Proceedings of the
Informing Science +
Information Technology
Education Conference

An Official Publication
of the Informing Science Institute
InformingScience.org

InformingScience.org/Publications

July 20 – 28, 2025

UNDERSTANDING THE ROLE OF TRAINING FOR TRUST IN AI-BASED CYBERSECURITY

Ruti Gafni * — The Academic College of Tel-Aviv Yaffo, Tel-Aviv, Israel — rutigafn@mta.ac.il

Itzhak Aviv — The Academic College of Tel-Aviv Yaffo, Tel-Aviv, Israel — itzhakav@mta.ac.il

* Corresponding author

ABSTRACT

Aim/Purpose — This paper investigates the impact of training on trust in AI-based cybersecurity solutions, addressing challenges related to skills development and trust dynamics.

Background — Implementing AI in cybersecurity has proven effective by enhancing threat identification, management, and prevention capabilities. Proper training and education facilitate comprehension of AI solutions and concepts, assisting cybersecurity professionals in utilizing the technology. Experts use AI-based cybersecurity systems based on their trust in these systems.

Methodology — A structured survey was conducted with 100 cybersecurity experts. Data were analyzed using multiple regression and structural equation modeling to explore the relationships between training, skills, perceived effectiveness, and trust.

Contribution — This study provides insights into how training influences trust through skill-building and perceived effectiveness. It contributes to better training programs' design and fosters trust in AI-driven cybersecurity solutions.

Findings — The study finds that training enhances skills, which in turn affects perceived effectiveness and trust, though highly skilled individuals may develop skepticism toward AI systems. Training significantly enhances AI-related skills. However, its impact on trust is indirect, as it primarily improves skills, which in turn influences perceived effectiveness and trust. Trust is mediated by improved perceptions of effectiveness driven by skill development. Advanced

Accepted by Editor Grandon Gill │ Received: December 17, 2024 │ Revised: April 13, 2025 │ Accepted: Ajpril 16, 2025.

Cite as: Gafni, R., & Aviv, I. (2025). Understanding the role of training for trust in AI-based cybersecurity. In M. Jones (Ed.), *Proceedings of InSITE 2025: Informing Science and Information Technology Education Conference*, Article 31. Informing Science Institute. https://doi.org/10.28945/5576

	skills may paradoxically reduce trust due to increased awareness of system limitations.
Recommendations for Practitioners	Training programs should integrate hands-on experiences and explainable AI techniques to balance skill-building with trust-enhancing strategies.
Recommendations for Researchers	Further investigation is needed into the trust-skills paradox and the cultural or contextual factors influencing trust in AI systems.
Impact on Society	Enhancing trust in AI-based cybersecurity systems promotes broader adoption, contributing to improved cybersecurity resilience.
Future Research	Future studies should focus on objective performance assessments, diverse user groups, and cultural factors affecting trust dynamics.

Keywords	AI-based cybersecurity, trust dynamics, training, skill development, perceived effectiveness.

INTRODUCTION

AI has emerged as a crucial tool for enhancing cybersecurity as cyber threats become increasingly complex and frequent. Many organizations are adopting AI technologies to enhance cybersecurity, recognizing the importance of protecting digital assets. AI can rapidly analyze large amounts of data to identify cyberattack risks, enabling organizations to avert such incidents (Camacho, 2024; Pendey, 2023). This adoption necessitates a paradigm shift that requires the organization to reimagine its current mechanisms and models to fully encompass the nature and potential of AI systems (Shoetan et al., 2024; Temara, 2024).

User trust in AI-based cybersecurity systems is crucial for their successful adoption. Trust influences both how widely these systems are used in organizations and how effectively they function. Consequently, this trust is established through various factors, such as organizational alignment, credibility, and perceived performance. For example, trust is significantly influenced by perceived efficacy, which is the user's perspective on the potential for AI systems to mitigate security risks (Udeh et al., 2024). Well-established theoretical frameworks from technology acceptance research, such as the adapted version of the Technology Acceptance Model (TAM) to AI (Zhang et al., 2025), offer insights into how perceived usefulness and perceived ease of use shape user acceptance and trust. Hoff and Bashir's (2015) model of trust in automation further illuminates how transparency, reliability, and user expectations influence users' willingness to rely on AI-driven solutions.

Moreover, human-computer interaction studies emphasize cognitive and behavioral factors underlying trust, suggesting that cultural and contextual variations, such as uncertainty avoidance and power distance, can moderate perceptions of AI trustworthiness (Oksanen & Savolainen, 2025). Including cross-cultural research in future investigations would thus broaden the relevance and applicability of AI-based cybersecurity solutions. This confidence is founded upon reliability, or consistent performance over time, which is a critical factor in selecting a vendor (Shoetan et al., 2024), mainly when an extensive range of commercial products are emerging for cybersecurity technical tasks (Gafni & Levy, 2024). Nevertheless, trust is most effectively comprehended in dynamic contexts contingent upon the user's perceived knowledge of AI capabilities and their skills. Wang (2024) and Temara (2024) discovered that the application of technology can be more confident when one has acquired better technology skills, but it can also reveal some of the technology's limitations.

This paper aims to determine the extent to which training, skills, perceived effectiveness, and trust influence the organization's adoption of AI-based cybersecurity systems. The study suggests that hands-on training is essential for improving AI skills, as it helps professionals feel more confident in

AI systems and their effectiveness. Drawing from the existing literature, the research endeavors to comprehend the intermediary mechanism and the impact of training on trust.

This work contributes to the current body of literature on adopting AI for cybersecurity matters and offers unique suggestions for developing training that fosters divided trust. In this scenario, the cognitive development of technical skills, the defects of AI, and ethical factors that influence user confidence can be significant in enhancing the optimum utilization of AI cybersecurity systems.

STATE OF THE ART

The implementation of AI in cybersecurity has proven effective by enhancing threat identification, management, and prevention capabilities. Artificial intelligence's strength is its capability to process and analyze huge quantities of data efficiently, automatically classify it, and recognize patterns while freeing up professionals from monotonous and repetitive tasks and enhancing their efficiency (Gafni & Levy, 2024). Artificial intelligence can automate processes, analyze data, and generate real-time threat intelligence, rendering it an essential tool in mitigating sophisticated cyber-attacks (Grassini, 2023; Urhobo, 2024). Effective deployment can be improved by addressing challenges such as limited high-quality training data, evolving cyber threats, and ethical concerns (Shahana et al., 2024; Villegas-Ch et al., 2024). Research distinguishes the utilization of supervised machine learning techniques, including decision trees, support vector machines, and neural networks, to improve AI-based cybersecurity. Although these algorithms have demonstrated efficacy in identifying and mitigating hazards, they also necessitate collaborative efforts to address inherent limitations (Ijiga et al., 2024; Shahana et al., 2024).

In addition to addressing general ethical concerns (e.g., privacy and regulatory compliance), future work should implement concrete bias-detection and fairness-assessment techniques. Fairness-aware machine learning models can help identify and mitigate biases that may disproportionately affect certain user groups or data segments. Proposing frameworks aligned with global data protection laws, such as GDPR and NIST AI Risk Management Framework guidelines, would ensure that AI-driven cybersecurity solutions meet both technical and ethical benchmarks.

This has highlighted the vulnerabilities posed by adversarial attacks on AI models. Attacks that mimic input data and yield erroneous outputs in AI systems necessitate robust detection mechanisms and continuous model retraining to maintain system reliability (Temara, 2024). Transitioning AI into cybersecurity frameworks requires compatibility and comprehension of conventional applications and systems, challenges often tied to commercial AI solutions (Villegas-Ch et al., 2024). Although these technical considerations are critical, human factors such as trust and proficiency can significantly affect successful AI adoption. This underscores the importance of training cybersecurity professionals whose skills and confidence in AI-based systems directly impact trust and utilization. Consequently, the following section explores the essential role of training in shaping trust dynamics in AI-driven cybersecurity.

THE ROLE OF TRAINING

Education is essential in bridging the gap between AI technology and its practical implementation in organizations. Programs and courses facilitate comprehension of AI solutions and concepts to assist cybersecurity professionals in utilizing the technology. Skills such as understanding algorithms, data management, and ethics enhance analytical and decision-making capabilities (Delcker et al., 2024; Pham et al., 2024). Therefore, the initial hypothesis (H1) asserts that:

H1: Training on AI-based cybersecurity solutions positively influences AI-related skills.

Practical exposure augments the understanding of AI applications and produces many factors that facilitate the assessment of the efficacy of these applications. The perceived effectiveness is crucial as it correlates with trust, a significant predictor of adopting AI technology (Shahana et al., 2024; Villegas-Ch et al., 2024). Consequently, the second hypothesis (H2) is formulated as:

H2: Training on AI-based cybersecurity solutions positively influences perceived effectiveness.

Training also contributes to familiarity with explainable AI (XAI) approaches that enhance comprehensibility and traceability. XAI alleviates prevailing apprehensions and uncertainties concerning AI operational processes, fostering confidence between users and training systems and aligning training objectives with organizational goals (Vemuri et al., 2023).

Skills and Trust Dynamics

AI skills pertain to the knowledge relevant to the objectives of artificial intelligence. Professionals and experts will evaluate AI's strengths and limitations while assessing the organization's ability to tackle cybersecurity challenges (Ijiga et al., 2024). Understanding the mechanisms of AI algorithms, data management, and potential negative implications of AI deployment will enhance the proficiency of experts in assessing the authenticity and responsibility of applied AI technologies (Familoni, 2024). The third hypothesis (H3) encapsulates this link:

H3: AI-related skills positively influence the perceived effectiveness of AI-based cybersecurity solutions.

The literature has emphasized the role of skills in influencing trust in technology. This competence influences trust, as users feel comfortable engaging with and evaluating the system due to their perceived effectiveness in their domain (Villegas-Ch et al., 2024; Wang, 2024). These factors are essential for enhancing the interpretability of tree outputs, identifying issues, and predicting behavior, thereby contributing to a sense of reliability and efficiency. This corroborates the fourth hypothesis (H4):

H4: AI-related skills positively influence trust in AI-based cybersecurity solutions.

To address this issue, AI training programs should include components that focus on strategies for confidence and resilience in the face of AI's inefficacy (Temara, 2024). The essential knowledge that should underpin honest system talks is the strengths and limitations of the established skills.

Perceived Effectiveness and Trust

The efficacy of an AI system is a significant factor influencing the degree of trust individuals have in the technology. It is essential for users to recognize and trust in the efficacy of these solutions to fulfill their security requirements. Only demonstrable outcomes that enhance danger detection and prevention may develop this confidence (Nadella et al., 2024). This results in the fifth hypothesis (H5):

H5: Perceived effectiveness positively influences trust in AI-based cybersecurity solutions.

Nevertheless, if an organization establishes measurable performance requirements for AI implementation and adheres to them, it might be perceived as exceptionally productive. These indicators provide tangible evidence of system reliability and efficacy, instilling user confidence and trust (Nadella et al., 2024; Olabanji et al., 2024).

Mediation Effects of Training

Trust is influenced both positively and negatively by training. It cultivates abilities that enhance perceived capability, offering a pathway for the Internet to promote trust (Mayeke et al., 2024). The aforementioned dynamics are evident in the sixth and seventh hypotheses:

H6: Training directly increases trust in AI-based cybersecurity solutions.

H7: Training indirectly increases trust by mediating AI-related skills and perceived effectiveness.

The sequential mediation effect necessitates the enhancement of advanced training activities. Theoretical knowledge and practical implementations of training models facilitate uniformity in the implementation of training programs across organizations, encompassing both technical proficiency and the cultivation of trust (Bhardwaj & Dave, 2024).

METHODOLOGY

This study used a structured survey to evaluate the associations between AI training, skills development, perceived efficacy of AI-based cybersecurity solutions, and trust in those solutions. The research hypotheses were derived from a literature review that facilitated the identification of factors that influence trust in AI systems. The constructs and survey questions were developed in response to the theoretical frameworks based on these hypotheses.

The survey comprised 24 questions to capture participants' attitudes, training experiences, and expertise with AI-driven cybersecurity solutions. In designing these questions, we mapped each item to specific theoretical constructs identified in the literature – primarily drawing from the Technology Acceptance Model (TAM), Hoff and Bashir's (2015) trust in automation framework, and human-computer interaction principles.

To ensure the survey captured relevant operational factors, we also considered recognized cybersecurity guidelines and standards, such as ISO/IEC 27001 and the NIST Cybersecurity Framework. These standards emphasize core aspects like risk management, incident response, and performance metrics, which can influence how AI solutions are adopted and perceived within organizations. For instance, questions assessing the perceived effectiveness of AI align with the performance metrics recommended under NIST's 'Identify' and 'Detect' functions, while items on trust incorporate the transparency and auditing principles advocated by ISO/IEC 27001.

Based on the identified hypotheses (H1–H7), we developed corresponding survey items covering five key constructs: (1) Training – measuring the extent, frequency, and perceived value of AI-focused training, (2) AI-Related Skills – gauging self-reported proficiency in machine learning, data analysis, and security operations, (3) Perceived Effectiveness – capturing the user's assessment of AI tools' reliability and threat detection capability, (4) Trust – derived from trust in automation literature, including transparency and reliability sub-dimensions, and (5) Adoption Intent – aligning with TAM-based measures of behavioral intent. Each question underwent face validity checks by subject matter experts and inter-observer reliability analysis to ensure clarity and consistency.

In total, 100 carefully selected subject matter experts participated in the survey, comprising cybersecurity professionals with diverse roles, such as information security administrators, analysts, and penetration testers. While we acknowledge the value of more extensive and diverse samples for broader applicability, the decision to focus on 100 experts was guided by several scientific considerations. First, this sample size meets established thresholds for reliable statistical analyses, including multiple regression and structural equation modeling (Hayes, 2022). Second, these participants were drawn from multiple organizations and varied professional backgrounds, thus offering a sufficiently heterogeneous viewpoint on AI-based cybersecurity practices. Finally, by concentrating on experts, the study captures in-depth knowledge and practical experience, which is critical for understanding nuanced factors like trust and perceived effectiveness in specialized contexts. The respondents asserted that they possess various academic and training qualifications and have an average of four years of cybersecurity professional experience (44% have more than seven years of experience in cybersecurity tasks).

Using SPSS (version 28), we performed multiple regression analyses and structural equation modeling (SEM) to evaluate both the direct and mediated models of the relationship between training and trust. Before conducting these analyses, we tested key assumptions such as normality, linearity, and homoscedasticity. Skewness and kurtosis values for continuous variables were within acceptable ranges (±2), indicating no severe deviations from normality. Next, we conducted a confirmatory factor analysis (CFA) on all latent constructs to assess measurement validity and reliability. Indicator loadings above 0.70, composite reliability (CR) above 0.70, and average variance extracted (AVE) above 0.50 suggested satisfactory convergent validity and reliability. Discriminant validity was tested by comparing the square root of AVE for each construct with the inter-construct correlations, confirming that each latent variable was distinct. We then used Model 6 of the PROCESS macro (Hayes, 2022) for our serial mediation analysis.

We employed 5,000 bootstrap samples to generate bias-corrected and accelerated (BCa) 95% confidence intervals for each regression path and indirect effect. The BCa bootstrap method adjusts for potential skewness and kurtosis in the sampling distribution, providing more accurate estimates than conventional normal-theory confidence intervals. Specifically, if the 95% BCa confidence interval for an indirect effect does not include zero, we conclude that the effect is statistically significant. Confidence intervals reflect the range of plausible values for the estimated effects. By using BCa intervals, we account for potential bias in the resampled distribution, enhancing the robustness of our inference. Consequently, any reported indirect or direct effect is deemed reliable when its confidence interval excludes zero. This approach reinforces the rigor and credibility of our analysis, ensuring that the results are valid and replicable.

For the SEM part, we evaluated model fit using multiple fit indices: the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). In this framework, 'Training' (X) served as the exogenous variable, with 'AI Skills' (M1) and 'Perceived AI Effectiveness' (M2) as sequential mediators and 'Trust in AI' (Y) as the outcome. All variables were measured through established scales and validated items aligned with prior technology acceptance and trust research (e.g., perceived effectiveness and trust items adapted from Hoff and Bashir (2015); skill-related items based on AI competency frameworks). This approach enabled us to rigorously assess both direct effects and indirect mediation paths among the constructs.

These methods allowed for the documentation of the impact of training on trust through skills and perceived training effectiveness and the provision of practical recommendations for training design.

Results

The analysis revealed significant insights into the relationships between training, skills, perceived effectiveness, and trust in AI-based cybersecurity solutions. The hypotheses were tested using multiple regression and mediation analysis, as summarized below.

The survey results indicated that training significantly enhances AI-related skills but does not directly increase perceived effectiveness or trust. While advanced skills improve users' ability to assess AI systems, they may also make professionals more aware of system limitations. This increased awareness could, in some cases, contribute to skepticism and lower trust in AI. Perceived effectiveness, however, strongly influences trust, highlighting the importance of measurable outcomes in fostering confidence. Table 1 presents the summary of the hypotheses tested and their results.

The mediation analysis provided deeper insights into the complex relationships between training, skills, perceived effectiveness, and trust. Table 2 summarizes the mediation pathways.

Table 1. The summary of hypotheses

Hypothesis	Statement	Result	Details
H1	Training on AI-based cybersecurity solutions positively influences AI-related skills.	Supported	Training significantly increases AI-related skills. Coefficient: β = .2591, p = .0077, CI [.0702, .4480].
H2	Training on AI-based cybersecurity solutions positively influences perceived effectiveness.	Not Supported	Training alone does not enhance perceived effectiveness. Coefficient: β = .0789, p = .3169, CI [-.0768, .2346].
H3	AI-related skills positively influence perceived effectiveness.	Supported	Higher AI-related skills lead to improved perceptions. Coefficient: β = .4765, p < .0001, CI [.3172, .6359].
H4	AI-related skills positively influence trust.	Not Supported	Higher skills slightly reduce trust. Coefficient: β = -.1977, p = .0024, CI [-.3233, -.0720].
H5	Perceived effectiveness positively influences trust.	Supported	Perceived effectiveness fosters trust. Coefficient: β = .3539, p < .0001, CI [.2178, .4900].
H6	Training directly increases trust in AI-based cybersecurity solutions.	Not Supported	Training does not directly impact trust. Coefficient: β = .0456, p = .3945, CI [-.0602, .1513].
H7	Training indirectly increases trust through skills and perceived effectiveness.	Partially Supported	Sequential mediation pathway (Skills → Effectiveness → Trust) validated. Indirect Effect: .0437, CI [.0059, .0872].

Table 2. Summary of mediation pathways

Pathway	Effect	Confidence Interval (CI)	Result	Interpretation
Training → Skills → Trust	-0.0512	[-0.1196, -0.0025]	Significant, Negative	Training improves AI-related skills, but higher skills reduce trust, possibly due to increased awareness of system limitations.
Training → Effectiveness → Trust	0.0279	[-0.0308, 0.1004]	Non-Significant	Training does not directly improve perceived effectiveness significantly, indicating other factors mediate this relationship.
Training → Skills → Effectiveness → Trust	0.0437	[0.0059, 0.0872]	Significant, Positive	Training enhances skills, which improves perceived effectiveness and subsequently increases trust. Highlights the importance of both mediators.

The relationships and effects are summarized visually in Figure 1, illustrating the mediation pathways and their statistical significance. Key findings summary:

- Training → Skills: Supported; training enhances AI-related skills.
- Skills → Perceived Effectiveness: Supported; skill development boosts perceptions of AI effectiveness.
- Perceived Effectiveness → Trust: Supported; higher perceived effectiveness leads to greater trust.

- Skills → Trust: Not supported; advanced skills can sometimes reduce trust, highlighting a ‘trust-skills paradox.’
- Training → Trust: Not supported; training does not directly raise trust but operates through skills and effectiveness.

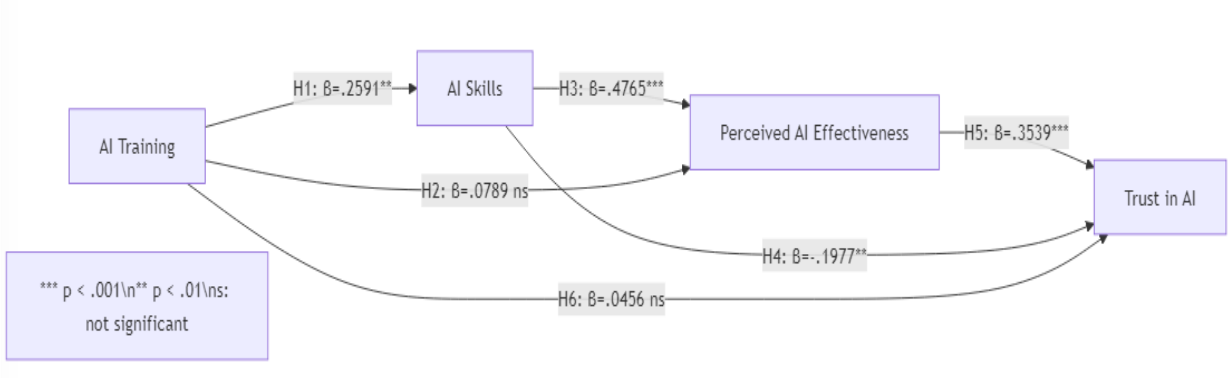

Figure 1. Research results

Having established these core findings, the following Discussion section interprets their implications in the context of existing literature and proposes directions for future research.

Discussion

The results of this study offer critical insights into the mechanisms through which training impacts trust in AI-based cybersecurity solutions. The findings illuminate the potential and limitations of current training approaches by examining the direct and mediated pathways. This discussion contextualizes the results within the broader literature, highlights practical implications, and suggests areas for future research.

Training and AI-Related Skills

The study confirmed that training significantly enhances AI-related skills (H1). This finding aligns with previous research emphasizing the role of structured training in skills development (Familoni, 2024; Rangaraju, 2023). By equipping professionals with technical knowledge and critical thinking abilities, training programs can address the digital and AI skills gaps prevalent in many organizations (Camacho, 2024; Pham et al., 2024). These skills are foundational for leveraging AI technologies effectively in cybersecurity, enabling professionals to understand and utilize underlying algorithms, data management techniques, and ethical frameworks (Camacho, 2024; Delcker et al., 2024).

The interaction between skills and trust offers a complex dynamic. The interaction between advanced skills and trust presents a complex dynamic. Higher skills and trust (H4) have negative associations since highly skilled experts are more suited to identify system weaknesses, thus fostering increased mistrust about dependability (Temara, 2024). While our findings suggest that deeper technical knowledge can highlight AI system vulnerabilities and heighten user skepticism, other factors may also be at play. For instance, highly skilled professionals might have prior negative experiences with AI tools that did not meet their expectations, or they may feel that current organizational processes do not fully support these technologies’ safe and transparent use. Additionally, cultural and contextual elements, such as fear of job displacement or concerns about the ethical use of AI, could further influence trust levels.

Because our study relied on a structured survey, we could not fully capture these alternative explanations through open-ended or qualitative means. Future research incorporating interviews or focus groups could more thoroughly explore the underlying reasons why advanced AI skills might correspond to lower trust, providing richer, context-specific insights into this counterintuitive finding.

From a cognitive perspective, more experience can result in the "risk sensitivity bias," in which deeper technical awareness raises worries about possible errors. Furthermore, aggravating uncertainty among advanced users is caused by organizational elements like poor transparency in AI system design or inadequate explanations of how algorithms operate. Future studies should combine theoretical views on cognitive load and trust building (e.g., Lyons & Guznov, 2019; Morita & Burns, 2014) to investigate how skill level interacts with perceived complexity and contextual elements in forming trust. This finding is consistent with the literature on cognitive biases, where greater expertise sometimes leads to skepticism or overcaution (Shehu et al., 2023; Temara, 2024). Training programs, therefore, need to balance skill-building with modules that address confidence and trust-building, incorporating transparent discussions about system limitations and strengths.

Training and Perceived Effectiveness

Contrary to expectations, training alone does not directly enhance the perceived effectiveness of AI-based cybersecurity solutions (H2). This result underscores the complexity of influencing perceptions, shaped by many factors beyond technical knowledge (Shahana et al., 2024; Villegas-Ch et al., 2024). Hands-on experience and contextual applications are likely critical for translating training content into perceived effectiveness. For instance, simulations and real-world scenarios can help participants observe tangible outcomes of their skills, bridging the gap between theoretical knowledge and practical efficacy (Grassini, 2023; Urhobo, 2024).

The strong positive relationship between AI-related skills and perceived effectiveness (H3) further supports this notion. When professionals possess robust skills, they are better equipped to evaluate and appreciate the capabilities of AI systems. This finding aligns with studies highlighting the importance of technical knowledge and critical thinking in enhancing perceptions of system effectiveness (Nadella et al., 2024; Olabanji et al., 2024). It also reinforces the need for skill-building as a prerequisite to improve perceptions of effectiveness.

Perceived Effectiveness and Trust

The study demonstrated a significant positive relationship between perceived effectiveness and trust (H5). This finding emphasizes the central role of demonstrable outcomes in fostering trust. Users are more likely to trust AI-based solutions when they perceive them as effective and reliable (Dash et al., 2022; van Bussel et al., 2022). Transparency, accountability, and the integration of robust ethical guidelines are essential for enhancing perceived effectiveness and, consequently, trust (Katrakazas & Papastergiou, 2024; Sarker, 2023).

The research highlights the importance of XAI techniques in improving transparency and accountability, which in turn fosters trust in AI-driven cybersecurity solutions (Capuano et al., 2022; Vemuri et al., 2023). Organizations should invest in XAI and other mechanisms that demonstrate the reliability and performance of their AI systems, addressing user concerns about potential biases or limitations.

Mediation Pathways: Training, Skills, Effectiveness, and Trust

The pathway "Training → Skills → Trust" showed a surprising result: higher AI skills were linked to lower trust. This suggests that as professionals become more skilled, they may also become more aware of system vulnerabilities, making them less trusting of AI systems. It reflects the need for training programs to address, apart from technical proficiency, trust-oriented perspectives, such as confidence-building and balanced assessments of AI capabilities (Shahana et al., 2024).

The pathway "Training → Effectiveness → Trust" was non-significant, suggesting that training alone does not immediately translate into perceptions of effectiveness. This finding aligns with the literature emphasizing the importance of experiential learning and contextual applications (Ijiga et al., 2024). Training programs can enhance perceived effectiveness by incorporating real-world use cases and measurable outcomes, thereby strengthening trust.

The pathway "Training → Skills → Effectiveness → Trust" emerged as the most robust mechanism. This sequential mediation underscores the cascading effects of training on trust through skill development and improved perceptions of effectiveness. It validates the importance of integrating comprehensive training strategies that align skill-building with demonstrable outcomes (Dash et al., 2022; Mayeke et al., 2024).

PRACTICAL IMPLICATIONS

Training programs should not only build skills but also show professionals how these skills make AI systems more effective. Incorporating real-world scenarios and hands-on experiences can enhance the practical applicability of training, helping participants connect theoretical knowledge with tangible outcomes. Addressing the trust-skills paradox is critical; advanced training should include modules that foster a balanced understanding of system limitations and strengths, reducing skepticism while building confidence. Experiential learning components, such as simulations and threat detection challenges, are effective tools for improving perceptions of effectiveness, directly influencing trust. Organizations should also invest in XAI techniques to enhance transparency and accountability, addressing user concerns about biases or limitations. Ethical considerations, including robust guidelines and privacy-preserving mechanisms, are essential in building trust and should be integral to both training and deployment strategies. Collaborative efforts between academia and industry can further strengthen training initiatives, fostering a skilled workforce capable of navigating the complex cybersecurity landscape.

LIMITATIONS AND FUTURE RESEARCH

Despite offering meaningful insights into how training, AI-related skills, and perceived effectiveness shape trust in AI-based cybersecurity solutions, this study has several limitations. First, the reliance on self-reported survey data raises potential issues of social desirability and recall bias. Future studies should integrate more objective metrics, such as performance-based assessments, practical exercises, or real-time monitoring of user interactions with AI-driven security tools. Such an approach would allow for more affluent, triangulated data and reduce the influence of self-reporting biases.

Second, the sample for this research comprised a relatively small group of cybersecurity professionals, potentially restricting the generalizability of the findings. While focusing on subject-matter experts provides in-depth, context-specific knowledge, broader participant pools, including policymakers, IT administrators, AI developers, and less technically oriented users, would offer a more holistic view of trust dynamics. Expanding to different industries, organizational sizes, and cultural contexts would further illuminate how the "trust-skills paradox" might manifest under varying regulatory, technological, or cultural conditions.

Third, our cross-sectional design captured user perceptions and trust levels at a single time. A longitudinal approach could reveal how training and trust evolve, especially as professionals gain experience with new AI tools or as cybersecurity threats become more sophisticated. By monitoring participants over multiple training phases and real-world applications, future research could identify critical tipping points where greater AI expertise either enhances or undermines trust.

Fourth, although our findings suggest that XAI features and real-world simulation exercises may boost perceived effectiveness and trust, these recommendations are based on existing literature rather than direct empirical verification within this study. Controlled experiments or case studies could explicitly test whether these interventions meaningfully shift user attitudes and behaviors. Such work would help practitioners refine training programs that both elevate technical competence and mitigate undue skepticism or overreliance.

The negative association between advanced AI skills and trust, which this study calls the "trust-skills paradox," warrants deeper exploration. Future investigations should use qualitative methods (e.g., interviews, focus groups) and experimental designs to unpack the cognitive processes, organizational policies, and cultural factors that might lead highly skilled professionals to become more skeptical of AI. Addressing these avenues would significantly enhance our understanding of how best to harness training for responsible and trusted AI-driven cybersecurity.

Conclusion

This study adds to the growing literature on AI trust by clarifying how training, skill acquisition, and perceived effectiveness influence users' confidence in AI-based cybersecurity solutions. While training does increase AI-related skills, the relationship with trust is not straightforward. Instead, our findings highlight a nuanced "trust-skills paradox" in which heightened technical awareness can sometimes amplify skepticism about AI capabilities. This underscores the importance of carefully designed training programs that not only impart technical proficiency but also address realistic concerns about AI limitations, algorithmic transparency, and ethical implications.

Moreover, we find that perceived effectiveness plays a pivotal mediating role: advanced skills bolster perceptions of effectiveness, fostering greater trust. Therefore, organizations seeking to implement AI-driven security tools should integrate real-world simulations, transparent performance metrics, and robust explanatory mechanisms into their professional development efforts. Doing so can bridge the gap between theoretical knowledge and tangible proof of AI's reliability, a step essential for enhancing trust.

Although our study draws on expert perspectives, its methodological and demographic scope is limited, indicating the need for more diverse samples and longitudinal designs. Future research could incorporate performance-based assessments and explore the impact of cultural or organizational differences, offering richer insights into the global landscape of AI adoption for cybersecurity. Additionally, investigating how hands-on learning experiences and XAI solutions specifically modify the trust-skills dynamic would have direct implications for training program design. Our results emphasize that trust-building is a multi-dimensional challenge requiring technical skill development, transparency, measurable outcomes, and an ethical framework. As organizations increasingly rely on AI to safeguard digital assets, attention to these facets can ensure that training initiatives yield better-skilled professionals and a more confident, critically engaged workforce capable of harnessing the true potential of AI-based cybersecurity.

Acknowledgments

We want to thank our students Noa Aharon, Rotem Asulin, and Shirel Zakai for helping us collect the data for this research.

References

Bhardwaj, S., & Dave, M. (2024). RAKSHAM: Responsive approach to Knock-off scavenging hackers and attack mitigation. *Transactions on Emerging Telecommunications Technologies*, *35*(1), e4904. https://doi.org/10.1002/ett.4904

Camacho, N. G. (2024). The role of AI in cybersecurity: Addressing threats in the digital age. *Journal of Artificial Intelligence General Science, 3*(1), 143–154. https://doi.org/10.60087/jaigs.v3i1.75

Capuano, N., Fenza, G., Loia, V., & Stanzione, C. (2022). Explainable artificial intelligence in cybersecurity: A survey. *IEEE Access, 10*, 93575-93600. https://doi.org/10.1109/ACCESS.2022.3204171

Dash, B., Ansari, M. F., Sharma, P., & Ali, A. (2022). Threats and opportunities with AI-based cybersecurity intrusion detection: A review. *International Journal of Software Engineering & Applications, 13*(5), 13–21. https://doi.org/10.5121/ijsea.2022.13502

Delcker, J., Heil, J., Ifenthaler, D., Seufert, S., & Spirgi, L. (2024). First-year students AI-competence as a predictor for intended and de facto use of AI-tools for supporting learning processes in higher education. *International Journal of Educational Technology in Higher Education*, *21*, Article 18. https://doi.org/10.1186/s41239-024-00452-7

Familoni, B. T. (2024). Cybersecurity challenges in the age of AI: Theoretical approaches and practical solutions. *Computer Science & IT Research Journal, 5*(3), 703-724. https://doi.org/10.51594/csitrj.v5i3.930

Gafni, R., & Levy, Y. (2024). The role of artificial intelligence (AI) in improving technical and managerial cybersecurity tasks' efficiency. *Information & Computer Security, 32*(5), 711-728. https://doi.org/10.1108/ICS-04-2024-0102

Grassini, S. (2023). Development and validation of the AI attitude scale (AIAS-4): A brief measure of general attitude toward artificial intelligence. *Frontiers in Psychology, 14*, 1191628. https://doi.org/10.3389/fpsyg.2023.1191628

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis* (3rd ed.). Guilford Press.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Ijiga, O. M., Idoko, I. P., Ebiega, G. I., Olajide, F. I., Olatunde, T. I., & Ukaegbu, C. (2024). Harnessing adversarial machine learning for advanced threat detection: AI-driven strategies in cybersecurity risk assessment and fraud prevention. *Research Journal of Science and Technology, 11*(1), 1–4. https://doi.org/10.53022/oarjst.2024.11.1.0060

Katrakazas, P., & Papastergiou, S. (2024). A stakeholder needs analysis in cybersecurity: A systemic approach to enhancing digital infrastructure resilience. *Businesses*, *4*(2), 225-240. https://doi.org/10.3390/businesses4020015

Lyons, J. B., & Guznov, S. Y. (2019). Individual differences in human-machine trust: A multi-study look at the perfect automation schema. *Theoretical Issues in Ergonomics Science, 20*(4), 440-458. https://doi.org/10.1080/1463922X.2018.1491071

Mayeke, N. R., Arigbabu, A. T., Olaniyi, O. O., Okunleye, O. J., & Adigwe, C. S. (2024). Evolving access control paradigms: A comprehensive multi-dimensional analysis of security risks and system assurance in cyber engineering. *Asian Journal of Research in Computer Science, 17*(5), 108-124. https://doi.org/10.9734/ajrcos/2024/v17i5442

Morita, P. P., & Burns, C. M. (2014). Understanding "interpersonal trust" from a human factors perspective: Insights from situation awareness and the lens model. *Theoretical Issues in Ergonomics Science, 15*(1), 88–110. https://doi.org/10.1080/1463922X.2012.691184

Nadella, G. S., Gonaygunta, H., Kumar, D., & Pawar, P. P. (2024). Exploring the impact of AI-driven solutions on cybersecurity adoption in small and medium enterprises. *World Journal of Advanced Research and Reviews, 22*(1), 1190-1197. https://doi.org/10.30574/wjarr.2024.22.1.1185

Oksanen, A., & Savolainen, I. (2025). Trust in emerging technologies and artificial intelligence. In K. J. Rotenberg, S. Petrocchi, A. Levante & F. Lecciso (Eds.), *Handbook of trust and social psychology* (pp. 184-205). Edward Elgar. https://doi.org/10.4337/9781803929415.00023

Olabanji, S. O., Oladoyinbo, O. B., Asonze, C. U., Oladoyinbo, T. O., Ajayi, S. A., & Olaniyi, O. O. (2024). Effect of adopting AI to explore big data on personally identifiable information (PII) for financial and economic data transformation. *Asian Journal of Economics Business and Accounting, 24*(4), 106–125. https://doi.org/10.9734/ajeba/2024/v24i41268

Pendey, B. (2023). Artificial intelligence and cyber security. *Journal Transnational Universal Studies, 1*(2), 93–99.

Pham, L.,O'Sullivan, B., Scantamburlo, T., & Mai, T. (2024). Addressing eigital and AI skills gaps in European living areas: A comparative analysis of small and large communities. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(21), 23119–23127. https://doi.org/10.1609/aaai.v38i21.30357

Rangaraju, S. (2023). AI Sentry: Reinventing cybersecurity through intelligent threat detection. *International Journal of Science and Engineering, 9*(3), 30-35. https://doi.org/10.53555/ephijse.v9i3.211

Sarker, I. H. (2023). Multi-aspects AI-based modeling and adversarial learning for cybersecurity intelligence and robustness: A comprehensive overview. *Security and Privacy, 6*(5), e295. https://doi.org/10.1002/spy2.295

Shahana, A., Hasan, R., Farabi, S. F., Akter, J., Al Mahmud, M. A., Johora, F. T., & Suzer, G. (2024). AI-driven cybersecurity: Balancing advancements and safeguards. *Journal of Computer Science and Technology Studies*, *6*(2), 76-85. https://doi.org/10.32996/jcsts.2024.6.2.9

Shehu, A. U., Umar, M., & Aliyu, A. (2023). Cyber Kill Chain analysis using artificial intelligence. *Asian Journal of Research in Computer Science*, *16*(3), 210-219. https://doi.org/10.9734/ajrcos/2023/v16i3357

Shoetan, P. O., Amoo, O. O., Okafor, E. S., & Olorunfemi, O. L. (2024). Synthesizing AI's impact on cybersecurity in telecommunications: A conceptual framework. *Computer Science & IT Research Journal, 5*(3), 594–605. https://doi.org/10.51594/csitrj.v5i3.908

Temara, S. (2024). Harnessing the power of artificial intelligence to enhance next-generation cybersecurity. *World Journal of Advanced Research and Reviews, 23*(2), 797–81. https://doi.org/10.36227/techrxiv.170785703.32137017/v1

Udeh, E. O., Amajuoyi, P., Adeusi, K. B., & Scott, A. O. (2024). The integration of artificial intelligence in cybersecurity measures for sustainable finance platforms: An analysis. *Computer Science & IT Research Journal, 5*(6), 1221–1246. https://doi.org/10.51594/csitrj.v5i6.1195

Urhobo, B. (2024). Understanding the role of artificial intelligence in enhancing GRC practices in cybersecurity. *World Journal of Advanced Research and Reviews, 22*(2), 269–274. https://doi.org/10.30574/wjarr.2024.22.2.1340

van Bussel, M. J. P., Odekerken-Schröder, G., Ou, C., Swart, R. R., & Jacobs, M. J. G. (2022). Analyzing the determinants to accept a virtual assistant and use cases among cancer patients: A mixed methods study. *BMC Health Services Research, 22*, Article 890. https://doi.org/10.1186/s12913-022-08189-7

Vemuri, N., Thaneeru, N., & Tatikonda, V. M. (2023). Securing trust: Ethical considerations in AI for cybersecurity. *Journal of Knowledge Learning and Science Technology 2*(2), 167-175. https://doi.org/10.60087/jklst.vol2.n2.p175

Villegas-Ch, W., Govea, J., & Ortiz-Garces, I. (2024). Developing a cybersecurity training environment through the integration of OpenAI and AWS. *Applied Sciences, 14*(2), Article 679. https://doi.org/10.3390/app14020679

Wang, M. (2024). Generative AI: A new challenge for cybersecurity. *Journal of Computer Science and Technology Studies, 6*(2), 13–18. https://doi.org/10.32996/jcsts.2024.6.2.3

Zhang, C., Hu, M., Wu, W., Kamran, F., & Wang, X. (2025). Unpacking perceived risks and AI trust influences pre-service teachers' AI acceptance: A structural equation modeling-based multi-group analysis. *Education and Information Technologies, 30*(2), 2645–2672. https://doi.org/10.1007/s10639-024-12905-7

AUTHORS

Ruti Gafni, PhD, is an Associate Professor, Dean, and establisher of the School of Information Systems at The Academic College of Tel Aviv Yaffo with BSc studies, including specialties in Cybersecurity, Digital Innovation, and Gamification, and MSc studies, including specialties in Data Science and Digital Transformation. She holds a PhD from Bar-Ilan University, Israel (in the Business Administration School), focusing on Information Systems, an MSc from Tel Aviv University in Information Systems Management, and a BA (Cum Laude) in Economics and Computer Science from Bar-Ilan University. She has more than 40 years of practical experience as a Software Project Manager and Analyst of information systems. Her research interests include cybersecurity, AI, and the adoption of new technologies.

Dr Itzhak Aviv, PhD, is a leading expert in cybersecurity, data science, quantum computing, and blockchain, with significant academic leadership roles. He is an affiliated researcher at the Vienna University of Economics and Business (Austria). He is an Assistant Professor and Head of the MSc Data Science program at the Academic College of Tel-Aviv Yaffo, and he serves as a PhD supervisor at the University of Haifa, both in Israel. He also chairs the IEEE RE4Web3 workshop devoted to requirements engineering for Web3 systems. His contributions to cybersecurity research have earned him recognition in academia and industry.