International Journal of Multidisciplinary Research of Higher Education (IJMURHICA)
http://ijmurhica.ppj.unp.ac.id/index.php/ijmurhica

Strengthening Security Awareness in the Era of Artificial Intelligence-Based Cyber Threats

Nusandika Patria[1], Sopian Amir[1], Dana Indra Sensuse[1], Sofian Lusa[1], Nur Indrawati[1], Nurcholis Ramlan[1]
[1]Universitas Indonesia, Indonesia
nusandika.patria@ui.ac.id *

Article Information:
Received February 8, 2026
Revised March 20, 2026
Accepted April 20, 2026

Keywords: *AI-based cyber threats, cybersecurity, human factors, security awareness*

Abstract

Artificial intelligence (AI) is developing rapidly and is increasingly being applied in various fields, including cybersecurity. However, this development also introduces new, more sophisticated threats that are difficult to detect. AI-based cyber threats, including adaptive malware, highly personalized phishing, deepfakes for identity manipulation, adaptive distributed denial of service (DDoS) attacks, and automated ransomware, are projected to escalate by 2025. These threats' complexity requires a security approach that relies not only on technology but also on human awareness as the first line of defense. In Indonesia, the 2024 Cybersecurity Landscape Report, published by the National Cyber and Crypto Agency (BSSN), shows that public and institutional awareness of information security is still relatively low. This study presents a systematic literature review of 30 articles to examine how security awareness is being strengthened in the context of AI-based cyber threats. The review identified six main categories of AI enabled threats social engineering and phishing, content manipulation and impersonation, malware and ransomware, attacks against machine learning models, service disruption, and automated AI-orchestrated attacks. In parallel, seven categories of awareness strategies were synthesized: education and training programs, gamification and simulation-based learning, policy and governance support, technical and system-level controls, collaboration and multi-stakeholder engagement, legal and ethical frameworks, and psychological or human-centric approaches. The findings highlight that strengthening security awareness requires an integrated and multidimensional approach that bridges technological, organizational, regulatory, and human-centered efforts.

How to cite: Patria, N., Amir, S., Sensuse, D. I., Lusa, S., Indrawati, N., Ramlan, N. (2026). Strengthening Security Awareness in the Era of Artificial Intelligence-Based Cyber Threats. *International Journal of Multidisciplinary of Higher Education (IJMURHICA)*, *9*(2), 329-353.

E-ISSN: 2622-741x
Published by: Islamic Studies and Development Center Universitas Negeri Padang

INTRODUCTION

In this interconnected digital age, cybersecurity has become a major concern for individuals, organizations, and governments alike. As technology advances, cyberattack tactics and techniques are also evolving, requiring organizations to remain vigilant and proactive in the face of ever-emerging threats (Jimmy, 2021).

The development of artificial intelligence (AI) brings great benefits, but at the same time poses new risks. Cyber attacks are no longer entirely driven by humans, but are increasingly powered by AI, which has become the tool of choice for criminals (Ghuge, 2024). In fact, AI is also a transformative force in improving situational awareness in cybersecurity (Edim et al., 2025). However, despite its benefits, AI can be used maliciously. AI-based attacks, including increasingly sophisticated phishing and malware, have increased significantly in recent years (Alanezi & Al-Azzawi, 2024; Guembe et al., 2022). In 2023, AI-powered cyberattacks increased by 238%, with global losses reaching more than USD 8.5 billion (Reddem, 2024).

Recent reports reinforce this trend. HiddenLayer asserts that AI is a technology that is highly vulnerable to exploitation, both in the form of attacks on AI-based systems and misuse to generate harmful content, deepfakes, and adaptive phishing attacks. A deepfake-based fraud case that caused losses of up to USD 25 million shows that the impact is not only financial but also has the potential to destabilize political and social stability(HiddenLayer, 2024). The complexity of this threat requires organizations to be fully prepared. Fortinet reports that 62% of organizational leaders believe their employees are at greater risk of becoming victims of attacks due to the use of AI by criminals. Ironically, 67% of decision makers believe their employees still lack security awareness, and nearly a third of organizations do not even have mechanisms in place to monitor their employees' use of AI (Fortinet, 2024).

In Indonesia, similar challenges are also evident. Public awareness of cyberattack threats, especially those based on AI, is still relatively low (Engkizar et al., 2025; Hermawan et al., 2023; Kassymova et al., 2025; Masoud & Almajri, 2025). The 2024 Cybersecurity Landscape report published by the National Cyber and Crypto Agency shows that information security awareness at both the public and institutional levels is still relatively low. In response to this, the government has passed Presidential Regulation No. 47 of 2023 on the National Cybersecurity Strategy and established BSSN Regulation No. 25 of 2024, which launched the 2024–2028 National Cybersecurity Action Plan. One of its priority programs is Voluntary Vulnerability Identification and Protection (VVIP), which includes a coordinated and sustainable agenda to increase cybersecurity awareness (Badan Siber dan Sandi Negara, 2024). These initiatives not only address the current gap in cybersecurity awareness, but also provide a strong foundation for accelerating Indonesia's Electronic-Based Government System (SPBE), in which effective cybersecurity safeguards are essential to achieving good governance, integrated public services, and long-term resilience against evolving cyber threats.

Given these circumstances, raising cybersecurity awareness plays a crucial role in reducing the likelihood of AI-driven attacks. This awareness involves keeping up with emerging threat patterns, recognizing suspicious messages or harmful content, and applying essential protective measures such as multi-factor authentication and regular system updates. In this context, the present study conducts a systematic literature review on security awareness related to AI-based cyber threats. Its objective is to identify practical strategies, educational efforts, and proven approaches that can enhance individuals' and

organizations' understanding of security, thereby contributing to stronger cyber resilience both in Indonesia and across the global landscape.

Acknowledging the growing need to enhance cybersecurity awareness, this study formulates specific research questions to explore the characteristics of AI-driven cyber threats and to promote a deeper understanding of how such risks can be prevented: RQ1: What are the types of AI-based cyber threats?. RQ2: What strategies can be implemented to enhance information security awareness against AI-based cyber threats?

To ensure a systematic flow of discussion, the structure of this article is organized as follows. Section 1 introduces the background of the study. Section 2 reviews the relevant literature. Section 3 describes the research methodology, including data collection and requirements. Section 4 presents the research results and discussion in relation to the research questions. Finally, Section 5 summarizes the key findings, highlights the study's limitations, and offers recommendations for future research.

This section reviews key theories relevant to understanding how security awareness must evolve in response to the rapid advancement of artificial intelligence and its growing impact on cybersecurity.

Cybersecurity is generally emphasize the coordinated use of resources and actions to prevent, detect, respond to, and recover from harmful events in cyberspace. Many authors frame cybersecurity as the protection of networks, systems, and data through the integration of people, processes, and technologies (Schiliro Francesco, 2023; Wilson & Kiy, 2014). Across studies, several core components consistently emerge. The primary objective is to safeguard assets such as data, systems, and services from compromise by deploying well-coordinated resources and processes (Schiliro Francesco, 2023). Technical controls, including firewalls, encryption, intrusion detection systems, secure architectures, and resilience mechanisms, are repeatedly cited as central pillars of cybersecurity (Wilson & Kiy, 2014). Organizational processes such as risk management, incident response, continuous monitoring, and lifecycle governance are also identified as essential layers of protection (Onwubiko & Ouazzane, 2022). In addition, human and legal dimensions encompassing personnel training, policy development, regulatory frameworks, and property rights are increasingly integrated into modern definitions of cybersecurity (Onwubiko & Ouazzane, 2022). At the core of all cybersecurity agendas lies the CIA triad, as shown in figure 1, which consists of Confidentiality, Integrity, and Availability, and serves as a guiding model for data security strategies. Confidentiality ensures that only authorized users can access sensitive information. Integrity restricts unauthorized modification or deletion of data. Availability guarantees that systems, services, and data remain accessible as required (Li & Liu, 2021; Subhani et al., 2023).

Fig 1. CIA Triad (Almass & Chowdhary, 2024)

Artificial Intelligence-Based Cyber Threats. AI-driven cyber threats represent a rapidly developing aspect of the cybersecurity landscape. Such threats make use of artificial intelligence technologies to increase the reach, accuracy, and overall effectiveness of attacks. Examples include AI-generated phishing attempts, self-adapting malware, and automated exploitation of vulnerabilities. These forms of attack are especially challenging to identify and

counter because of their complexity, learning capability, and constant evolution (Ghuge, 2024). A key challenge is that AI-powered attacks often mimic legitimate user behavior, thereby blending seamlessly into normal network traffic and complicating detection efforts (Sharma, 2024). Moreover, threat actors ranging from cybercriminals to state-sponsored groups are increasingly employing AI to advance forms of cyber warfare, frequently motivated by financial or strategic gains (Kovaci, 2024). Beyond the threats themselves, the integration of AI into cybersecurity raises additional concerns related to data quality, algorithmic bias, and ethical implications, which must be carefully addressed to ensure that AI-based defense mechanisms remain both effective and equitable (Okdem & Okdem, 2024; Sharma, 2024). As illustrated in figure 2, AI-powered malware can be categorized into several key dimensions: evasion techniques that help malware avoid detection, autonomous capabilities enabling smart and machine-speed decisions, AI against itself where adversarial methods are used to exploit or deceive AI systems, and bio-inspired & swarm intelligence approaches that allow malware to evolve, mutate, and coordinate attacks. This taxonomy highlights how AI not only strengthens traditional forms of malware but also introduces entirely new paradigms of cyber threats that demand innovative countermeasures (Thanh & Zelinka, 2019).

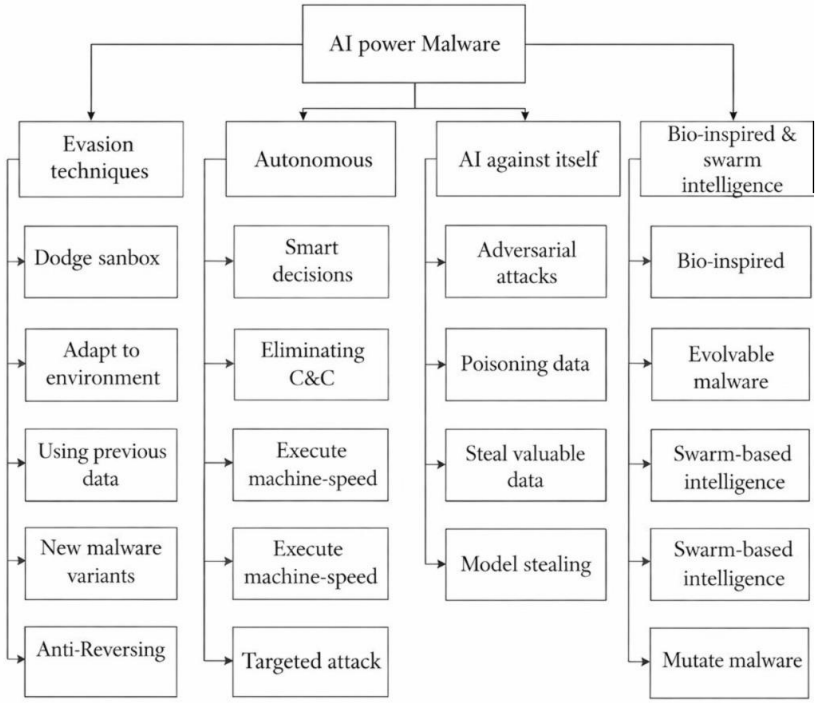

Fig 2. Taxonomy of AI techniques in malware
(Thanh & Zelinka, 2019)

Information security awareness (ISA) is widely recognized in the literature as a distinct construct that primarily reflects an attitudinal and attentional state predisposing individuals to recognize and respond to security concerns. ISA is understood as an individual's passive engagement combined with growing attentiveness to security-related issues that gradually fosters a security-oriented mindset (Chua et al., 2021). Hsu and Zhou emphasize the importance of distinguishing awareness from training and education. They frame awareness as an organizational and individual-level construct designed to sensitize people to potential security risks and motivate appropriate security behaviors (Hu et al., 2022). In line with this perspective, Parsons et al. introduced the Human Aspects of Information Security Questionnaire (HAISQ) as a validated instrument to measure ISA across multiple dimensions

of security behavior. The HAISQ emphasizes not only knowledge but also the practical application of security practices in everyday contexts, thereby operationalizing ISA into measurable factors. This framework is among the most commonly used models for evaluating security awareness, largely because it connects awareness with measurable behavior instead of viewing it purely as a matter of cognition. The HAISQ highlights several essential dimensions that encompass the range of human factors influencing information security. Through these dimensions, it provides a holistic perspective on how awareness is reflected in practical, security-oriented actions (Parsons et al., 2014).

METHODS

This study's systematic literature review methodology was based on Prisma International Standards. PRISMA, which stands for "Preferred Reporting Items for Systematic Reviews and Meta-Analyses," is an updated guideline used for reporting systematic reviews (SLR) (Damri et al., 2023; E Engkizar et al., 2024; 2018; Iskandar et al., 2023; Moher et al., 2009; Page, McKenzie, et al., 2021; Sauer & Seuring, 2023). The PRISMA 2020 framework for implementing SLR consists of four stages: Identification, Screening, Eligibility, and Included (Aryasutha et al., 2025; Page, Moher, et al., 2021).

The purpose of this research was to collect and analyze studies related to security awareness in the context of AI-driven cyberattacks. To meet this objective, the study presents the criteria used in selecting relevant literature and explains the methods applied to gather and evaluate the data and publications.

Research Questions and Review Process

In conducting SLR, the research questions (RQ) were developed using the PICOC framework Population, Intervention, Comparison, Outcome, and Context. Table 1 outlines the PICOC elements applied in this study.

Table 1. PICOC Framework

Element	Explanation
(P)	Individuals and/or organizations exposed to AI-based cyber threats.
(I)	Strategies or programs that aim to improve information security awareness.
(C)	-
(O)	Enhanced information security awareness and practical skills improve users' ability to detect phishing, deepfakes, and AI-based attacks, fostering safer behavior and lowering attack success rates.
(C)	AI-based cyber threats

Based on the PICOC framework, the final RQs were derived and summarized in table 2. This structured justification provides a clear foundation for identifying relevant studies and developing the search criteria.

Table 2. Research Question

Research question	Rationale
RQ1: What are the types of AI-based cyber threats?	To identify the various forms of AI-based cyber threats, and determine their mechanism.
RQ2: What strategies can be implemented to enhance information security awareness against AI-based cyber threats?	To explore existing security awareness measures that can strengthen individuals and organizations capacity to detect, respond, and prevent AI-based cyber threats

Search Criteria and Identification of Studies

The search criteria were developed based on the RQs presented in table 1. The process of constructing the search strategy involved several stages. First, keywords were generated from the RQs. Second, keywords used in relevant literature were identified. Third, distinct synonyms and alternative spellings of the keywords were recognized. Fourth, the primary keywords and concepts were combined using Boolean operators.

The final search string applied was: ("security awareness") AND ("artificial intelligence" OR "AI") AND ("threat" OR "risk" OR "attack" OR "cybercrime"). Comprehensive searches were conducted across leading academic databases with broad scholarly coverage, including Scopus, ScienceDirect, IEEE Xplore, Taylor & Francis, ProQuest, and Sage Journals. The reference lists of the included studies were screened to identify additional relevant works.

Eligibility Criteria

The review process was conducted in four sequential stages: initiation, title and abstract screening, full-text screening, and quality assessment. At each stage, predefined inclusion and exclusion criteria were applied. For the quality assessment, six guiding questions were used to evaluate the selected studies. A summary of the stages along with their corresponding inclusion and exclusion criteria is presented in table 3.

Table 3. Stages of Review Process

Stages	Inclusion Criteria	Exclusion Criteria
Initiation	Search strategy: Boolean search Publication period: 2021–2025 Language: English Disciplinary scope: Information Technology and related domains Accessibility: Open access sources	Duplicate literature Study type: SLR, Literature Review, Conference Notes, Speaker Notes
Title and abstract selection	The study focuses on types of AI-based cyber threats and strategies to enhance information security awareness against AI-based cyber threats.	The topic is irrelevant to AI-based cyber threats and not discuss information security awareness
Full-Text Selection	The study explains the types of AI-based cyber threats. The study discusses the strategy to enhance information security awareness against AI-based cyber threats.	The article is not available in full text. The study lacks sufficient data for synthesis or fails to meet quality criteria

RESULT AND DISCUSSION

During a comprehensive search across six databases, 1232 journal articles were identified. 3 duplicate articles were found at this stage. At the initiation stages, 791 articles were excluded because they did not meet the inclusion and exclusion criteria. In the title and abstract selection, the screening focused on studies addressing AI-based cyber threats and strategies for strengthening information security awareness. This process reduced 320 articles form 498 articles report sought for retrieval to 178 articles, which were then examined in full-text selection to ensure relevance. From this review, 148 articles were screened based on the quality criteria, as shown in table 4. Finally, the assessment confirmed that 30 of these studies met the required quality standards and were subsequently included in the review. The figure 4 illustrates the overall selection

process.

Table 4. Quality Criteria

Checklist	Checklist Statement
QC1	Are the problems and solutions clearly stated?
QC2	Are the research objectives clearly stated?
QC3	Is there identification of the types of AI-based cyber threats?
QC4	Does the study discuss the factors of information security awareness against AI-based cyber threats?
QC5	Are the research results clearly explained?
QC6	Does the article's conclusion answer the research question?

In this study, data are drawn from the selected articles, encompassing details such as titles, publication years, authors, countries of origin, contextual case studies, research aims, categories of AI-related cyber threats, and approaches for improving information security awareness. The collected information is then carefully examined through a data-driven synthesis process. This method allows the integration of findings from multiple studies, producing a structured and evidence-based understanding of the topic. In addition, it helps reveal recurring patterns and emerging trends that appear across the analyzed body of literature.

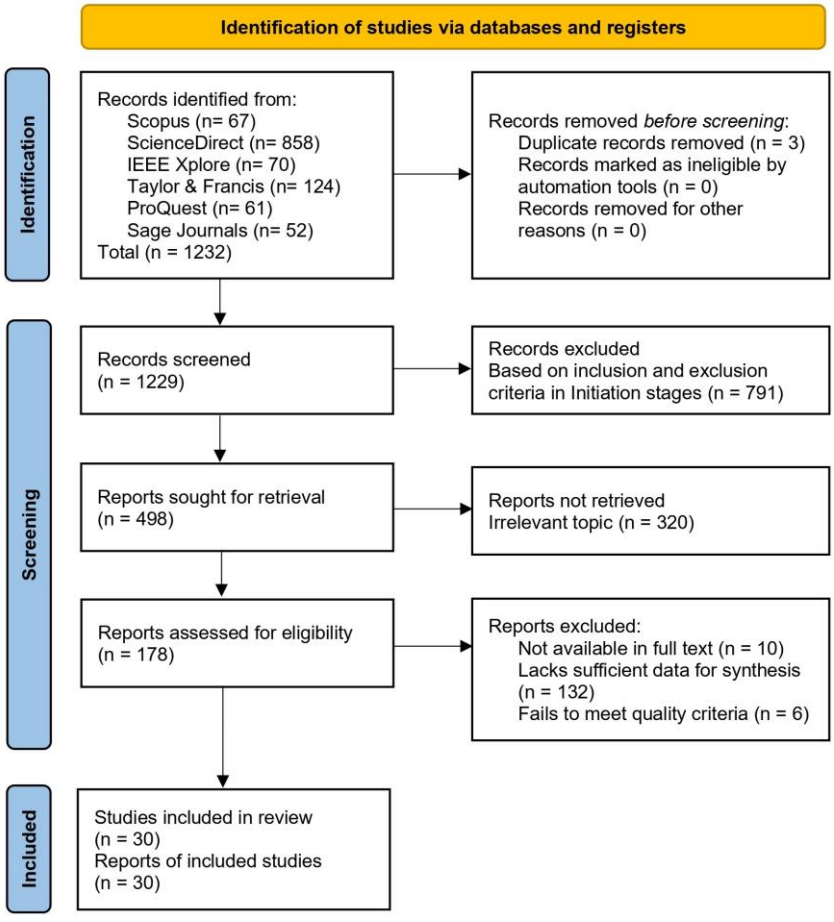

Fig 3. PRISMA flowchart of systematic review process and article selection

This section outlines the results of the systematic literature review, which investigates how security awareness is being reinforced in response to AI-driven cyber threats. The findings are arranged according to the established research questions and encompass two main aspects: the types of AI-based cyber threats discussed in the literature, and the strategies proposed to improve awareness among individual users and organizations. By synthesizing insights from the

selected studies, this section highlights emerging patterns, thematic clusters, and research gaps that provide a comprehensive understanding of the current state of knowledge in this field.

Country Analysis

The distribution of studies across countries indicates varying levels of research related to the topic. As illustrated in figure 4.

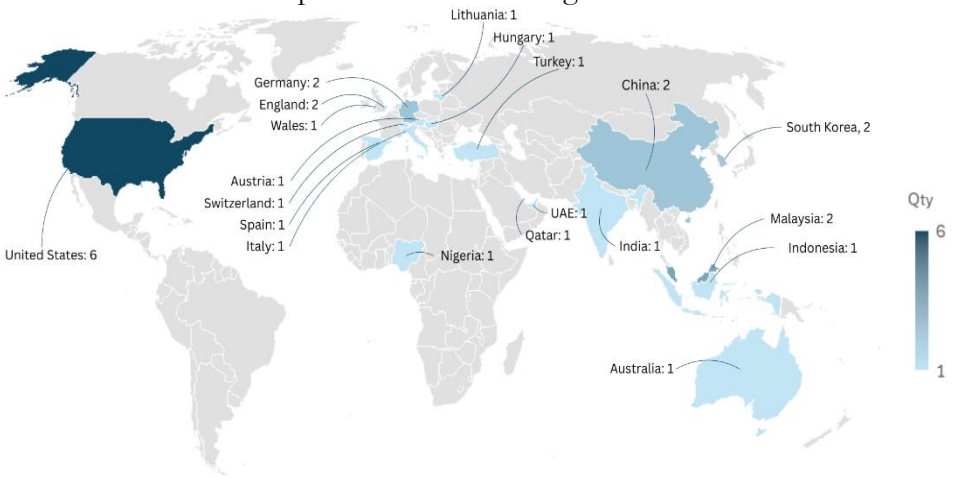

Fig 4. Country distribution of primary studies

The United States leads with the highest number of publications with 6 study, reflecting its strong research capacity and active engagement in addressing issues within this domain. China follows with 1 study, while other Asian countries such as South Korea, Malaysia, and Indonesia have also contributed, although at a smaller scale. In Europe, contributions are spread across several countries including Germany, England, Switzerland, Spain, Lithuania, Austria, Hungary, Turkey, and Wales, each producing at least 1 publication. Additionally, studies from the Middle East (Qatar and the United Arab Emirates) and Africa (Nigeria) highlight the global scope of interest in this area, despite relatively lower contributions compared to North America and Europe. This distribution suggests that while the research field has gained worldwide attention, there remains an imbalance, with most of the work concentrated in a few leading countries, leaving room for broader participation from underrepresented regions.

Published Year Analysis

Figure 5 illustrates the distribution of primary studies included in the SLR across the years 2021 to 2025.

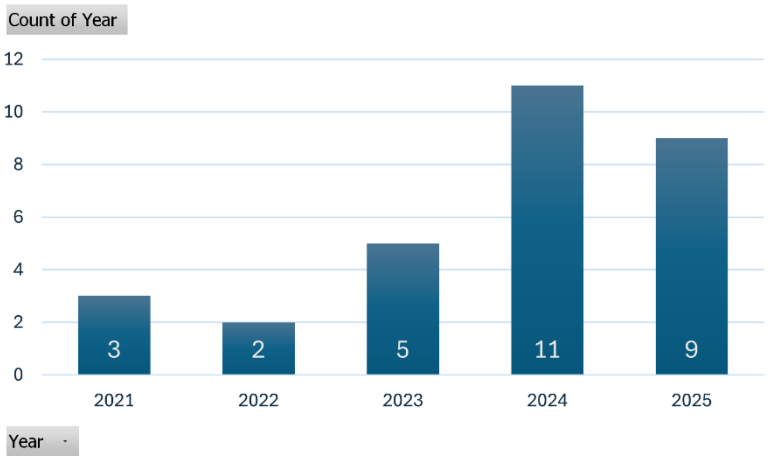

Fig 5. Published Year Analysis

In 2021, a total of three studies were identified, followed by a decline to two studies in 2022. The number of publications then increased to five studies

in 2023 and reached its peak in 2024 with eleven studies. In 2025, the number of studies slightly decreased to ten, though it remained substantially higher compared to the earlier years. Overall, the trend indicates a growing research interest in the reviewed topic, particularly in the last three years (2023–2025), reflecting its increasing relevance in contemporary scholarly discourse.

Types of AI-based Cyber Threats

After reviewing 30 selected articles, researchers identified several techniques of AI-based cyber threats. Figure 6 visualizes these techniques with a thematic word cloud based on the frequency of cyber threat words appearing in the articles, reflecting how artificial intelligence is increasingly exploited by malicious actors to compromise digital systems. The types then collected into six main categories that highlight the diverse and evolving ways in which AI is weaponized, emphasizing the urgent challenges faced by organizations, governments, and individuals in enhancing their security awareness.

Fig 6. AI-Based cyber threats techniques

Table 5. Types of AI-Based Cyber Threat

No	Types of AI-Based Cyber Threat	Example Techniques / Methods	References
1	Social engineering & phishing (human-targeting)	General phishing, spear phishing, smishing/vishing, business email compromise (BEC), social networking service (SNS) phishing, and LLM/AI-powered or lateral phishing.	(Alahmed et al., 2024; Angafor et al., 2024; Baltuttis & Teubner, 2024; Banire et al., 2021; Bethany et al., 2025; Bitrián et al., 2024; Calvo et al., 2025; Gallo et al., 2024; Jagadeesan et al., 2024; Luh et al., 2025; Park & Kim, 2025; pratama et al., 2025; Qin et al., 2025; Riskhan et al., 2025; Soon et al., 2024; Stylianou et al., 2025; Tawalbeh & Muheidat, 2023; Tinubu et al., 2023; Yoo & Cho, 2022; Zhang et al., 2025)
2	Content manipulation & impersonation	deepfake pornography and misuse, AI bots for deception, and synthetic media.	(Ghazi-Tehrani & Pontell, 2021; Laczi & Poser, 2024; Serrano, 2025; Soon et al., 2024)
3	Malware & ransomware (incl.	AI-based malware, ransomware, bring-	(Chinmaya et al., 2023; Ghazi-Tehrani & Pontell, 2021;

	IoT/BYOD)	your-own-device (BYOD) malware risk, and web-based malware.	Rawindaran et al., 2021; Riskhan et al., 2025; Wani et al., 2024)
4	Attacks against ML models (Adversarial ML)	Evasion or adversarial examples, poisoning, model stealing, and membership inference.	(Grosse et al., 2023)
5	Service disruption	DoS/DDoS and Informational DoS (IDoS), which exploits the limitations of human attention.	(Huang & Zhu, 2022; Rawindaran et al., 2021)
6	Automated/advanced AI-orchestrated attacks	APT, MITM, automated AI attacks	(Pratama et al., 2025; Serrano, 2025)

Table 5 further illustrates this categorization by presenting detailed examples of techniques and methods associated with each threat type. For instance, phishing threats encompass approaches such as general phishing, spear phishing, and AI-powered phishing, while content manipulation includes deepfake misuse and AI bots for deception. Similarly, AI-based malware, adversarial machine learning attacks, and service disruptions highlight the technical sophistication of these threats. By mapping these examples to their corresponding references, Table 5 not only validates the identified categories but also provides a comprehensive overview of how these threats manifest in practice across literature.

Social engineering & phishing (human-targeting)

Social engineering is a cyber-attack method that manipulates human psychology to deceive individuals into divulging confidential information or performing specific actions (Alahmed et al., 2024; Banire et al., 2021). The emergence of AI, particularly Large Language Models (LLMs), has significantly amplified this threat by enabling attackers to generate highly targeted, personalized, and automated attacks that are convincing and free of common errors like typos (Bethany et al., 2025). These AI-driven tools can create content that mimics human communication, making it easier for attackers to exploit trust and create a sense of urgency (Alahmed et al., 2024; Gallo et al., 2024).

Social Engineering and Phishing includes: 1) Lateral Phishing: This involves using a compromised internal email account to send phishing messages to other employees (Bethany et al., 2025). 2) Targeted & Personalized Content: AI used to craft highly individualized and persuasive messages. These can be tailored to specific individuals by leveraging publicly available information, such as job roles, organizational news, or current events like an upcoming solar eclipse, to make the phishing attempt more credible (Bethany et al., 2025; Gallo et al., 2024). 3) Impersonation and Credential Harvesting: A common tactic involves attackers masquerading as legitimate entities, such as a company's IT department or a supervisor, to gain unauthorized access to accounts and confidential data (Bethany et al., 2025). AI has made social engineering and phishing attacks more sophisticated and harder to detect. These attacks often leverage psychological manipulation and exploit internal trust structures.

Content manipulation & impersonation

AI-powered tools can generate highly convincing and personalized phishing messages, emails, and even deep-fake audio or video content that mimics a person's voice and appearance, making it easier for attackers to deceive individuals and bypass traditional security measures (Alahmed et al., 2024; Soon et al., 2024). AI-based attacks utilize machine learning to gather and analyze large volumes of public data, including social media posts and professional profiles. This information is then used to craft personalized phishing messages tailored to specific individuals or groups, a technique known as spear phishing (Alahmed et al., 2024; Soon et al., 2024). For instance, a common tactic involves an email seemingly from a supervisor to their direct report, leveraging the inherent trust and power dynamics of the relationship to compel the recipient to act (Bethany et al., 2025).

Malware & ransomware

Malware and ransomware are significant cyber threats that leverage vulnerabilities in digital systems. Ransomware is a specific type of malicious software, or malware, that holds a user's data or device hostage, typically through encryption, until a ransom is paid (Chinmaya et al., 2023). There are several categories of ransomware, including Crypto Ransomware (encrypts documents), Locker Ransomware (encrypts the entire system), Scareware (displays pop-up warnings), and Doxware (steals data and threatens publication) (Pratama et al., 2025). This is a broad category of software designed to cause damage or conduct malicious actions on a device without the user's knowledge (Gangone et al., 2023). These threats exploit vulnerabilities in devices and human psychology to achieve malicious objectives such as data theft, system compromise, and financial extortion (Alahmed et al., 2024).

Attacks against ML models

Adversarial Machine Learning (AML) refers to a category of cyber threats where attackers intentionally circumvent or exploit machine learning (ML) models. This is achieved by manipulating the data that the model processes, either during its training phase or when it is operational, to cause it to make incorrect predictions or classifications. These attacks can compromise a model's performance, leak private information from its training data, or allow an attacker to copy the model's intellectual property. Concerns about AML are significant, particularly in the IT security sector, where practitioners show heightened concern about threats like backdoors, evasion, and membership inference attacks (Grosse et al., 2023).

Service disruption

Service disruption in the context of AI-based cyber threats refers to incidents where critical system operations are interrupted or compromised due to malicious activities, often leveraging artificial intelligence. These disruptions can manifest as system downtime, unavailability of services, or degradation of performance, impacting an organization's ability to function. For example, Distributed Denial of Service (DDoS) attacks, a common form of service disruption, can overwhelm a system's resources, making it unavailable to legitimate users (Bayesh & Jahan, 2025). The failure of a single, seemingly minor component can trigger cascading effects across a network, leading to widespread operational failure (Bayesh & Jahan, 2025). Human-caused issues, such as IT bugs or infrastructure failures, can also lead to service disruptions (Setyawan et al., 2020).

Automated/advanced AI-orchestrated attacks

Automated and advanced AI-orchestrated attacks are sophisticated cyber threats where generative AI systems are used to create and execute attacks with

unprecedented scale, speed, and personalization. These attacks leverage machine learning and AI to mimic human communication, exploit psychological triggers, and bypass traditional security measures (Alahmed et al., 2024). Attackers use AI to automate the creation of convincing and tailored messages, emails, or even deepfakes, making it easier to manipulate individuals into divulging sensitive information or performing malicious actions (Alahmed et al., 2024). AI-powered attacks employ adversarial machine learning to subtly alter phishing content, making it difficult for traditional security systems to detect. These tactics exploit vulnerabilities in security algorithms to evade identification and prevention measures (Soon et al., 2024).

Strategies to Enhance Information Security Awareness Against AI-based Cyber Threats

To effectively counter AI-based cyber threats, organizations and stakeholders must adopt comprehensive strategies that strengthen information security awareness across multiple dimensions. As summarized in table 6, these strategies encompass seven key categories, each addressing different aspects of the challenge. Education and training programs remain central, emphasizing comprehensive user training, tailored curricula, and the integration of gamification and continuous learning methods. Complementing this, gamification and simulation-based learning, AI-driven simulations, and LLM-based training scenarios offer interactive and engaging approaches to enhance user resilience against phishing and social engineering. Policy, governance, and management support further ensure that security awareness is embedded within organizational structures through initiatives like BYOD frameworks, leadership engagement, and cost-effective governance practices. At the technical level, system-oriented controls such as AI-powered detection, anomaly monitoring, and adaptive authentication reinforce defenses while raising awareness of emerging risks. Equally significant is collaboration across academia, industry, government, and other stakeholders, including partnerships with financial institutions and social media platforms, to foster shared responsibility. Legal, regulatory, and ethical frameworks also play a vital role by establishing clear guidelines for issues such as deepfake misuse and generative AI governance. Finally, psychological and human-centric approaches address cognitive vulnerabilities, leverage motivational models, and implement reinforcement mechanisms to encourage lasting secure behavior. Taken together, these strategies highlight that enhancing security awareness against AI-enabled threats requires an integrated approach that combines education, governance, technical safeguards, collaboration, regulation, and psychological engagement.

Education & Training Programs

Educational and training programs are crucial strategies for mitigating the risks associated with phishing attacks, though their effectiveness varies based on design and implementation. Many organizations utilize security awareness training, which can include generic guidelines, induction training, online cybersecurity modules, phishing simulations, and periodic training events (Wani et al., 2024). However, generic training often has a minimal impact on improving clinician awareness and user behavior, as different clinical specialties have unique workflows and requirements. Research shows that people can be trained to recognize phishing attempts, with education being particularly effective when it emphasizes conceptual knowledge and provides direct feedback (Ghazi-Tehrani & Pontell, 2021). Gamification has emerged as a promising approach to enhance the effectiveness of these programs. Gamified e-training systems that incorporate elements like challenges, clear goals, feedback, and narrative context can increase employee awareness and their perception of having the ability to

perform appropriate security behaviors (Bitrián et al., 2024). Such systems have been shown to improve employees' actual security behaviors, leading to a significant decrease in the percentage of clicks on fake links in phishing simulations and an increase in the reporting of suspicious emails.

Another method involves regular simulated phishing campaigns that target staff to measure training effectiveness, with those who fall for the attempt receiving additional training (Ghazi-Tehrani & Pontell, 2021). While some suggest that adding an element of public shaming to these tests could improve outcomes, the morality of such a solution is questionable. Despite these efforts, the adequacy of current cybersecurity education is often questioned, with many participants in one study believing it to be inadequate. The most common delivery method for ransomware is phishing, underscoring the importance of effective training. The sophistication of phishing attacks, particularly with the use of LLMs, necessitates that training programs evolve. These programs should educate employees to discern subtle cues beyond simple grammatical errors, such as the overall context of the email and the nature of the request (Bethany et al., 2025). Ultimately, a consensus exists that proper training, including awareness of what phishing is and how to recognize it, could reduce the vast majority of successful phishing attacks.

Table 6. Strategic Approaches for Strengthening Information Security Awareness

No	Strategy Category	Example Strategies	References
1	Education & Training Programs	Comprehensive training, user education programs, curriculum integration, tailored training, gamification, SAEC (Security Awareness Education Campaigns), e-learning, continuous training, opinion-dynamics based SAT	(Angafor et al., 2024; Baltuttis & Teubner, 2024; Bethany et al., 2025; Bitrián et al., 2024; Calvo et al., 2025; Chinmaya et al., 2023; Gallo et al., 2024; Ghazi-Tehrani & Pontell, 2021; Grosse et al., 2023; Is, 2024; Jagadeesan et al., 2024; Kruger & Kearney, 2006; Park & Kim, 2025; Pratama et al., 2025; Qin et al., 2025; Rawindaran et al., 2021; Riskhan et al., 2025; Soon et al., 2024; Stylianou et al., 2025; Tawalbeh & Muheidat, 2023; Tinubu et al., 2023; Wani et al., 2024; Yoo & Cho, 2022)
2	Gamification & Simulation-based Learning	Game-based learning (PHISHGEM, intelligent game for tweens), simulated phishing, penetration testing, AI-driven simulations, interactive scenarios, LLM-based training scenarios	(Angafor et al., 2024; Baltuttis & Teubner, 2024; Bethany et al., 2025; Bitrián et al., 2024; Brilingaitė et al., 2025; Is, 2024; Luh et al., 2025; Park & Kim, 2025; Pratama et al., 2025; Riskhan et al., 2025; Tinubu et al., 2023; Zhang et al., 2025)
3	Policy, Governance & Management Support	BYOD framework, security culture promotion, organizational effort, management incentives, policy	(Alahmed et al., 2024; Angafor et al., 2024; Baltuttis & Teubner, 2024; Banire et al., 2021; Qin et al., 2025; Rawindaran et al., 2021; Stylianou et al., 2025; Tawalbeh & Muheidat, 2023; Wani et al., 2024)

		alignment, cost-effective SAT policies, leadership support	
4	Technical & System-level Controls	Visual risk indicators, UEBA (User and Entity Behavior Analytics), email tagging, anomaly detection, AI-based security assistants, adaptive authentication, alert management (RADAMS)	(Baltuttis & Teubner, 2024; Bethany et al., 2025; Calvo et al., 2025; Huang & Zhu, 2022; Is, 2024; Soon et al., 2024; Wani et al., 2024; Yoo & Cho, 2022)
5	Collaboration & Multi-stakeholder Engagement	Collaboration academia–industry–government, financial institutions partnership, social media/platform responsibility, responsive institutions	(Alahmed et al., 2024; Banire et al., 2021; Serrano, 2025; Soon et al., 2024; Wani et al., 2024)
6	Legal, Regulatory & Ethical Frameworks	Legal references in training, regulatory frameworks for deepfake misuse, ethical guidelines, generative AI governance	(Alahmed et al., 2024; Laczi & Poser, 2024; Qin et al., 2025; Riskhan et al., 2025)
7	Psychological & Human-centric Approaches	Addressing cognitive vulnerabilities, behavioral psychology, motivational models (PMT), reinforcement through authority/urgency simulations, reward–punishment strategy	(Bitrián et al., 2024; Gallo et al., 2024; Riskhan et al., 2025; Stylianou et al., 2025; Tawalbeh & Muheidat, 2023)

Gamification & Simulation-based Learning

Gamification and simulation-based learning are increasingly recognized as effective pedagogical tools for enhancing cybersecurity awareness and training across various age groups. These methods leverage interactive and immersive elements to make learning complex concepts more engaging and memorable than traditional approaches (Bitrián et al., 2024; Tinubu et al., 2023). Gamification applies game mechanics like points, challenges, and leaderboards to non-game contexts to foster psychological outcomes such as enjoyment and satisfaction, ultimately encouraging desired behaviors (Bitrián et al., 2024). Simulation-based learning, particularly through serious games, allows participants to actively engage with security concepts in a dynamic, risk-free environment, practicing decision-making and exploring the consequences of their actions firsthand (Luh et al., 2025). Research shows that gamified e-training can significantly improve employees perceptions of information quality and system quality, which in turn boosts perceived usefulness and satisfaction

(Bitrián et al., 2024). These positive perceptions enhance security self-efficacy an employee's belief in their ability to perform security-related tasks (Bitrián et al., 2024).

For younger users, such as tweens, intelligent game-based simulations that mimic real-world threats like spam and malware through interactive pop-ups have proven effective. These systems often use a reward-punishment strategy, where safe online behavior is rewarded with virtual incentives and unsafe actions trigger simulated attacks followed by educational feedback (Riskhan et al., 2025). Studies have demonstrated that these methods not only increase knowledge retention but also enhance motivation and engagement (Riskhan et al., 2025; Tinubu et al., 2023). For example, a mobile game called PHISHGEM educates users on various phishing techniques through a five-level structure, demonstrating high rates of user satisfaction and knowledge acquisition (Tinubu et al., 2023). Similarly, desktop-based simulations for tweens have shown a significant increase in cybersecurity awareness after just one month of use (Riskhan et al., 2025). Both gamification and simulation-based learning offer powerful, scalable, and adaptable frameworks for cybersecurity education. By creating interactive and engaging learning environments that mimic real-world scenarios, these approaches effectively improve security behaviors, enhance user knowledge, and boost confidence in dealing with cyber threats across different demographics, from corporate employees to young children.

Policy, Governance & Management Support

Effective cybersecurity hinges robust policy, governance, and management support, which together create a structured and secure operational environment. Policies must be dedicated and specific, rather than generic, to provide clear guidance and avoid ambiguity for users (Rawindaran et al., 2021; Wani et al., 2024). For instance, in a Bring-Your-Own-Device (BYOD) context, many clinicians report a lack of dedicated BYOD policies from their hospitals, leading to unsafe practices (Wani et al., 2024). Furthermore, policies and security processes must align with the practical workflow requirements of staff; otherwise, they risk being perceived as too strict or impractical and are often circumvented (Wani et al., 2024). Management support is crucial for reinforcing security behaviors, particularly when direct control over user activities is limited, such as with personal devices. This support can manifest through mandating regular, practical training, allocating dedicated time for it, and incentivizing participation to foster a positive security culture (Wani et al., 2024).

Governance structures, such as national cybersecurity centers, play a vital role by providing guidelines and a single point of contact for organizations, particularly Small and Medium Enterprises (SMEs) (Rawindaran et al., 2021). However, a significant barrier for SMEs is the complexity and cost of implementing industry standards, which are often designed with large organizations in mind (Rawindaran et al., 2021). This disparity highlights a need for non-traditional, tailored solutions and government support through grants and subsidies to bridge the funding gap (Rawindaran et al., 2021). A partnership culture between technical departments and users is also essential, where the requirements of both are acknowledged to maintain a balance between security and usability (Wani et al., 2024). This collaborative approach, combined with strong policies and active management, is key to proper implementation and enforcement of security measures (Wani et al., 2024).

Technical & System-level Controls

Technical and system-level controls are essential for protecting

information systems and can be categorized into detection, prevention, and response mechanisms, which together form the three pillars of information security defense (Luh et al., 2025). Detection mechanisms, such as Security Information and Event Management (SIEM) systems and supervisory computers, monitor log files and physical readings to generate alerts containing device-level information that is later mapped to system-level metrics through alert triage, thereby supporting timely decision-making (Huang & Zhu, 2022). Additional tools, including detonation chambers, API monitors, and intrusion detection systems (IDS), enhance threat visibility by analyzing files, messages, traffic, and user behavior (Luh et al., 2025). Preventive mechanisms are designed to block attacks before they take place, utilizing methods such as URL scanning, email reputation assessment, and sender authentication (Luh et al., 2025). Advanced defense tools such as intrusion prevention systems (IPS), packet filters, and honeypots serve to further enhance the overall resilience of a network. In the context of email security, protocols such as SPF, DKIM, and DMARC are used to prevent message spoofing and misuse; however, their effectiveness largely depends on correct configuration and enforcement by both the sending and receiving mail servers (Gallo et al., 2024). Hardening practices that focus on securing systems, applications, and networks add an extra layer of preventive defense against potential attacks (Luh et al., 2025). When an attack manages to bypass existing defenses, response mechanisms become critical. Steps such as isolating affected systems, removing intruders, and restoring data or assets from backups are essential for minimizing the overall impact and preventing further compromise. For instance, restoring data from backups can significantly lessen the impact of ransomware attacks by recovering encrypted systems and minimizing downtime. More sophisticated approaches, such as the Resilient and Adaptive Data-Driven Alert and Attention Management Strategy (RADAMS), enhance operator performance by prioritizing alerts, reducing cognitive workload, and improving the accuracy of incident responses (Huang & Zhu, 2022). Effective information security depends on the seamless integration of detection, prevention, and response mechanisms. Together, tools such as SIEM-based monitoring, preventive email authentication, and adaptive frameworks like RADAMS create a multilayered defense that strengthens organizational resilience against increasingly sophisticated cyber threats.

Collaboration & Multi-stakeholder Engagement

One of the studies reviewed, which examined lateral phishing involving Large Language Models (LLMs), highlighted the value of a collaborative, multi-stakeholder framework within a large university environment. The research was carried out in coordination with the university's cyber operations team, which oversaw the phishing awareness infrastructure, maintained the anonymization of personally identifiable information (PII), and provided both quantitative and qualitative datasets for analysis (Bethany et al., 2025). Ethical oversight was ensured through IT governance approval and a trusted agents committee, which included representatives from risk management, legal affairs, academic operations, human resources, and student affairs. This committee reviewed all research materials and templates to reduce potential risks and ensure responsible implementation (Bethany et al., 2025). This coordinated approach demonstrates how research expertise, operational capabilities, and institutional oversight can work together to enable large-scale phishing simulations that are both effective and ethically sound.

Legal, Regulatory & Ethical Frameworks

Legal, regulatory, and ethical frameworks play a vital role in addressing

emerging cybersecurity threats such as the misuse of genetic information, deepfake manipulation, and advanced phishing campaigns. These frameworks are designed to safeguard privacy, ensure informed consent, prevent discrimination, and promote the responsible use of new technologies, thereby protecting both individuals and organizations from potential harm (Brilingaitė et al., 2025; Laczi & Poser, 2024; Luh et al., 2025).

When genetic data are used to study or better understand the cybersecurity workforce, important ethical challenges emerge. Mishandling such data can lead to discrimination or create conditions that expose certain groups to increased vulnerability and bias (Brilingaitė et al., 2025). To minimize these risks, a strong ethical framework is essential. Such a framework should ensure that any form of genetic screening in the workplace complies with legal, medical, and ethical standards, emphasizing the importance of obtaining informed consent and protecting individual rights. Informed consent serves as a fundamental safeguard, requiring that individuals receive clear and comprehensive information about the potential risks, benefits, and implications of genetic testing including who will have access to the data and how it will be used. Existing regulations, such as the Genetic Information Nondiscrimination Act (GINA) in the United States and the Working Document on Genetic Data in the European Union, prohibit the use of genetic information as a basis for discrimination in areas such as health insurance and employment. Regulatory oversight bodies play a crucial role in ensuring compliance and in responding effectively to any violations of privacy or confidentiality (Brilingaitė et al., 2025).

With respect to deepfake technology, the potential for misuse and harm calls for a combination of legal and educational responses. The rapid circulation of sexually explicit or deceptive content often involving minors underscores the urgency of establishing a global regulatory framework capable of addressing, deterring, and penalizing the creation and dissemination of harmful deepfakes (Laczi & Poser, 2024). Legislation should promote ethical practices in both media consumption and content creation, supported by accessible reporting mechanisms that allow users to flag suspicious or inappropriate material (Laczi & Poser, 2024). The ethical debate surrounding synthetic media has also gained attention, as illustrated by a South Korean court case that recognized the creation of abusive images involving virtual children not only real individuals as a punishable offense. In the context of phishing simulations used for employee training, ethical concerns center on safeguarding employee privacy and reducing potential psychological harm. Such exercises must adhere to relevant privacy laws and regulations, including the CCPA, FERPA, and TSPA, to ensure that training practices remain both effective and respectful of individual rights (Bethany et al., 2025). Tackling the complex challenges arising from rapid advancements in cybersecurity demands a multifaceted approach that integrates robust legal and regulatory oversight with a strong commitment to ethical principles. Frameworks such as GINA and the GDPR, supported by institutional review boards and ethics committees, establish the foundation for conducting responsible research and ensuring the ethical application of technology. Ultimately, the aim is to strike a balance between technological innovation and the protection of fundamental human rights, including privacy, autonomy, and overall well-being.

Psychological & Human-centric Approaches

Recent academic work indicates a growing shift away from purely technical cybersecurity solutions toward more integrative, human-centered approaches that take psychological and behavioral factors into account. These approaches aim to better understand and reduce threats such as phishing by

examining human behavior, cognitive biases, and decision-making patterns rather than depending exclusively on automated defenses. This perspective acknowledges that while human users often represent the weakest point in the security chain, they can also be empowered to serve as an active and effective line of defense (Baltuttis & Teubner, 2024). Human-centered approaches to phishing mitigation emphasize improving users' awareness and ability to recognize suspicious emails. These approaches involve examining the human factors that influence susceptibility to phishing, including socio-demographic characteristics, personality traits, and contextual conditions. Theoretical frameworks such as the Heuristic-Systematic Model (HSM) and Protection Motivation Theory (PMT) are often applied to explain how individuals interpret phishing messages and decide how to respond to potential threats (Baltuttis & Teubner, 2024).

An essential element of the human-centered paradigm is the integration of technology designed to assist rather than replace human decision-making. This support is often implemented through explicit cues, such as visual risk indicators or contextual warning messages embedded within email interfaces, which help users recognize and respond appropriately to potential threats (Baltuttis & Teubner, 2024). These tools offer users technology-assisted risk assessments while leaving the ultimate decision in their hands a principle often referred to as user empowerment.

Contemporary discussions in cybersecurity increasingly promote a hybrid approach that combines technological innovation with insights from human psychology. By examining cognitive vulnerabilities, decision-making patterns, and the behavioral traits of both end users and security professionals, human-centered strategies seek to develop more adaptive and resilient defenses against threats such as phishing.

The discussion section explores the broader implications of the findings by addressing the research gaps revealed in the reviewed studies. It emphasizes two key areas of concern. First, it identifies several underexplored dimensions within AI-driven cyber threat categories that require deeper examination. Second, it highlights the limitations of existing security awareness strategies, which hinder their effectiveness in responding to the fast-evolving landscape of cyber threats.

Research Gaps in AI-based Cyber Threat Categories

The rapid emergence of artificial intelligence has introduced a variety of cyber threats, each targeting different aspects of digital systems. This subchapter focuses on identifying the research gaps in AI-based cyber threat categories, emphasizing which threats have been extensively explored and which remain underrepresented in the current body of literature. By highlighting these gaps, the discussion aims to provide direction for future studies that can address overlooked yet critical dimensions of AI-enabled attacks.

Table 7. Heatmap Of AI-Based Cyber Threat Type

AI-Based Cyber Threat	Frequency
Social engineering & phishing (human-targeting)	21
Malware & ransomware (incl. IoT/BYOD)	5
Content manipulation & impersonation	4
Service disruption	2
Automated/advanced AI-orchestrated attacks	2
Attacks against ML models (Adversarial ML)	1

As shown in the table 7, the majority of studies concentrate on social

engineering and phishing (21 occurrences), demonstrating that human-targeted attacks remain the most researched domain in AI-related security. In contrast, categories such as malware and ransomware (5), content manipulation and impersonation (4), and especially adversarial attacks on machine learning models (1) receive limited attention. This imbalance indicates a critical gap, particularly regarding the resilience of AI and machine learning models themselves, as well as the evolving nature of automated AI-orchestrated attacks. Addressing these underexplored areas is essential to achieving a more comprehensive understanding of the threat landscape and guiding the development of balanced security awareness initiatives.

Research Gaps in Security Awareness Strategies

While numerous strategies have been proposed to enhance security awareness against AI-based cyber threats, not all approaches receive equal attention in the literature. This subchapter examines the distribution of strategies across existing studies, identifying areas that are heavily researched and those that remain underexplored. Such analysis is critical for revealing imbalances in current research efforts and highlighting opportunities for advancing a more comprehensive and multidimensional approach to security awareness.

Table 8. Heatmap Of Security Awareness Strategy

Security Awareness Strategy	Frequency
Education & Training Programs	23
Gamification & Simulation-based Learning	12
Policy, Governance & Management Support	9
Technical & System-level Controls	8
Collaboration & Multi-stakeholder Engagement	5
Psychological & Human-centric Approaches	5
Legal, Regulatory & Ethical Frameworks	4

Table 8 demonstrates that education and training programs dominate the discussion (23 occurrences), underscoring their central role in shaping awareness. Similarly, gamification and simulation-based learning (12) and policy, governance, and management support (9) have attracted notable attention. However, strategies such as technical and system-level controls (8), collaboration and multi-stakeholder engagement (5), psychological and human-centric approaches (5), and especially legal, regulatory, and ethical frameworks (4) are comparatively less emphasized. This uneven focus suggests that while skill development and awareness training are well studied, broader systemic, legal, and human-centered dimensions are not yet sufficiently explored. Addressing these gaps can strengthen the resilience of security awareness initiatives by integrating organizational, regulatory, and psychological perspectives alongside technical training.

CONCLUSION

This study addresses its research objectives by systematically mapping the landscape of AI-based cyber threats and identifying strategies to enhance information security awareness in response to those threats. The findings indicate that the growing sophistication of AI-enabled attacks necessitates a shift from fragmented awareness efforts toward integrated, multidimensional security awareness programs that combine education, governance, technical safeguards, and human-centered interventions. In practical terms, organizations and policymakers can apply these insights by designing awareness initiatives that go beyond conventional training, incorporating

simulation-based learning, cross-sector collaboration, and policy alignment with emerging AI risks. Furthermore, the identified research gaps highlight the need for future studies to focus on underexplored AI-driven attack vectors and the effectiveness of advanced technical, psychological, and regulatory measures. By aligning awareness strategies with the evolving nature of AI-enabled cyber threats, this research provides a foundation for developing more adaptive, sustainable, and context-aware cybersecurity awareness frameworks.

REFERENCES

Alahmed, Y., Abadla, R., & Ansari, M. J. Al. (2024). Exploring the Potential Implications of AI-generated Content in Social Engineering Attacks. *2024 International Conference on Multimedia Computing, Networking and Applications, MCNA 2024*, 64–73. https://doi.org/10.1109/MCNA63144.2024.10703950

Alanezi, M., & Al-Azzawi, R. M. A. (2024). AI-Powered Cyber Threats: A Systematic Review. *Mesopotamian Journal of CyberSecurity*, *4*(3), 166–188. https://doi.org/10.58496/MJCS/2024/021

Almass, S., & Chowdhary, S. K. (2024). Comprehensive Study on Cyber Security and Cyber Attacks. *Proceedings - 1st International Conference on Electronics, Communication and Signal Processing, ICECSP 2024*, 1–6. https://doi.org/10.1109/ICECSP61809.2024.10698540

Angafor, G. N., Yevseyeva, I., & Maglaras, L. (2024). Securing the remote office: reducing cyber risks to remote working through regular security awareness education campaigns. *International Journal of Information Security*, *23*(3), 1679–1693. https://doi.org/10.1007/s10207-023-00809-5

Aryasutha, R., Azizah Ria Kusrini, N., Nurul Ulya, J., & Syamsiah Septiani, N. (2025). Opportunities and Challenges for Islamic Education Teachers in Using Artificial Intelligence in Learning. *Muaddib.Intischolar.Id*, *2*(1), 43. https://muaddib.intischolar.id/index.php/muaddib/article/view/6

Badan Siber dan Sandi Negara. (2024). Lanskap Keamanan Siber Indonesia. In *Id-SIRTII /CC* (Vol. 70, Issue 70, pp. 1–107). Id-SIRTII /CC. bit.ly/44bzpHM

Baltuttis, D., & Teubner, T. (2024). Effects of visual risk indicators on phishing detection behavior: An eye-tracking experiment. *Computers and Security*, *144*. https://doi.org/10.1016/j.cose.2024.103940

Banire, B., Al Thani, D., & Yang, Y. (2021). Investigating the experience of social engineering victims: Exploratory and user testing study. *Electronics (Switzerland)*, *10*(21). https://doi.org/10.3390/electronics10212709

Bayesh, M., & Jahan, S. (2025). Embedding Security Awareness in IoT Systems: A Framework for Providing Change Impact Insights. *Applied Sciences (Switzerland)*, *15*(14). https://doi.org/10.3390/app15147871

Bethany, M., Galiopoulos, A., Bethany, E., Bahrami Karkevandi, M., Beebe, N., Vishwamitra, N., & Najafirad, P. (2025). Lateral Phishing With Large Language Models: A Large Organization Comparative Study. *IEEE Access*, *13*, 60684–60701. https://doi.org/10.1109/ACCESS.2025.3555500

Bitrián, P., Buil, I., Catalán, S., & Merli, D. (2024). Gamification in workforce training: Improving employees' self-efficacy and information security and data protection behaviours. *Journal of Business Research*, *179*. https://doi.org/10.1016/j.jbusres.2024.114685

Brilingaitė, A., Bukauskas, L., Domarkienė, I., Rančelis, T., Ambrozaitytė, L., Pirta, R., Lugo, R. G., & Knox, B. J. (2025). Towards projection of the individualised risk assessment for the cybersecurity workforce. *Computer*

Standards and Interfaces, *93*. https://doi.org/10.1016/j.csi.2024.103962

Calvo, A., Escuder, S., Ortiz, N., Escrig, J., & Compastié, M. (2025). RBD24 : A labelled dataset with risk activities using log application data. *Computers and Security*, *150*. https://doi.org/10.1016/j.cose.2024.104290

Chinmaya, B. J., Kudtarkar, S. A., & Mohana. (2023). Targeted Ransomware Attacks and Detection to Strengthen Cybersecurity Strategies. *2nd International Conference on Automation, Computing and Renewable Systems, ICACRS 2023 - Proceedings*, 1039–1044. https://doi.org/10.1109/ICACRS58579.2023.10404203

Chua, H. N., Teh, J. S., & Herbland, A. (2021). Identifying the Effect of Data Breach Publicity on Information Security Awareness Using Hierarchical Regression. *IEEE Access*, *9*, 121759–121770. https://doi.org/10.1109/ACCESS.2021.3107426

Damri, D., Amalia, R., Engkizar, E., Efendi, E., Ramadhani, R., & Asril, Z. (2023). Improving Students' Dyscalculia Numeracy Ability Using Learning Media Colored Bead Montessori. *Al-Ta Lim Journal*, *30*(2), 116–124. https://doi.org/10.15548/jt.v30i2.751

Edim, B. E., Udofot, A. I., & Omotosho M. O. (2025). AI-augmented cyber security threat intelligence – enhancing situational awareness. *International Journal of Science and Research Archive*, *14*(1), 890–897. https://doi.org/10.30574/ijsra.2025.14.1.2650

Engkizar, E., Jaafar, A., Alias, M. F. B., Guspita, R., & Albizar, A. (2025). Utilisation of Artificial Intelligence in Qur'anic Learning: Innovation or Threat? *Journal of Quranic Teaching and Learning*, *1*(2), 1–17. https://joqer.intischolar.id/index.php/joqer/index

Engkizar, E, Jaafar, A., Sarianto, D., Ayad, N., Rahman, A., Febriani, A., Oktavia, G., Puspita, R., & Rahman, I. (2024). Analysis of Quran Education Problems in Majority Muslim Countries. *International Journal of Islamic Studies Higher Education*, *3*(1), 65–80. https://doi.org/10.24036/insight.v3i1.209

Engkizar, Engkizar, Muliati, I., Rahman, R., & Alfurqan, A. (2018). The Importance of Integrating ICT Into Islamic Study Teaching and Learning Process. *Khalifa: Journal of Islamic Education*, *1*(2), 148. https://doi.org/10.24036/kjie.v1i2.11

Fortinet. (2024). *Security Awareness and Training Global Research Report*. https://www.fortinet.com/content/dam/fortinet/assets/reports/report-2024-security-awareness-and-training.pdf

Gallo, L., Gentile, D., Ruggiero, S., Botta, A., & Ventre, G. (2024). The human factor in phishing: Collecting and analyzing user behavior when reading emails. *Computers and Security*, *139*. https://doi.org/10.1016/j.cose.2023.103671

Gangone, A., Bala, B., Gangone, S., & Bharat Kumar, G. J. (2023). The Deep Learning and Machine Learning Methods for Botnet Identification in the Internet of Things. *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2023*, 435–441. https://doi.org/10.1109/IC3I59117.2023.10397881

Ghazi-Tehrani, A. K., & Pontell, H. N. (2021). Phishing Evolves: Analyzing the Enduring Cybercrime. *Victims and Offenders*, *16*(3), 316–342. https://doi.org/10.1080/15564886.2020.1829224

Ghuge, G. D. (2024). The AI-Powered Attack Vector: Evaluating the Potential Impact and Feasibility of AI-Generated Cyber Threats. *International Journal of Advanced Research in Science, Communication and Technology*, 226–231. https://doi.org/10.48175/ijarsct-22631

Grosse, K., Bieringer, L., Besold, T. R., Biggio, B., & Krombholz, K. (2023). Machine Learning Security in Industry: A Quantitative Survey. *IEEE Transactions on Information Forensics and Security*, *18*, 1749–1762. https://doi.org/10.1109/TIFS.2023.3251842

Guembe, B., Azeta, A., Misra, S., Osamor, V. C., Fernandez-Sanz, L., & Pospelova, V. (2022). The Emerging Threat of Ai-driven Cyber Attacks: A Review. *Applied Artificial Intelligence*, *36*(1). https://doi.org/10.1080/08839514.2022.2037254

Hermawan, A. Z., Anggoro, M. N., Lozera, D., & Faroqi, A. (2023). Studi Literatur: Ancaman Serangan Siber Artificial Intelligence (Ai) Terhadap Keamanan Data Di Indonesia. *Prosiding Seminar Nasional Teknologi Dan Sistem Informasi*, *3*(1), 581–591. https://doi.org/10.33005/sitasi.v3i1.363

Hu, S., Hsu, C., & Zhou, Z. (2022). Security Education, Training, and Awareness Programs: Literature Review. *Journal of Computer Information Systems*, *62*(4), 752–764. https://doi.org/10.1080/08874417.2021.1913671

Huang, L., & Zhu, Q. (2022). RADAMS: Resilient and adaptive alert and attention management strategy against Informational Denial-of-Service (IDoS) attacks. *Computers and Security*, *121*, 102844. https://doi.org/10.1016/j.cose.2022.102844

Is, H. (2024). LLM-Driven SAT Impact on Phishing Defense: A Cross-Sectional Analysis. *12th International Symposium on Digital Forensics and Security, ISDFS 2024*. https://doi.org/10.1109/ISDFS60797.2024.10527274

Iskandar, M. Y., Bentri, A., Hendri, N., Engkizar, E., & Efendi, E. (2023). Integrasi Multimedia Interaktif Berbasis Android dalam Pembelajaran Agama Islam di Sekolah Dasar. *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, *7*(4), 4575–4584. https://doi.org/10.31004/obsesi.v7i4.5021

Jagadeesan, S., Sameer, Singh, D., Ojha, R., Ibrahim, R. K., & Alazzam, M. B. (2024). *Implementation of an Artificial Intelligence with Cyber Security in E-Learning-Based Education Management System* (pp. 01–05). https://doi.org/10.1109/iccakm58659.2023.10449611

Jimmy, F. (2021). Emerging Threats: The Latest Cybersecurity Risks and the Role of Artificial Intelligence in Enhancing Cybersecurity Defenses. *International Journal of Scientific Research and Management (IJSRM)*, *9*(02), 564–574. https://doi.org/10.18535/ijsrm/v9i2.ec01

Kassymova, G. K., Talgatov, Y. K., Arpentieva, M. R., Abishev, A. R., & Menshikov, P. V. (2025). Artificial Intelligence in the Development of the Theory and Practices of Self-Directed Learning. *Multidisciplinary Journal of Thought and Research*, *1*(3), 66–79. https://mujoter.intischolar.id/index.php/mujoter/article/view/19

Kovaci, P.-D. (2024). Threat Actors Seeking To Exploit Ai Capabilities. Types and Their Goals. *Strategic Impact*, *89*(4), 53–63. https://doi.org/10.53477/1842-9904-23-21

Kruger, H. A., & Kearney, W. D. (2006). A prototype for assessing information security awareness. *Computers and Security*, *25*(4), 289–296. https://doi.org/10.1016/j.cose.2006.02.008

Laczi, S. A., & Poser, V. (2024). Impact of Deepfake Technology on Children: Risks and Consequences. *SISY 2024 - IEEE 22nd International Symposium on Intelligent Systems and Informatics, Proceedings*, 215–220. https://doi.org/10.1109/SISY62279.2024.10737593

Li, Y., & Liu, Q. (2021). A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments. *Energy Reports*,

7, 8176–8186. https://doi.org/10.1016/j.egyr.2021.08.126

Luh, R., Eresheim, S., Tavolato, P., Petelin, T., Gmeiner, S., Holzinger, A., & Schrittwieser, S. (2025). Gamifying information security: Adversarial risk exploration for IT/OT infrastructures. *Computers and Security*, *151*. https://doi.org/10.1016/j.cose.2024.104287

Masoud, M., & Almajri, S. (2025). The use of artificial intelligence-based translation tools for language department students. *Journal of Arabic Literature, Teaching and Learning*, *1*(3), 76–92. https://jaliter.intischolar.id/index.php/jaliter

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J. A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J. J., Devereaux, P. J., Dickersin, K., Egger, M., Ernst, E., Gøtzsche, P. C., … Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7). https://doi.org/10.1371/journal.pmed.1000097

Okdem, S., & Okdem, S. (2024). Artificial Intelligence in Cybersecurity: A Review and a Case Study. *Applied Sciences (Switzerland)*, *14*(22), 10487. https://doi.org/10.3390/app142210487

Onwubiko, C., & Ouazzane, K. (2022). Multidimensional Cybersecurity Framework for Strategic Foresight. *International Journal on Cyber Situational Awareness*, *6*(1), 46–77. https://doi.org/10.22619/ijcsa.2021.100137

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Bmj*, *372*. https://doi.org/10.1136/bmj.n71

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., Mcdonald, S., … Mckenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. In *The BMJ* (Vol. 372). https://doi.org/10.1136/bmj.n160

Park, J. Y., & Kim, T. S. (2025). An Automated Scenario Generation Model for Anti-phishing using Generative AI. *Proceedings of the IEEE International Conference on Big Data and Smart Computing, BIGCOMP*, *2025*, 368–370. https://doi.org/10.1109/BigComp64353.2025.00073

Parsons, K., McCormac, A., Butavicius, M., Pattinson, M., & Jerram, C. (2014). Determining employee awareness using the Human Aspects of Information Security Questionnaire (HAIS-Q). *Computers and Security*, *42*, 165–176. https://doi.org/10.1016/j.cose.2013.12.003

Pratama, I. P. A. E., Widyantara, I. M. O., Linawati, Gunantara, N., Suakanto, S., & Irawan, E. T. (2025). Technology-Enhanced Learning for User Security Awareness Using AI-based Naive RAG: A Design and Prototype. *ICADEIS 2025 - 2025 International Conference on Advancement in Data Science, E-Learning and Information System: Integrating Data Science and Information System, Proceeding*. https://doi.org/10.1109/ICADEIS65852.2025.10933283

pratama, R. Y., Muhith, A., & Turmudi, I. (2025). Advancing Educational Practices: Implementing Think-Pair-Share to Achieve Learning Achievement in Islamic Education. *International Journal of Islamic Studies*

Higher Education, *4*(1), 59–67. https://doi.org/10.24036/insight.v4i1.219

Qin, Y., Yang, X., Yang, L. X., & Huang, K. (2025). Mitigating Social Engineering Attacks Through Cost-Effective Security Awareness Training Policy. *IEEE Transactions on Network Science and Engineering*, *12*(4), 3145–3158. https://doi.org/10.1109/TNSE.2025.3556927

Rawindaran, N., Jayal, A., & Prakash, E. (2021). Machine learning cybersecurity adoption in small and medium enterprises in developed countries. *Computers*, *10*(11), 150. https://doi.org/10.3390/computers10110150

Reddem, P. R. (2024). The Rise of AI-powered Cybercrime: a Data-driven Analysis of Emerging Threats. *International Journal For Multidisciplinary Research*, *6*(6), 1–12. https://doi.org/10.36948/ijfmr.2024.v06i06.30744

Riskhan, B., Saleem, M. T., Hussain, K., Jhanjh, N. Z., & Abul Hasanat, M. H. (2025). Intelligent Game-Based Simulation in Cyber Security Awareness for Tweens/Preadolescents. *ICoICC 2025 - 3rd International Conference on Intelligent and Cloud Computing*. https://doi.org/10.1109/ICoICC64033.2025.11052071

Sauer, P. C., & Seuring, S. (2023). How to conduct systematic literature reviews in management research: a guide in 6 steps and 14 decisions. *Review of Managerial Science*, *17*(5), 1899–1933. https://doi.org/10.1007/s11846-023-00668-3

Schiliro Francesco. (2023). Towards a Contemporary Definition of Cybersecurity. In *Towards a Contemporary Definition of Cybersecurity* (pp. 1–18). https://www.researchgate.net/publication/368304689

Serrano, W. (2025). CyberAIBot: Artificial Intelligence in an intrusion detection system for CyberSecurity in the IoT. *Future Generation Computer Systems*, *166*. https://doi.org/10.1016/j.future.2024.107543

Setyawan, A., Giri Sucahyo, Y., & Gandhi, A. (2020). Design of disaster recovery plan: State university in indonesia. *2020 5th International Conference on Informatics and Computing, ICIC 2020*. https://doi.org/10.1109/ICIC50835.2020.9288543

Sharma, S. K. (2024). AI-Enhanced Cyber Threat Detection and Response Systems. *Shodh Sagar Journal of Artificial Intelligence and Machine Learning*, *1*(2), 43–48. https://doi.org/10.36676/ssjaiml.v1.i2.14

Soon, J. N. P., Chan, R. Q., Lee, Q. H., En Loke, D., Chun, S. L. H., & Yuen, P. K. (2024). User perceptions of artificial intelligence powered phishing attacks on Facebook's resilient infrastructure. *International Journal of Advances in Applied Sciences*, *13*(4), 878–886. https://doi.org/10.11591/ijaas.v13.i4.pp878-886

Stylianou, I., Bountakas, P., Zarras, A., & Xenakis, C. (2025). Suspicious minds: Psychological techniques correlated with online phishing attacks. *Computers in Human Behavior Reports*, *19*. https://doi.org/10.1016/j.chbr.2025.100694

Subhani, A., Khan, I. A., & Ahmad, U. (2023). Importance of Conducting Cyber Security Awareness Sessions among Undergraduate Students. *Journal of Advanced Research in Social Sciences and Humanities*, *8*(2). https://doi.org/10.26500/jarssh-08-2023-0202

Tawalbeh, L. A., & Muheidat, F. (2023). Factors that Motivate Defense Against Social Engineering Attacks Across Organizations. *Procedia Computer Science*, *224*, 75–82. https://doi.org/10.1016/j.procs.2023.09.013

Thanh, C. T., & Zelinka, I. (2019). A survey on artificial intelligence in malware as next–generation threats. *Mendel*, *25*(2), 27–34. https://doi.org/10.13164/mendel.2019.2.027

Tinubu, C. O., Falana, O. J., Oluwumi, E. O., Sodiya, A. S., & Rufai, S. A. (2023). PHISHGEM: a mobile game-based learning for phishing awareness. *Journal of Cyber Security Technology*, *7*(3), 134–153. https://doi.org/10.1080/23742917.2023.2167276

Wani, T. A., Mendoza, A., & Gray, K. (2024). BYOD security behaviour and preferences among hospital clinicians – A qualitative study. *International Journal of Medical Informatics*, *192*. https://doi.org/10.1016/j.ijmedinf.2024.105606

Wilson, K. S., & Kiy, M. A. (2014). Some fundamental cybersecurity concepts. *IEEE Access*, *2*, 116–124. https://doi.org/10.1109/ACCESS.2014.2305658

Yoo, J., & Cho, Y. (2022). ICSA: Intelligent chatbot security assistant using Text-CNN and multi-phase real-time defense against SNS phishing attacks. *Expert Systems with Applications*, *207*. https://doi.org/10.1016/j.eswa.2022.117893

Zhang, J., Wu, P., London, J., & Tenney, D. (2025). Benchmarking and Evaluating Large Language Models in Phishing Detection for Small and Midsize Enterprises: A Comprehensive Analysis. *IEEE Access*, *13*, 28335–28352. https://doi.org/10.1109/ACCESS.2025.3540075