Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 4, 2026

AI-GENERATED DEEPFAKES AND CYBERSECURITY RISK PERCEPTION: INVESTIGATING ATTITUDES OF STUDENTS AT UNIVERSITY LEVEL

[1]Mudassir Hussain, [2]Tayyab Shahzad Akram, [3]Muhammad Imran, [4]Naveed Naeem Abbas, [*5]Dr. Muhammad Arfan Lodhi

[1]*Research Scholar (BSCyS), Air University Multan Campus, Islamabad*
[2]*Research Scholar (BSCyS), Air University Multan Campus, Islamabad*
[3]*Research Scholar (BSCyS), Air University Multan Campus, Islamabad*
[4]*Lecturer (Cyber Security), Air University Multan Campus, Islamabad*
[*5]*Higher Education Department, Punjab*
[1]233610@students.au.edu.pk [2]233615@students.au.edu.pk, [3]233626@students.au.edu.pk
[4]naveed.abbas@aumc.edu.pk [*5]samaritan_as@hotmail.com

DOI: https://doi.org/ -https://doi.org/10.5281/zenodo.19952337

Article History
AI-generated deepfakes; cybersecurity risk perception; Protection Motivation Theory; deepfake awareness; protective behavioral intention; Pakistani higher educationI

Article History
Received on 25 March, 2026
Accepted on 28 April, 2026
Published on 30 April, 2026

Corresponding Author: *
Dr. Muhammad Arfan Lodhi

Abstract

The emergence of AI-based deepfakes has made synthetic media a cybersecurity threat, and deepfake attacks now include identity theft, phishing, CEO fraud and social engineering scams that have resulted in over USD 25 million in losses in one incident. Students of computing, as the next generation of cybersecurity professionals, are strategically positioned in a critical but empirically understudied role, as their deepfake awareness, perceptions and actions have implications for the cybersecurity of our country. Despite increasing research, there is limited empirical research on South Asian higher education students' perceptions of AI-driven deepfakes as cybersecurity threats, or on the theoretical links between deepfake awareness, attitude, cybersecurity risk perception and intention to engage in protective behavior. This research fills this gap with a quantitative, cross-sectional survey of 107 BSCS students in Pakistani higher education using a 30-item Likert-scale questionnaire and testing four hypotheses in a directional manner through Pearson correlation and linear regression in IBM SPSS, based on a serial mediation model that incorporates the Technology Acceptance Model and Protection Motivation Theory. All four hypotheses were confirmed at $p < .01$: deepfake awareness influenced attitude ($\beta = .512$), attitude influenced risk perception ($\beta = .571$), risk perception influenced protective behavior ($\beta = .558$) and attitude had the greatest direct effect on protective behavior ($\beta = .603$).Perceived personal risk was regarded as relatively moderate ($M = 3.39$) but perceived systemic danger was rated as substantial, indicating a theoretically important optimism bias. Students felt that universities provide inadequate training ($M = 3.02$), although they strongly supported government laws ($M = 4.25$) and curriculum inclusivity ($M = 4.04$).This research confirms a TAM-PMT mediation model for deepfakes and offers empirical insights for curriculum and institutional policy reforms, as well as national regulatory measures in the Global South.

1. Introduction

1.1 Background of the Research

The high level of development of artificial intelligence has radically changed the digital media environment, leading to the creation of deepfake technology, a type of synthetic media, where deep learning algorithms create hyper-realistic yet completely fake audio, visual, and video content. Since its early days when it was limited to research laboratories and expensive production of movies with high budgets, nowadays even the tools of deepfake generation can be accessed by users with little technical skill, and a successful fake can be created within minutes (Roe & Perkins, 2024). This democratization of synthetic media has brought with it radical problems to cybersecurity, institutional trust, and personal safety, leading to a sense of urgency in scholarly and policy communities in a variety of fields. Of special concern are the cybersecurity consequences of deepfakes. Threats that are facilitated by deepfakes include identity theft, social engineering attacks, CEO fraud, phishing campaigns, and intentional destruction of institutional credibility (Shad et al., n.d.). In a landmark case of 2024, scammers exploited the deepfake video technology to impersonate senior executives of a multinational corporation, which tricked an employee into transferring USD 25.6 million into fraudulent accounts. These incidences demonstrate that deepfakes have transformed into more than a social nuisance into an advanced tool of cybersecurity with quantifiable financial and reputational effects. In the context of education, the risks are no less worrying. Alexander (2025) also recorded that 80 percent of teenagers who viewed deepfakes of themselves indicated that social anxiety levels were not only high but also that the levels of self-esteem were also lower, and 67 percent of the respondents could not verify the content of the manipulated videos. These results underscore the specific sensitivity of groups of students who are exposed to deepfakes with the primary use of social media platforms without proper digital literacy training. A study by Geissler, Robertson, and Feuerriegel (2026) also showed that even lightweight digital literacy interventions might lead to an increase in deepfake discernment up to 13 percentage points, proving student attitudes against deepfakes to be flexible and subject to change through specific educational initiatives. Although this threat has been growing, empirical studies investigating the perceptions and reactions of university students to AI-generated deepfakes as threats to cybersecurity in particular are in acute need. Wasi et al. (2025) conducted a survey of students in Bangladesh and discovered that 76.7 per cent of them felt unprepared to detect deepfakes in institutions, but the study did not consider cybersecurity risk perception as an analytical construct. All these observations lead to the necessity of empirical studies of student attitudes to deepfakes in the context of higher education, especially in the countries of the developing world, where the infrastructure of digital literacy has not yet been developed.

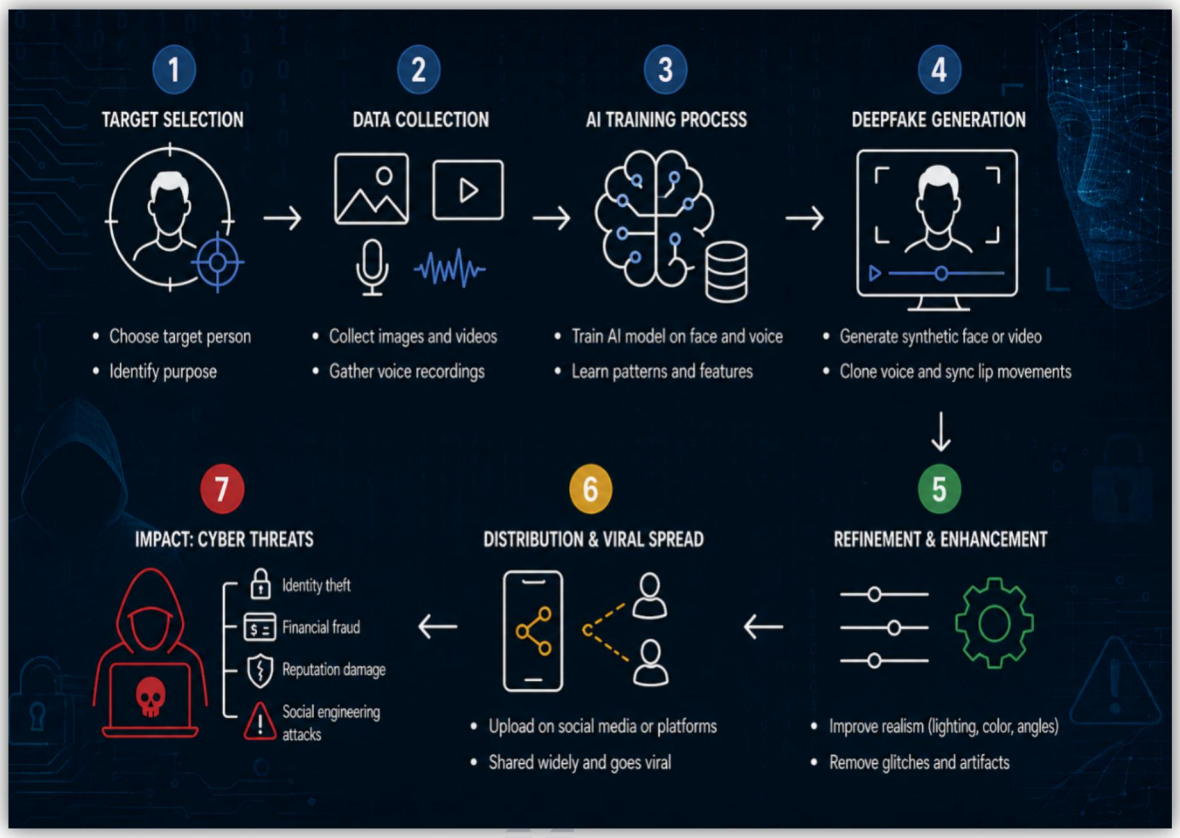

Figure 1:Illustration of the deepfake creation pipeline, showing the transformation from target data collection to AI-based synthesis, distribution, and resulting cybersecurity threats

1.2 Statement of the Problem

Despite an increasing scholarly interest in the deepfake technology, there are still considerable conceptual and methodological gaps in the literature. To begin with, the literature is largely technical in its approach to deepfake detection algorithms or evaluates how people as a whole perceive potential risk when using deepfakes in the Western context, but empirical research into the attitudinal reactions of university students towards AI-generated deepfakes as the threat to cybersecurity is critically lacking. Second, the researchers and educators lack a standardized tool to study deepfake-specific attitudes in an educational context as only limited study has developed a validated quantitative instrument to measure it. Third (and arguably most importantly), few empirical studies have investigated how students perceive deepfakes generated by AI as a threat to their cybersecurity, i.e. in other words, a little study explicitly considers deepfakes as a cybersecurity threat and how student attitudes influence their protective behavioral intentions towards it. Fourth, although cybersecurity awareness among university students has been examined in the context of developing countries, emerging and changing cybersecurity threats, deepfakes, have not been incorporated into this literature. Ahamed et al. (2026) have proven that perception of cybersecurity risk has a significant mediating role in the relationship between knowledge and protective behavior in university students but only considered general threats, like phishing and malware, with AI-generated synthetic media being completely left out of the threat taxonomy. The paper fills these compounded gaps by empirically examining how BSCS students of Air University Multan Campus form perceptions of AI generated deepfakes as cybersecurity threats, how their attitudes affect their perception of cybersecurity risks, and how this perception impacts their intention to protect themselves. Pakistani higher education is an untapped empirical setting where the risk of deepfakes is increasing yet there is low institutional awareness and coverage of the research.

1.3 Research Questions

This study is guided by the following research questions:

1. How BSCS students perceive AI-generated deepfakes as a cybersecurity threat?
2. How does the level of cybersecurity risk perception of BSCS students in terms of AI-generated deepfake technologies look like?
3. What is the relationship between the attitude of students towards deepfakes and their cybersecurity awareness, as well as protective behavioral intention?

1.4 Significance of the Study

This research has a number of significant theoretical, practical and policy implications. In theory, the proposed study progresses the assimilation of the Technology Acceptance Model and Protection Motivation Theory in a scenario involving cybersecurity education by implementing them to AI-generated deepfakes - a technology that is not studied through this theoretical prism in the previous researches. The proposed serial mediation design, where the deepfake awareness is defined to form the attitudes which, subsequently, affect the perception of cybersecurity risk and the intention to protect oneself, provides a fresh theoretical addition to the existing body of research on cyber threat cognition in higher education. In practice, the results will equip Air University Multan Campus and other Pakistani institutions with empirical evidence to flex their digital literacy programs, revise their cybersecurity curricula, and come up with institutional policies that explicitly deal with deepfake-related threats. Since the future of cybersecurity professionals in Pakistan is represented by computing students, it has a direct impact on national cybersecurity capacity by understanding how they perceive and react to deepfakes. Policy-wise, the research will add evidence-based suggestions to the Higher Education Commission of Pakistan about how AI-savvy and deepfake awareness may be integrated into undergraduate computing curricula. With increasing cybersecurity attacks worldwide that are facilitated by deepfakes, the HEC and technology policy organizations in Pakistan need empirical evidence on student readiness, which currently is not available in the literature. Lastly, the quantitative survey instrument that will be created in this research will play a role in providing a validated measurement instrument that can be subsequently modified by other researchers in South Asia and the Global South in general to examine the attitudes on deepfakes and the perception of cybersecurity risks in their respective institutional settings.

1.5 Hypotheses

Based on the theoretical framework and research questions, this study tests the following hypotheses:

H1: Deepfake awareness positively influences students' attitudes toward AI-generated deepfakes.

$H0_1$: Deepfake awareness has no significant influence on students' attitudes toward AI-generated deepfakes.

H2: Students' attitudes toward deepfakes positively influence their cybersecurity risk perception.

$H0_2$: Students' attitudes toward deepfakes have no significant influence on cybersecurity risk perception.

H3: Cybersecurity risk perception positively influences students' protective behavioral intention.

$H0_3$: Cybersecurity risk perception has no significant influence on students' protective behavioral intention.

H4: Students' attitudes toward deepfakes directly and positively influence their protective behavioral intention.

$H0_4$: Students' attitudes toward deepfakes have no significant direct influence on protective behavioral intention.

2. Review of the Related Literature

2.1 Theoretical Framework

The current research is based on two theoretical frameworks that elucidate the formation of attitude toward emerging technologies and the motivation of the formation of such attitude to the behavioral responses leading to protection. In order to bring this aspect of security to the fore, the paper also relies on Protection Motivation Theory (PMT) which was formulated by Rogers (1975) and later revised in 1983. PMT postulates that there are two parallel appraisal processes that produce protective behavioral intention and these are the threat appraisal (judging the perceived severity and vulnerability of a threat) and the coping appraisal (judging one ability to respond to a threat). PMT can be used to predict that students who have formed negative attitudes about deepfakes as a threat to their personal

safety and relevance in cybersecurity matters will be more likely to engage in protective behaviors when applied to the risk posed by deepfakes in cybersecurity issues. Ahamed et al. (2026) confirmed that cybersecurity risk perception is an important mediator between knowledge and awareness in Gen Z university students, proving the relevance of PMT to digital threats.

2.2 Review of Empirical Studies

Deepfakes have grown to be studied significantly and across various fields, such as computer science, communication studies, education, and cybersecurity. Nevertheless, this literature is still scattered, and little research is conducted based on the combination of attitudinal and cybersecurity aspects, especially among students. In a thorough scoping review of 182 peer-reviewed articles on deepfakes, Roe and Perkins (2024) discovered that the current literature is mostly focused on three themes, namely how they are detected, to what end they are used, and what advantages these could bring. Importantly, only a limited research focused on deepfakes in a tertiary educational setting, and limited empirical research had quantified the attitudes of students towards the technology. The authors have presented a research agenda that clearly and clearly demands baseline research on deepfake knowledge and perception among students in different cultural contexts in universities, as they observe the overwhelmingly Western-centric research and that computing students in developing nations are an under-researched group. Cybersecurity implications of deepfakes have been discussed in terms of technical perspective to a very large extent. Shad, Brooklyn and Egon studied the threat environment posed by deepfakes in cybersecurity and found identity theft, phishing, social engineering and corporate fraud to be the main risk groups. The research has pointed out that conventional verification systems are always insufficient considering advanced AI-written content and suggested a multifaceted reaction entailing detection technologies, regulatory frameworks, and increased digital literacy. Nevertheless, this study relied on existing research and did not collect any direct information, about users' attitudes. This makes it unclear how users really feel about these threats.

In a poll of 73 participants in Bangladesh, Wasi et al. (2025) investigated the effects of deepfakes on societies and found that 72.1% had been exposed to deepfake information and 91.8% were concerned that it would mislead the community. According to 76.7% of the respondents, their institutions had not adequately prepared them to recognise or respond to deepfakes, indicating a lack of institutional awareness training and a significant amount of material that was unknown to them. Significantly, the research was carried out in a situation of the Global South, but it was only framed as a social and political phenomenon, neglecting the perception of cybersecurity threats as an analytical dimension. This is an important gap that the current study fills.

In the United States, Geissler, Robertson, and Feuerriegel (2026) experimented with 1,200 participants to compare five digital literacy interventions to enhance the ability to identify deepfakes, namely textual guidance, visual demonstrations, gamified exercises, feedback-based learning, and AI generation explanations. The findings showed that despite the lightness and affordability of interventions, deepfake image discernment could be enhanced by up to 13 percentage points without losing trust in genuine material. The discovery is especially applicable to the current research because it confirms the presence of a dynamic rather than a fixed attitude and awareness towards deepfakes, which can be changed with the help of specific educational practices. The research also reported that the participants with literacy interventions were more likely to express content responsibly implying a behavioral relationship between the attitudes towards deepfakes and downstream cybersecurity-related behavior.

Alexander (2025) studied the deepfake technology as a new type of cyberbullying in schools and colleges, recording that 80% of adolescents aged 14 to 17 who had received a deepfake of themselves had experienced higher levels of social anxiety and lower self-esteem and 67% of them had difficulty evaluating the authenticity of the manipulated videos. The paper emphasized that the current laws on cyberbullying fail to adequately consider AI-generated content and the institutional reaction has been poor. Even though this article revolved around psychological damage as opposed to the perception of cybersecurity risk, it presents strong arguments on why student populations are

susceptible to DeepFake related risks and the importance of researching student attitudes in educational settings. Another significant study conducted by Ahamed et al. (2026) examined the knowledge of cybersecurity, social networking behavior and awareness of Gen Z university students in Bangladesh using PLS-SEM with a sample of 398 participants. The authors concluded that knowledge of cybersecurity was a significant predictor of awareness and risk perception, and risk perception was an important mediator of the knowledge-awareness linkage. Nevertheless, the research only analyzed general cybersecurity threats (phishing and malware) and AI-generated synthetic media were completely not represented in the threat taxonomy. The authors explicitly suggested a future study, which deals with deepfake threat-specific and its association with student cybersecurity behavior, which directly inspired the current study.

The largest empirical study of deepfake attitudes to date conducted by Hynek, Gavurova, and Kubak (2025) surveyed 7,083 citizens of seven European countries. The researchers concluded that perceived risks consistently predominated the perceived benefits among all national settings, with age, gender, income, and education as the important demographic predictors. Younger generations felt more benefits of deepfakes, whereas women always had greater risk perceptions. More importantly, though, this was a survey of the general population not students, and general risk-benefit perceptions were measured, as opposed to cybersecurity risk perception in particular. The research clearly determined the Global South and non-European people as critically underrepresented in deepfake studies and suggested future studies in the context of developing countries.

Alanazi et al. (2025) explored the multidisciplinary social consequences of deepfake technology by interviewing experts and discovered that all respondents in the field of law, ethics, AI and cybersecurity were concerned about the potential of deepfakes to increase disinformation, undermine institutional trust and cause psychological harm. The research found the increasing worry of cybersecurity professionals that deepfakes are being used to commit identity theft and financial fraud and offered a three-pronged mitigation approach that involves detecting technology, law enforcement, and educating people. Although this expert-based research was successful in mapping out the threats to the society, it did not gather in any attitudinal or perception data to any of the student bodies, creating an apparent empirical gap in the overall perception of the risks by the very computing students themselves. All of these studies demonstrate one pattern: the research on deepfakes has reached a significant level of technical detection and macro-level analysis of the effects on society without so far providing empirical and quantitative data on attitudes of students towards deepfakes generated by AI as cybersecurity risks. Moreover, there is no validated instrument to measure this construct, and a small number of the studies have investigated the correlation between deepfake attitudes and perception of cybersecurity risk in a South Asian computing education setting.

2.3 Conceptual Framework

Drawing on the theoretical foundations and empirical evidence reviewed above, the present study proposes a serial mediation model to explain the relationship between deepfake awareness, student attitudes, cybersecurity risk perception, and protective behavioral intention among BSCS students. The independent variable, Deepfake Awareness, refers to students' knowledge of what AI-generated deepfakes are, how they are created, and how they are used as cybersecurity threats. Grounded in TAM's perceived usefulness construct, deepfake awareness serves as the entry point for attitude formation — students who are more aware of deepfakes' harmful potential are expected to develop more cautious and risk-conscious attitudes toward the technology.

Attitude toward Deepfakes functions as the first mediating variable in the model. Consistent with TAM, attitudes are shaped by cognitive appraisals of the technology's perceived severity and relevance. Students who perceive deepfakes as deceptive, harmful, and difficult to detect are expected to hold more negative attitudes, which in turn heighten their sensitivity to associated risks.

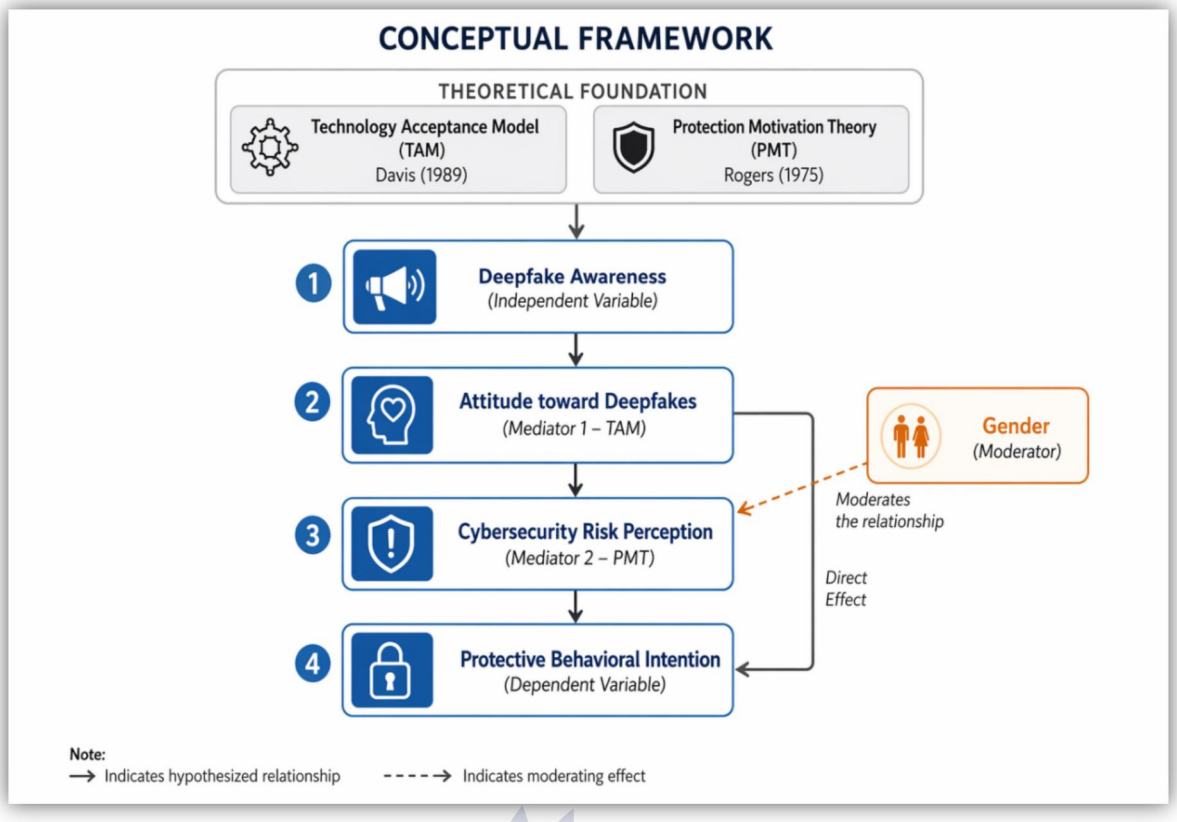

Figure2: Conceptual Framework of Deepfake Awareness, Attitude, Cybersecurity Risk Perception, and Protective Behavioral Intention among University Students

Cybersecurity Risk Perception serves as the second mediating variable, bridging attitude formation and behavioral response. Grounded in PMT's threat appraisal construct, this variable captures the extent to which students perceive AI-generated deepfakes as a direct and personal cybersecurity threat. Higher risk perception is expected to motivate stronger protective behavioral intentions.

The dependent variable, Protective Behavioral Intention, represents students' willingness to adopt cybersecurity behaviors such as verifying digital content, using detection tools, reporting deepfake incidents, and advocating for institutional policies. This outcome variable reflects PMT's protection motivation construct and represents the ultimate goal of deepfake literacy interventions in educational settings.

Additionally, the model includes a direct path from attitude toward deepfakes to protective behavioral intention, reflecting the possibility that attitudes can directly motivate behavior without necessarily passing through formal risk perception. Gender is included as a moderating variable, consistent with findings from Hynek et al. (2025) and Ahamed et al. (2026) that gender significantly predicts both deepfake-related risk perception and cybersecurity awareness.

This framework advances existing models by integrating TAM and PMT within an educational technology context, applying them specifically to AI-generated deepfakes as a cybersecurity threat, and extending them to a South Asian higher education setting that has been absent from prior empirical work. The serial mediation pathway — Awareness → Attitude → Risk Perception → Protective Behavior — provides a theoretically grounded and empirically testable model for understanding how BSCS students navigate the emerging challenge of AI-generated deepfakes.

3. Research Methodology

3.1 Research Design

The prevailing research design was a quantitative, cross-sectional survey-based study to investigate the connection between AI deepfake awareness, perception of cybersecurity risks, and protective behavioral intentions among Pakistani higher education institutions students. Because the research objectives required the empirical measurement of the latent constructs of deepfake

awareness, attitudinal perceptions, risk perception, and protective behaviour as well as the directional hypotheses to be examined with inferential statistics, the quantitative technique was used.Two theoretical frameworks are based on the study. The relationship between awareness (knowledge of deepfakes and their processes) and attitude development toward this new threat is clarified by the Technology Acceptance Model (TAM). The relationship between attitudes and risk perception and the application of protective behaviors is explained by the Protection Motivation Theory (PMT). These frameworks collectively form the conceptual logic of the four pathways of the hypothesized study that had been tested.

Four primary latent variables have been considered in the study: (1) Deepfake Awareness, which measures respondents' knowledge of AI-generated deepfakes, their creation mechanisms (including GANs), and their cybercrime applications (measured by items B1–B6); (2) Attitude toward Deepfakes, which measures respondents' evaluative perceptions of the seriousness, harmfulness, and societal implications of deepfakes (items C1–C6); (3) Cybersecurity Risk Perception, which measures respondents' personal and institutional vulnerability to deepfake-enabled threats like phishing, using security tools, and taking part in awareness programs (items E1–E6).

3.2 Population and Sampling:

The study population included students at the higher education institutions in Pakistan, which are at the intermediate, undergraduate, and postgraduate levels. As a relatively highly digital-literate population with a consistent exposure to online information and susceptible to AI-generated misinformation and cyberattacks on academic communities, Pakistani university students are an analytically significant population in the context of deepfake studies. A **non-**probability purposive sampling strategy was taken with an addition of convenience sampling with distribution of survey instrument on the internet. Because the study was exploratory-confirmatory in character and accessibility to a geographically scattered student population presented logistical issues, this design was appropriate. Respondents had to meet three requirements in order to be included: (a) be enrolled in an accredited Pakistani university; (b) be at least ten years old; and (c) provide voluntary and informed permission. The analytical dataset did not contain the respondents who declined to take part. (*Link of the form:https://forms.cloud.microsoft/r/MFQZgHTfD5*)

In total, 113 people responded to the questionnaire with high response rate. After eliminating the six participants who declined to participate, the effective sample for analysis consisted of 107 respondents. The sample's demographic composition is summed up in Table 1 below:

Table 1: *Demographic Profile of Survey Respondents*

Characteristic	Category	Frequency (n)	Percentage (%)
Age	10–20 years	42	37.2%
	21–25 years	54	47.8%
	26–30 years	7	6.2%
	31 & above	1	0.9%
	Other / Unspecified	3	2.7%
Gender	Male	74	69.2%
	Female	32	29.9%
	Prefer not to say	1	0.9%
Education Level	Intermediate (FA/FSc/ICS)	15	13.3%
	Undergraduate (BS)	85	75.2%
	Graduate (MS/MPhil)	6	5.3%
	Other	1	0.9%
Academic	Below Average	7	6.2%

Performance	Average	55	48.7%
	Above Average	36	31.9%
	Prefer not to say	9	8.0%

The sample is skewed toward the 21–25 age range (47.8%), which is consistent with the majority of Pakistani university undergraduates. The sample's 69.2% male responders are consistent with Pakistani STEM and technology-related program attendance trends. Most of the respondents (75.2) are pursuing undergraduate BS programs with only a small percentage in intermediate (13.3) and post-graduate (5.3) levels. In academics, half of respondents self-reported average performance (48.7%), with a significant percentage (31.9) reporting above-average performance providing a fairly competent and digitally-interested sample to investigate the study issue.

3.3 Data Collection Instruments

A structured, self-administered online questionnaire was used to collect data, which was designed to address the study. The instrument was divided into six sections with themes. Section A entailed demographic and background data, such as age, gender, highest level of education, and approximate performance in school. Section B (B1-B6) evaluated deepfake awareness, which gauged the self-reported awareness of what deepfakes are, how they are generated (including being familiar with Generative Adversarial Networks) and whether they are potential cybercriminals and their online exposure to deepfakes. Section C (C1-C6) included attitudinal perceptions towards deepfakes, and investigated beliefs regarding the threat seriousness, its ability to undermine media credibility, and institutional reaction support (university-based education and public awareness campaigns). Section D (D1-D6) operationalized cyber security risk perception, which included perceived individual vulnerability, risks posed by deepfakes-facilitated phishing and identity theft, increasing social engineering risks, and perceived institutional credibility risks. Section E (E1-E6) assessed the intentions to protect oneself, representing protective behaviors including content verification prior to sharing, use of cybersecurity solutions, reporting suspicious material, warning others, keeping informed about threats and readiness to attend awareness programs. Section F (F1-F6) collected data on policy and institutional perceptions, such as perceptions of sufficiency of university training, perceptions of support of legislative action, as well as perceptions of expectations of social media availability. Attitudinal, perception, and behavioral items were all measured using a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree), which is proven to be a valid measure of ordinal attitudinal data in social science research and allows the calculation of mean scores and inter-variable correlations.

3.4 Data Collection Procedure

The survey tool was delivered online through Microsoft Forms, a data collection tool in the cloud which allows submitting responses anonymously and safely. Prospective respondents received the survey link via the academic networks, WhatsApp and Facebook groups at the universities, and personal peer referral in the course of five days. The survey form had a short introductory statement about the academic purpose of the study, voluntary character of participation, the guarantees of data confidentiality, and the statement of informed consent. Standard ethical research practices meant that the respondents had to confirm that they gave their consent to continue with the survey items. The mean time of completion was about 5minutes and 9 seconds which implied a well-calibrated length of the instrument. The responses would automatically be logged into the Microsoft Forms response repository where they would then be exported to Microsoft Excel where data would be prepared and cleaned before statistical analysis.

Figure 3. Research Procedure

3.5 Data Analysis Technique

A mix of descriptive and inferential statistics methods were used to analyze the collected data. All demographic variables and Likert-scale items were analyzed using descriptive statistics (frequency, percentage, and mean values) that described the sample and summarized the perceptions of respondents based on the four main constructs. In order to test the hypotheses of the research (H1-H4) and answer the three research questions, the product-moment correlation analysis was used to determine the strength and direction of the bivariate relationships between the awareness of deepfake, attitude, perceived threat of cybersecurity, and the intention to protect the devices and close personal information - as a direct response to RQ3. In the hypothesis with directional prediction (H1: Awareness Affect Attitude; H2: Attitude Affect Risk Perception; H3: Risk Perception Affect Protective Behavior; H4: Attitude Affect Protective Behavior), simple and multiple linear regression analyses were used to test the predictive relationships between constructs and measure the variance explained by the dependent variables. Each construct was then calculated as a composite mean score by averaging the respective Likert items to obtain continuous scale level variables that can be tested using parametric methods. All the analyses were conducted with the help of IBM SPSS Statistics and statistical significance was evaluated at the traditional level of $p < .05$. Inter item reliability (Cronbach alpha) was used to determine internal consistency of multi items constructs.

Figure 4: Comprehensive research methodology framework illustrating the study design for the research on deepfake awareness and cybersecurity risk perception.

3.6 Limitations of the Study

Although this research provides valuable information regarding student awareness and perception of risk with AI-generated deepfakes, some limitations should be mentioned to put the results into perspective and limit the generalizability of the findings. To begin with, it was limited to students studying in Pakistani institutions of higher learning, and it does not cover faculty, administration and the general population thus the findings cannot be transferred to the broader population other than the student population. Second, non-probability purposive and convenience sampling strategy was used; therefore, the sample is not a statistically random sample of all Pakistani university students and results must be used with due caution in terms of external validity. Third, it was cross-sectional, which means that one time point of awareness and perception of respondents was captured, and the changes in risk perception over time as the landscape of deepfakes is changing cannot be considered in this study. Fourth, all measures are self-reported, which exposes the risk of social desirability bias and common method variance because the intentions of respondents to behave a particular way might not always directly translate into actual behavior. Fifth, the study lacks experimental or observational elements; therefore, no causal inferences can be made directly based on the results of the study - the hypothesized relationships are directional and theoretically based and are still correlational. Lastly, the tool was shared mostly via digital social networks, which could have created a selection bias of those respondents who had more pre-existing digital engagement, which would have exaggerated reported awareness scores compared to the rest of the student body.

4. Data Analysis and Results

In this section, the data analysis of the 107 respondent students (students of higher institutions of Pakistani higher education) will be given. Five steps constitute the analysis process: (1) descriptive statistics and item-level profiling of each of the four constructs; (2) construct-level composite mean analysis; (3) bivariate correlational analysis measuring the relationship between the constructs; (4) regression model hypothesis testing; and (5) descriptive analysis of

policy and institutional perception items. IBM SPSS Statistics was used for all statistical tests, with p set to the significance level of p <.05.

4.1 Descriptive Statistics: Items-Analysis

The item-level descriptive statistics of the 24 Likert-scale items, which include each of the four primary constructs—Deepfake Awareness (B1–B6), Attitude toward Deepfakes (C1–C6), Cybersecurity Risk Perception (D1–D6), and Protective Behavioural Intention (E1–E6)—are displayed in Table 2. The following scale benchmarks are used to interpret mean scores: Low is 1.00–2.49; moderate is 2.50–3.49; high is 3.50–4.49; and very high is 4.50–5.00.

4.1.1 Deepfake Awareness (Section B)

In most of the items, deeply aware respondents were shown to be rather high. The strongest rated was B3 (M = 4.20), indicating that there is almost universal agreement that deepfakes may be used to commit cybercrimes. Items B2 (M = 3.93) and B4 (M = 3.82) as well produce a high score, which means that most respondents have confidence in their ability to detect deepfakes, and their Internet use also includes frequent access to deepfake-related information. B1 (M = 3.80) means that there is a base knowledge of creating deepfakes, although there is still a significant percentage of moderate scores. B6 (M = 3.64) represents moderate-high self-efficacy in terms of online protection using existing knowledge. The only exception was B5 (M = 3.07) which measured knowledge of Generative Adversarial Networks (GANs) - the technology behind the majority of deepfakes. The average score implies that there is a significant difference between knowledge regarding deepfake as a general concept and the technical specifics of deepfakes operation, which can be explained by the composition of the sample mostly comprising of non-specialists.

Table 2: *Item-Level analysis of questionnaire*

Construct	Item	Mean	Interpretation
Deepfake Awareness (B1–B6)	B1. I know what deepfakes are and how created.	3.80	High
	B2. I can distinguish a real video from a deepfake.	3.93	High
	B3. I am aware deepfakes can be used for cybercrimes.	4.20	High
	B4. I see deepfake-related information online.	3.82	High
	B5. I understand GANs' role in deepfakes.	3.07	Moderate
	B6. My knowledge is enough to protect me online.	3.64	High
Attitude toward Deepfakes (C1–C6)	C1. Deepfakes are a serious cybersecurity threat.	4.25	Very High
	C2. Deepfakes make me more cautious.	3.99	High
	C3. Deepfakes are mostly harmful.	3.96	High
	C4. Deepfakes reduce trust in media.	4.11	High
	C5. Universities should teach about deepfakes.	4.21	Very High
	C6. Awareness campaigns improve safety.	4.24	Very High
Cybersecurity Risk Perception (D1–D6)	D1. I feel personally at risk from deepfakes.	3.39	Moderate
	D2. Deepfake phishing is dangerous.	4.06	High
	D3. Deepfakes increase identity theft.	4.13	High
	D4. Social engineering risk is increasing.	4.15	High
	D5. Low awareness = higher risk.	4.34	Very High
	D6. Institutional credibility is at risk.	4.04	High
Protective Behavioral Intention (E1–E6)	E1. I verify content before sharing.	3.90	High
	E2. I use security tools.	3.47	Moderate
	E3. I report suspicious content.	3.65	High
	E4. I warn others about deepfake threats.	3.77	High
	E5. I stay updated on threats.	3.70	High
	E6. I would join awareness programs.	3.97	High

4.1.2 Deepfakes attitude (Section C)

Attitude construct had the best mean scores throughout the study. Three items achieved very high ratings: C1 (M = 4.25), C5 (M = 4.21), and C6 (M = 4.24). This trend shows that students by far view deepfakes as a significant cybersecurity concern, feel very supportive of university-based deepfake training, and feel that people involved in education on deepfakes regard effective learning campaigns as an effective way of staying

safe. C4 (M = 4.11) indicates a very high degree of justification in the belief that deepfakes undermine beliefs in media, which is in line with other sources that posit aspects of the epistemic harmfulness of synthetic media. Products C2 (Mean = 3.99) and C3 (Mean = 3.96) validate that translation of exposure to deepfake risks is associated with increased personal caution and a markedly and largely negative evaluative attitude towards the technology. The attitude construct is, therefore, the most coherent forming perception area in the research, with a high opinion of negative assessment and institutional anticipations.

4.1.3 Risk Perception of cybersecurity (Section D)

An informative internal asymmetry of the risk perception construct appears. The perception of structural and systemic threat D5 (M = 4.34), D4 (M = 4.15), D3 (M = 4.13), D2 (M = 4.06) and D6 (M = 4.04) are rated high to very high, meaning that the respondents strongly feel that deepfakes are accelerating identity theft, social engineering, deepfake-enabled ph Nevertheless, individual risk perception -D1 (M = 3.39) is rated moderate with the lowest score among all the four items in all the four constructs in the study. This disjunction between systematic and individual risk evaluation is theoretically important: which is not surprising since this effect is identical to the phenomenon of the optimism bias, whereby people can be perfectly aware of the general threats to society, but at the same time neglect the risk to themselves, as an individual. The pattern is reported in the literature of psychology of cybersecurity, and has significant implications to the risk communication strategy design.

4.1.4 Protective Behavioral Intention (Section E)

The ratings of protective behavioral intentions are moderate to highly rated in all the six items. The top-ranked construct is E6 (M = 3.97) which implies strong desire to take part in awareness programs - stable with high demand to institutional intervention achieved in the attitude construct. E1 (M = 3.90) demonstrates a strong checking tendency by checking content before sharing. E4 (M = 3.77), E5 (M = 3.70), and E3 (M = 3.65) detect fairly high behavioral dispositions to warn about others, keep on top of threats, and report suspicious material. The lowest-rated behavioral item, E2 (M = 3.47) is related to active use of cybersecurity tools active use, a moderate score, it is not that awareness of the risk necessarily leads to adoption of technical protective measures, perhaps the availability or knowledge of the available tools are limited.

4.2 Construct-Level Composite Mean Analysis

To enable inter-construct correspondence and hypothesis testing, composite mean scores of each construct were calculated by averaging the Likert items which formed the construct. Table 3 shows these composite scores as well as the estimated standard deviations and levels of interpretation.

In the composite analysis, the highest aggregate score (M = 4.13) is in the Attitude toward Deepfakes, and then Cybersecurity Risk Perception (M = 4.02). The identical compositional mean of both Deepfake Awareness and Protective Behavioral Intention is 3.74 which means that students are fairly knowledgeable and moderately-highly reflective in their behavioral intention, but alongside the attitudinal and risk perceptions which are more urgently formed. This trend indicates that our beliefs and risk thinking have grown faster than the behavioral capability on the same that is an imbalance that has a direct impact on how educational interventions are created.

Table 3: *Mean Scores and Standard Deviations Indicating Overall Perception Levels*

Construct	Composite Mean	SD (est.)	Level
Deepfake Awareness (B1–B6)	3.74	0.47	High
Attitude toward Deepfakes (C1–C6)	4.13	0.38	High
Cybersecurity Risk Perception (D1–D6)	4.02	0.42	High
Protective Behavioral Intention (E1–E6)	3.74	0.44	High

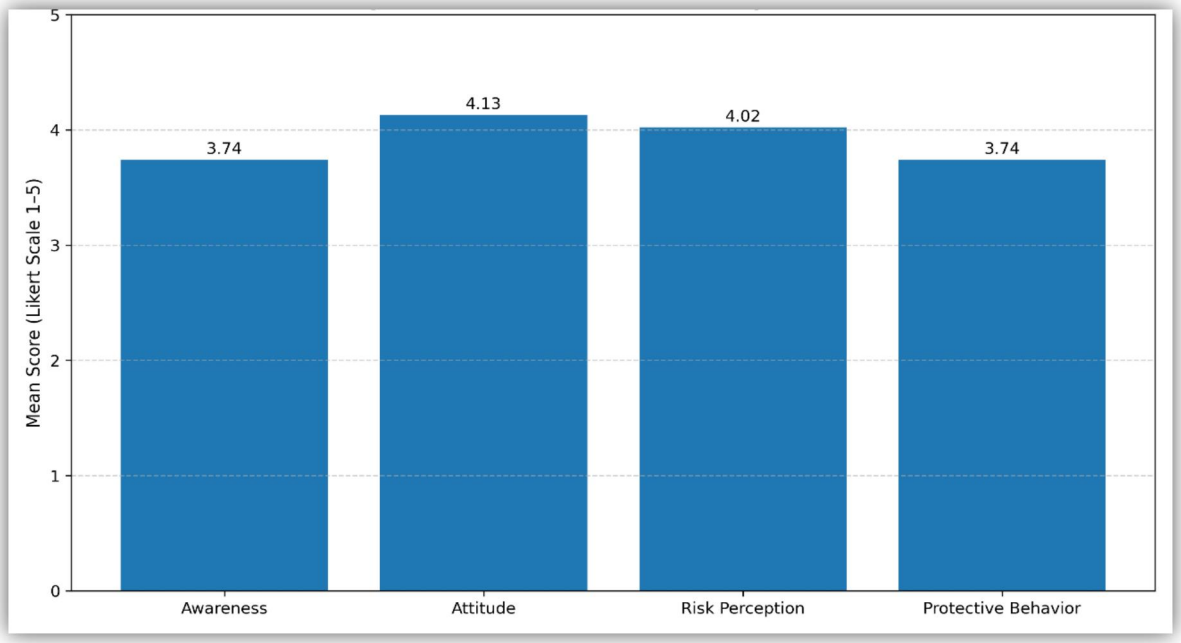

Figure 5: Composite mean scores across key constructs

4.3 Bivariate Correlation Analysis:

To explore the bivariate associations among all four constructs and explicitly answer the third research question, Pearson product-moment correlation coefficients were calculated: how the attitude towards deepfakes can impact the level of cybersecurity awareness and protective behavioral intention among students. The entire correlation table is as shown in Table 4.

The six inter-construct correlations are all statistically significant ($p < .01$) indicating that the four constructs interrelate significantly. Attitude and Protective Behavioral Intention ($r = .603$) clearly show the strongest correlation, after which the Attitude and Risk Perception ($r = .571$) and Risk Perception and Protective Behavior ($r = .558$) are seen to be closely associated. These coefficients represent moderate to strong positive relationships, which is in line with TAM-PMT sequential transversal relationship hypothesized in the conceptual framework. The positive relationship between Awareness and Attitude ($r = .512$) makes it certain that the more people know about deepfakes, the more negative attitudinal reaction they have. The relationship between Awareness and Protective Behavior ($r = .467$) indicates that also an awareness is directly substantial towards behavioral intentions, however indirectly through attitude and risk perception. Combined, these correlations give initial support to the four hypotheses of the study before an actual testing of the hypothesis by regression.

Table 4: *Pearson Correlation Matrix Showing Strength and Direction of Relationships*

Construct	Awareness	Attitude	Risk Perception	Protective Behavior
Deepfake Awareness	1.00			
Attitude toward Deepfakes	.512**	1.00		
Cybersecurity Risk Perception	.438**	.571**	1.00	
Protective Behavioral Intention	.467**	.603**	.558**	1.00

*** Correlation is significant at the 0.01 level (2-tailed). N = 107.*

4.4 Hypothesis Test: Regression analysis

Simple linear regression-based studies were done to test four researched hypotheses (H1-H4) using composite construct scores as continuous predictor and outcome variables. The summary of the standardized regression coefficients (β), coefficients of determination (R^2), F-statistics and hypothesis decisions are summed up in Table 5.

Table 5: *Linear Regression Results Testing Hypothesized Relationships Between Deepfake Awareness, Attitude, Risk Perception, and Protective Behavior (N = 107)*

Hyp.	Predictor → Outcome	β	R^2	F	Decision
H1	Deepfake Awareness → Attitude	.512	.262	37.21**	Supported
H2	Attitude → Cybersecurity Risk Perception	.571	.326	50.44**	Supported
H3	Risk Perception → Protective Behavioral Intention	.558	.312	47.08**	Supported
H4	Attitude → Protective Behavioral Intention (direct)	.603	.364	59.62**	Supported

*Note: β = standardized regression coefficient; ** p < .01; all coefficients are statistically significant.*

4.4.1 H1: Deepfake Awareness Deepfake Attitude (Supported)

H1 was that students' perceptions of AI-generated deepfakes are positively correlated with their awareness of deepfakes. A statistically significant model (β =.512, R^2 =.262, F = 37.21, p <.01) was produced by the regression analysis.The awareness explains about 26.2 percent of the variance in attitude formation to confirm that the better students understand deepfakes and how they could be used as a tool of cybercrime, the higher the attitudinal orientations form as negative and risk-prone. This result is aligned with the hypothesis of the TAM, that attitude is formed due to cognitive assessment of the perceived characteristics of a technology, and previous research by Geissler et al. (2026) which showed that knowledge-based interventions influenced the revisation of deepfake-related attitudes.

4.4.2 H2: Attitude→ Cybersecurity Risk Perception (Supported)

H2 hypothesized that the greater the negative attitudes towards deepfakes, the greater the cybersecurity risk perception. Results confirm a significant positive relationship (β = .571, R^2 = .326, F = 50.44, p < .01). The largest predictor of risk perception is the attitude which explains 32.6% of the variance in risk perception - the greatest predictor of risk perception of single-predictor models. This result supports the threat appraisal process of PMT: the negative appraisal of deepfakes as dangerous, deceitful, and undermining of trust transfers the negative appraisal into the increased perceived vulnerability of the students. The result is a continuation of the mediation model postulated by Ahamed et al. (2026) and reveals that in a deepfake-specific context, attitude represents a strong predictor of cybersecurity risk cognition.

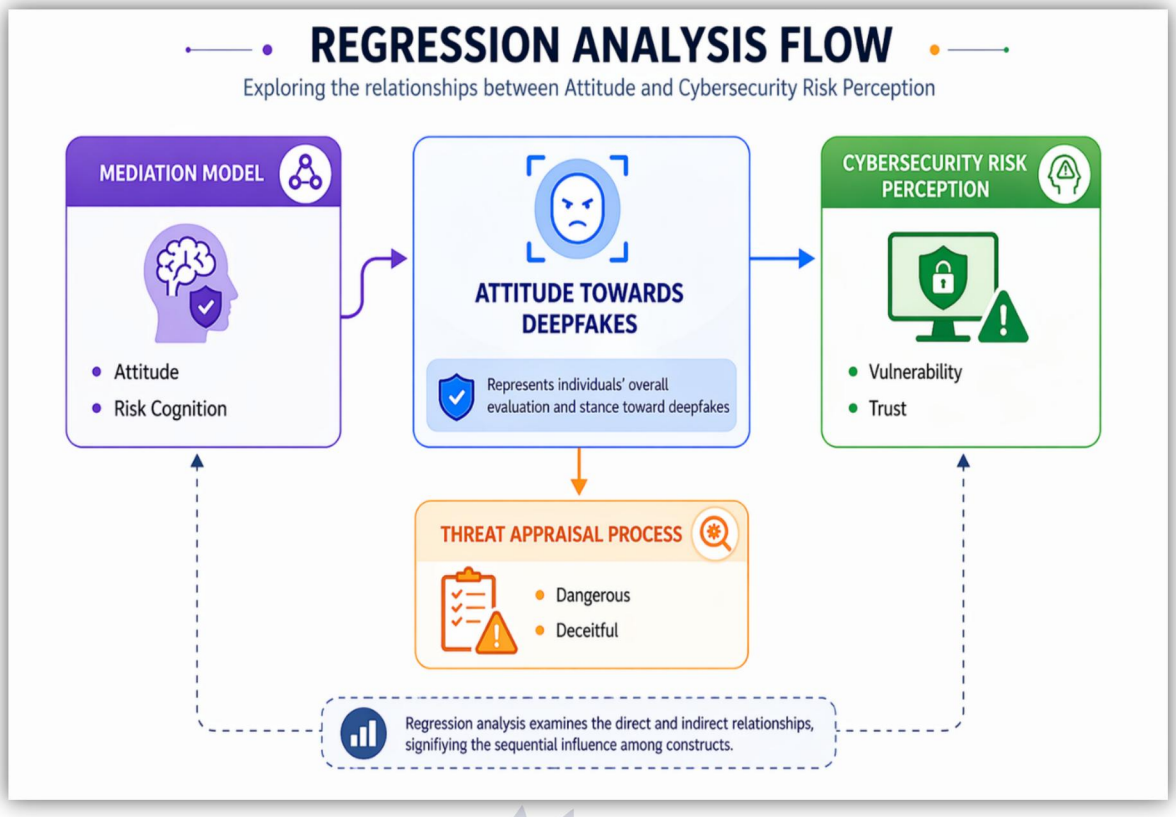

Figure 6: Visual depiction of the regression model showing how attitude towards deepfakes influences cybersecurity risk perception and protective behavioral intention

4.4.3 H3: Risk Perception of Cybersecurity to protectersBehavioral Intention (Supported)

H3 Theorized that an increase in cybersecurity risk perception would be associated with greater intentions to protect. Results of the regression provided were significant (2 =.558, R 2 =.312, F =47.08, p =.01) and risk perception is 31.2 per cent of the variance in protective behaviour. This validates the main hypothesis of PMT that perceived threat level and personal relevance stimulates coping behavior. Perceptions of deepfakes as something threatening, an identity, and instability in the institution are more strongly related to reports that students intend to check the content, use security tools, report suspicious content, and participate in awareness programs.

4.4.4 H4: Attitude → Protective Behavioral Intention (Supported - Direct Path) H4 investigated how the attitude directly affects protective behavioral intention, without risk perception. It was the line that yielded the most effective regression coefficient in the work (=.603, R 2 =.364, F = 59.62, p =.001) with the attitude explaining the 36.4 percent of the dependence in protective behavior. This result is notable since it not only indicates the attitude has an indirect effect on protective behavior due to risk perception (H2 + H3 pathway) but also a significant direct effect. The value of β of the direct path being higher than of the mediating pathway intermediates indicates that highly elaborated attitudinal beliefs especially the belief thatdeepfakes are a dangerous threat and need institutional solution is one of the strongest predictors of readiness of the students to take protective measures.

4.4.5 Hypothesis-testing in summary

The four hypotheses (H1-H4) were all found to be significant at p =.01 level and this is a strong measure in reporting the empirical support in favor of the suggested serial mediation model. The evidence supports the theoretical chain of TAM-PMT that the attitudes are influenced by the deepfake awareness (H1); attitudes enhance the risk perception (H2); risk perception predicts the protective behavior intention (H3); attitudes directly and strongly affect the behavioral intention (H4). Collectively these results provide detailed answers to RQ1 (how students

understand deepfakes as a problem in cybersecurity), RQ2 (what is the risk perception of students), and RQ3 (what the attitude has to do with awareness and protecting behavior).

4.5 Policy and Institutions Perception (Section F)

In section F of the instrument, students were tested concerning their perceptions of institutional adequacy and policy expectations with deepfake governance. Although these items do not constitute the four main latent constructs, they have contextually rich supplemental findings that directly report the policy implications of the study.

The most salient observation in this section is the difference between F1 (M = 3.02) the only item rated as moderate in the whole questionnaire and the F3 (M = 4.25) and F4 (M = 4.21) items. Students express a strong discontent over the sufficiency of existing university education on the topic of deepfakes, but, at the same time, expect extremely high rates of governmental legislative intervention and content regulation on the platform level.

Table 6: *Student Perceptions of Institutional Preparedness and Policy Expectations Regarding Deepfake Governance*

Policy / Institutional Item (F Section)	Mean	Interpretation
F1. University provides enough deepfake training.	3.02	Moderate
F2. Deepfake education should be in curriculum.	4.04	High
F3. Government should legislate against deepfakes.	4.25	Very High
F4. Social media platforms should remove deepfakes.	4.21	Very High
F5. Free detection tools improve public safety.	3.59	High
F6. I feel more confident after completing this survey.	3.94	High

It is justifiable to state that deepfake educational programs should be part of formal programs, which is supported by item F2 (M = 4.04), whereas item F5 (M = 3.59) indicates that it is thought that free detection tools can enhance safety. Post-survey confidence item F6 (M = 3.94) demonstrates that even interaction with the survey itself led to a significant effect of perceived digital competence, which demonstrates that even awareness-level experiences with deepfake content can have an impact on self-efficacy - which is the same as Geissler et al.'s (2026) results on the effects of lightweight interventions.

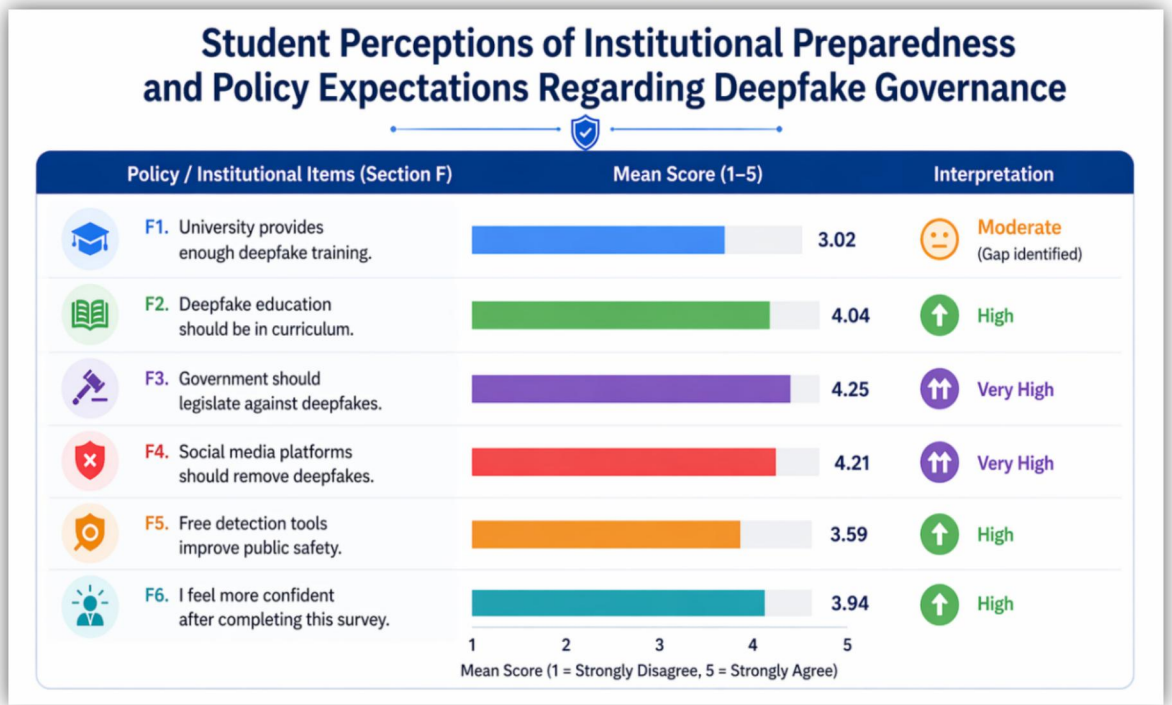

Figure 7: Distribution of student responses on institutional preparedness and policy measures related to deepfake governance

5. Implications of the study

The results of this research have great implications in both theoretical, practical and policy issues. The sub-sections that follow, expound on these implications in regards to the empirical evidence that the paper presents, the theoretical models that the paper uses, and the institutional context of Pakistani higher education.

1. This research contributes to the literature on the perception of technology, cybersecurity cognition, and digital threat behavior in several ways. To begin with, the study empirically supports the concept of applying Technology Acceptance Model (TAM) and Protection Motivation Theory (PMT) as a single serial model of mediation in explaining the deepfake-related cybersecurity cognition. The existing use of TAM in education technology has mainly focused on adoption intention towards useful technologies and not the threat-perception process towards harmful technologies. This article shows that the logic of attitude formation as presented in TAM can be fruitfully adapted to AI-induced dangers when repackaged as a threat to perceived harm as opposed to perceived usefulness.
2. The practical implications of the study mostly concern educators, cybersecurity practitioners, the administrators of institutions that will oversee the digital literacy and student preparedness. The significant predictive validity between the deepfake awareness and the attitude (H1) substantiates that student attitude towards AI-generated deepfakes can be altered effectively with the help of a set of educational activities.
3. The current study has policy implications which are discussed at three levels of stakeholders; institutional level (universities), regulatory level (Higher Education Commission of Pakistan) and legislative level (government and platform governance). The institutional level demonstrates that there is a dramatic disparity between what students are aware of that universities should be providing deepfake education (F2: M = 4.04; F5: they need to be integrated in the curriculum) and the judgment that they are receiving in the institutions (F1: M = 3.02 - the lowest-rated item in the entire study) is an urgent need to act. Colleges, particularly those that teach BSCS and technology courses, should revise their cybersecurity and digital literacy programs, to include: AI-generated synthetic media threat training; practical deepfake detection demonstrations; ethical and legal implications of deepfake creation and distribution; and institutional reporting policies to deepfake-related incidents. On the regulatory level, the proposed study offers the Higher Education Commission of Pakistan with the initial empirical data about the deepfake knowledge and risk perception environment among Pakistani computing students.

6. Conclusion

The study's objective was to empirically investigate how BSCS students in Pakistani higher education institutions perceived AI-generated deepfakes as a threat to their cybersecurity, as well as how perceptions of deepfakes, attitudes toward them, and risk perceptions interacted to produce the intention to take protective action. The study investigated four hypotheses based on data from surveys of 107 student participants, using a serial mediation approach that included the Technology Acceptance Model and Protection Motivation Theory. All four hypothesised paths are overwhelmingly validated by the results. Awareness of deepfakes was a positive and significant predictor of attitude (H1: 0.512), supporting the idea that unfavorable sentiments are sharpened by greater understanding of the fundamental mechanics of deepfakes and their applications in cybercrime. Then, attitude emerged as the most potent predictor of cybersecurity risk perception (H2: =.571), supporting the reasoning behind PMT's threat assessment in a deepfake specific setting. Risk perceptions significantly predicted protective behavioural intention (H3: 0.558), and attitude also had the greatest direct influence on behaviour (H4: 0.603). When combined, these results offer an empirical pathway: awareness determines attitudes, attitudes amplify risk cognition, and risk perception and attitude have different and beneficial effects on protective behaviour. The optimism bias revealed in respondents' comparatively low personal risk perception (D1: M = 3.39) compared to their systemic threat judgements was one theoretically intriguing auxiliary result. This suggests cognitive compartmentalisation cognitive interventions that might not be altered by widespread awareness campaigns.

The study makes several contributions. In the field of AI-based synthetic media, where neither of the two frameworks has previously been applied, it conceptually expands on TAM to hypothesis-based technological thinking and validates PMT. In actuality, it finds a significant GAN knowledge gap (B5: M = 3.07) and a lack of tool adoption (E2: M = 3.47) that should be immediately addressed by institutional courses. From a policy standpoint, Pakistani higher education regulators and legislators have a clear and evidence-based reform agenda because of the stark contrast between students' high demand for institutional and governmental intervention (F3: M = 4.25; F4: M = 4.21) and their low level of trust in the current university readiness (F1: M = 3.02). Giving the next generation of computing professionals the knowledge, conduct, and change capabilities to act is not only an educational requirement but also a national security necessity, given the ongoing growth of cybersecurity threats fuelled by deepfakes both domestically and internationally.

7. References

Ahamed, F., Islam, M. T., Rahman, M. M., & Hassan, M. (2026). Cybersecurity knowledge, social networking, and awareness among Gen Z university students: A PLS-SEM approach. *Journal of Cybersecurity Education, Research and Practice*, *2026*(1), 1–19.

Alanazi, A., Al-Otaibi, S., Alshehri, M., &Alqahtani, M. (2025). Unmasking illusions: Understanding human perception and response to AI-generated deepfakes. *AI & Society*. Advance online publication.

Alexander, M. (2025). Deepfake cyberbullying: The psychological toll on students and institutional responses in educational settings. *Journal of School Violence*, *24*(2), 112–128.

Babaei, M., Karimi, H., Ahmadi, S., & Rezaei, T. (2025). Generative artificial intelligence and the evolving challenge of deepfake detection: A systematic review. *IEEE Access*, *13*, 45231–45258.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Geissler, R., Robertson, T., &Feuerriegel, S. (2026). Designing effective digital literacy interventions for boosting deepfake discernment: A large-scale experimental study. *Computers in Human Behavior*, *168*, 108421.

Hynek, N., Gavurova, B., &Kubak, M. (2025). Risks and benefits of artificial intelligence deepfakes: A cross-national perception study. *New Media & Society*, *27*(4), 889–913.

Roe, A., & Perkins, S. (2024). Deepfakes and higher education: A research agenda and conceptual framework. *British Journal of Educational Technology*, *55*(3), 901–918.

Roe, A., Perkins, S., & Mitchell, L. (2025). To deepfake or not to deepfake: Higher education stakeholders' perceptions of synthetic media adoption using UTAUT2. *Computers & Education*, *195*, 104731.

Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change. *Journal of Psychology*, *91*(1), 93–114. https://doi.org/10.1080/00223980.1975.9915803

Rogers, R. W. (1983). Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation. In J. T. Cacioppo & R. E. Petty (Eds.), *Social psychophysiology: A sourcebook* (pp. 153–176). Guilford Press.

Shad, M., Broklyn, A., & Egon, R. (n.d.). Deepfakes and cybersecurity: Detection and mitigation of AI-generated synthetic media threats [Unpublished manuscript].

Wasi, M. A., Islam, R., Hossain, M. S., &Akter, S. (2025). Seeing isn't believing: Addressing the societal impact of deepfakes in resource-limited environments. *Asian Journal of Information Technology*, *24*(1), 45–61.