AI AND CYBERCRIME: A POLICY FRAMEWORK FOR ORGANIZATIONAL DEFENSE

Luca Knecht, Petra Maria Asprion, and Bettina Schneider

University of Applied Sciences and Arts Northwestern Switzerland FHNW, School of Business, Institute for Information Systems, Basel, Switzerland

ABSTRACT

Artificial Intelligence (AI) is a transformative technology with significant potential and accompanying threats. While AI fosters innovation, it is increasingly exploited for criminal activities, especially cybercrime. This research examines the criminal elements of AI, focusing on its role in enabling sophisticated cyber threats. Addressing AI requires organizations to adopt measures as traditional cybersecurity frameworks struggle to keep pace with rapidly evolving AI-driven threats. This research identifies and categorizes the criminal elements of AI, with particular attention to their application in cybercrime. Requirements derived from expert interviews are the foundation for developing a conceptual policy framework to provide organizations with measures to combat AI-driven threats. The proposed framework, refined through expert feedback, offers targeted, actionable recommendations to enhance organizational resilience against AI-driven threats. By providing a systematic structure and evaluated measures, the framework supports cybersecurity professionals, IT managers, and risk officers in mitigating the dual-use risks of AI. The research results contribute to bridging the gap in cybersecurity literature by addressing the evolving nature of AI-related threats and presenting a forward-looking policy framework tailored to middle-to-large organizations.

KEYWORDS

Artificial Intelligence, AI-Driven Threats, Policy Framework, Cybersecurity, Cybercrime

1 INTRODUCTION

1.1 Relevance and Problem Statement

The dual-use nature of artificial intelligence (AI)—its capacity for both beneficial and malicious applications—has significantly contributed to the rise in cyberattacks, particularly in the areas of social engineering, deepfakes, and autonomous hacking (Bueermann & Rohrs, 2024; Loh et al., 2024; Brundage et al., 2018). The World Economic Forum's *Global Risk Report 2024* identifies AI-generated misinformation and disinformation as the second-highest global risk, with far-reaching implications such as destabilizing governments, inciting unrest, and enabling terrorism (Bueermann & Rohrs, 2024). Cybercrime has surged globally, with incidents increasing by 600% between 2020 and 2023 (Rao et al., 2023). In Switzerland, cybercrime rose by 31.5% in 2023 compared to the previous year, threatening financial stability, reputational integrity, and business continuity (Federal Statistical Office, 2024).

Organizations face increasing difficulty in detecting and mitigating AI-driven threats, as conventional cybersecurity frameworks lag behind the evolving threat landscape (Melaku, 2023). AI-enhanced social engineering further exploits human vulnerabilities, increasing susceptibility to attacks (Melaku, 2023). The absence of globally harmonized regulations facilitates cross-border misuse of AI, while the rapid pace of AI development continues to outstrip existing security protocols (Evang, 2022). Moreover, AI's reliance on large datasets

raises privacy concerns, as anonymized data remains vulnerable to inference attacks (Admass et al., 2024). Given AI's transformative yet hazardous potential, organizations must establish a comprehensive understanding of its criminal applications and implement timely, actionable countermeasures (Admass et al., 2024; Maurya, 2023).

1.2 Research Goal and Related Questions

This study aims to identify and categorize the criminal elements of AI, with a particular focus on AI-enabled cyber threats. To bridge the gap between the malicious use of AI and the growing need for cybersecurity resilience, it proposes a conceptual policy framework that provides organizations with structured, actionable mitigation strategies. The research is guided by the following key research questions (RQ):

RQ1: What are the current criminal elements associated with AI-driven threats?

RQ2: How can these criminal elements be systematically categorized?

RQ3: What are the key requirements for developing a policy framework to address AI-driven threats

RQ4: How can these requirements be translated into a coherent conceptual policy framework?

RQ5: How useful is the resulting framework in supporting organizations against AI-enabled cyber threats?

2 LITERATURE REVIEW

As part of the research design, first a literature review was carried out with focus on the criminal use of AI in cybersecurity, examining how cybercriminals exploit AI to automate attacks and evade detection, as well as identifying countermeasures to strengthen security frameworks. The literature review covers sources from 2020 to 2024, in English and German, reflecting the rapid evolution of AI. It begins with a review of grey literature from industry leaders (e.g., Palo Alto Networks, the Big Four, WEF) to establish key terminology and trends. Academic insights were gathered from databases such as Google Scholar, ScienceDirect, Web of Science, and Semantic Scholar. AI tools like ChatGPT 4o and Scholar GPT supported the search process. Of 132 sources identified, 88 were selected for in-depth analysis based on their recency, credibility, and relevance to RQs. The following sections summarize the key findings in a concise manner.

2.1 General Classification of AI

A common generic classification of AI is developed from Saghiri et al. (2022). They classify AI in (1) Artificial Narrow Intelligence (performs specialized tasks), (2) Artificial General Intelligence (seeks to replicate human intelligence), and (3) Artificial Super Intelligence (a hypothetical form exceeding human intelligence). Currently, AI progresses primarily through machine learning (ML) and deep learning (DL), which use neural networks to analyze large datasets. ML models improve with experience, while DL processes unstructured data such as images or speech through multi-layered algorithms (IBM, 2024). These technologies power AI's role in cybersecurity, automation, and fraud detection—but also introduce new avenues for misuse (Guembe et al., 2022). Understanding these misuse risks is vital for mitigating AI-related cyber threats (Admass et al., 2024).

2.2 Criminal Elements of AI

To provide structure and clarity, we categorize the criminal elements of AI into six distinct domains based on, but extended from Blauth et al. (2022), including (1) bias in AI decision-making, (2) autonomous weapons and AI warfare, (3) AI-driven social engineering, (4) AI in cybercrime, (5) intellectual property theft and AI manipulation, and (6) privacy invasion and mass surveillance. This categorization allows for a more comprehensive analysis of the diverse ways in which AI can be misused in the context of cybersecurity.

(1) Bias in AI Decision-Making: AI systems can reflect and amplify bias through flawed algorithms, biased training data, or human input (Tabassi, 2023). In cybersecurity, this can result in inaccurate threat

detection or unfair profiling (IBM, 2023). Mitigating these risks requires responsible AI practices, including transparency, explainability, and adherence to emerging regulations like the EU AI Act (ISACA, 2024).

(2) Autonomous Weapons and AI Warfare: Autonomous Weapons Systems (AWS) use AI to identify and attack targets without direct human control (Blauth et al., 2022). AI is also deployed in cyberwarfare, espionage, and surveillance for its speed and data-processing capabilities (Yamin et al., 2021). These uses raise ethical concerns around accountability and escalation risks (European Union Agency for Cybersecurity, 2023).

(3) AI-Driven Social Engineering: AI enables more personalized and deceptive social engineering attacks via generative tools and language models (Falade, 2023). By analyzing online data, attackers craft effective phishing and impersonation campaigns. Broader access to AI lowers the barrier to launching such sophisticated attacks (Schmitt & Flechais, 2023).

(4) AI in Cybercrime: Cybercriminals automate attacks using AI, increasing scale and precision (Malatji, 2023). AI-driven malware evades detection by adapting code or execution paths. Criminals also use AI to execute large-scale DDoS attacks, targeting networks and critical infrastructure (Guembe et al., 2022).

(5) Intellectual Property Theft and AI Manipulation: AI challenges IP law by reducing human input in content creation (Hilty et al., 2021; Kokane, 2021). Legal ambiguity around AI-generated outputs increases IP theft risks. Cybercriminals exploit AI to steal proprietary data, highlighting the need for global legal harmonization (Nnamdi et al., 2023; Pavis, 2021).

(6) Privacy Invasion and Mass Surveillance: AI's need for extensive data raises privacy risks through inference attacks and misuse (Admass et al., 2024; Shahriar et al., 2023). Users face the personalization-privacy paradox (Meurisch & Mühlhäuser, 2022). Technologies like facial recognition facilitate mass surveillance, with risks of bias, misidentification, and authoritarian misuse (Clarke, 2022; Maphosa, 2023).

2.3 Cybercrime – Patterns and Trends

2.3.1 AI-enabled Cybercrime

Advances in AI have transformed cybercrime by enabling automation, personalization, and large-scale operations (Malatji, 2023). AI-driven threats include phishing, deepfakes, malware, and ransomware, which target vulnerabilities in both individuals and organizations (Guembe et al., 2022). We identified four attack types:

- Distributed Denial-of-Service (DDoS) Attacks: AI enhances botnet coordination, allowing faster, more adaptive DDoS assaults (Aslan et al., 2023; Humayun et al., 2020).
- Malware and Ransomware: AI enables malware to mutate dynamically, evading traditional defenses through evolving behaviors (Aslan et al., 2023; Humayun et al., 2020).
- Phishing and Spear-Phishing: natural language processing NLP-driven tools craft personalized messages that adapt to targets in real (Aslan et al., 2023; Humayun et al., 2020).
- AI-enhanced Identity Theft: Deepfakes of voices, images, and videos facilitate impersonation for fraud and espionage (Aslan et al., 2023; Humayun et al., 2020).

2.3.2 Motivations for Cyberattacks

Cyberattacks are driven by financial gain, identity theft, espionage, and sabotage (Mijwil et al., 2023; Li, 2017; Adlakha et al., 2019). Criminals steal data for profit or resale, while state actors seek strategic intelligence. Some attacks aim to disrupt systems or damage reputations (Aftab et al., 2022). Other motives include cyberterrorism, misinformation campaigns, and resource hijacking for botnets or cryptomining (Li, 2017; Mijwil et al., 2023). AI enables diverse threat actors to automate exploitation processes and adapt attack strategies in real time. Common actors behind AI-driven attacks: (1) Organized Crime Groups: Use AI for large-scale fraud, laundering, and ID theft (Edwards et al., 2022), (2) State-Sponsored Actors: Employ AI for espionage, sabotage, political interference (Hylender et al., 2023), (3) Individual Hackers: Pre-built AI tools enable attacks with minimal expertise (Edwards et al., 2022) and (4) Insiders: Leverage insider knowledge and AI tools to manipulate or extract data (Hylender et al., 2023).

2.4 Cybersecurity – AI Enhancements

As AI advances cyber threats, organizations increasingly adopt AI-enabled cybersecurity to improve strategic, operational, and technical defenses, including enhanced threat detection, faster incident response, and proactive prevention, boosting digital infrastructure resilience.

Strategic Defense Measures. Strategic cybersecurity reduces risks and aligns with standards like ISO/IEC 27001 and NIST CSF for risk management and protection (Alshar'e, 2023; Evang, 2022). The MITRE ATT&CK framework aids threat hunting and response (MITRE Corporation, 2024). However, these frameworks can be costly, complex, and less flexible against evolving AI threats (Melaku, 2023; Alshar'e, 2023). Regulatory compliance is crucial: GDPR governs EU data protection (Wolff et al., 2023), the AI Act introduces risk-based AI governance from 2024 (European Parliament & Council of the European Union, 2024), and Switzerland enforces the Federal Act on Data Protection (The Federal Assembly of the Swiss Confederation, 2020). Human error remains a major vulnerability; thus, Security Education Training and Awareness (SETA) programs are vital (MIT Technology Review, 2021; Dash & Ansari, 2022).

Operational Defence Measures. These include daily tools like access control (MFA, least privilege) (Hu et al., 2017), firewalls, intrusion detection systems (NIST, 2020), encryption, and backups (Plaka, 2022). Incident response plans, patching, antivirus, and SIEM systems support threat management and monitoring (Souppaya & Scarfone, 2022; Palo Alto Networks, 2024; Ban et al., 2023). Employee training reinforces threat awareness and best practices (Dash & Ansari, 2022; MIT Technology Review, 2021).

Technical Defence Measures. AI boosts cybersecurity via automation and intelligent detection, identifying anomalies, phishing, and malware variants (Maurya, 2023; Shanthi et al., 2023; Mohamed, 2023). AI-driven intrusion detection monitors traffic, while automated remediation and SOAR platforms speed responses (Mohamed, 2023; Shanthi et al., 2023). AI also supports vulnerability analysis, penetration testing, threat intelligence, user behavior profiling, and zero-day exploit prediction (Maurya, 2023; Mohamed, 2023). Additional uses include biometric authentication, cloud and IoT security, and tailored cybersecurity training (Shanthi et al., 2023; Mohamed, 2023). Challenges persist regarding data privacy, bias, and AI transparency (Adewale & Segun, 2024; Zhang et al., 2022; Shanthi et al., 2023).

2.5 Research Gap

Despite advances in AI-powered cybersecurity, organizations struggle to adapt to rapidly evolving AI threats. Existing frameworks like ISO/IEC 27001/2 and NIST CSF do not fully address AI's fast-paced changes, leaving vulnerabilities exposed (Evang, 2022). Traditional measures target known threats, but AI-driven attacks continuously evolve, bypassing standard controls (Malatji, 2023). Cybersecurity training remains insufficient for AI-specific threats, and employees often lack skills to detect AI-generated phishing, deepfakes, and automated attacks. Therefore, AI-focused awareness programs are needed to boost human resilience against AI-enabled deception (Dash & Ansari, 2022; Alshar'e, 2023). Another key gap is the absence of standardized AI security frameworks; while the EU AI Act sets ethical guidelines, no universal policy framework governs AI's cybersecurity role (Wolff et al., 2023). This regulatory uncertainty hampers enforcement and enables AI misuse. To address these gaps, we propose a conceptual policy framework integrating strategic, operational, and technical AI cybersecurity measures to enhance organizational resilience and mitigate AI-driven risks.

3 RESEARCH APPLICATION

3.1 Design and Process

This research was based on a pragmatic research philosophy focused on real-world solutions to AI-driven cybersecurity threats (adhered to Saunders et al., 2019). Employing an inductive approach, we explored emerging AI threats beyond existing theories by deriving an artefact from literature and qualitative expert interviews (based on recommendations from Saunders et al., 2019). Using qualitative methods and coding, we extracted key themes from cybersecurity and AI experts categorizing challenges and solutions (adhered to vom Brocke

et al., 2020). To enhance methodological transparency, this research involved semi-structured interviews with eight Swiss cybersecurity experts selected for their roles in IT governance, risk management, and AI deployment. Interviews followed a standardized protocol focused on AI threat categorization and mitigation strategies. Thematic coding was conducted manually to identify recurring patterns and validate framework components. The literature review (Section 2) laid the theoretical groundwork and identified research gaps, while interviews provided nuanced perspectives on threats and mitigation. Data were collected cross-sectionally to capture the current AI threat landscape, with future longitudinal studies recommended. As leading methodology the Design Science Research (DSR) approach by Kuechler & Vaishnavi (2004) seemed most preferably applicable with its four iterative phases:

1. **Problem Awareness:** An extensive literature review analyzed AI cybercrime, threats, and frameworks, addressing RQs 1 and 2 and identifying gaps.
2. **Suggestion:** Using the Double Diamond Model (Meinel et al., 2011), semi-structured interviews with Swiss experts informed AI-centric policy framework requirements. Thematic analysis and standards review guided development, addressing RQ3.
3. **Development:** Requirements were translated into an initial AI-driven policy framework (Version 0.1), refined iteratively with expert feedback to Conceptual Framework 1.0, addressing RQ4.
4. **Evaluation:** Cybersecurity experts assessed the framework's strengths and limitations through interviews, enabling refinements to finalize Conceptual Framework 1.0, ensuring responsiveness to evolving threats and addressing RQ5.

3.2 Artifact Development

The conceptual policy framework evolved iteratively to provide a structured, adaptable approach to AI-driven cybersecurity threats. It began with an initial structure incorporating essential cybersecurity measures aligned with best practices. The framework includes an overview outlining its purpose, target audience, and key principles, offering a structured approach to AI threats with risk assessments, AI-specific security measures, and adaptability for various organizations. Version 0.1 emphasized modularity, allowing users to apply measures comprehensively or selectively. Security measures were refined and categorized by best practices, focusing on AI-augmented defenses, human factors, legal compliance, and technical controls. Each measure details implementation steps, risk assessments, and use cases. The initial version was systematically compared to requirements to ensure completeness before evaluation.

3.3 Artifact Evaluation

Evaluation assessed the framework's effectiveness, scalability, robustness, and usability through expert interviews. Experts validated its practical applicability and suggested refinements for clarity, such as clearer goals, separating categories into distinct frameworks, and adding practical examples. The framework's scalability was praised, though experts recommended tailoring advice to different risk levels with guiding questions and case studies. For robustness, clearer terminology and hierarchical categorization were advised, along with consolidating overlapping categories to reduce redundancy. Usability was commended for simplicity, with suggestions to include actionable examples and simplified workflows. Table 1 summarizes adaptations made from expert feedback:

Table 1 Adaptations resulting from the artefact evaluation.

#	Conceptional Policy Framework - Artefact Adaptions
1	Expanded goals and purpose clarify the framework's mission and target organizations, replacing a brief original statement.
2	Baseline cybersecurity reaffirmed by emphasizing traditional controls (e.g., NIST CSF, ISO 27001), with AI as an added layer.
3	"How to use it" section enhanced with guiding questions to help practitioners navigate the framework.
4	Terminology reviewed for consistency, eliminating confusing or duplicate terms.
5	Overlapping categories and measures consolidated, streamlining taxonomy and reducing redundant activities.

4 RESULTING POLICY FRAMEWORK

The final conceptual policy framework 1.0 (Figure 1) integrates theoretical foundations with practical strategies to help organizations mitigate AI-driven cybersecurity threats. Designed with a modular structure, it enables flexible application- either holistically or focused on specific areas - and complements existing cybersecurity strategies. Serving IT teams, compliance officers, policymakers, researchers, and educators, the framework addresses both traditional security needs - such as authentication, patch management, and training - and emerging AI-specific risks requiring adaptive responses, ethical oversight, and regulatory compliance.
Organized into four main categories (Figure 1, detailed in Table 2), the framework presents targeted measures with clear objectives, implementation steps, and references to standards like ISO/IEC 27001 and the NIST CSF.

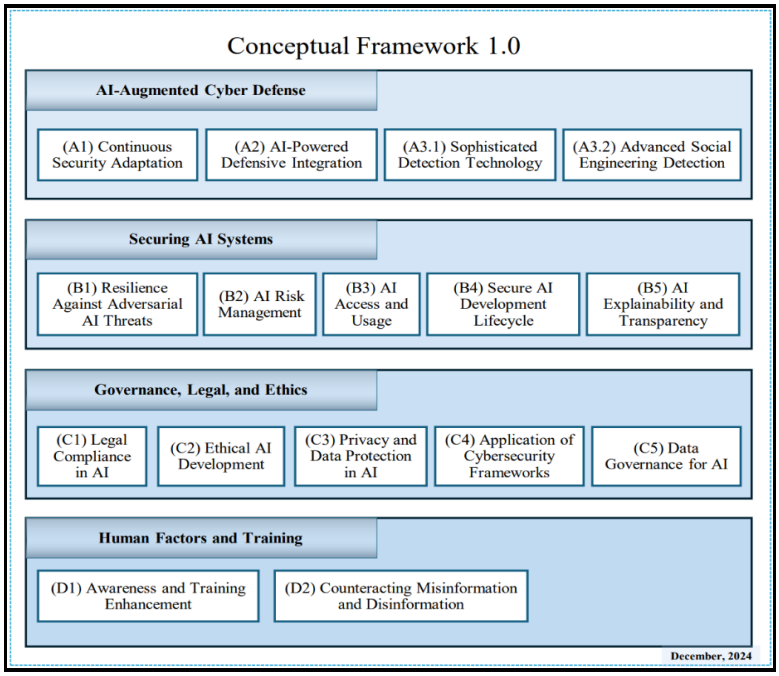

Figure 1. Conceptual Framework 1.0.

Organizations are encouraged to prioritize these measures based on their risk profiles, resource availability, and strategic goals. Its scalable design supports continuous improvement and focuses on critical domains such as social engineering and secure AI development. By integrating technical, governance, and human-focused components, the policy framework aligns with key regulations, including the GDPR, Swiss FADP, the upcoming EU AI Act, and management standards such as ISO/IEC and the NIST AI RMF.

Table 2. Conceptional Framework 1.0 with four main categories, measures and detailed context.

##	**Conceptual Policy Framework Main Categories**		**Detailed Description**
A	AI-Augmented Cyber Defense		Enhancing defense via real-time detection, automated responses, and predictive analytics.
	Measure		
	A1	Continuous Security Adaptation	Implementing security systems that adjust in real-time to emerging threats using AI and machine learning.
	A2	AI-Powered Defensive Integration	Integrating AI technologies into existing cybersecurity defenses to enhance capabilities in threat detection, response, and prevention.
	A3.1	Sophisticated Detection Technology	Utilizing advanced technologies like AI and machine learning to detect complex and evolving cyber threats that traditional methods might miss.
	A3.2	Advanced Social Engineering Detection	Using AI and behavioral analytics to identify and prevent social engineering attacks like phishing and spear-phishing.

B	Securing AI Systems		Protecting AI from threats with dedicated security measures.
	Measure		
	B1	Resilience Against Adversarial AI Threats	Building systems robust against attacks that exploit AI systems, such as adversarial machine learning attacks.
	B2	AI Risk Management	Identifying, assessing, and mitigating risks associated with the use of AI technologies within an organization.
	B3	AI Access and Usage	Managing who has access to AI systems and how they are used to prevent unauthorized use or abuse.
	B4	Secure AI Development Lifecycle	Incorporating security best practices throughout the AI development lifecycle to prevent vulnerabilities.
	B5	AI Explainability and Transparency	Ensuring that AI systems are transparent, and their decision-making processes are understandable to humans.
C	Governance, Legal, and Ethics		Ensuring that AI complies with laws, ethics, and strong governance.
	Measure		
	C1	**Legal Compliance in AI**	Ensuring that AI technologies comply with relevant laws, regulations, and ethical guidelines.
	C2	**Ethical AI Development**	Ensuring that AI systems are designed and deployed according to ethical principles like fairness and respect for human rights.
	C3	**Privacy and Data Protection in AI**	Ensuring that AI systems handle personal and sensitive data in compliance with data protection laws like GDPR.
	C4	**Application of Cybersecurity Frameworks**	Updating existing cybersecurity frameworks to incorporate AI considerations and address new technological challenges.
	C5	**Data Governance for AI**	Establishing robust data governance frameworks to manage the quality, security, and ethical use of data in AI systems.
D	Human Factors and Training		Emphasizes awareness and education to reduce AI-driven risks.
	Measure		
	D1	Awareness and Training Enhancement	Enhancing employee awareness and training regarding cybersecurity, with a focus on AI-related threats and tools.
	D2	Counteracting Misinformation and Disinformation	Strategies and tools to detect, analyze, and mitigate the spread of false or misleading information.

5 CONCLUSION AND OUTLOOK

This research examined the dual-use nature of AI, highlighting its growing misuse in advanced cyberattacks such as automated phishing, AI-driven malware, deepfakes, and data manipulation. While AI enhances cybersecurity resilience, it also presents new vulnerabilities exploited by cybercriminals. The main contribution is a conceptual policy framework developed through expert interviews and literature review, targeting organizations in Europe and Switzerland. Designed for cybersecurity professionals, IT managers, and risk officers, the framework supports the management of AI-related threats in alignment with regulatory standards such as the EU AI Act and GDPR.

The findings emphasize the urgency of adaptive, proactive cybersecurity strategies and validate the relevance of bridging theoretical insight with practical application. However, limitations include the regional scope and the need for iterative refinement. Compared to existing models such as ISO/IEC 27001, NIST CSF, and the EU AI Act, the proposed policy framework offers a more targeted approach to AI-specific threats. While ISO and NIST provide general cybersecurity guidance, they lack granularity in addressing AI misuse. The EU AI Act focuses on ethical governance but does not offer operational mitigation strategies. This conceptual policy framework bridges these gaps by integrating strategic, operational, and technical measures tailored to AI-enabled cybercrime.

Regarding limitations, it should not be concealed that the policy framework is grounded in a regional and sectoral context, drawing primarily from qualitative insights provided by eight Swiss cybersecurity experts. Although informative, this scope may constrain global applicability and overlook broader industry nuances. Furthermore, the policy framework has yet to undergo empirical validation or real-world implementation.

Future research should therefore broaden the sample base, engage cross-sectoral case studies, and apply longitudinal and human-centered approaches - such as AI-driven training for social engineering awareness - to ensure broader relevance and impact. Additional alignment with international legal frameworks would further

strengthen global adoption. Ultimately, the findings underscore the urgency of adaptive, forward-looking cybersecurity strategies and affirm the importance of translating theoretical insights into actionable tools for practitioners navigating the evolving landscape of AI-enabled threats.

REFERENCES

1. Adewale, D. S., & Segun, V. S. (2024). The intersection of Artificial Intelligence and cybersecurity: Challenges and opportunities. World Journal of Advanced Research and Reviews, 21(2), 1720–1736. https://doi.org/10.30574/wjarr.2024.21.2.0607
2. Adlakha, R., Sharma, S., Rawat, A., & Sharma, K. (2019). Cyber Security Goal's, Issue's, Categorization & Data Breaches. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 397–402. https://doi.org/10.1109/COMITCon.2019.8862245
3. Admass, W. S., Munaye, Y. Y., & Diro, A. A. (2024). Cyber security: State of the art, challenges and future directions. Cyber Security and Applications, 2, 100031. https://doi.org/10.1016/j.csa.2023.100031
4. Aftab, R. M., Ijaz, M., Rehman, F., Ashfaq, A., Sharif, H., Riaz, N., Hussain, S., Arslan, M., & Maqsood, H. (2022). A Systematic Review on the Motivations of Cyber-Criminals and Their Attacking Policies. 2022 3rd International Conference on Innovations in Computer Science & Software Engineering (ICONICS), 1–6. https://doi.org/10.1109/ICONICS56716.2022.10100569
5. Alshar'e, M. (2023). Cyber Security Framework Selection: Comparison of NIST and ISO 27001. Applied Computing Journal, 245–255. https://doi.org/10.52098/acj.202364
6. Aslan, Ö., Aktuğ, S. S., Ozkan-Okay, M., Yilmaz, A. A., & Akin, E. (2023). A Comprehensive Review of Cyber Security Vulnerabilities, Threats, Attacks, and Solutions. Electronics, 12(6), 1333. https://doi.org/10.3390/electronics12061333
7. Ban, T., Takahashi, T., Ndichu, S., & Inoue, D. (2023). Breaking Alert Fatigue: AI-Assisted SIEM Framework for Effective Incident Response. Applied Sciences, 13(11), 6610. https://doi.org/10.3390/app13116610
8. Blauth, T. F., Gstrein, O. J., & Zwitter, A. (2022). Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI. IEEE Access, 10, 77110–77122. https://doi.org/10.1109/ACCESS.2022.3191790
9. Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., & others. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. https://doi.org/10.48550/ARXIV.1802.07228
10. Bueermann, G., & Rohrs, M. (2024). Global Cybersecurity Outlook 2024. World Economic Forum. https://www.weforum.org/publications/global-cybersecurity-outlook-2024/
11. Clarke, R. (2022). Responsible application of artificial intelligence to surveillance: What prospects? Information Polity, 27(2), 175–191. https://doi.org/10.3233/IP-211532
12. Dash, B., & Ansari, M. F. (2022). An effective cybersecurity awareness training model: First defense of an organizational security strategy. International Research Journal of Engineering and Technology (IRJET), 9(4), 1–6. https://www.irjet.net/archives/V9/i4/IRJET-V9I401.pdf.
13. Edwards, M., Williams, E., Peersman, C., & Rashid, A. (2022). Characterizing Cybercriminals: A Review. https://doi.org/10.48550/ARXIV.2202.07419
14. ENISA. (2023). Artificial intelligence and cybersecurity research: ENISA research and innovation brief. Publications Office. https://data.europa.eu/doi/10.2824/808362
15. European Parliament & Council of the European Union. (2024, July 12). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L 2024/1689, 1–144.
16. European Union Agency for Cybersecurity. (2023). ENISA threat landscape 2023: July 2022 to June 2023. Publications Office. https://data.europa.eu/doi/10.2824/782573
17. Evang, J. M. (2022). ISO 27001 as a Tool for Availability Management. 2022 International Conference on Advanced Enterprise Information System (AEIS), 82–85. https://doi.org/10.1109/AEIS59450.2022.00018
18. Falade, P. V. (2023). Decoding the Threat Landscape: ChatGPT, FraudGPT, and WormGPT in Social Engineering Attacks. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 185–198. https://doi.org/10.32628/CSEIT2390533

19. Federal Statistical Office. (2024). Digitale Kriminalität. https://www.bfs.admin.ch/bfs/de/home/statistiken/kriminalitaet-strafrecht/polizei/digitale-kriminalitaet.html, last accessed 2025/04/28
20. Guembe, B., Azeta, A., Misra, S., Osamor, V. C., & Pospelova, V. (2022). The Emerging Threat of AI-Driven Cyber Attacks: A Review. Applied Artificial Intelligence, 36(1), 2037254. https://doi.org/10.1080/08839514.2022.2037254
21. Hilty, R. M., Hoffmann, J., & Scheuerer, S. (2021). Intellectual Property Justification for Artificial Intelligence. In Artificial Intelligence and Intellectual Property. Oxford University Press. https://doi.org/10.1093/oso/9780198870944.003.0004
22. Hu, V. C., Kuhn, R., & Yaga, D. (2017). Verification and Test Methods for Access Control Policies Models (NIST SP 800-192). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-192
23. Humayun, M., Niazi, M., Jhanjhi, N., Alshayeb, M., & Mahmood, S. (2020). Cyber Security Threats and Vulnerabilities: A Systematic Mapping Study. Arabian Journal for Science and Engineering, 45(4), 3171–3189. https://doi.org/10.1007/s13369-019-04319-2
24. Hylender, C. D., Langlois, P., Pinto, A., & Widup, S. (2023). 2023 Data Breach Investigations Report. (16th ed.). Verizon. https://www.verizon.com/business/resources/reports/2023-data-breach-investigations-report-dbir.pdf
25. IBM. (2023). What is AI Bias? https://www.ibm.com/think/topics/ai-bias?mhsrc=ibmsearch_a&mhq=bias , last accessed 2025/05/01
26. IBM. (2024). What is AI? https://www.ibm.com/topics/artificial-intelligence, last accessed 2025/04/28
27. ISACA. (2024). Achieving Trustworthy, Ethical AI (Journal Volume 1).
28. Kokane, S. (2021). The Intellectual Property Rights of Artificial Intelligence-Based Inventions. Journal of Scientific Research, 65(02), 116–119. https://doi.org/10.37398/JSR.2021.650223
29. Kuechler, W., & Vaishnavi, V. (2004). Design Research in Information Systems (Version of January 20, 2004; online resource, last updated December 20, 2017). DESRIST—Design Science Research in Information Systems and Technology. https://www.desrist.org/design-research-in-information-systems/.
30. Li, X. (2017). A Review of Motivations of Illegal Cyber Activities. Kriminologija & socijalna integracija, 25(1), 110–126. https://doi.org/10.31299/ksi.25.1.4
31. Loh, P. K. K., Lee, A. Z. Y., & Balachandran, V. (2024). Towards a Hybrid Security Framework for Phishing Awareness Education and Defense. Future Internet, 16(3), 86. https://doi.org/10.3390/fi16030086
32. Malatji, M. (2023). Offensive Artificial Intelligence: Current State of the Art and Future Directions. 2023 International Conference on Digital Applications, Transformation & Economy (ICDATE), 1–6. https://doi.org/10.1109/ICDATE58146.2023.10248780
33. Maphosa, V. (2023). Artificial Intelligence and State Power. 2023 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), 1–5. https://doi.org/10.1109/icABCD59051.2023.10220459
34. Maurya, R. (2023). Analyzing the Role of AI in Cyber Security Threat Detection & Prevention. International Journal for Research in Applied Science and Engineering Technology, 11(11), 514–519. https://doi.org/10.22214/ijraset.2023.56510
35. Meinel, C., Leifer, L., & Plattner, H. (Eds.). (2011). Design Thinking. Springer. https://doi.org/10.1007/978-3-642-13757-0
36. Melaku, H. M. (2023). A Dynamic and Adaptive Cybersecurity Governance Framework. Journal of Cybersecurity and Privacy, 3(3), 327–350. https://doi.org/10.3390/jcp3030017
37. Meurisch, C., & Mühlhäuser, M. (2022). Data Protection in AI Services: A Survey. ACM Computing Surveys, 54(2), 1–38. https://doi.org/10.1145/3440754
38. Mijwil, M., Unogwu, O. J., Filali, Y., Bala, I., & Al-Shahwani, H. (2023). Exploring the Top Five Evolving Threats in Cybersecurity: An In-Depth Overview. Mesopotamian Journal of Cyber Security, 57–63. https://doi.org/10.58496/MJCS/2023/010
39. MIT Technology Review. (2021, April 8). Preparing for AI-enabled cyberattacks. https://www.technologyreview.com/2021/04/08/1021696/preparing-for-ai-enabled-cyberattacks/
40. MITRE Corporation. (2024). MITRE ATT&CK. https://attack.mitre.org/, last accessed 2025/04/28
41. Mohamed, N. (2023). Current trends in AI and ML for cybersecurity: A state-of-the-art survey. Cogent Engineering, 10(2), 2272358. https://doi.org/10.1080/23311916.2023.2272358
42. NIST. (2020). Security and Privacy Controls for Information Systems and Organizations (Rev. 5). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-53r5

43. Nnamdi, N., Oniyinde, O. A., & Abegunde, B. (2023). An Appraisal of the Implications of Deepfakes: The Need for Urgent International Legislations. American Journal of Leadership and Governance, 8(1), 43–70. https://doi.org/10.47672/ajlg.1540
44. Palo Alto Networks. (2024). What is Endpoint Security Antivirus? What is Endpoint Security Antivirus? https://www.paloaltonetworks.com/cyberpedia/what-is-endpoint-security-antivirus, last accessed 2025/05/01
45. Pavis, M. (2021). Rebalancing our regulatory response to Deepfakes with performers' rights. Convergence: The International Journal of Research into New Media Technologies, 27(4), 974–998. https://doi.org/10.1177/13548565211033418
46. Plaka, R. (2022). Backup & Data Recovery in Cloud Computing: A Systematic Mapping Study. Ingenious, 2(1), 94–113. https://doi.org/10.58944/pwhk4843
47. Rao, G. R. K., Battu, V. V., Anupama, V., Allada, A., Krishna, S. V. R., & Hema, C. (2023). Modern Progressive Pitfalls of Cyber Attacks on the Digital World. 2023 2nd International Conference on Edge Computing and Applications (ICECAA), 244–248. https://doi.org/10.1109/ICECAA58104.2023.10212303
48. Onwubiko, C. (2020). CyberOps: Situational Awareness in Cybersecurity Operations. International Journal on Cyber Situational Awareness, 5(1), 82–107. https://doi.org/10.22619/IJCSA.2020.100134
49. Saghiri, A. M., Vahidipour, S. M., Jabbarpour, M. R., Sookhak, M., & Forestiero, A. (2022). A survey of artificial intelligence challenges: analyzing the definitions, relationships, and evolutions. Applied Sciences, 12(8), 4054. https://doi.org/10.3390/app12084054
50. Saunders, M. N. K., Lewis, P., & Thornhill, A. (2019). Research Methods for Business Students (8th ed.). Pearson. Print ISBN 978-1-292-20878-7; eText ISBN 978-1-292-20880-0.
51. Schmitt, M., & Flechais, I. (2023). Digital deception: Generative artificial intelligence in social engineering and phishing (arXiv:2310.13715) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2310.13715
52. Shahriar, S., Allana, S., Hazratifard, S. M., & Dara, R. (2023). A survey of privacy risks and mitigation strategies in the artificial intelligence life cycle. IEEE Access, 11, 61829–61854. https://doi.org/10.1109/ACCESS.2023.3287195
53. Shanthi, R. R., Sasi, N. K., & Gouthaman, P. (2023). A new era of cybersecurity: The influence of artificial intelligence. 2023 International Conference on Networking and Communications (ICNWC), 1–4. https://doi.org/10.1109/ICNWC57852.2023.10127453
54. Souppaya, M., & Scarfone, K. (2022). Guide to enterprise patch management planning: Preventive maintenance for technology (NIST SP 800-40 Rev. 4). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-40r4
55. Tabassi, E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1
56. The Federal Assembly of the Swiss Confederation. (2020). Federal Act on Data Protection (FADP). https://www.fedlex.admin.ch/eli/cc/2022/491/en, last accessed 2025/04/28
57. Vom Brocke, J., Hevner, A., & Maedche, A. (Eds.). (2020). Design Science Research: Cases. Springer. https://doi.org/10.1007/978-3-030-46781-4
58. Wolff, J., Lehr, W., & Yoo, C. S. (2023). Lessons from GDPR for AI Policymaking. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4528698
59. Yamin, M. M., Ullah, M., Ullah, H., & Katt, B. (2021). Weaponized AI for cyber attacks. Journal of Information Security and Applications, 57, 102722. https://doi.org/10.1016/j.jisa.2020.102722
60. Zhang, Z., Ning, H., Shi, F., Farha, F., Xu, Y., Xu, J., Zhang, F., & Choo, K.-K. R. (2022). Artificial intelligence in cyber security: Research advances, challenges, and opportunities. Artificial Intelligence Review, 55(2), 1029–1053. https://doi.org/10.1007/s10462-021-09976-0