent (*n*=57) who were aware of the technology, the overall familiarity was moderately high, with a median score of 5.0 on our Likert scale.

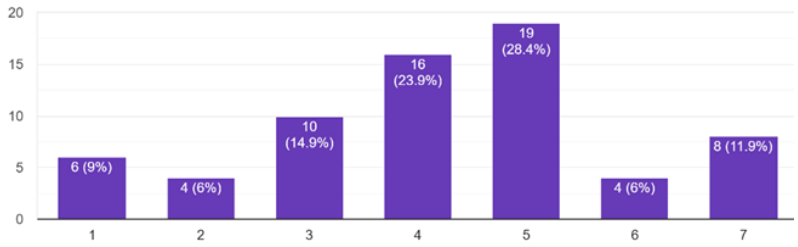

Figure 1. Respondent distribution regarding familiarity with the concept of deepfakes.

Despite this minority unawareness, a substantial portion of the community reported active exposure. Approximately 22.4 percent of the respondents indicated that they had encountered a significant amount of deepfake content online or through various media channels with a median score of 4.0.

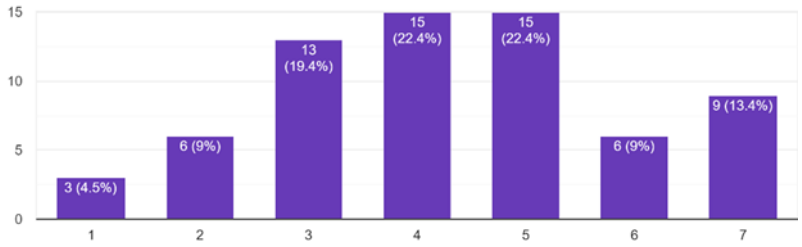

Figure 2. Respondent distribution regarding the frequency of encountering deepfake content.

However, recognizing these manipulations proved to be a critical challenge for the majority. The survey indicated that around 85.1 percent of the participants experienced a hard time distinguishing between authentic media files and those that had been artificially edited using deepfake algorithms. This finding highlighted a significant psychological vulnerability, suggesting that mere exposure did not equate to successful visual detection.

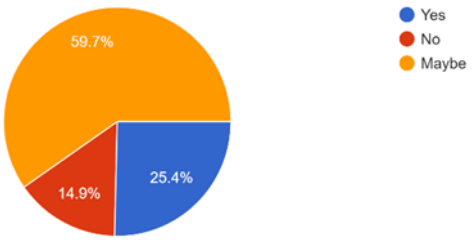

Figure 3. Respondent distribution on the ability to differentiate between edited and original media.

This vulnerability is highly consistent with the technical realities pointed out by previous study [3], [7], who explain that the rapid evolution of generative algorithms makes manual visual detection incredibly difficult even for tech-savvy individuals. It also underscores the "human element" gap raised [24], suggesting that digital natives are highly vulnerable to visual deception despite their daily internet exposure.

B. Perceived Threats to Cybersecurity and Digital Trust

The study further investigated the psychological impact of deepfakes on the public's perception of security and information integrity. The empirical data indicated a severe decline in digital trust among internet users (61.2 percent). At least half of the participants reported that the mere existence of deepfakes had directly reduced their trust in online information.

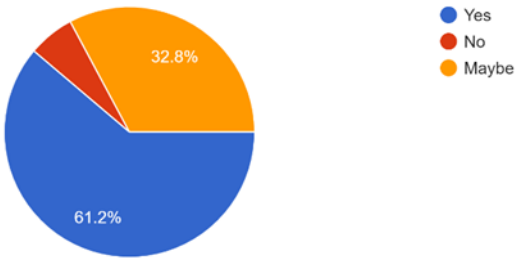

Figure 4. Respondent distribution on whether deepfakes reduced trust in online information.

This skepticism translated into high levels of anxiety regarding cyber safety. Majority of the respondents expressed deep concern over the potential impact of deepfakes on the broader cybersecurity landscape. Giving it a high rating of 5 to 7 with a median concern score of 6.0

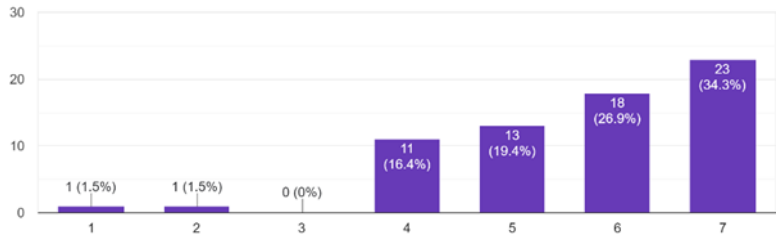

Figure 5. Respondent distribution on the level of concern regarding cybersecurity impacts.

When evaluating the severity of the risk, the survey revealed a unanimous consensus. Approximately 97 percent of the participants agreed that deepfakes posed a significant and imminent threat to both individuals and large-scale organizations.

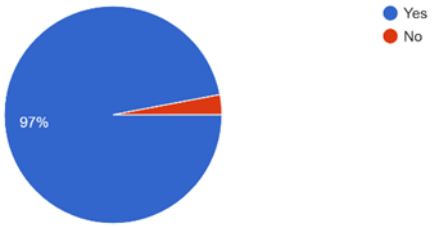

Figure 6. Respondent consensus on deepfakes as a significant threat to individuals and organizations.

Specifically concerning the technical domain, 46.3 percent of the respondents highly rated synthetic media as a direct threat to cybersecurity infrastructures. On a more intimate level, 64.2 percent of the surveyed group felt that their personal lives, identities, and privacy were directly threatened by this rapidly evolving technology.

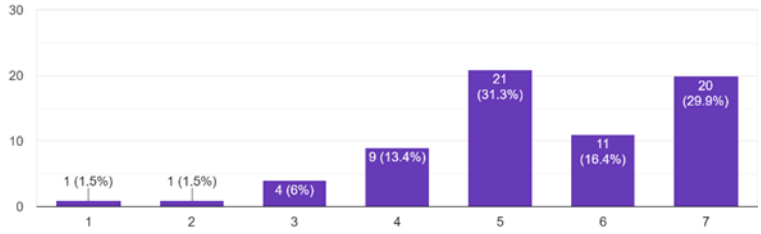

Figure 7. Respondent ratings on deepfakes as a specific cybersecurity threat.

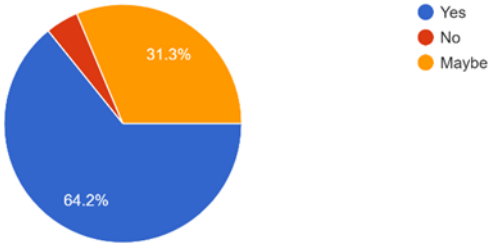

Figure 8. Respondent distribution on deepfakes posing a threat to personal life.

These findings mirror the concerns identified [4],[11], who argue that deepfakes introduce a deep "epistemic uncertainty" that threatens public discourse. The high corporate and personal concern also aligns with warnings identify in previous study [16], [17] about deepfakes being weaponized for social engineering and enterprise fraud, showing that the public is highly aware of these emerging dangers.

C. *Perspectives on Mitigation and Countermeasures*

The final section of the evaluation focused on the participants believe are the most effective ways to combat the deepfake threat. We found an interesting paradox: despite acknowledging how hard it is to visually spot fakes, 76.1 percent of respondents still believe that current cybersecurity software and detection algorithms are highly effective at identifying manipulated media.

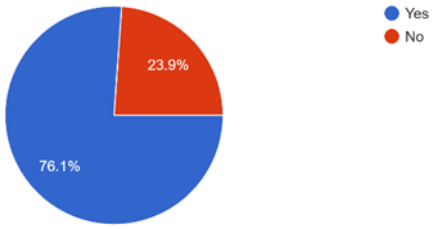

Figure 9. Respondent belief in the effectiveness of current detection technologies.

However, the findings strongly suggested that the community favored proactive, systemic interventions over mere reliance on automated detection tools. The data demonstrated a 92.5 percent agreement among respondents that governing authorities needed to establish and strictly enforce specialized legal frameworks to prevent the malicious misuse of synthetic media.

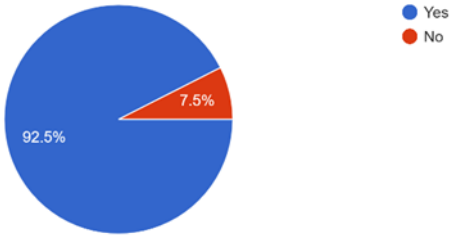

Figure 10. Respondent consensus on the necessity of legal frameworks by authorities.

The survey highlighted unanimous consensus, with 100 percent of the participants strongly advocating for comprehensive public education. Respondents strongly believe that schools, universities, and organizations must offer public training to help everyday citizens critically analyze and handle manipulated media.

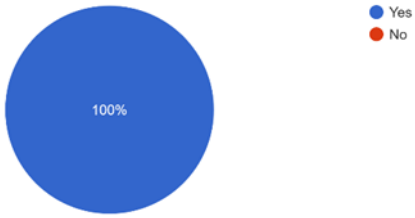

Figure 11. Respondent consensus on the need for public education and training.

Lastly, the demand for accessible protective tools was highly evident. At least 71.6 percent of the respondents supported the development of dedicated, user-friendly software applications designed specifically to prevent personal media from being irresponsibly manipulated by malicious actors.

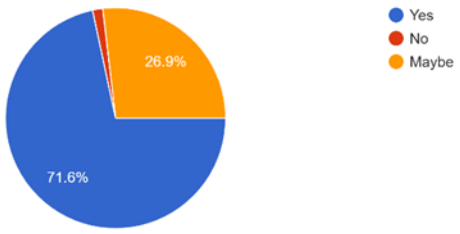

Figure 12. Respondent distribution on the need for preventative software applications.

D. Discussion Summary

Collectively, our empirical results address our core research questions. The findings indicate that while digital natives are highly aware of deepfakes, their actual ability to spot them is dangerously low. This vulnerability has triggered a major decline in digital trust and a massive surge in cyber anxiety. The public's unanimous demand for strict laws and widespread digital literacy training strongly supports our core argument: we cannot solve the deepfake problem with technology alone. As highlighted by previous study [2],[6], while automated detection is vital, it must be paired with user empowerment, public education, and legal policies to build a truly resilient digital society.

V. CONCLUSIONS

In conclusion, this study highlights how deepfakes threaten modern cybersecurity and examines how prepared everyday users are to handle deceptive online media. While our findings show that digital natives generally understand what deepfakes are, the rapid evolution of artificial intelligence makes it incredibly difficult for individuals to manually tell the difference between real and altered content [29], [30]. This high level of visual deception has serious consequences, as bad actors can weaponize synthetic media to spread disinformation, commit identity theft, and manipulate public opinion, which severely erodes digital trust in our media ecosystems [29], [31].

Based on the survey data, participants expressed a strong consensus that relying solely on technical, automated detection software may not be enough to protect everyday users [32], [33]. Rather than expecting technology to catch every fake, we need a more proactive, multi-layered defense. The unanimous agreement among our respondents highlights an urgent need for structural interventions; specifically, governing authorities must enforce strict legal frameworks, and educational institutions must launch public digital literacy programs to empower citizens against cyber fraud [34].

To safeguard the future of the digital landscape, it is highly recommended that further development and follow-up research focus on the creation of accessible, user-friendly detection applications specifically designed for everyday netizens [35], [36]. Additionally, future studies should investigate the integration of advanced digital watermarking and real-time authentication protocols across major media platforms [37]. Ultimately, continuous collaborative efforts across the technological, legal, and educational sectors are imperative to preserve information integrity and prevent the disastrous consequences of deepfake exploitation.

While these findings provide valuable empirical insights into the human side of cybersecurity, several key limitations should be noted. Our dataset is limited to a small sample of 67 respondents, consisting mostly of female university students aged 18 to 23 recruited through online convenience sampling. This demographic and geographic narrowness means our results reflect the views of a specific youth group and may not capture how older generations or different professional groups perceive deepfake risks. Furthermore, our reliance on self-

reported survey responses introduces potential subjective biases regarding participants' actual visual detection skills. To build on this work, future studies should recruit larger, more demographically balanced participant pools across Malaysia and combine subjective surveys with practical visual detection tests to measure real-time accuracy

References

[1] Á. F. Gambín, A. Yazidi, A. Vasilakos, H. Haugerud, and Y. Djenouri, *Deepfakes: current and future trends*, vol. 57, no. 3. Springer Netherlands, 2024. doi: 10.1007/s10462-023-10679-x.

[2] I. Amerini *et al.*, "Deepfake Media Forensics: Status and Future Challenges," *J. Imaging*, vol. 11, no. 3, pp. 1–42, 2025, doi: 10.3390/jimaging11030073.

[3] A. Heidari, N. Jafari Navimipour, H. Dag, and M. Unal, "Deepfake detection using deep learning methods: A systematic and comprehensive review," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 14, no. 2, pp. 1–45, 2024, doi: 10.1002/widm.1520.

[4] C. Vaccari and A. Chadwick, "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News," *Social Media and Society*, vol. 6, no. 1, 2020, doi: 10.1177/2056305120903408.

[5] S. He *et al.*, "Identity Deepfake Threats to Biometric Authentication Systems: Public and Expert Perspectives," 2025.

[6] R. Zia, M. Rehman, A. Hussain, S. Nazeer, and M. Anjum, "Improving synthetic media generation and detection using generative adversarial networks," *PeerJ Comput. Sci.*, vol. 10, pp. 1–20, 2024, doi: 10.7717/peerj-cs.2181.

[7] R. Mubarak, T. Alsboui, O. Alshaikh, I. Inuwa-Dutse, S. Khan, and S. Parkinson, "A Survey on the Detection and Impacts of Deepfakes in Visual, Audio, and Textual Formats," *IEEE Access*, vol. 11, pp. 144497–144529, 2023, doi: 10.1109/ACCESS.2023.3344653.

[8] L. Y. Gong and X. J. Li, "A Contemporary Survey on Deepfake Detection: Datasets, Algorithms, and Challenges," *Electronics (Switzerland)*, vol. 13, no. 3, 2024, doi: 10.3390/electronics13030585.

[9] S. H. Al-Khazraji, H. H. Saleh, A. I. Khalıd, and I. A. Mıshkhal, "Impact of Deepfake Technology on Social Media: Detection, Misinformation and Societal Implications," *Eurasia Proceedings of Science, Technology, Engineering and Mathematics*, vol. 23, pp. 429–441, 2023, doi: 10.55549/epstem.1371792.

[10] S. Singh and A. Dhumane, "Unmasking digital deceptions: An integrative review of deepfake detection, multimedia forensics, and cybersecurity challenges," *MethodsX*, vol. 15, p. 103632, 2025, doi: 10.1016/j.mex.2025.103632.

[11] J. Twomey, D. Ching, M. P. Aylett, M. Quayle, C. Linehan, and G. Murphy, "Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine," *PLoS One*, vol. 18, no. 10 October, pp. 1–22, 2023, doi: 10.1371/journal.pone.0291668.

[12] M. Masood, M. Nawaz, K. M. Malik, A. Javed, A. Irtaza, and H. Malik, "Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward," *Applied Intelligence*, vol. 53, no. 4, pp. 3974–4026, 2023, doi: 10.1007/s10489-022-03766-z.

[13] S. M. Qureshi, A. Saeed, S. H. Almotiri, F. Ahmad, and M. A. A. Ghamdi, "Deepfake forensics: a survey of digital forensic methods for multimodal deepfake identification on social media," *PeerJ Comput. Sci.*, vol. 10, pp. 1–40, 2024, doi: 10.7717/PEERJ-CS.2037.

[14] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A Survey of face manipulation and fake detection," *Information Fusion*, vol. 64, pp. 131–148, Dec. 2020, doi: 10.1016/J.INFFUS.2020.06.014.

[15] Y. Mirsky and W. Lee, "The Creation and Detection of Deepfakes," *ACM Comput. Surv.*, vol. 54, no. 1, 2021, doi: 10.1145/3425780.

[16] N. A. Suleiman, "Deepfake-as-a-Service: The Next Challenge for Enterprise Cybersecurity," *Journal of Technology and Systems*, vol. 7, no. 3, pp. 47–59, 2025, doi: 10.47941/jts.2752.

[17] K. T. Pedersen, L. Pepke, T. Stærmose, M. Papaioannou, G. Choudhary, and N. Dragoni, "Deepfake-Driven Social Engineering: Threats, Detection Techniques, and Defensive Strategies in Corporate Environments," *Journal of Cybersecurity and Privacy*, vol. 5, no. 2, pp. 1–32, 2025, doi: 10.3390/jcp5020018.

[18] S. Verma, "Synthetic Identities and Deepfake Attacks: The Next Frontier in Enterprise AI Security," *European Modern Studies Journal*, vol. 9, no. 5, pp. 176–189, 2025, doi: 10.59573/emsj.9(5).2025.17.

[19] A. Dehghani and H. Saberi, "Generating and Detecting Various Types of Fake Image and Audio Content: A Review of Modern Deep Learning Technologies and Tools," Jan. 2025, Accessed: Aug. 16, 2026. [Online]. Available: https://arxiv.org/pdf/2501.06227

[20] O. Giudice, L. Guarnera, and S. Battiato, "Fighting deepfakes by detecting gan dct anomalies," *J. Imaging*, vol. 7, no. 8, pp. 1–17, 2021, doi: 10.3390/jimaging7080128.

[21] H. El-Taj, F. Alammari, J. Alkhowaiter, L. Bogari, and R. Essa, "Deepfake Detection Based on Visual Lip-sync Match and Blink Rate," *International Journal of Computational and Experimental Science and Engineering*, vol. 11, no. 1, pp. 886–898, 2025, doi: 10.22399/ijcesen.755.

[22] A. Kaur, A. Noori Hoshyar, V. Saikrishna, S. Firmin, and F. Xia, *Deepfake video detection: challenges and opportunities*, vol. 57, no. 6. Springer Netherlands, 2024. doi: 10.1007/s10462-024-10810-6.

[23] R. Ratnawita, "Cybersecurity in the AI Era Measures Deepfake Threats and Artificial Intelligence-Based Attacks," *Journal of the American Institute*, vol. 2, no. 2, pp. 180–189, 2025, doi: 10.71364/s3emxx77.

[24] Subandri, Muhammad Asep, Agus Tedyyana, and I. Gusti Agung Putu Mahendra. "Implementation of AES-128 Encryption for Fingerprint Template Protection in ESP32-Based Biometric Ticketing System." *Journal of Innovation and Technology Polbeng Series on Informatics (INOVTEK Polbeng-Seri Informatika)* 11.1 (2026): 370-380.

[25] I. Etikan, "Comparison of Convenience Sampling and Purposive Sampling," *American Journal of Theoretical and Applied Statistics*, vol. 5, no. 1, p. 1, 2016, doi: 10.11648/J.AJTAS.20160501.11.

[26] J. W. Creswell and D. Creswell, "Research design: Qualitative, quantitative & mixed methods approaches (5th, international student ed.)," *SAGE.*, p. 267, 2018, Accessed: Aug. 16, 2026. [Online]. Available: https://uk.sagepub.com/en-gb/asi/research-design-international-student-edition/book258102

[27] A. Joshi, S. Kale, S. Chandel, and D. Pal, "Likert Scale: Explored and Explained," *Br. J. Appl. Sci. Technol.*, vol. 7, no. 4, pp. 396–403, Jan. 2015, doi: 10.9734/BJAST/2015/14975.

[28] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika 1951 16:3*, vol. 16, no. 3, pp. 297–334, Sep. 1951, doi: 10.1007/BF02310555.

[29] L. Verdoliva, "Media Forensics and DeepFakes: An Overview," *IEEE Journal on Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910–932, 2020, doi: 10.1109/JSTSP.2020.3002101.

[30] T. T. Nguyen *et al.*, "Deep learning for deepfakes creation and detection: A survey," *Computer Vision and Image Understanding*, vol. 223, p. 103525, Oct. 2022, doi: 10.1016/J.CVIU.2022.103525.

[31] M. Mustak, J. Salminen, M. Mäntymäki, A. Rahman, and Y. K. Dwivedi, "Deepfakes: Deceptions, mitigations, and opportunities," *J. Bus. Res.*, vol. 154, p. 113368, Jan. 2023, doi: 10.1016/J.JBUSRES.2022.113368.

[32] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake Detection: A Systematic Literature Review," *IEEE Access*, vol. 10, pp. 25494–25513, 2022, doi: 10.1109/ACCESS.2022.3154404.

[33] M. Taeb and H. Chi, "Comparison of Deepfake Detection Techniques through Deep Learning," *Journal of Cybersecurity and Privacy*, vol. 2, no. 1, pp. 89–106, 2022, doi: 10.3390/jcp2010007.

[34] A. Naitali, M. Ridouani, F. Salahdine, and N. Kaabouch, "Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions," *Computers*, vol. 12, no. 10, pp. 1–26, 2023, doi: 10.3390/computers12100216.

[35] S. Tariq, S. S. Woo, P. Singh, I. Irmalasari, S. Gupta, and D. Gupta, *From Prediction to Explanation: Multimodal, Explainable, and Interactive Deepfake Detection Framework for Non-Expert Users*, vol. 1, no. 1. arXiv, 2025. doi: 10.1145/3746027.3755786.

[36] D. Ghiurău and D. E. Popescu, "Distinguishing Reality from AI: Approaches for Detecting Synthetic Content," *Computers*, vol. 14, no. 1, 2025, doi: 10.3390/computers14010001.

[37] A. D. Samuel-Okon, O. I. Akinola, O. O. Olaniyi, O. O. Olateju, and S. A. Ajayi, "Assessing the Effectiveness of Network Security Tools in Mitigating the Impact of Deepfakes AI on Public Trust in Media," *Archives of Current Research International*, vol. 24, no. 6, pp. 355–375, 2024, doi: 10.9734/acri/2024/v24i6794.

Acknowledgements

The authors would like to express their deepest gratitude to all individuals and institutions that have provided the necessary academic platforms and resources to conduct this research. Special appreciation is extended to the 67 respondents who willingly participated in the online survey; their valuable time and insightful feedback were crucial in formulating the empirical findings of this study. The authors also wish to acknowledge the continuous collective effort, dedication, and strong teamwork among all group members in successfully completing this paper.