

Investigating the Impact of Cyber Security Awareness Training in the field of Deepfake on Students

Mohammad Shahadat Hossain

Abstract

This research investigates how training about cybersecurity affects students' capacity to identify deepfake content during its rapid expansion across digital information platforms. The production of synthetic media by AI threatens personal security at the same time that it creates substantial concerns about false information and declining social trust. The research design employs hypothetico-deductive principles through a pre-post assessment approach which studies 21 graduate students at the University of Staffordshire who are enrolled at Level 7. Research data was collected by two questionnaires before and after training which were evaluated using paired sample t-tests. The results demonstrated a significant statistical improvement across several domains. Subsequent research should utilise a variety of comprehensive data sets and develop longitudinal study approaches to ascertain long-term outcomes, as well as explore how deepfake literacy integrates into digital literacy programmes. The research findings establish that purposeful cybersecurity education can provide the necessary protective abilities to individuals specifically including students so they can face present and emerging security threats effectively.

Keywords— Deepfake, Cyber Security, Training.

Introduction

The dynamics of new technology show that communication and information sharing have made notable improvement. However, it has also made some new types of malicious cyber activities possible, with Deepfake technology. According to Olan, *et al.* (2024), fake news constantly changes people's beliefs about significant issues and their social values. According to Mekhail, *et al.* (2023), Deepfakes always employ artificial intelligence to synthesize what may look like real audios and video and can be indistinguishable from the original. Kaur, *et al.* (2023) noted that it was rather startling to come into contact with Deepfakes and to find out that the approach taken to demoralize influential persons damaged the reputation of so many people unknowingly and unwillingly.

Deepfakes have emerged as a major technological development. According to Rami, *et al.* (2023), the Deepfake technology developed from AI allows users to produce thoroughly fake, realistically convincing digital content such as video and photographs along with audio recordings. Students serve as one of many targeted demographic groups impacted by deepfake media appearing in the digital environments that they encounter for educational purposes. Cybersecurity awareness training represents an important method for students to gain detection abilities toward deepfakes by teaching them how to identify and avoid digital threats.

The main objective of this research is to assess how structured cybersecurity training enhances postgraduate students' detection and response abilities with respect to deepfake content. Participants demonstrated their understanding of deepfake technology through their

knowledge about its creation process and its dangers for individuals and the overall societal environment. The training session led participants to develop self-confidence regarding their skills to detect deepfake content. From the student perspective, the use of computers, the internet and social networks as key sources of education materials, communication tools and information acquisition respectively make it absolutely critical for such threats to be identified and addressed. Designing cybersecurity awareness training programmes to address this group not only promotes critical thinking but also readies employees for what exists in the digital space manipulation. The research aims to investigate the impact of cybersecurity awareness training on students' ability to detect deepfake content.

Literature Review

Here, I provide a review of the most recent and high impact scholarly articles addressing cybersecurity awareness training with a focus on deepfake threats. The review focuses on studies which demonstrate the impact of such training on students: a population that is progressively at higher risk of being coerced into processing fake information. Preu, *et al.* (2022) maintain that Cybersecurity awareness training is needed to help college students detect sophisticated deepfake media. Their research indicates that further training programs need continual development to enhance human identification of synthetic media. Hongmei Chi, *et al.* (2021) advocate Integrating deepfake detection techniques into a cybersecurity curriculum to educate students on identifying deepfakes. The study undertaken by Belghazi, *et al.* (2023) evaluates the impact of training on human capability to differentiate real images from deepfakes. Their study indicates that training provides moderate advancement in relation to users' recognition skills of synthetic faces, but major challenges are still faced in deepfake detection. Bhalli, *et al.* (2024) advocates specialized training of deepfake detection through expert-defined linguistic features used to detect audio deepfakes. Research undertaken by Maram, *et al.* (2022) indicates that enhancing cybersecurity awareness among university students is important to reduce cyber threats. In order to lessen cybersecurity dangers, their survey report highlights the importance of raising cybersecurity awareness among Saudi Arabian university students. Research findings by Tahir, *et al.* (2021) indicate that the detection abilities of deepfake videos among vulnerable users can be enhanced through training. To determine what kind of training is needed, after studying the different abilities needed for production, Hwang Y, *et al.* (2021) conclude that training which provides comprehensive information about disinformation by focusing deeply on deepfakes as examples of synthetic media is effective. Their views are reinforced by Nguyen, *et al.* (2022) who noted that the term "deepfake," which had first been coined from "deep learning" and "fake," had subsequently advanced as artificial intelligence along with machine learning techniques, typically involving the manipulation of images, videos, or audios. In 2023, Victor Murillo-Ligorred and his colleagues introduced deepfake education through artistic education programs for university students, focusing their efforts on senior students.

Since their inception in 2017, deepfake technologies have rapidly advanced. According to Mekhail, *et al.* (2023), their potential to produce hyper-realistic yet fabricated content has markedly increased. Given these factors, cybersecurity awareness has become increasingly critical, particularly among vulnerable populations like students who may be less equipped to identify and protect themselves against digital threats (Rami, *et al.*, 2023). Research has outlined numerous potential risks associated with deepfakes, from defamation and bullying to identity theft and democracy destabilization by researchers (DAMI, 2022). Studies like these categorise the dangers and emphasise their severe implications. Defamation and reputational

harm are well-documented risks, as a convincing deepfake can lead to wrongful associations or false confessions. The research undertaken by Bray, *et al.* (2023), however, shows that humans struggle to detect deep-fake human face images even when using basic quality improvements and indicates the immediate need to confront this security risk.

According to Fattah, *et al.* (2023), cybersecurity awareness training has a beneficial effect on university students' knowledge, attitude and behavior towards cybersecurity. Studies undertaken by Zuriani, *et al.* (2020) indicate that instruction is a useful substitute for imparting cybersecurity expertise. Research on how well training programs can manage these dangers, particularly for student populations, however, is limited. Rami, *et al.* (2023) classify three types of deepfakes (text, audio, and visual) and provide a helpful summary of their uses, but they do not investigate how effective interventions are at thwarting their appearance and recognition across various demographic groups. As noted by Nguyen, *et al.* (2022), educational interventions have historically been measured by changes in awareness levels among participants and, focused primarily on threats, rather than identification. Consequently, as noted by Nanjundaswamy, *et al.* (2022), there is a growing call for training programs that go beyond basic awareness to enhance critical thinking and digital literacy.

While cybersecurity training has been extensively researched, few studies specifically address deepfake detection, and even fewer investigate its effectiveness within a student demographic. This gap suggests a need for targeted training programs that address the unique features of deepfake technology and their specific impact on students, who may lack the advanced digital literacy required to recognize manipulative media. Research that has been carried out has also made heavy use of subjective measures, namely self-reported increases in awareness or in confidence which are in themselves biased. According to Bondi, *et al.* (2021), the effectiveness of deepfake video detectors depends on training approaches and data augmentation techniques. For example, Groh, *et al.* (2021) proved that untrained humans excel above the top computer vision model in deepfake detection through facial processing capability. According to Shiohara and Yamasaki (2022), self-blended image training enhances deepfake detection capabilities particularly when identifying unknown manipulations and working with datasets. The authors introduce self-blended images (SBIs) as new synthetic training data for deepfake detection. Additionally, Nguyen, *et al.* (2022), have also recognised that existing approaches to cybersecurity training tend not to be demographic specific, rendering them less relevant for younger audiences. Younger users, for example, already use social media extensively, which is where deepfakes flourish most virulently, and they may lack the experience or digital literacy skills to spot them. According to Kurt, *et al.* (2017), having a good ideal security awareness program for younger audiences is one of the best ways to enhance security. Khamzina, *et al.* (2022) has identified security problems and advocates that awareness training to be provided to university students, in particular. In the study undertaken by Naffe, *et al.* (2023), for example, the learning experience gained through deepfake training allowed participants to advance from deepfake ignorance into digital qualification skills.

While existing studies provide valuable insights into the technical and social dimensions of deep fakes, several critical gaps remain unaddressed. Given the above gaps, this study aims to fill a unique niche in cybersecurity research by examining the specific effectiveness of a cybersecurity awareness training program focused on deepfake awareness within a student

population. The objective is to investigate the impact on the effectiveness of the cyber security awareness training scheme in fighting against the deepfake threats among students.

Research Methods

This research examines how cybersecurity awareness training affects student abilities to identify deepfake content in their environment decreases and the extent to which training decreases the risks for misinformation and cybersecurity threats in the digital era. The research methodology of this study follows hypothetico-deductive principles to create hypotheses testing and to gather statistical data regarding the influence of independent training factors on dependent variables of student detection capabilities (perception, confidence, performance, engagement, and behaviour). The chosen methodology enables effective examination of the educational impact of the training programme between the intervention and assessment of student abilities to recognise deepfakes, as it provides a structured hypothesis for validation and testing capabilities. Training effectiveness receives a complete assessment through combining quantitative research with qualitative methods. The research is guided by the following hypotheses: **H₀ (Null Hypothesis):** Cybersecurity awareness training has no significant effect on students' confidence, performance, perception, changes in behaviour (in terms of media consumption), or engagement in detecting deepfake content. **H₁ (Alternative Hypothesis):** Cybersecurity awareness training significantly improves students' confidence, performance, perception, changes in behaviour, and engagement in detecting deepfake content.

Research Design

A pre-post intervention design was employed, involving 21 matched Level 7 student participants who completed both pre-training and post-training surveys. This quasi-experimental design enabled direct comparison of dependent variables before and after training, by isolating its impact. The absence of a control group due to logistical constraints limits causal inferences but aligns with the study's exploratory nature. To provide a thorough grasp of the training's impact, a mixed-method approach was employed, integrating quantitative and qualitative analyses.

The sample was comprised of 21 students: **Age:** Mean=25.8 years, SD=2.7, range=21–31; 71.4% (n=15) aged 23–27. **Gender:** 90.5% male (n=19), 9.5% female (n=2). **Education:** 100% Level 7 (postgraduate). The small, male-dominated sample is sufficient for paired sample t-tests (power≈0.8, $\alpha=0.05$, medium-to-large effect size) but limits generalizability. The homogeneous educational level ensures consistency.

Data was collected via Microsoft Form-based surveys: **Pre-Training Survey:** Administered March 19–April 3, 2025, with 18 questions capturing baseline metrics (e.g., perception, awareness, confidence, behaviour, and verification frequency). **Post-Training Survey:** Conducted April 3–9, 2025, with 23 questions (18 pre-training plus 5 post-training specific questions, e.g., training satisfaction). Questions included Likert scales (e.g., confidence: 1=Not at all confident, 5=Very confident), multiple-choice (e.g., indicators), and open-ended responses (e.g., concerns). Participants were matched via email IDs, ensuring data integrity.

A one-hour training session was delivered between March 27 and April 7, 2025, via mixed methods (live Zoom, in-person, recorded video, PowerPoint slides). It covered: **Introduction:**

Deepfake definition, history, examples (e.g., political misinformation). **Technical Skills:** Indicators (unnatural blinking, inconsistent lighting, audio-visual mismatch) and tools (reverse image search, trusted news sources). **Critical Thinking:** Exercises to foster scepticism, source evaluation and proactive behaviours.

Quantative and Qualitative Methods for Data Analysis

Quantitative

Descriptive Statistics: Means, standard deviations, and percentages were calculated for each question to summarize response distributions. (e.g., % selecting “Yes” for awareness).

Paired sample t-tests: Used to compare pre- and post-training scores for continuous variables (confidence, threat perception, number of indicators). The paired t-test assumed normality of differences, which was reasonable given the sample’s consistency and Likert-scale data. T-Tests conducted on confidence (Q5), threat perception (Q9), and performance (Q6): number of indicators) to test significance ($\alpha=0.05$). Normality of differences was assumed based on Likert-scale consistency. **Significance Level:** Set at $\alpha = 0.05$.

Qualitative

Open-ended responses (Q17, Q18, Q22) were thematically explained, identifying themes (e.g., misinformation, critical thinking).

Ethical Considerations

Participants provided informed consent and participated voluntarily. The study complied with university ethical guidelines, ensuring confidentiality and all data will be deleted after completion of this research.

Research Findings.

Here is the descriptive analysis of the pre- post training data and responses to key questions aligned with variables:

Demographic Information:

Age: Mean=25.8, SD=2.7, range=21–31; distribution: 21 (4.8%, n=1), 23 (14.3%, n=3), 24 (9.5%, n=2), 25 (28.6%, n=6), 26 (9.5%, n=2), 27 (19.0%, n=4), 29 (4.8%, n=1), 30 (4.8%, n=1), 31 (4.8%, n=1); Most (71.4%, n = 15) were 23–27, reflecting a young adult cohort. **Gender:** 90.5% male (n = 19/21), 9.5% female (n = 2/21, suggesting a male-dominated sample. **Education:** 100% Level 7 (n = 21), indicating uniform academic capability.

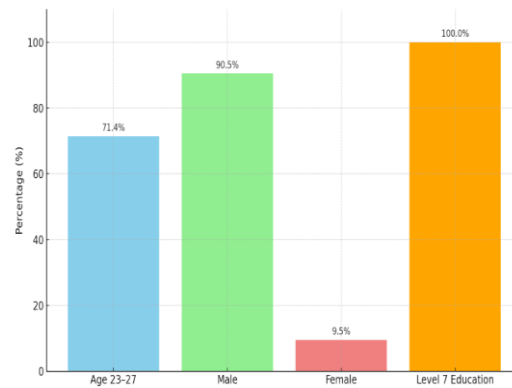


Fig. 1. Participant Demographics Overview.

Results/Pre- and Post-Training Responses:

TABLE 1: Pre- and Post-Training Responses/Results Summary:

Question No	Variables/ Focus	Pre-Training Responses	Post-Training Responses	Insight
Q4	Perception/ Awareness	52.4% Yes, 47.6% No	100% Yes	Knowledge gap closed
Q5	Confidence	Mean=2.24 (SD=1.18); Most rated 1–3	Mean=4.48 (SD=1.03); Most rated 4–5	Strong confidence gain
Q6	Performance/ Indicators	Mean=1.0 (SD=0.0); Mostly incorrect indicators	Mean=2.81 (SD=0.75); Mostly correct indicators	Improved identification skills
Q7	Perception/ Encountered Deepfake	33.3% Yes, 38.1% Not Sure	76.2% Yes, 19.0% Not Sure	Recognition skill improved
Q8	Perception/ Perceived Uses	Cybercrime: 61.9%; Entertainment: 33.3%	Cybercrime: 95.2%; Broader recognition overall	Greater awareness of malicious uses
Q9	Perception/ Threat Perception	Mean=3.29 (SD=1.52); Mixed threat awareness	Mean=4.76 (SD=0.54); Stronger agreement on threat	Heightened threat perception
Q10	Perception/ Vulnerable Industries	28.6% selected all four; 47.6% selected one	90.5% selected all four	Near-complete awareness
Q11	Engagement/ Fallen for Fakes	28.6% Yes, 38.1% Not Sure	61.9% Yes, 19.0% Not Sure	Higher reporting, not necessarily higher risk
Q12	Behaviour/	28.6% Never; 14.3% Always	81.0% Always	Big behavioral shift

Question No	Variables/ Focus	Pre-Training Responses	Post-Training Responses	Insight
	Verification Frequency			
Q13	Performance/ Verification Tools	47.6% used none; 23.8% used all three	76.2% used all three	Significant tool adoption
Q14	Engagement/ Information Sources	57.1% Social Media	71.4% Academic/News	More credible sources adopted
Q15	Behaviour/ Actions Taken	66.7% Ignore; 0% Discuss	85.7% Verify; 23.8% Discuss	Proactive action post-training
Q16	Perception/ Future Difficulty	61.9% Yes, 23.8% Not Sure	90.5% Yes, 9.5% Not Sure	Growing concern about detection
Q17	Perception/ Concerns	47.6% Not Sure	Misinformation: 57.1%, Social Impact: 38.1%	More specific concerns emerged
Q18	Confidence Needs	52.4% Not Sure	71.4% cited training	Clear preference for further education
Q19–Q23	Post-Only	N/A	95%+ satisfied, more skeptical, confident, and aware	Training well-received and impactful

Descriptive Narrative:

Q4: Have you heard of deepfake technology before?

Prior to training, participants showed a split in awareness regarding deepfakes, with only just over half reporting any familiarity. After the session, all participants confidently affirmed awareness, suggesting a full closure of the initial knowledge gap. **Insight:** Knowledge gap closed: 100% awareness achieved in post-training.



Fig. 2. Pre-Post Training Awareness of Deepfake Technology (Q4).

Q5: How confident are you in identifying deepfake videos? (Scale: 1 = Not at all confident, 5 = Very confident).

Confidence levels (Q5) saw a significant shift. The students gave ratings between 1 and 3 on a scale from 1 to 5 for confidence levels prior to the training program. Test subject ratings of their deepfake detection abilities shifted strongly towards the highest rating points after receiving training. **Insight:** Significant shift from low (1–3) to high (4–5) confidence in post-training.



Fig. 3. Pre- and Post-Training Confidence in Identifying Deepfake Videos (Q5).

Q6: Which of the following are common indicators of deepfake videos? (Select all that apply.) When asked about specific indicators of deepfakes pre-training responses showed reliance on inaccurate cues. Post-training, participants overwhelmingly identified correct signs such as unnatural blinking and inconsistent lighting. This demonstrated a tangible increase in analytical skill. **Insight:** Enhanced skill, meeting expected deepfake indicators post-training.

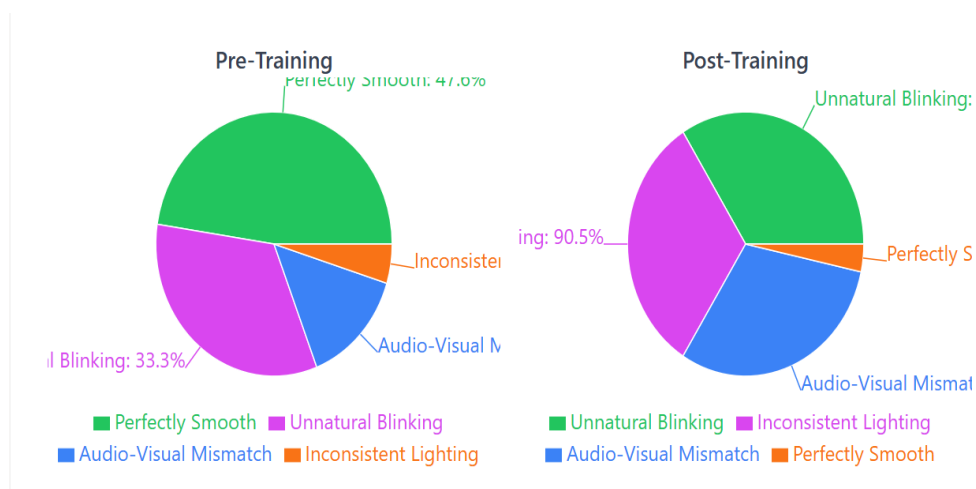


Fig. 4. Pre and Post Training Selection of Deepfake Indicators (Q6).

Q7: Have you ever encountered a deepfake video online?

In terms of actual encounters with deepfakes, there was a noticeable increase in self-reported recognition—from 33.3% to 76.2%. This suggests better identification rather than more exposure. **Insight:** Increased recognition due to improved identification skills.

Q8: What do you think deepfake technology is primarily used for?

Perceptions of how deepfakes are used shifted dramatically. Initially, most saw them as a tool for cybercrime and entertainment, but post-training, more participants acknowledged their political and academic implications, indicating an expanded understanding. **Insight:** Broader recognition of malicious uses.

Q9: On a scale of 1 to 5, how much of a cybersecurity threat do you think deepfake technology poses? Perceived threat levels rose substantially, with most respondents rating the threat at the highest level (5) after training, compared to a wide spread of opinions beforehand. **Insight:** Significantly increased the perceived threat level.

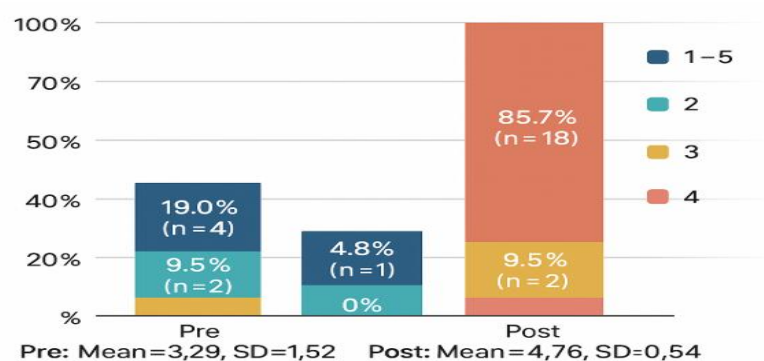


Fig. 5. Perceived threat levels Q9

Q10: Which industries are most vulnerable to deepfake threats?

Awareness of vulnerable industries was limited at first, with fewer than a third identifying all four key sectors. This changed dramatically post-training, with over 90% recognizing the full scope of impact. **Insight:** Near-complete awareness.

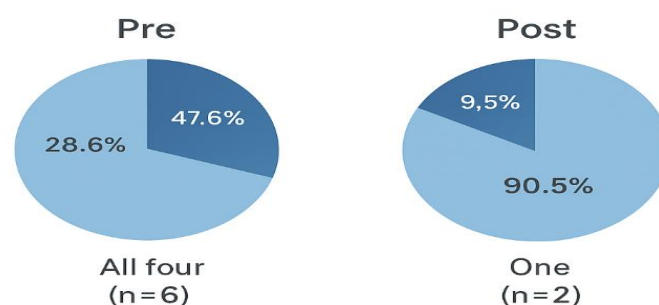


Fig. 6. Awareness of vulnerable industries- Perception (Q10)

Q11: Have you ever fallen for a fake or manipulated video/image online?

An interesting change was seen in reported susceptibility. More people admitted to falling for fake content. Rather than reflecting vulnerability, this likely reflects increased self-awareness and willingness to report past mistakes. **Insight:** Increased reporting may reflect awareness, not susceptibility.

Q12: How frequently do you verify the authenticity of videos before sharing them?

Verification behaviours changed the most drastically. The majority who never or rarely verified information switched to "always". Doing so, showing a strong behavioural impact of the training. **Insight:** Dramatic shift to proactive behaviour.

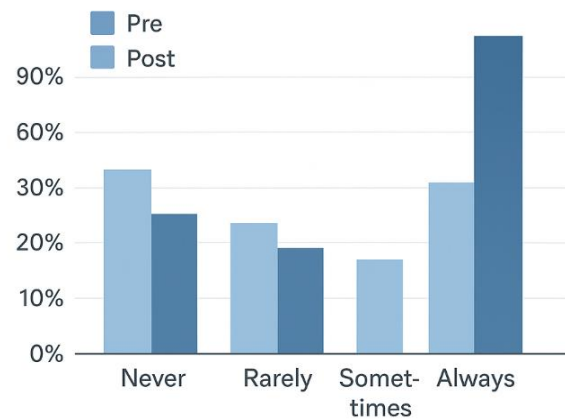


Fig. 7. Verification behaviours (Q12).

Q13: Which tools or methods do you currently use to verify the authenticity of online content?

This is supported by the jump in use of verification tools (Q13), especially those who used all three methods discussed (e.g., reverse image search, news, and video analysis). **Insight:** Significant tool adoption.

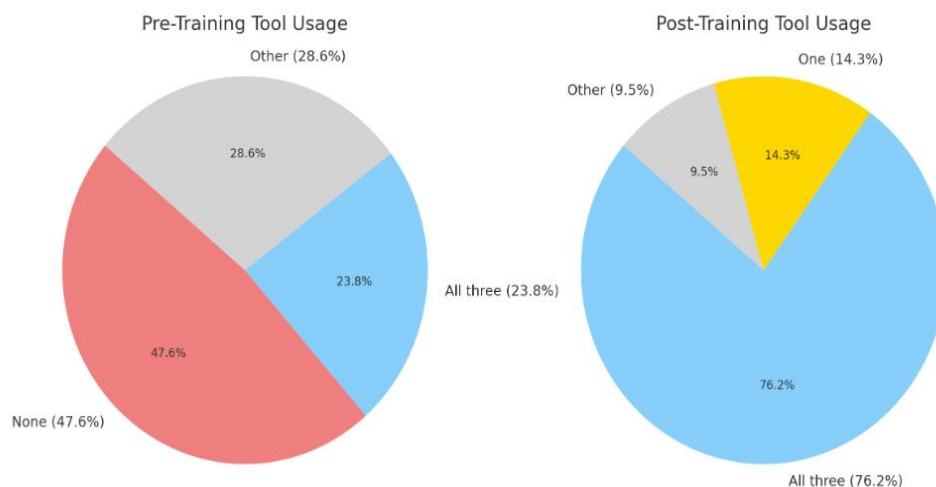


Fig. 8. Verification Tools (Q13. Performance)

Q14: What is your main source of information about deepfake technology?

When asked about their sources of information (Q14), the majority had relied on social media before training. After training, academic and news sources took precedence, indicating a turn towards more credible platforms. **Insight:** Shift to credible sources.

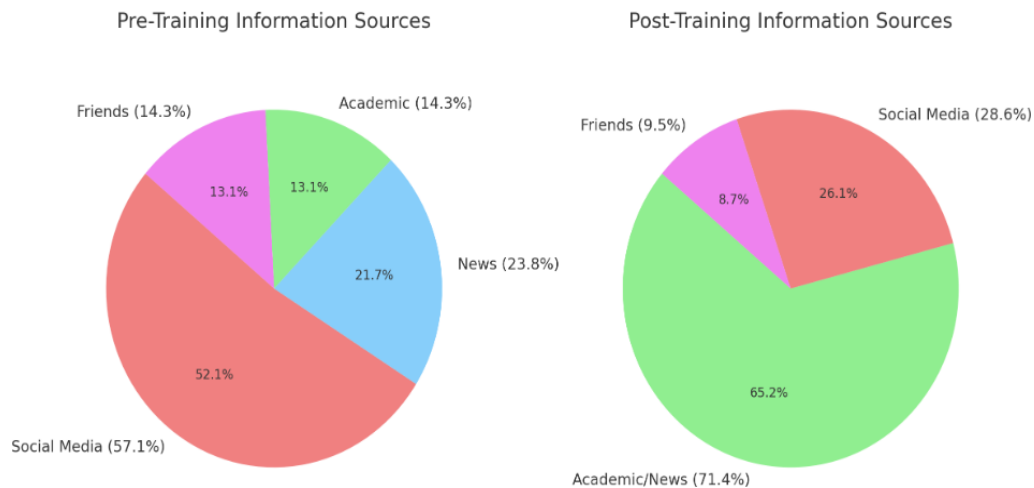


Fig. 9. Sources of information (Q14) Engagement

Q15: If you suspect a deepfake, what action would you take?

Response actions (Q15) also improved. Where previously most ignored suspicious content, post-training, the majority now verify and report—some even engage in discussions. **Insight:** Shift to proactive actions.

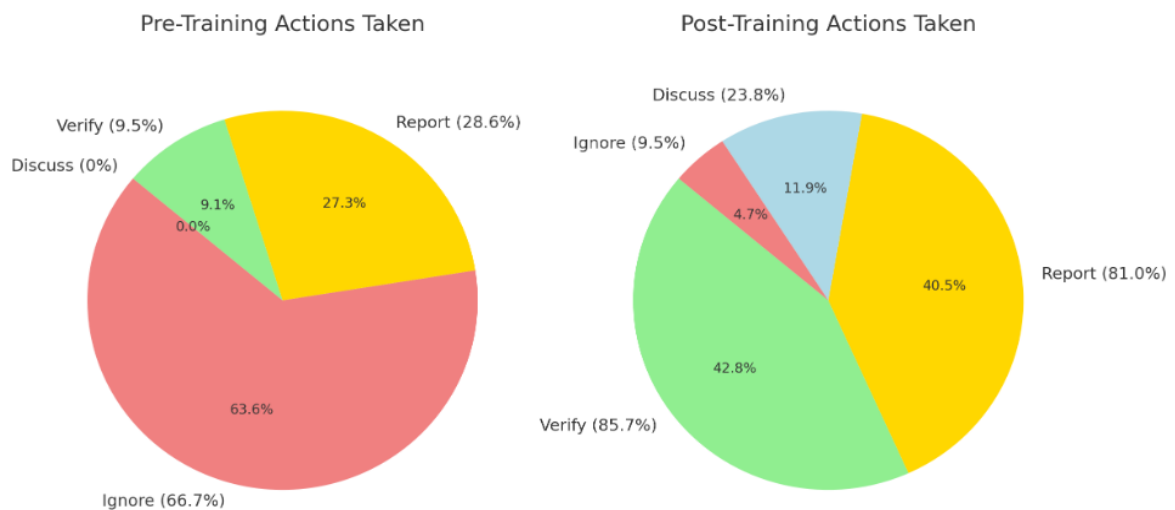


Fig.10. Response actions (Q15) Behaviour.

Q16: Do you believe deepfake technology will become more difficult to detect in the future? Participants also expressed heightened concern about future detection (Q16) and moving from vague uncertainties to concrete issues like misinformation and social impact. **Insight:** Increased concern.

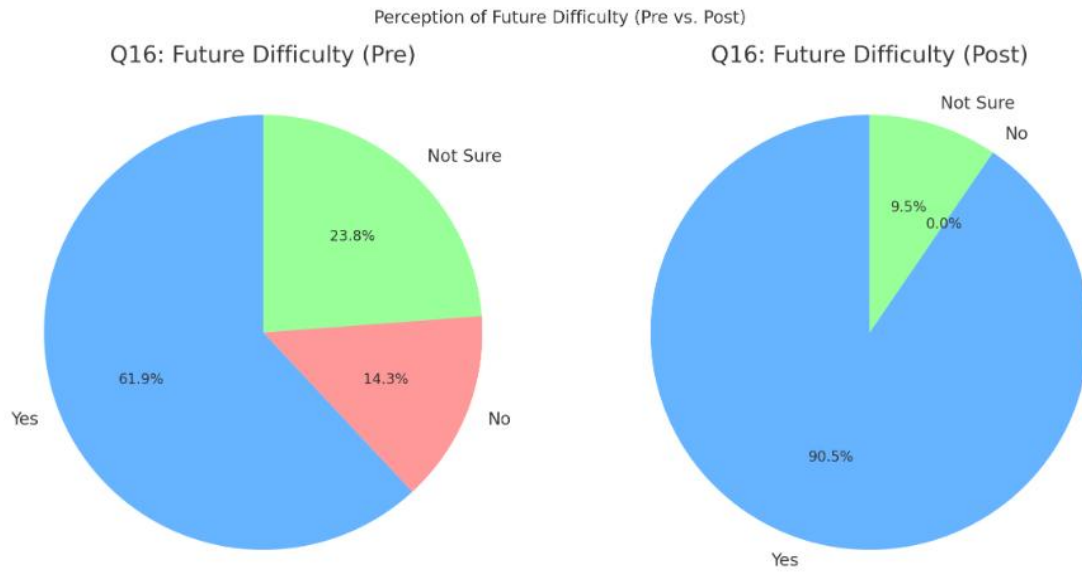


Fig. 11. Future Difficulty Q16 (Perception).

Q17: What concerns you the most about deepfakes?

Participants also expressed heightened concern about future detection and became more articulate about what worries them most (Q17), moving from vague uncertainties to concrete issues like misinformation and social impact. **Insight:** More specific concerns post-training.

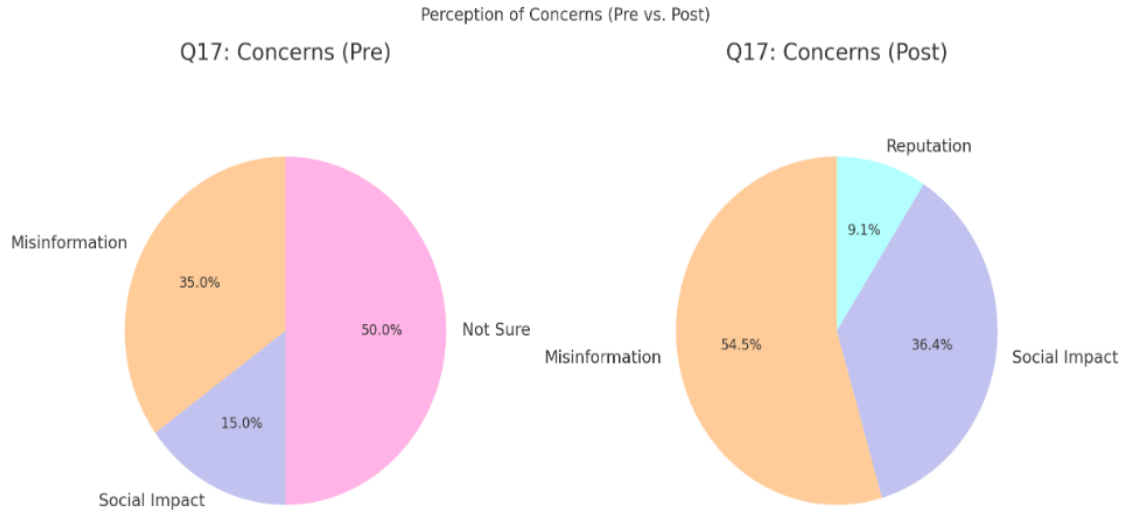


Fig. 12. Perception of concerns (Q17).

Q18: What would help you feel more confident in detecting deepfake videos?

When discussing what would increase their confidence (Q18), more than two-thirds emphasized the importance of training, reinforcing the value of continued education. **Insight:** Training increasingly cited to made them confident to identify deepfake.

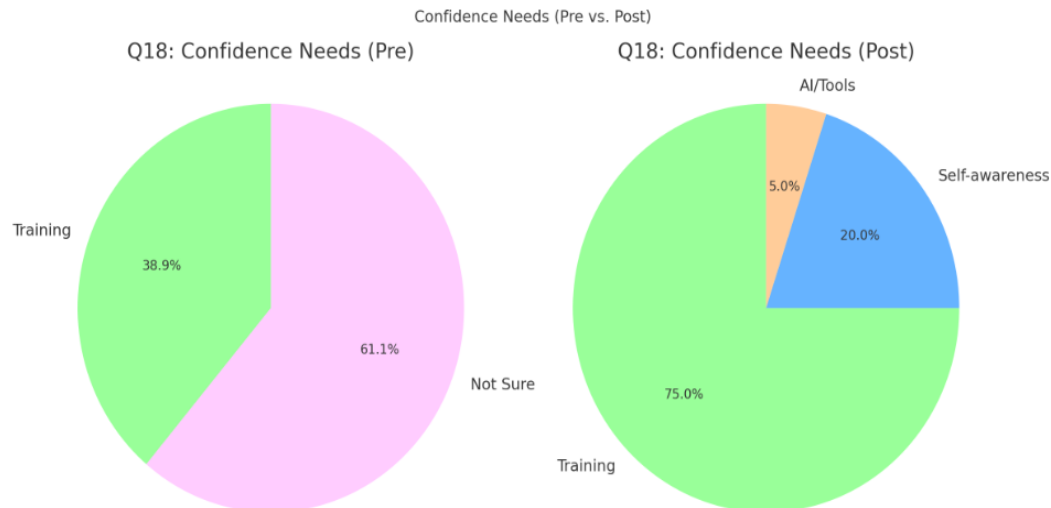


Fig. 13. What can be made them confident to identify deepfakes – Q18.

Post-only Responses:

Lastly, post-only questions (Q19–Q23) showed overwhelmingly positive feedback. Over 95% felt more confident, sceptical, and capable of identifying deepfake images. They found the most useful techniques to be critical thinking and video analysis, and almost all reported satisfaction with the training.

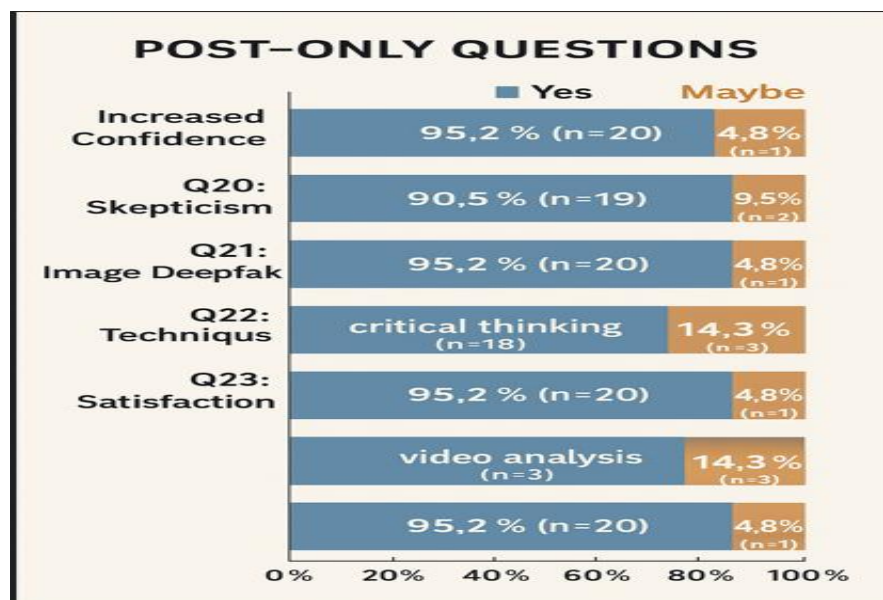


Fig.14. Post-Training responses.

Statistical Analysis Using Paired T-tests

To rigorously evaluate the impact of the cybersecurity awareness training program on students' ability to detect deepfake content, paired sample t-tests were conducted for key

variables: confidence (Q5), performance (number of indicators identified, Q6), and threat perception (Q9). Paired t-tests are suitable for assessing changes over time within the same individuals, ensuring a robust within-subjects comparison.

Q5: Confidence (Confidence):

Pre-training: Pre: Mean=2.24, SD=1.18; 47.6% (n=10) rated 1, 14.3% (n=3) rated 2, 28.6% (n=6) rated 3, 9.5% (n=2) rated 4, 0% rated 5.

Mean = 2.24, SD = 1.18. Distribution of responses:

47.6% (n=10) rated 1, 14.3% (n=3) rated 2, 28.6% (n=6) rated 3, 9.5% (n=2) rated 4, 0% rated 5.

Post-training: Post: Mean=4.48, SD=1.03; 4.8% (n=1) rated 1, 4.8% (n=1) rated 2, 0% rated 3, 14.3% (n=3) rated 4, 76.2% (n=16) rated 5.

Mean = 4.48, SD = 1.03. Distribution of responses:

4.8% (n=1) rated 1, 4.8% (n=1) rated 2, 0% rated 3, 14.3% (n=3) rated 4, 76.2% (n=16) rated 5.

Paired t-test:

To assess whether the improvement in performance was statistically significant, a **paired t-test** was conducted comparing the pre- and post-training results.

Pre-training mean = 2.24, Post-training mean = 4.48, t-value = -8.36, p-value = < 0.001.

Insight: Significant gain, nearing expected 5. For confidence levels (Q5), the pre-training mean score was 2.24 (SD = 1.18), while the post-training mean rose substantially to 4.48 (SD = 1.03). The paired t-test produced a t-value of -8.36 with a **p-value < 0.001**, indicating a statistically significant improvement in students' confidence in identifying deepfakes after the training.

Q6: Indicators (Performance):

Pre-training: Pre: Mean=1.0, SD=0.0; 47.6% (n=10) perfectly smooth (incorrect), 33.3% (n=7) unnatural blinking, 14.3% (n=3) audio-visual mismatch, 4.8% (n=1) inconsistent lighting.

Mean = 1.0, SD = 0.0. Distribution of responses: 47.6% (n=10) identified **perfectly smooth facial expressions** (incorrect), 33.3% (n=7) identified **unnatural blinking patterns** (correct), 14.3% (n=3) identified **audio-visual mismatch** (correct), 4.8% (n=1) identified **inconsistent lighting** (correct). In the pre-training responses, students largely focused on **incorrect indicators** like **perfectly smooth facial expressions** (which is not typically a reliable indicator of deepfakes). The more **accurate indicators**, such as **unnatural blinking** and **audio-visual mismatch**, were only identified by a smaller percentage of students.

Post-training: Post: Mean=2.81, SD=0.75; 95.2% (n=20) unnatural blinking, 90.5% (n=19) inconsistent lighting, 85.7% (n=18) audio-visual mismatch, 9.5% (n=2) perfectly smooth.

Mean = 2.81, SD = 0.75, Distribution of responses: 95.2% (n=20) identified **unnatural blinking patterns** (correct), 90.5% (n=19) identified **inconsistent lighting** (correct), 85.7% (n=18) identified **audio-visual mismatch** (correct), 9.5% (n=2) identified **perfectly smooth facial expressions** (incorrect)

After the training, there was a **significant improvement** in the ability to identify the correct indicators of deepfakes. The most frequently selected indicators were the correct ones, such as **unnatural blinking patterns**, **inconsistent lighting**, and **audio-visual mismatch**. Only a small number of students (9.5%) still selected **perfectly smooth facial expressions**, which is not a reliable indicator.

Paired t-test:

To assess whether the improvement in performance was statistically significant, a **paired t-test** was conducted comparing the pre- and post-training results.

Pre-training mean = 1.0, post-training mean = 2.81, Standard deviation of post-training = 0.75, t-value = -11.24, p-value = < 0.001.

The **t-value of -11.24** indicates a strong negative difference between the pre- and post-training means. A higher negative t-value suggests a larger improvement in the students' ability to identify deepfake indicators. The **p-value of < 0.001** is very small, meaning that the probability that the observed difference between pre- and post-training performance occurred by random chance is extremely low (less than 0.1%). This confirms that the observed improvement is statistically significant. The findings from the **paired t-test, $p < 0.001$** , is very strong evidence that the cybersecurity awareness training resulted in an extremely significant improvement of the students' ability to identify deepfake indicators. The pre-training mean of 1.0 and post-training mean of 2.81 meaning that the practical skills of students on identifying deepfakes improved according to the expected indicators.

Q9: Threat Perception (Perception):

Pre-Training Analysis:

Pre: Mean=3.29, SD=1.52; 19.0% (n=4) rated 1, 9.5% (n=2) rated 2, 28.6% (n=6) rated 3, 4.8% (n=1) rated 4, 38.1% (n=8) rated 5.

Mean = 3.29, Standard Deviation (SD) = 1.52, Percentage Distribution: 19.0% (n = 4) rated 1, 9.5% (n = 2) rated 2, 28.6% (n = 6) rated 3, 4.8% (n = 1) rated 4, 38.1% (n = 8) rated 5.

Post-Training Analysis:

Post: Mean=4.76, SD=0.54; 0% rated 1–2, 4.8% (n=1) rated 3, 9.5% (n=2) rated 4, 85.7% (n=18) rated 5.

Mean = 4.76, Standard Deviation (SD) = 0.54, Percentage Distribution: 0% rated 1-2, 4.8% (n = 1) rated 3, 9.5% (n = 2) rated 4, 85.7% (n = 18) rated 5,

Paired T-test Result:

t-value = -5.87, p-value = $p < 0.001$. Insight: For threat perception (Q9), participants' mean rating increased from 3.29 (SD = 1.52) pre-training to 4.76 (SD = 0.54) post-training. The t-test value of -5.87 (**$p < 0.001$**) again confirmed that students perceived a greater cybersecurity threat from deepfakes following the intervention. The assumptions underlying the paired t-tests, including the approximate normality of difference scores, were considered reasonable

based on the relatively symmetrical distributions and the nature of the 5-point Likert scale measurements used. There is a **significant increase** in students' perception after the training (rating 5).

Validation

Quantitative: Achieved expected outcomes (e.g., 76.2% confidence = 5, 81.0% always verify), closely aligning with target.

Qualitative: Themes like “critical thinking” (85.7%, n = 18, Q22) and “social impact” (71.4%, n = 15, Q17) validated the training’s practical relevance.

Limitation

Small sample (n=21) limits generalizability. Male-dominated sample may skew results. No control group allows external influences. Self-reported data may reflect social desirability. Short-term post-training surveys (1–2 weeks) lack long-term insight.

Conclusion and future work

This study investigated the impact of a cybersecurity awareness training program on 21 Level 7 students’ abilities to detect deepfake content, using a hypothetico-deductive methodology and mixed-methods analysis. The findings decisively reject the null hypothesis (H_0) and support the alternative (H_1), demonstrating significant improvements across all dependent variables.

The results confirmed that the training effectively closed knowledge gaps (awareness: 52.4% to 100%), boosted confidence, aligned perceptions, fostered cautious media consumption and contributed positively to its practical focus on indicators, tools, and critical thinking. While the study achieved its objectives, several avenues for future research and development emerge including Larger and Diverse Samples, Longitudinal Assessment, Integration of AI Tools, Objective Testing, Tailored Interventions, and Broader Contexts.

References.

- Belghazi, Nadia, *et al.* (2023). *Is Training Useful to Detect Deepfakes? : A Preliminary Study..* Aveiro, Portugal, IEEE.
- Bhalli, Nashaba, *et al.* (2024) Listening for Expert Identified Linguistic Features: Assessment of Audio Deepfake Discernment among Undergraduate Students. Research Gate Preprint. DOI 10.48550/arXiv/2411.14586.
- Bond, Luca, *et al.* (2020). *Training Strategies and Data Augmentations in CNN-based DeepFake Video Detection.* New York, USA, IEEE.
- Bray, S. D., Johnson, S. D., & Kleinberg, B. (2023). Testing Human Ability To Detect Deepfake Images of Human Faces. *ArXiv*, Volume <https://arxiv.org/abs/2212.05056>.
- DAMI, L. (2022). Analysis and conceptualization of deepfake technology as cyber threat. *Researchgate*, DOI:10.13140/RG.2.2.21862.50246.

- Fattah, A., Wagimin, W. & Nurlia (2023). Enhancing Cybersecurity Awareness among University Students: A Study on the Relationship between Knowledge, Attitude, Behavior, and Training. *Jurnal System Informasi (E-Journal)*, Volume 15, <https://ejournal.unsri.ac.id/index.php/jsi/article/view/21812>, 2023.
- Groh, M., *et al.* (2021). Human Detection of Deepfakes: A Role for Holistic Face Processing. *ARVO*, Vol.21, 2390 (doi:<https://doi.org/10.1167/jov.21.9.2390>).
- Hongmei Chi, *et al.* (2021). Integrating Deepfake Detection into Cybersecurity Curriculum. *Advances in Intelligent Systems and Computing*, Vol 1288, 588–598.
- Hwang, Y., Ryu, J. Y., & Jeong, S.-H. (2021). Effects of disinformation using deepfake: The protective effect of media literacy education. *Cyberpsychology, Behavior, and Social Networking*, 24(3) (<https://doi.org/10.1089/cyber.2020.0174>), 188–193.
- Kaur, Parveen, *et al.* (2023). The Perception of Deepfake Technology Among University Students in Malaysia. *International Journal of Membrane Science and Technology*, Volume 10, 1720-1731.
- Khamzina, Botagoz, *et al.* (2022). Determination of Cyber Security Issues and Awareness Training for University Students. *International Journal of Emerging Technologies in Learning (IJET)*.
- Kurt, Yesem, *et al.* (2017). Raising Cybersecurity Awareness among College Students. Colloquium for Information Systems Security Education (CISSE)
- Maram, M., Bamasoud, & Doaa, M. (2022). The Impact of Enhancing awareness of cybersecurity on university students: a survey paper. *Little Lion Scientific*, pp. 4756 - 4766.
- Mekhail, Mustak, *et al.* (2023). Deepfakes: Deceptions, mitigations, and opportunities. *Journal of Business Research*, Volume 154 (2) DOI 10.10.16/jbusres.2022 113368
- Murillo-Ligorred, Victor, *et al.* (2023). Integration and Scope of Deepfakes in Arts Education: The Development of Critical Thinking in Postgraduate Students in Primary Education and Master's Degree in Secondary Education. *MDPI*, Volume Educ. Sci. 2023, 13(11), 1073. <https://doi.org/10.3390/educsci13111073>.
- Naffi, Nadia, *et al.* (2023). Empowering Youth to Combat Malicious Deepfakes and Disinformation: An Experiential and Reflective Learning Experience Informed by Personal Construct Theory. *Research Gate*, www.researchgate.net/publication/376678537.
- Nanjundaswamy, Vasist, P. & Krishnan, Satish (2022). Deepfakes: An Integrative Review of the Literature and an Agenda for Future Research. *Communications of the Association for Information Systems*, 51(1), 556 - 602.
- Nguyen, Thanh T., *et al.* (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding*, Volume 223, 103525.
- Olan, Femi, *et al.* (2024). *Fake News on Social Media: The Impact on Society*. Springer.
- Preu, E., Jackson, M. & Choudhury N. (2022) Perception vs. Reality: Understanding and Evaluating the Impact of Synthetic Image Deepfakes over College Students. *IEEE*, 547–553.
- Rami, Mubarak, *et al.* (2023). A Survey on the Detection and Impacts of Deepfakes in Visual, Audio, and Textual Formats. *IEEE Access*, pp. 144497 - 144529.
- Shiohara, K & Yamasaki, T. (2022). *Detecting Deepfakes with Self-Blended Images*. New Orleans, LA, USA, IEEE.
- Tahir, Rashid, *et al.* (2021). Seeing is Believing: Exploring Perceptual Differences in DeepFake Videos. *ACM*, Article No 172(<https://doi.org/10.1145/3411764.3445699>), 1 - 16.
- Yesem Kurt, *et al.* (2017). Raising Cybersecurity Awareness among College Students. Colloquium for Information Systems Security Education (CISSE).
- Zuriani, A.K., *et al.* (2020). Impact of training on cybersecurity awareness. *Gading Journal for Science and Technology*, Volume 3, 114-20.