

Original Research Article

Cyber Security Awareness and Responsible AI Adoption: A Mediation Analysis of AI Risk Perception

Dr. Sachin Joshi

Assistant Professor, HOD (Law), M.L. Dahanukar College of Commerce (Autonomous) Vile-Parle (East) Mumbai 400057 Maharashtra, India

Received: 10 September, 2026 | Accepted: 24 September, 2026 | Published: 5 October, 2026

Abstract:

The rapid integration of artificial intelligence into organisational workflows has introduced unprecedented cybersecurity challenges, necessitating a deeper understanding of how users perceive and respond to AI-related risks. This study investigates the mediating role of AI Risk Perception in the relationship between Cyber Security Awareness and Secure AI Usage Behaviour. Drawing on Protection Motivation Theory and employing a cross-sectional survey design, data were collected from 300 respondents across diverse industries using a structured questionnaire measuring three latent constructs. Mediation analysis, conducted using the PROCESS macro with bootstrap confidence intervals, revealed that Cyber Security Awareness has a strong positive effect on AI Risk Perception (path a: $\beta = 0.712$, $p < 0.001$), which in turn positively influences Secure AI Usage Behaviour (path b: $\beta = 0.444$, $p < 0.001$). The direct effect of Cyber Security Awareness on Secure AI Usage Behaviour remained significant (path c': $\beta = 0.413$, $p < 0.001$), while the indirect effect through AI Risk Perception was also significant ($\beta = 0.316$, 95% CI [0.229, 0.404]). These findings support partial mediation, indicating that AI Risk Perception accounts for approximately 43.3% of the total effect. The results underscore the importance of integrating cybersecurity education with AI risk awareness programmes to foster responsible AI adoption.

Keywords: Cyber Security Awareness, AI Risk Perception, Secure AI Usage Behaviour, Mediation Analysis, Responsible AI Adoption, Protection Motivation Theory

1. Introduction:

The proliferation of artificial intelligence technologies across professional and personal domains has fundamentally transformed the digital landscape, introducing both transformative opportunities and novel security vulnerabilities. As organisations increasingly deploy AI-powered tools for data processing, decision-making, and customer interaction, the intersection of cybersecurity and AI has emerged as a critical area of concern. The National Institute of Standards and Technology (NIST) AI Risk Management Framework emphasises that trustworthy AI requires deliberate attention to security, privacy, and risk governance throughout the AI lifecycle (NIST, 2023). Despite this growing recognition, the human dimension of AI security remains underexplored, particularly regarding how individual-level awareness of cybersecurity principles translates into safer AI usage practices.

Cyber security awareness, defined as an individual's knowledge and understanding of potential threats and protective measures in digital environments, has long been established as a significant predictor of security-compliant behaviour (Parsons et al., 2017). Research consistently demonstrates that employees with higher cybersecurity awareness exhibit fewer risky behaviours and greater compliance with organisational security policies (Hadlington & Parsons, 2017). However, the emergence of AI introduces a distinct category of risk that may not be fully captured by traditional cybersecurity awareness. AI systems present unique vulnerabilities, including data exposure through prompt injection, model hallucination, adversarial attacks, and the potential misuse of AI-generated content by malicious actors. Consequently, a new construct—AI Risk Perception—may serve as a cognitive bridge between general cybersecurity awareness and the specific behavioural adaptations required for secure AI usage.

Protection Motivation Theory (PMT), originally proposed by Rogers (1975), provides a robust theoretical lens for understanding how individuals respond to perceived threats. PMT posits that threat appraisal and coping appraisal jointly determine protective behaviour. In the context of AI adoption, individuals who are more aware of cybersecurity threats may develop heightened perceptions of AI-specific risks, which in turn motivates them to adopt more secure usage practices. This study examines this mediating mechanism through a simple mediation model, testing whether AI Risk Perception partially or fully mediates the relationship between Cyber Security Awareness and Secure AI Usage Behaviour.

2. Literature Review and Theoretical Framework: 2.1 Cyber Security Awareness

Cyber security awareness encompasses an individual's understanding of common digital threats, knowledge of protective measures, and the motivation to apply such measures consistently. Parsons et al. (2017) developed and validated the Human Aspects of Information Security Questionnaire (HAIS-Q), demonstrating that information security awareness can be reliably measured across dimensions of knowledge, attitude, and self-reported behaviour. Their findings indicated that employees who score higher on security awareness scales are significantly more likely to comply with organisational security policies and engage in protective behaviours. Similarly, Hadlington and Parsons (2017) found that positive attitudes towards cybersecurity in business environments are negatively associated with risky cybersecurity behaviours, underscoring the protective value of awareness.

The concept of cybersecurity awareness has expanded beyond traditional threats such as phishing and malware to encompass emerging technological risks. Research has revealed that employees continue to inadvertently expose sensitive information through AI tools, with studies indicating that approximately 4.7% of employees had pasted confidential data into public AI platforms such as ChatGPT as of mid-2023 (Cyberhaven, 2023). This finding highlights a critical gap: even technically aware individuals may not perceive the specific risks

introduced by AI systems, suggesting that traditional cybersecurity awareness alone may be insufficient to guarantee secure AI usage.

2.2 AI Risk Perception

AI Risk Perception refers to an individual's subjective assessment of the potential threats, harms, and vulnerabilities associated with the use of AI technologies. Unlike general technology risk perception, AI-specific risk perception encompasses concerns about data privacy breaches, algorithmic bias, misinformation from AI-generated content, and the weaponisation of AI by cybercriminals. Fildor et al. (2026) conducted a comprehensive study differentiating dimensions of AI risk perception, identifying privacy, accuracy, and misuse as distinct but interrelated concern domains. Their research demonstrated that individuals do not perceive AI risks as a monolithic construct but rather evaluate them across multiple dimensions, each with distinct behavioural implications.

Kerstan et al. (2024) developed the AI Risks and Benefits Scale, providing a validated psychometric instrument for measuring risk and benefit perceptions of AI in applied settings. Their findings indicated that both risk and benefit perceptions emerge as strong predictors of preferences for AI adoption, reinforcing the notion that balanced risk appraisal is essential for informed technology acceptance. The Financial Stability Board (2024) further emphasised that responsible AI adoption requires organisations to proactively identify and manage AI-specific risks, including model opacity, data quality issues, and third-party dependencies. At the individual level, heightened AI risk perception may function as a motivational catalyst, prompting users to exercise greater caution when interacting with AI systems.

2.3 Secure AI Usage Behaviour

Secure AI Usage Behaviour describes the specific protective actions individuals take when interacting with AI tools and platforms. These behaviours include avoiding the input of confidential or proprietary information into public AI systems, verifying AI-generated outputs before acting upon them, adhering to organisational AI usage policies, and reporting suspicious AI-related incidents. The construct extends traditional information security behaviour by incorporating AI-specific considerations, such as reviewing AI privacy settings, understanding the limitations of AI-generated content, and recognising the potential for adversarial manipulation.

The NIST AI Risk Management Framework (NIST, 2023) identifies several categories of AI risk, including harmful bias, security vulnerabilities, and privacy concerns, all of which necessitate user-level mitigation strategies. Microsoft's responsible AI principles similarly emphasise the importance of user agency, recommending that individuals maintain awareness of how AI systems process their data and exercise informed judgment when relying on AI-generated outputs (Microsoft, n.d.). Despite these guidelines, empirical research on the determinants of secure AI usage behaviour remains limited, particularly regarding the psychological mechanisms that translate general security awareness into AI-specific protective actions.

2.4 Mediation Framework

Mediation analysis provides a statistical framework for testing whether the effect of an independent variable on a dependent variable operates through an intervening mechanism (Baron & Kenny, 1986). The classic causal steps approach requires establishing significant relationships between the independent variable and the mediator (path a), the mediator and the dependent variable (path b), and the independent variable and the dependent variable (path c), followed by demonstrating that the direct effect (path c') is reduced when the mediator is included.

Modern approaches, particularly the bootstrap method advocated by Preacher and Hayes (2008) and implemented through the PROCESS macro (Hayes, 2018), provide more powerful and flexible tests of indirect effects without relying on the distributional assumptions of the traditional approach.

In the present study, Cyber Security Awareness is posited as the independent variable, AI Risk Perception as the mediator, and Secure AI Usage Behaviour as the dependent variable. Drawing on Protection Motivation Theory, it is hypothesised that cybersecurity awareness enhances individuals' perception of AI-specific risks, which subsequently motivates the adoption of secure AI usage practices. Partial mediation is anticipated, as cybersecurity awareness may also influence secure AI usage behaviour through pathways not involving risk perception, such as habitual security practices or normative compliance.

3. Research Methodology

3.1 Research Design and Sample

A cross-sectional survey design was employed to collect quantitative data from individuals who actively use AI tools in professional or personal contexts. A total of 300 valid responses were obtained through online distribution across multiple industries, including information technology (14.7%), finance (15.0%), education (12.0%), healthcare (10.7%), manufacturing (12.7%), retail (14.3%), and government (10.3%). The sample comprised 148 males (49.3%), 147 females (49.0%), and 5 respondents identifying as other (1.7%). The age distribution was concentrated in the 26–35 age group (37.3%), followed by 18–25 (26.7%), 36–45 (23.0%), 46–55 (8.0%), and 56 and above (5.0%). The majority of respondents held postgraduate (43.7%) or undergraduate (43.3%) qualifications, and 63.3% reported daily AI usage of one to four hours.

3.2 Instrument Development

The measurement instrument consisted of three constructs, each measured using six items on a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). Cyber Security Awareness (CSA) was assessed with items covering knowledge of common threats, password management practices, multi-factor authentication adoption, website verification habits, and awareness of risks in sharing confidential information with AI tools. AI Risk Perception (AIR) was measured using items addressing concerns about confidential information exposure, AI-generated content inaccuracy, privacy implications, exploitation by cybercriminals, increased cybersecurity challenges, and misuse of AI-generated information. Secure AI Usage Behaviour (SAU) captured behavioural intentions including avoiding confidential data entry into public AI tools, verifying AI-generated responses, following organisational policies, responsible AI use, reporting suspicious incidents, and reviewing AI privacy settings.

3.3 Data Analysis Procedure

Mediation analysis was conducted using the PROCESS macro (Model 4) for SPSS, following the bootstrap approach recommended by Hayes (2018) and Preacher and Hayes (2008). The analysis estimated three key effects: the total effect of Cyber Security Awareness on Secure AI Usage Behaviour (path c), the direct effect after accounting for the mediator (path c'), and the indirect effect through AI Risk Perception (path a $\times$ path b). Bootstrap confidence intervals with 5,000 resamples were computed to assess the statistical significance of the indirect effect, with an interval not containing zero indicating significant mediation.

4. Results

4.1 Descriptive Statistics

The demographic profile of the 300 respondents is presented in Table 1. The sample was relatively balanced in terms of gender, with near-equal representation of male and female participants. The largest age cohort was 26–35 years (37.3%), and the majority of respondents possessed postgraduate or undergraduate qualifications. AI usage was prevalent, with 69.0% of respondents using AI tools for one to four hours daily, indicating substantial engagement with AI technologies.

Table 1: Demographic Characteristics of Respondents (N = 300)

Variable	Category	n	%
Gender	Female	147	49.0
	Male	148	49.3
	Other	5	1.7
Age	18–25	80	26.7
	26–35	112	37.3
	36–45	69	23.0
	46–55	24	8.0
	56+	15	5.0
Education	Undergraduate	130	43.3
	Postgraduate	131	43.7
	Doctorate	28	9.3
	Other	11	3.7
AI Experience	<6 months	64	21.3
	6–12 months	78	26.0
	1–3 years	110	36.7
	>3 years	48	16.0
Daily AI Usage	<1 hour	59	19.7
	1–2 hours	105	35.0
	2–4 hours	102	34.0
	>4 hours	34	11.3

4.2 Mediation Analysis Results

The mediation analysis results are summarised in Table 2. The total effect of Cyber Security Awareness on Secure AI Usage Behaviour was significant ($\beta = 0.730$, $p < 0.001$), confirming that higher cybersecurity awareness is associated with greater secure AI usage behaviour. After including AI Risk Perception as a mediator, the direct effect remained statistically significant ($\beta = 0.413$, $p < 0.001$), indicating that cybersecurity awareness continues to influence secure AI usage independently of risk perception.

The indirect effect through AI Risk Perception was positive and statistically significant ($\beta = 0.316$, 95% CI [0.229, 0.404]), with the bootstrap confidence interval not crossing zero. This confirms that Cyber Security

Awareness enhances Secure AI Usage Behaviour partly by increasing individuals' perception of AI-related risks. Path a (CSA → AIR) yielded $\beta = 0.712$ ($p < 0.001$), and path b (AIR → SAU) yielded $\beta = 0.444$ ($p < 0.001$).

Table 2: Mediation Effects

Effect	Estimate	95% CI	p-value
Indirect Effect ($a \times b$)	0.316	[0.229, 0.404]	<0.001
Direct Effect (c')	0.413	[0.289, 0.538]	<0.001
Total Effect (c)	0.730	[0.622, 0.838]	<0.001

Table 3: Path Coefficients

Path	Relationship	β	p-value
a	CSA → AIR	0.712	<0.001
b	AIR → SAU	0.444	<0.001
c'	CSA → SAU	0.413	<0.001

The percentage of mediation was calculated as 43.3% indirect and 56.7% direct, confirming partial mediation. Since both the direct and indirect effects are statistically significant, with the direct effect remaining substantial after accounting for the mediator, the findings support the presence of partial rather than full mediation.

4.3 Hypothesis Testing

Table 4: Hypothesis Testing Results

Hypothesis	Statement	Result
H1	Cyber Security Awareness positively influences AI Risk Perception.	Supported
H2	AI Risk Perception positively influences Secure AI Usage Behaviour.	Supported
H3	Cyber Security Awareness positively influences Secure AI Usage Behaviour.	Supported
H4	AI Risk Perception mediates the relationship between CSA and SAU.	Supported (Partial Mediation)

5. Discussion

The findings of this study provide empirical evidence for the mediating role of AI Risk Perception in the relationship between Cyber Security Awareness and Secure AI Usage Behaviour. All four hypotheses were supported, contributing to a more nuanced understanding of the psychological mechanisms that underpin responsible AI adoption. The strong positive relationship between Cyber Security Awareness and AI Risk Perception (H1; $\beta = 0.712$) suggests that individuals who are knowledgeable about general cybersecurity threats are well-positioned to recognise and evaluate the specific risks posed by AI technologies. This finding aligns with the threat appraisal component of Protection Motivation Theory, which posits that awareness of potential threats enhances individuals' sensitivity to related dangers (Rogers, 1975). The substantial magnitude of this path coefficient indicates that cybersecurity awareness is a powerful antecedent of AI risk perception, suggesting that existing cybersecurity training programmes may serve as an effective foundation for building AI risk literacy.

The significant positive effect of AI Risk Perception on Secure AI Usage Behaviour (H2; $\beta = 0.444$) demonstrates that individuals who perceive higher AI-related risks are more likely to adopt protective behaviours when using AI tools. This result is consistent with the coping appraisal dimension of PMT and corroborates findings by Fildor et al. (2026), who identified distinct risk perception dimensions that drive cautious technology engagement. The practical implication is that fostering accurate AI risk perception—rather than either amplifying or minimising risks—may be an effective strategy for promoting secure AI usage.

The persistence of a significant direct effect of Cyber Security Awareness on Secure AI Usage Behaviour (H3; $\beta = 0.413$) after accounting for the mediator indicates that cybersecurity awareness promotes secure behaviour through additional pathways beyond risk perception. These may include habitual security practices, normative influences, or direct knowledge application. The partial mediation finding (H4) reveals that AI Risk Perception accounts for 43.3% of the total effect, while 56.7% operates through direct channels. This suggests that while risk perception is an important mechanism, it is not the sole pathway through which cybersecurity awareness influences secure AI usage.

6. Limitations and Future Research

Several limitations should be acknowledged. First, the cross-sectional design precludes causal inference; longitudinal or experimental designs would strengthen conclusions about the directional relationships among the constructs. Second, the reliance on self-reported behavioural measures introduces the possibility of social desirability bias, where respondents may overstate their secure AI usage practices. Future research could incorporate objective behavioural data or observational measures to validate self-report findings. Third, the sample, while diverse in industry representation, was limited to 300 respondents and may not be fully representative of broader populations, particularly in regions with lower AI adoption rates. Fourth, the study examined a simple mediation model with a single mediator; future investigations could explore additional mediators, such as self-efficacy or organisational culture, or moderating variables such as AI literacy or industry sector. Finally, the rapidly evolving nature of AI technology means that the specific risk perceptions and behavioural responses captured in this study may require periodic reassessment as new AI capabilities and associated threats emerge.

7. Conclusion

This study empirically demonstrates that Cyber Security Awareness significantly enhances Secure AI Usage Behaviour both directly and indirectly through AI Risk Perception. The partial mediation model, supported by bootstrap analysis, reveals that approximately 43.3% of the influence of cybersecurity awareness on secure AI usage is transmitted through heightened AI risk perception. These findings carry practical implications for organisations seeking to promote responsible AI adoption: integrating AI-specific risk education into existing cybersecurity awareness programmes may substantially improve users' protective behaviours. As AI technologies continue to permeate professional and personal domains, fostering a security-conscious mindset that encompasses both traditional cyber threats and AI-specific vulnerabilities will be essential for safeguarding digital ecosystems.

References

- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173>

- Cyberhaven. (2023). Data exposure report: ChatGPT in the enterprise. Cyberhaven. <https://cyberhaven.com/>
- Fildor, D., Škrinjarić, B., & Budak, J. (2026). Anatomy of AI risk perception: Differentiating privacy, misinformation, and technical failures. *International Journal of Human–Computer Interaction*. Advance online publication. <https://doi.org/10.1080/10447318.2025.2610439>
- Financial Stability Board. (2026). *Sound practices for responsible adoption of artificial intelligence (AI): Consultation report*. FSB. <https://www.fsb.org/2026/06/sound-practices-for-responsible-adoption-of-artificial-intelligence-ai-consultation-report/>
- Hadlington, L., & Parsons, K. (2017). Human factors in cybersecurity: Examining the link between Internet addiction, impulsivity, attitudes towards cybersecurity, and risky cybersecurity behaviours. *Heliyon*, 3(7), e00345. <https://doi.org/10.1016/j.heliyon.2017.e00345>
- Hayes, A. F. (2018). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (2nd ed.). The Guilford Press.
- Kerstan, A., Bienefeld, N., & Grote, G. (2024). AI Risks and Benefits Scale. APA PsycTests. <https://doi.org/10.1037/t93153-000>
- Microsoft. (n.d.). Responsible AI. <https://www.microsoft.com/en-us/ai/responsible-ai>
- National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
- Parsons, K., Calic, D., Pattinson, M., Butavicius, M., McCormac, A., & Zwaans, T. (2017). The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies. *Computers & Security*, 66, 1–12. <https://doi.org/10.1016/j.cose.2017.01.004>
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891. <https://doi.org/10.3758/BRM.40.3.879>
- Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change. *The Journal of Psychology*, 91(1), 93–114.