Lier, Sarah K.; Eppers, Tjelve M.; Gerlach, Jana; Müller, Pascal; Breitner, Michael H.

An iterative five-phase process model to successfully implement AI for cybersecurity in a corporate environment

Electronic Markets

Provided in Cooperation with:
Springer Nature

Suggested Citation: Lier, Sarah K.; Eppers, Tjelve M.; Gerlach, Jana; Müller, Pascal; Breitner, Michael H. (2025) : An iterative five-phase process model to successfully implement AI for cybersecurity in a corporate environment, Electronic Markets, ISSN 1422-8890, Springer, Berlin, Heidelberg, Vol. 35, Iss. 1,
https://doi.org/10.1007/s12525-025-00802-x

Electronic Markets (2025) 35:56
https://doi.org/10.1007/s12525-025-00802-x

RESEARCH PAPER

An iterative five-phase process model to successfully implement AI for cybersecurity in a corporate environment

Sarah K. Lier[1] · Tjelve M. Eppers[1] · Jana Gerlach[1] · Pascal Müller[1] · Michael H. Breitner[1]

Received: 30 August 2024 / Accepted: 21 May 2025

Abstract

While traditional cybersecurity approaches effectively address static or well-known threats, they often struggle to keep pace with the rapidly evolving threat landscape. New research highlights that increasing sophistication and dynamism in cyberattacks require adaptive and proactive measures, such as artificial intelligence (AI) applications and services, to complement conventional methods. AI for cybersecurity is needed to respond efficiently and reliably to threats and attacks, to detect dynamic threats faster, to analyze more precisely, and to enable adaptive protective measures that outperform conventional approaches. We identified research needs for AI in cybersecurity that need to be addressed by implementing respective AI applications and services. Companies and organizations need further research and company-centric approaches. We address AI for cybersecurity through a literature review and semi-structured expert interviews in a design science research–oriented framework. We identify typical implementation steps, deduce critical process phases, and develop a new process model to successfully implement AI for cybersecurity, including five process phases and 19 process steps. Our iterative five-phase process model provides a structured framework that is flexible to adapt to specific and general requirements, focuses on iterative evaluations; addresses cost, functional requirements, certifications, and environmental impact; facilitates early risk identification; and strengthens resilience against cyberattacks. Furthermore, we deduce seven key performance indicators to support a quantitative assessment of AI's efficiency and effectiveness, allow benchmarking, and develop best practices. Finally, we provide limitations and a further research agenda.

Keywords Cybersecurity · Cyberspace process model · Artificial intelligence · Key performance indicators · Corporate environment

JEL Classification L210 · M150

Introduction

Cybersecurity ventures expect the cost resulting from global cybercrime to grow by 15% annually, up to 10.5 trillion USD in 2025 (Morgan, 2020). Due to the growing volume and complexity of cyber threats confronting stakeholders, e.g., organizations and cybersecurity analysts, they are overwhelmed by the number of emerging threat reports (Aloqaily et al., 2022). Due to threat-type landscapes, companies and organizations must evolve their cybersecurity for long-term security. Artificial intelligence (AI) is a frequently mentioned solution due to its efficiency and reliability. A study by IBM (2024), which investigated the cost of data breaches, analyzed 604 companies about the implementation status of AI technology in their cybersecurity, surveyed by the Ponemon Institute in MI, USA. The overall global percentage of

Responsible Editor: Emma Gritt

✉ Sarah K. Lier
lier@iwi.uni-hannover.de

Tjelve M. Eppers
tjelve.eppers@stud.uni-hannover.de

Jana Gerlach
gerlach@iwi.uni-hannover.de

Pascal Müller
pascal.mueller3@stud.uni-hannover.de

Michael H. Breitner
breitner@iwi.uni-hannover.de

[1] Institute of Information Systems, Leibniz University Hannover, Königsworther Platz 1, Hannover 30167, Germany

Published online: 19 June 2025

companies fully implementing AI in cybersecurity is about 24%. Despite clear evidence of cost and efficiency benefits, this low adoption rate shows that organizations face substantial implementation barriers that current research and practice do not sufficiently address. Companies surveyed by IBM (2024) reported an average 45% reduction in data breach cost after implementing new AI-driven security measures. Roshanaei et al. (2024) used historical data from real-world security incidents, simulated attack scenarios, and the performance of AI-based security solutions in controlled environments. After introducing AI-based strategies, the average detection time was reduced by 94%, the false positive rate by 75%, the threat response time by 96%, and the number of undetected attacks by 70%. The IBM study (2024) and Roshanaei et al. (2024) highlight the advantages of implementing AI in cybersecurity.

Gerlach et al. (2022b) created a taxonomy for AI in cybersecurity and the corresponding decision tree DETRAIS. DETRAIS supports decision-makers in their efforts to select a suitable AI-driven cybersecurity business model and service. Their paper calls for developing process or maturity models. Dangi et al. (2023) highlighted the need for AI in cybersecurity based on elements, e.g., cost pressure, social engineering attacks, speed of technology adaptation, and one-size-fits-all security technologies. Vegesna (2023) highlighted AI's potential "to revolutionize threat detection, response, and resilience in the digital landscape" (p. 2) related to cybersecurity. While several studies provide a foundation for understanding AI applications, they do not offer actionable, step-by-step implementation guidance and do not support performance monitoring or iterations during deployment (e.g., Amershi et al., 2019; Reim et al., 2020; Tolido et al., 2019).

AI analyzes large amounts of data and reaches real-time decisions and analyses (Familoni, 2024; Kühl et al., 2022). Cybersecurity problems such as the manual processing of large amounts of data, dangers and threats from attacks, and cost factors for preventing threats are barriers for many companies and organizations. AI can provide supportive applications and services to cybersecurity issues (Dangi et al., 2023; Familoni, 2024; Jawhar et al., 2024; Rampášek et al., 2025; Salem et al., 2024). It is necessary to implement AI in cybersecurity (Dangi et al., 2023; Gerlach et al., 2022b). Among the significant benefits of implementing AI cybersecurity solutions, companies and organizations struggle to integrate AI cybersecurity applications and services into their current infrastructure. Ansari et al. (2022) identify a lack of knowledge, insufficient planning, and high implementation costs as key barriers to adopting AI-based cybersecurity solutions. Jayathilaka and Wijayanayake (2025) suggest implementing guidelines and less complex cybersecurity frameworks as particular obstacles preventing companies from implementing AI cybersecurity applications and services. Process models reduce the complexity and cost to implement new technologies. These process models assist in identifying relationships and barriers during the implementation and provide recommendations to minimize them (Nilsen, 2020). Therefore, we develop a process to implement AI for cybersecurity, following the research question (RQ):

RQ1: *How can an efficient and reliable process of implementing AI for cybersecurity be modeled?*

To address RQ1, we conduct a systematic and efficient literature review, including an extension to AI- and graphic-based tools. We analyze several publications and define severe problems. Based on this, we highlight the advantages and disadvantages of AI in cybersecurity. We expanded our literature-based knowledge base through expert interviews and interviewed nine experts. Therefore, we have a theoretical and practical knowledge base that eliminates limitations like subjectivity. We provide a status quo about current process models, adapt some phases, and highlight some critical process phases addressing RQ2:

RQ2: *What are the critical phases in the implementation process?*

Implementing AI in cybersecurity increases cyber risk management's efficiency and reliability (Jawhar et al., 2024). Critical phases highlighted the indispensable need to reach this efficiency and reliability. Key performance indicators (KPIs) are defined based on our process model. KPIs are specific metrics used to monitor and evaluate the performance of processes, support the successful implementation of AI in cybersecurity, and enable continuous improvements. We address RQ3 through a systematic analysis and assessment of our process model.

RQ3: *How can KPIs be derived from the process model to optimize the efficiency and reliability of the implementation of AI for cybersecurity?*

First, we describe the theoretical background of AI for cybersecurity. Then, we ground our research design. We follow a research design oriented by design science research (DSR) based on vom Brocke et al. (2020) and Hevner and Chatterjee (2010). DSR is characterized "by the flexibility to constantly change literature, a practice- and solution-oriented view, and innovation potential" (Lier et al., 2024, p. 6848). We developed a process model based on our systematic and efficient literature review and semi-structured expert interviews. The application of our DSR-oriented framework highlights the practical relevance of our research, offering a scientific foundation through recognized scientific methods and research frameworks, flexible adaptation to the cybersecurity threat landscape, triangulation increasing the validity of our results related to quantitative and qualitative research, and scalability of our process model through adaptability to

its environment (Gregor et al., 2020; Hevner & Chatterjee, 2010; Lier et al., 2023; Venkatesh et al., 2013; vom Brocke et al., 2020). Afterward, we deduce KPIs and discuss our results, findings, and implications. Our work contributes to research analyzing AI integration into security processes, identifying optimization opportunities, and proposing new RQs to enhance AI solutions. In practice, we enable targeted AI implementations that balance benefits and risks, ensure compliance through KPIs, and support better decision-making. Based on limitations and expert insights, we propose a research agenda to advance AI for cybersecurity, allowing companies and organizations to adapt our iterative five-phase process model and foster further scientific exploration. We provide limitations, a further research agenda, and recommendations for research and practice.

Theoretical background

Cybersecurity for organizations, governments, individuals, and other stakeholders is essential to protect against the growing threats from cyberspace as digitalization and connectivity increase. This includes practical protection against viruses, malware, phishing, ransomware, and other cyberattacks and threats (Li & Lui, 2021). All systems must ensure key security objectives and principles, including confidentiality, integrity, and availability (Gunduz & Das, 2020; Li & Lui, 2021; Sarker et al., 2021).

- Confidentiality: Ensuring that information can only be viewed by authorized individuals or systems
- Integrity: Ensuring data remains accurate, unaltered, authentic, and protected from unauthorized changes
- Availability: Ensuring that systems and data are accessible to authorized users when needed

Another objective and principle include authentication, which is the verification of the identities of users, devices, or systems to ensure that access to resources is restricted to authorized persons or processes (Gunduz & Das, 2020). Similarly, Aftergood (2017) and Craigen et al. (2014) define cybersecurity as tools, policies, or technologies that protect networks, programs, and data from attack, damage, and unauthorized access. Thus, cybersecurity is considered to prevent and protect systems from various cyberattacks and threats (Sarker et al., 2021). Therefore, it can be divided into security types, e.g., network, information, cloud, and user security (Li & Lui, 2021; Sarker et al., 2021). Network security protects computer networks from disruption, attack, or unauthorized access. It is critical to a company's and organization's security strategy, as networks are central to communication and data exchange (Zhang, 2021). Information security protects digital data, physical documents, human resources, and other essential elements in an organization's information exchange and business operations (Ogbanufe, 2021). Cloud security ensures the security of data, applications, and infrastructure stored or running in cloud services. It includes identity and access management, encryption, and security monitoring (Krishnasamy & Venkatachalam, 2021). User cybersecurity training is critical to security risk mitigation, as human factors often cause security incidents. It involves security awareness, phishing awareness, and password management (Krishnasamy & Venkatachalam, 2021).

Cyber threats describe security breaches to exploit a vulnerability in a system, whereas cyberattacks describe intentional and unauthorized actions on a system (Sarker et al., 2021). Most cyberattacks and threats include denial of services (DoS) (Jang-Jaccard & Nepal, 2014), distributed denial-of-service (DDoS) (Saghezchi et al., 2022), phishing (Alsayed & Bilgrami, 2017; Jang-Jaccard & Nepal, 2014), ransomware (McIntosh et al., 2019), malware (Moghimi et al., 2019; Tong & Yan, 2017), man-in-the-middle (MitM) (Kügler, 2003), and supply chain attack (Eggers, 2021; Ohm et al., 2020). DoS attacks are the most common and are distinguished as attackers gaining access to data, devices, and network resources and denying access to users. The aim is to disrupt the availability of data, devices, and networks to cause damage (Dangi et al., 2023; Jang-Jaccard & Nepal, 2014; Rangrez et al., 2024; Salem et al., 2024). In the extension of the DoS attack, DDoS, the attacker uses several infected computers or devices to flood the target system with data traffic simultaneously to create a coordinated overload through bot networks (Saghezchi et al., 2022). Phishing attacks determine the identity of users (Alsayed & Bilgrami, 2017; Dangi et al., 2023; Jang-Jaccard & Nepal, 2014). Ransomware is malware that aims to encrypt or block access to a victim's data. In contrast to DoS attacks, ransomware demands a ransom (McIntosh et al., 2019). Malware is a generic term for malicious software designed to damage, disrupt, or gain unauthorized access to computers, networks, and devices. Malware includes a variety of threats, such as viruses, worms, trojans, spyware, and ransomware; each type has different damage mechanisms and attack targets (Dangi et al., 2023; Moghimi et al., 2019; Tong & Yan, 2017). In MitM attacks, an attacker intervenes between two communicating parties, e.g., user and website, and intercepts and manipulates the communication so that information such as passwords or credit card details is intercepted (Kügler, 2003). Supply chain attacks affect the entire supply chain of products and aim to exploit vulnerabilities in the supply chain to achieve large-scale compromises (Eggers, 2021; Ohm et al., 2020).

Cybersecurity defense strategies protect networks and systems from cyberattacks and threats. They prevent data breaches or security incidents (Familoni, 2024; Khraisat et al., 2019; Sarker et al., 2021). Some classic cybersecurity

defense strategies include access control (Qi et al., 2018), anti-malware (Xue et al., 2017), sandbox (Hunt et al., 2018), and security information and event management (Irfan et al., 2016). Attacks on cyber-physical systems (CPS) are increasing with the rise of AI (Charanarur et al., 2025). Attacks affect multiple actors, and attackers identify different types of vulnerabilities. Primarily, malicious actions in cyberspace and monetizing cybercrime are serious threats (de Azambuja et al., 2023; Kaloudi & Li, 2020). Although AI is a reason for more cyberattacks, AI can also identify the vulnerabilities of CPS and implement defensive measures (de Azambuja et al., 2023; Familoni, 2024; Novikov, 2018). Intelligent cybersecurity defense strategies are based on AI or use AI applications and services to generate intelligent decision-making of cyber applications. Sarker et al. (2021) list several of the most commonly used AI-based techniques, including clustering for intrusion detection analysis, k-nearest neighbor for network intrusion detection, decision trees for malicious behavior analysis, intrusion detection systems (IDS) for anomaly detection, neural networks, deep learning (DL) for classification of anomaly intrusion detection attacks and malware traffic, and reinforcement learning for malicious activity and intrusion detection (Rangrez et al., 2024; Salem et al., 2024; Sarker et al., 2021). Rampášek et al. (2025) analyze the development of cybersecurity for digital products, AI technology, and AI-based products and call for increased standardization and certification of best practices, more comprehensive regulation, and a future-oriented approach that includes AI in the field of cybersecurity. Most of these techniques are based on machine learning (ML) and DL, which were expanded through AI (Ozkan-Okay et al., 2024; Salem et al., 2024; Sarker et al., 2021).

ML is a subcategory of AI and a method to implement AI algorithms for data analysis. ML algorithms are divided into supervised, unsupervised, and reinforcement learning. Supervised learning is an ML process in which the model is trained from a labeled data set to predict new, unseen data (de Azambuja et al., 2023; Ozkan-Okay et al., 2024). Unsupervised learning is a method in which the model discovers patterns and structures from an unlabeled dataset without specific input for the output (de Azambuja et al., 2023; Ozkan-Okay et al., 2024). Reinforcement learning is an approach in which an agent learns by interacting with its environment, receiving rewards for good actions and penalties for bad ones to develop optimal behavior (de Azambuja et al., 2023; Kühl et al., 2022; Ozkan-Okay et al., 2024). ML is often used to assist cybersecurity. Chen et al. (2020) developed a model for fraud detection platforms based on ML to detect fraud in Industry 4.0 transactions. A framework developed by Le et al. (2019) uses ML for real-time analytics to protect automation networks and system operations. Saghezchi et al. (2022) use ML to identify the non-standard behavior of natural networks and develop an ML model to detect DDoS attacks from CPS. Ozkan-Okay et al., (2024) summarized the benefits of ML usage in cybersecurity as improved accuracy, faster detection, automation, and scalability. The limitations of using ML predominantly correspond to the benefits through dependencies of high qualitative data, high complexity, and hostile attacks (Ozkan-Okay et al., 2024). Nevertheless, AI and ML contribute to cybersecurity and cyberspace (Ozkan-Okay et al., 2024; Roshanaei et al., 2024; Salem et al., 2024; Sarker et al., 2021).

Research design

Design science research

To address our RQs, we follow a DSR-oriented framework proposed by vom Brocke et al. (2020) and Hevner and Chatterjee (2010). DSR prioritizes problem-solving and application-focused strategies, making it suitable to analyze dynamic topics and incorporate current literature and research developments. DSR facilitates the development of artifacts that address research challenges and fosters a foundational understanding of research subjects by continuously refining and enhancing design artifacts (Hevner & Chatterjee, 2010). DSR is inherently iterative, involving repeated cycles of refinement and enhancement of a design artifact. It supports collaboration with various stakeholders and enables the integration of theoretical and practical insights during an artifact's creation and evaluation phases (vom Brocke et al., 2020). Through this process, we contribute a process model and corresponding KPIs as DSR artifacts at Level 2 (nascent design theory), as Gregor and Hevner (2013) outlined. Level 2 contributions emphasize generalizable artifacts such as methods, models, or design principles. Gregor's and Hevner's (2013) framework highlights a broad categorization of artifacts, further detailed by Lee et al. (2015), who distinguish between technological artifacts (e.g., software), informational artifacts (e.g., messages), and social artifacts (e.g., altruistic actions). Lowry et al. (2017) expanded on this categorization, identifying additional types of information system artifacts and advocating for more precise classifications. In this context, we develop a domain-specific artifact through our process model for AI in cybersecurity. DSR allows us to build upon current processes and create innovative linkages within the field. In doing so, we design our artifact with a strong focus on user activities. The iterative nature of our DSR-oriented process is illustrated in Fig. 1. In our initial step, we identify limitations of current cybersecurity process models and deduce research needs within the relevance cycle. This cycle establishes the relevance of DSR by grounding our artifact's application in a specific context. The application context provides requirements and acceptance criteria to evaluate

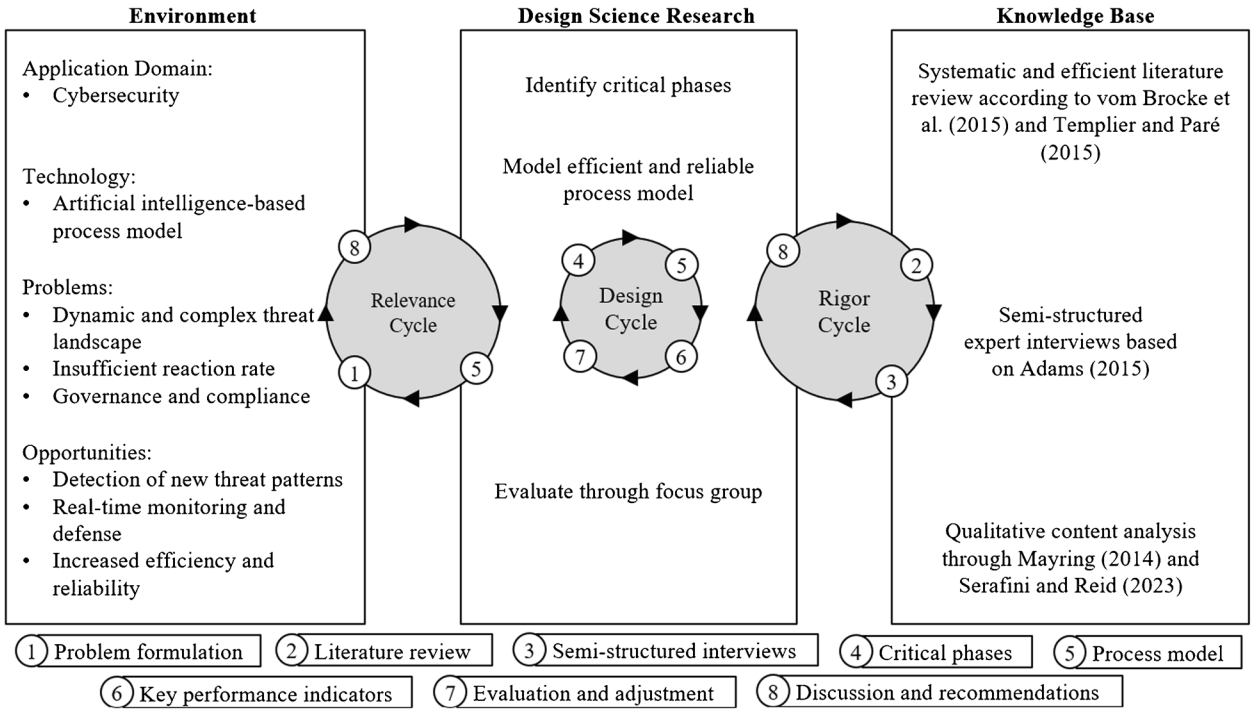

Fig. 1 Research design based on Hevner & Chatterjee (2010)

outcomes and ensure consistency with environmental needs (Hevner & Chatterjee, 2010). Our second step involves building our knowledge base through a systematic and efficient literature review, followed by confirmatory qualitative expert interviews in our third step. Triangulating insights from these steps, we enhance the validity and reliability of our findings, enabling a comprehensive understanding of our research topic through theoretical integration (Venkatesh et al., 2013). Steps 2 and 3 correspond to the rigor cycle, which focuses on the leverage of available research knowledge, including models and methods, to ensure our artifact's applicability and generalizability (Hevner & Chatterjee, 2010). In our fourth step, we identify critical process phases to develop our efficient and reliable process model in our fifth step. Our development aligns with the design cycle, which encapsulates our iterative creation and validation of artifacts (Hevner & Chatterjee, 2010). Building on our insights gained from our literature review and expert interviews, we define KPIs in our sixth step. Finally, our findings are evaluated and adjusted through a focus group discussion. We conclude by discussing our results, findings, and recommendations.

Literature review

In Step 2, we conducted a keyword-based database search to identify relevant literature according to vom Brocke et al. (2015) and Templier and Paré (2015). The systematic and efficient literature review includes a six-step literature search process (Templier & Paré, 2015) and recommendations (vom Brocke et al., 2015). The databases "SpringerLink," "AIS eLibrary," "IEEE Xplore," "ACM Digital Library," and "JSTOR" were used based on the following search string: ("artificial intelligence" OR "AI" OR "machine learning" OR "ML") AND ("cybersecurity" OR "cyber security") AND ("implementation" OR "integration") in the period from January 2017 to June 2024. After formulating the problem and searching for literature, we selected 120 papers. In order to exclude the possibility of missing relevant publications in the databases due to unavailability, we used the AI- and graphic-based research tools Semantic Scholar,[1] ORKG ASK,[2] and VosViewer[3] for a citation network analysis, and COnnecting REpositories (CORE)[4] for scientific literature reviews with the same search strings to increase the efficiency. This extension allowed us to add 16 publications to the search that were unavailable in the databases. We use this option solely as an add-on for increased efficiency, as these tools are partially incomplete, dependent on the metadata, and do not usually consider the differences in quality in the journals and publications. Knowledge graphs support

[1] https://www.semanticscholar.org/
[2] https://ask.orkg.org/
[3] https://www.vosviewer.com/
[4] https://core.ac.uk/

the visualization of relationships between publications and the discovery of new insights (Schröder et al., 2024). The tools we use increase research efficiency and facilitate understanding of complex publication landscapes. Schröder et al. (2024) applied cyber mapping with knowledge graphs to the German financial sector, significantly expanding their analysis scope and providing new methodological perspectives for analyzing and reviewing the literature. We screened the full text of our scientific literature database and excluded 62 papers. In line with Webster and Watson (2002), we added 12 papers in backward, forward, author, and Google Scholar similarity searches. Finally, we included 86 papers in our final data set. During this process, we considered each paper's quality and contribution based on the quality of the outlet, e.g., white papers are not included, impact factors are considered, the novelty of research results, and citations, e.g., on Google Scholar.

Severe problems

Bérubé et al. (2021) used a ranking-based Delphi study with 18 panelists. They identified organizational barriers to AI implementation. The barriers extend from data acquisition through lack of data governance, problems in understanding AI at the management level, and lack of experts who can prepare the data and build appropriate AI models. Hamm and Klesel (2021) provide a systematic overview of success factors to adopt AI in companies and organizations using systematic literature research. In total, thirteen papers were screened in their paper (Hamm & Klesel, 2021). They divided the success factors into organizational, technological, and environmental dimensions. The compatibility was found the most in the dimension of technological success factors. Another critical aspect was data availability and quality regarding the model's training. Top management support (technical) competencies, and resources were found in 8 papers as the highest influential success factors for the organizational dimension. Resources must be available with sufficient data, budget, and employees. In the environmental dimension, they identified competition, industry pressure, and governmental regulation as the most impactful success factors (Hamm & Klesel, 2021).

Advantages

Siroya and Mandor (2021) discussed reducing work for cybersecurity teams. Due to the current threat landscape, cybersecurity analysts require much time to deal with all the issues in detail. It allows for the concentration of the most crucial threats and the dealing of more strategic-related tasks or other critical security issues (Kairu & Shin, 2022; Mughal, 2018; Rampášek et al., 2025; Roshanaei et al., 2024). AI can handle data and analyze real-time security events (Mughal, 2018). It also allows the correlation of the activities in a company's and organization's network "to billions of events per day" (Kairu & Shin, 2022, p. 2). Due to this correlation, AI increases reaction times in recognizing alerts from various sources compared to familiar security solutions (Sharma, 2021). Therefore, the cybersecurity teams are notified quickly and can solve the problems proactively. Another advantage is the identification of anomalies, as the AI can find unusual activities indicating a security breach. It analyzes data sources like system logs, network traffic, and user behavior (Das & Sandhane, 2021; Kairu & Shin, 2022; Mughal, 2018). AI provides the ability to keep up with current threats, which aims to use AI technology on the defender's and attacker's sides to launch more sophisticated attacks (Sharma, 2021; Siroya & Mandot, 2021). AI replaces traditional approaches, increasing detection rates (Berman et al., 2019; Segal, 2020). Berman et al. (2019) refer to DL approaches. Depending on the programming of the AI, it can detect cyberattacks, inform the security staff, and respond autonomously, blocking the attacker from the system or shutting down the system to prevent further damage (Familoni, 2024; Kairu & Shin, 2022; Mughal, 2018). The key advantages are summarized in Table 1, which includes the corresponding references.

Table 1 Advantages of AI for cybersecurity

Advantage	References
Work facilitation for cybersecurity teams	Ali et al. (2022); Fay and Trenholm (2019); Kairu and Shin (2022); Mughal (2018); Rampášek et al. (2025); Siroya and Mandot (2021)
High-volume data analysis	Ali et al. (2022); Fay and Trenholm (2019); Kairu and Shin (2022); Mughal (2018); Siroya and Mandot (2021)
Quicker detection and responses	Ali et al. (2022); Kairu and Shin (2022); Sharma (2021); Siroya and Mandot (2021)
Identifying anomalies	Das and Sandhane (2021); Kairu and Shin (2022); Mughal (2018); Siroya and Mandot (2021)
Keeping up with actual threats	Familoni (2024); Sharma (2021); Siroya and Mandot (2021)
Better detection rates	Berman et al. (2019); Segal (2020)
Automated responses	Familoni (2024); Kairu and Shin (2022); Mughal (2018)

Table 2 Disadvantages of AI for cybersecurity

Disadvantage	References
AI manipulation (adversarial attacks)	Badhwar (2021); Fay and Trenholm (2019); Hamon et al. (2020); Handa et al. (2019); Hu et al. (2021); Machado et al., (2021); Martins et al. (2020); Rosenberg et al. (2021); Sen et al. (2022); Shrestha and Mahmood (2019); Zhang et al. (2022)
False classification	Badhwar (2021); Handa et al. (2019); Segal (2020); Shaukat et al. (2020)
AI-generated attacks	Hu et al. (2021); Segal (2020); Sharma (2021)
Data protection	Hu et al. (2021); Mughal (2018); Sharma (2021)
Black box problem	Berman et al. (2019); Hu et al. (2021); Zhang et al. (2022)

Disadvantages

The most crucial disadvantage of AI in literature is its proneness to adversarial attacks, discussed in eleven papers. Machado et al. (2021) and Rosenberg et al. (2021) developed a taxonomy of adversarial attacks in the cybersecurity domain. Adversarial attacks can be understood as attacks on the ML model to manipulate performance (Martins et al., 2020). They can be categorized as causative/poisoning or evasive/exploratory attacks (Machado et al., 2021). Hu et al. (2021) make another differentiation for the poisoning attacks and divide them into availability and integrity attacks. Availability attacks aim for misclassification and model performance degradation (Hu et al., 2021; Martins et al., 2020). In evasive or exploratory attacks, the adversary extracts information about the model's architecture, parameters, and cost function (Machado et al., 2021). The false classification of the AI models is another disadvantage in scientific literature. These are false positives and false negatives (Badhwar, 2021).

Data privacy remains an open question for AI applications and services. It bears the risk that trained data gets misused (Mughal, 2018), where it gets stored, and who has access to it (Sharma, 2021). Compliance with general data protection or other regulations is necessary (Mughal, 2018; Sharma, 2021). Another disadvantage is the black box problem of AI, which is mentioned in three papers. Especially in the decision-making process, an often-used argument is the lack of transparency in AI-based results (Zhang et al., 2022). Table 2 provides an overview of the disadvantages.

Expert interviews

We conducted nine semi-structured interviews (SSI) based on Adams (2015) to expand our knowledge base in the third step of our DSR-oriented framework. SSIs provide the advantage of analyzing different perspectives of individuals and thus gathering a more comprehensive overview. It is allowed to ask open and closed questions to deepen the knowledge of the scientific literature (Adams, 2015). We created an interview guide with several guiding questions based on our identified literature. The SSIs with the individual experts were conducted in a video conference and lasted between 30 and 60 min. The SSIs were recorded and transcribed. The selection of experts was based on a Crunchbase[5] search and divided into three groups. The first group consists of companies and organizations offering AI for cybersecurity. The second group contains companies and organizations that are customers and have implemented AI in cybersecurity. Regarding the RQs, those using AI in cybersecurity are preferred interview partners, as they have already completed the implementation process. The third group comprises companies and organizations involved in consulting and has supported several companies and organizations in implementing AI. All companies and organizations involved in AI and cybersecurity were filtered in Crunchbase, and general company and organization information and their active or closed status were listed. Companies and organizations outside of the European Union were excluded to avoid potential problems such as time differences and general accessibility and to achieve a higher success rate for accepted interviews. As reference customers are often listed on the respective company's and organization's website to present successful projects to interested customers or other stakeholders, they were requested for an interview first. We provided the experts with our guiding questions in advance in order to prepare for the interview. Table 3 contains the experts' profiles, including their assigned group (Group 1, companies and organizations offering AI for cybersecurity; Group 2, companies and organizations implementing AI in cybersecurity; Group 3, consulting companies and organizations). We added the experts' company and organization descriptions to provide transparency and robustness. The position describes the experts' jobs and assignments in the corresponding company and organization.

The core of our content analysis is text segmentation, where the coding unit (the smallest text segment to be categorized), the context unit (the most significant text component), and the recording unit (all text parts confronted with a category system) are identified. Mayring (2014) outlines three qualitative content analysis techniques: summarization

[5] https://www.crunchbase.com/

Table 3 Expert (E) profiles

Expert	Group	Company and organization description	Position
E1	1	Cybersecurity vendor (AI-based products and services)	Chief scientist/senior vice president
E2	2	Energy supplier	The lead of the IT infrastructure
E3	3	Consulting	Cybersecurity consultant, former Chief Information Security Officer (CISO) at an insurance company
E4	2	Retail and service company	Lead of a security monitoring team
E5	3	Cybersecurity services and consulting	Head for cybersecurity
E6	2	Bank	Head of the company, threat detection and response
E7	3	EY-Audit/Consulting	Partner and lead of the cybersecurity business
E8	1	Cybersecurity services	Red team solution lead
E9	2	Software development and consulting	Head of business development

(distilling key content), explication (clarifying ambiguous passages), and structuring (organizing text by predetermined criteria), each technique containing its process model. The analysis commences with determining the material, its origin, and its formal characteristics. RQs and theoretical background guide the creation of primary and subcategories, followed by a coding guideline with specific rules. The material is coded, and adjustments are made if necessary. After coding, the entire material is reviewed to ensure consistency. The final steps involve interpreting the results and applying quality criteria such as objectivity, reliability, and validity to ensure the analysis's rigor and credibility (Mayring, 2014; Serafini & Reid, 2023). We have chosen a content structuring analysis. Therefore, the text passages were screened for either advantages or disadvantages of AI in cybersecurity and the process phases that were asked during the SSIs. Challenges in the process, involved parties, KPIs, and guiding questions were selected as categories, see Table 6 in the appendix. This ensures that the topics and content relevant to the RQs can be filtered out. After 30% of the material, no revision of the coding categories was necessary. MAXQDA[6] software was used to codify the literature and the SSI transcripts. MAXQDA is a software for qualitative data analysis (VERBI Software, 2021).

Similar process models

Creating process models to introduce new solutions, services, and procedures is part of implementation science, particularly implementation support. In implementation science, there is a focus on comprehending relevant factors and how they influence the implementation processes. These process models enhance elements such as speed, quantity, and quality for the implementation process (Lobb & Colditz, 2013). Implementing support aims to employ tailored tools or training to enhance implementation processes, improving intervention outcomes (Durlak & DuPre, 2008). This paper focuses on process models for software acquisition and implementation. The process model for software application acquisition and implementation was developed based on expert interviews and serves as a roadmap for companies and organizations. The process of acquisition and implementation is spread into several process phases and criteria. The goal of process models for software application acquisition and implementation is to increase the perceived quality of the application while reducing the risk and cost compared to an intern development (Shin & Lee, 1996). Developing a process model for acquiring and implementing AI in cybersecurity is essential to support the technology adaptation for companies and organizations.

Tolido et al. (2019) display the implementation process in a six-step roadmap. As a first step, they mark the creation of a data platform and highlight the importance of keeping data up-to-date to generate a high-quality output of the AI algorithm. In the second step, companies and organizations must select suitable AI use cases with significant benefits. Then, a collaboration with threat researchers, security professionals, and peers recommends staying current on the industry's threats to improve AI algorithms. Tolido et al. (2019) suggest deploying a security orchestration, automation, and response (SOAR) as the fourth step. SOAR allows it to "collect security data and alerts from different sources" (p. 18), which assists in defining, prioritizing, and driving "standardized incident response activities" (Tolido et al., 2019, p. 18). The penultimate step marks the training of cyber analysts. According to half of their survey respondents, cyber analysts cannot improve the logic behind the algorithms. Another proposition is creating an accurate interface for the operation of incident alerts and AI applications and services to increase the efficiency of cyber experts. Installing governance for AI in cybersecurity displays the last step in

[6] https://www.maxqda.com/

Tolido et al.'s (2019) roadmap. This step includes defining the responsibilities and roles of the analysts, tracking AI outputs, and implementing KPIs.

Welte et al. (2020) discuss implementing ML for predictive maintenance in small- and medium-sized enterprises. Their process model does not deal with cybersecurity. Nevertheless, some aspects will be used in our process model. They divide the process into four phases. The first phase is the preparation phase. This phase starts with defining the use case, containing a problem analysis and an objective. The next step involves selecting an approach that determines the external partners. After deciding on the ML infrastructure (local IT software, hardware, or cloud solution), the project tasks must be assigned. The design phase follows as a second phase. This phase involves checking the availability of the data and preparing a training data set. After selecting the ML algorithm, the model is created and evaluated. The next phase describes the implementation phase of operating the ML model. As a final phase, a monitoring phase is proposed to review the ML model for refinement and apply KPIs to compare the model's performance.

Amershi et al. (2019) studied how different Microsoft software teams develop AI solutions. They discuss a nine-stage ML workflow (model requirements, data collection, data cleaning, data labeling, feature engineering, model training, model evaluation, model deployment, and model monitoring). Feedback loops are integrated into the fifth, seventh, and ninth stages. Another model is the Cross Industry Standard Process for Data Mining (CRISP-DM) (Saltz, 2021), which is still a common approach in data mining projects (Schröer et al., 2021). The iterative model consists of business understanding, data understanding, data preparation, modeling, evaluation, and deployment, including several feedback loops (Saltz, 2021).

Several process phases were identified during the literature review and general phases, which we adopted in our process model. We summarize the most frequently mentioned process phases in Table 4 and identify critical implementation phases. Additionally, we use the expert interviews to increase the robustness of identified critical phases, thus addressing RQ2.

Results and findings

Process model of AI implementation for cybersecurity

We identified critical phases to address RQ2 in Step 4 of our research design. Therefore, we combine experts' statements and frequently mentioned statements from the scientific literature. We conducted an iterative five-phase process model, as illustrated in Fig. 2. The first phase of the model is called *Requirement Analysis*. Security operations centers (SOCs) recognize a problem using signature-based IDSs since they do not detect many new threats or the company has recently been the victim of a cyberattack (Chen, 2023). The first step is identifying the company's and organization's cybersecurity problem (Aichele & Schönberger, 2015). Analyzing the suitability of AI-driven applications and services for the problem structure is crucial. Depending on the company's and organization's structure, the person responsible for cybersecurity measures, e.g., the CISO, must be involved in the problem (Chen, 2023). If the responsible person decides to address the problem, a project team will be initiated in Step 2 (Welte et al., 2020). According to the experts' statements, the following stakeholders must be included: the data privacy and compliance department, operational IT, information security with the responsible person, workers' council, procurement, strategic IT responsible for the IT portfolio

Table 4 Process steps identified in the literature

Process phase	References
Requirement analysis	Aichele and Schönberger (2015); Roohparvar (2023); Sikkel (2014); Wieczorrek and Mertens (2010)
Planning	Aichele and Schönberger (2015); Brewer and Dittman (2018); Harvard Business Review (2016); Wieczorrek and Mertens (2010)
Employees' training	Das and Sandhane (2021); Harvard Business Review (2016); Roohparvar (2023); Ross (2018)
Resources coordination	Aichele and Schönberger (2015); Brewer and Dittman (2018); Harvard Business Review (2016); Welte et al. (2020)
Setting goals	Aichele and Schönberger (2015); Harvard Business Review (2016); Welte et al. (2020)
Project monitoring	Brewer and Dittman (2018); Harvard Business Review (2016); Roohparvar (2023)
Implementation	Harvard Business Review (2016); Roohparvar (2023); Sikkel (2014)
Problem definition	Harvard Business Review (2016); Welte et al. (2020)
Project evaluation	Harvard Business Review (2016); Wieczorrek and Mertens (2010)
Closing	Brewer and Dittman (2018); Harvard Business Review (2016)
AI selection	Roohparvar (2023); Welte et al. (2020)

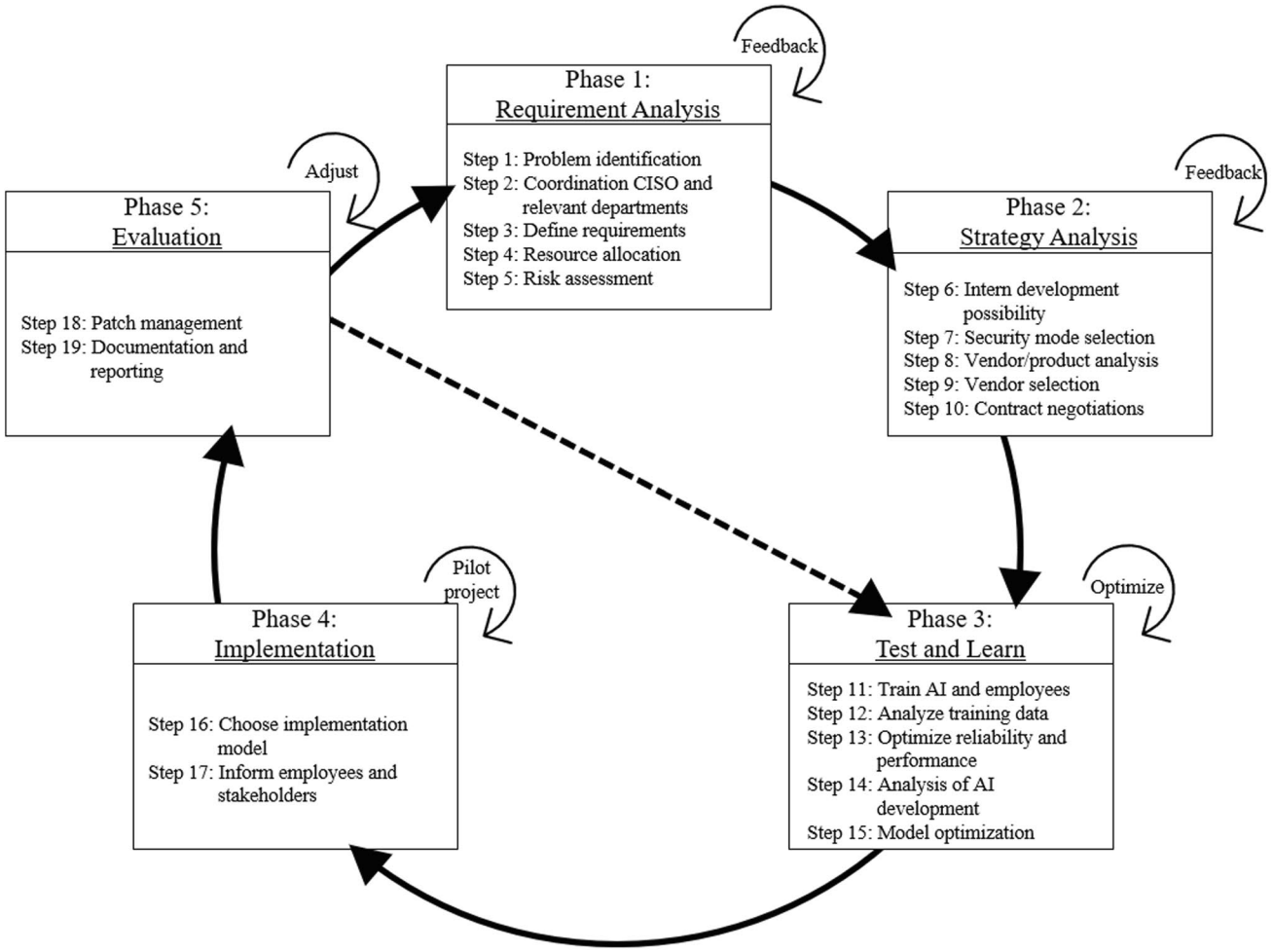

Fig. 2 Iterative five-phase process model

management, and a project manager (E3, E5, E6, E7, and E8). Responsibilities must be defined according to the operational area of the AI applications and services. Based on the findings of the problem, requirements can be developed (Aichele & Schönberger, 2015). The project team defines minimum requirements to speed up the evaluation of solution and service proposals in the third step. Those requirements include functional (e.g., what to track and model) and non-functional requirements (e.g., required reliability and efficiency) and features (e.g., report extraction) necessary for the efficient operation of the new applications and services to be developed (Aichele & Schönberger, 2015). Including stakeholders in developing requirements ensures that different stakeholder groups fulfill various requirements. In Step 4, resources must be allocated. Resource allocation includes budget planning and defining the project's deadline and future behavior (Aichele & Schönberger, 2015; Welte et al., 2020). For budget planning, a budget must be established to acquire and implement the new cybersecurity application and service and for the long-term cost. In addition to the long-term cost, the continuation of operations and further development must be determined (Jadhav & Sonar, 2011; Welte et al., 2020). The price of AI and license fees must be compared to the budget. The costs are compared with the expected benefits in both states, according to E3 and E5. Project scope planning covers the period from the first step of problem identification to successful implementation (Aichele & Schönberger, 2015). It is essential to allocate infrastructure and data to train and test the model. Dedicated servers or cloud capabilities are needed to test and train the model in a secure environment (E9). The company and organization must verify their human resources and capabilities to implement or develop an AI-based application and service. The resources are reviewed, including performance and data storage (Nortje & Grobbelaar, 2020). Determining the data to be used to train the model is crucial. The final step in the first phase involves risk assessment. Potential risks within the implementation process and the product itself are identified. Returning to the problem statement, the project team has to define a set of project targets (Aichele

& Schönberger, 2015). Additionally, the group develops an evaluation guideline consisting of criteria to assess the performance and reliability of the product. For this, they must choose KPIs to evaluate the solution's effectiveness. This phase includes a feedback loop-based approach as an iterative process. The final requirements are evaluated by top management and stakeholders. Based on feedback, new adaptations (e.g., to the requirements, cybersecurity strategy, and problem definition) and the addition and removal of requirements can result in a new iteration of the processes within the phase.

The second phase, *Strategy Analysis*, involves finding a solution. Step 6 includes an assessment of the possibility of an intern development or an external vendor's product development. Suppose no resources are available or other arguments against internal development exist. Implementing new cybersecurity measures requires identifying a suitable AI for cybersecurity from external vendors (Aichele & Schönberger, 2015). The search process requires scanning the market for current applications and services. In Step 7, the security mode is chosen and developed (Yaseen, 2023). That includes classifying threats in which an AI for cybersecurity can respond autonomously to threats and intrusions and those it has only the permission to detect and indicate threats (Kairu & Shin, 2022; Mughal, 2018). In this step, the type of AI-driven security measures to be deployed must be defined. Examples of these modes are ML-based anomaly detection, capable of detecting anomalies in network data streams based on training with standard business data, and natural language processing-based threat detection, detecting anomalies and potential threats in emails or chats (Yaseen, 2023). The project team conducts a product and vendor analysis for external applications and services in Step 8. Vendor selection in Step 9 assumes that a company and organization decide to purchase an AI for cybersecurity. Some vendors request suggestions after the company and organization have concentrated on current applications and services (E4). For vendor selection, the vendors identified during market screening present their applications and services. After introducing the applications and services, suggestions are compared with the requirements list (Jadhav & Sonar, 2011). The project team analyzes if the proposed applications and services satisfy the minimum requirements and provide the functions determined by the team (Jadhav & Sonar, 2011). Analyzing the applications and services is crucial to solving current problems and preparing for future cybersecurity risks; it is also essential to analyze the vendors' business and functional requirements by analyzing the vendor's corporate structure (Jadhav & Sonar, 2011). This analysis includes the financial and innovation performance. Companies and organizations aiming to implement long-term solutions and partnerships must analyze the vendor's stability, supported by E2. It is necessary to identify and assess the financial risks to the vendor that can result in the cancellation of the implementation or the termination of use. Since requirements and cyber threats change over time, the AI for cybersecurity must be updated in the future to stay up-to-date and ensure a level of service. E4 states that the trustworthiness and reliability of the vendor and its innovational capabilities must be assessed. To ensure trustworthiness, it is also necessary to verify the certifiability of the product and the selected company to ensure high qualitative results (Thiebes et al., 2021). There may already be independent certifications from authorities or auditing companies that support this. Data protection must take priority. The management and the project team decide which data will be processed and where, supported by E5, E6, and E8. A solution of data processing methods is necessary (Lier et al., 2023). E9 mentioned two possibilities for AI usage. The first possibility is to transfer the data to a cloud infrastructure. The second possibility is the deployment "on an edge device like a server" (E9), which will be placed behind the core switch. To ensure data safety, data storage must be regulated. In the literature and expert interviews, adversarial attacks were cited as the most severe challenge for AI in cybersecurity. The project team must evaluate whether a proposed AI meets all the (minimum) requirements or exceeds the requirements (Jadhav & Sonar, 2011). If the AI monitors networks and clouds and can autonomously respond, a legal question is a responsibility in case of a failure of the AI (Lier et al., 2023; Meske et al., 2022; Thiebes et al., 2021). In addressing these issues, a company and organization choose different AIs for cybersecurity to test and evaluate. In Step 10, contract negotiations will occur, and a proof of concept will be agreed upon to demonstrate the functionality (Jadhav & Sonar, 2011). The phase uses feedback loops for in-phase iterations to minimize risk error repetitions and to further integrate stakeholders. Finally, the make-or-buy decision is made.

Test and Learn describe Phase 3. The AI will be deployed into the infrastructure reserved for AI testing. The AI learns how ordinary users and participants in a network behave (Yaseen, 2023). It includes knowing how many endpoints, servers, access points, subnets, and IP addresses a company owns, so the AI has a learning process in Step 11. After the AI has the required knowledge of the typical network behavior, the AI for cybersecurity is tested in a controlled and corporate environment, e.g., a sandbox environment for only a few users. The AI model's performance can be tested by selecting a team to attack the company and organization infrastructure to determine the AI's response and duration to send an alert (E4). Therefore, to analyze the results, explain the AI model and how adaptation can be made if there are false positive or negative mitigations (Step 12). The AI must be supervised to analyze these issues. The training of employees for the

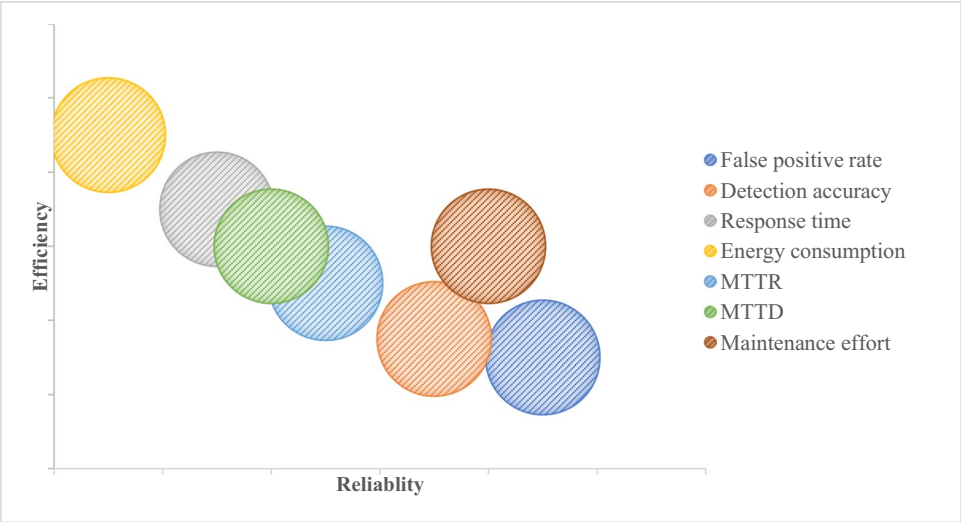

Fig. 3 Key performance indicators distribution

operation and maintenance of the AI for cybersecurity begins in conjunction with the training of the AI model. After testing the AI, it is necessary to optimize its reliability and performance (Step 13). The chosen KPIs can evaluate the AI solution's performance, e.g., false positive rate and detection accuracy (Yaseen, 2023). A final assessment occurs if the AI development analyzes the initially determined project targets (Step 14) (Aichele & Schönberger, 2015; Welte et al., 2020). This phase ends by optimizing the AI in multiple iterations until the reliability and confidence in the model exceed the requirements (Step 15). The steps in Phase 3 can be repeated based on monitoring the performance of the AI models during the training phase. If the AI results do not correspond to the requirements, Step 12 has to be restarted. The training data can be analyzed and adjusted to refine the AI for cybersecurity in the following steps.

An AI for cybersecurity must be selected in Step 16 of the *Implementation* phase. Implementation is required after selecting a suitable AI. A model must be created for the implementation to determine the integration of the selected AI for cybersecurity into the IT infrastructure. Stakeholders are informed about the risks and potential misconduct of the implemented AI in Step 17 (Welte et al., 2020). This provides stakeholders with a deeper understanding of the solution's behavior in detecting and preventing potential threats, thus resulting in increased confidence in the new AI for cybersecurity (E1 and E9). The *Implementation* phase is complemented by a pilot project loop to analyze and decide on implementation strategies (Welte et al., 2020). The employees using the new AI for cybersecurity will be trained in the pilot project. The phase is complete when the stakeholders consider the pilot project successful (Brewer & Dittman, 2018). During the implementation of the new AI in the pilot phase, there are factors arising from the needs of the employees and other needs and requirements of the stakeholders involved in the pilot phase that require a reassessment of the new AI and the implementation method. This leads to an iterative process in the implementation phase itself.

The *Test and Learn* and *Implementation* phase can be evaluated by measuring the new AI's cybersecurity performance. Monitoring the performance of the new AI for cybersecurity and generating reports is part of the Phase 5 *Evaluation*. Patch management is specified in Step 18. To avoid disrupting business operations by ensuring the availability of the most recent cybersecurity application and service releases, patch cycles and roll-out strategies are developed and implemented in the maintenance processes after the new cybersecurity measures are implemented (Cavusoglu et al., 2008). Reporting and monitoring guidelines are developed to assess operations and monitor the reliability and effectiveness of the new application (Welte et al., 2020). This ensures the evaluation of the product processes and the long-term assessment of the reliability and efficiency of the new security measures (Step 19). Iterative adjustments based on iterative evaluation can address user-friendliness, user interface, autonomy, and possible applications. If significant changes are required, there can be interaction with Phase 3. If Phases 1 and 2 are suitable for AI in cybersecurity, the first two phases can be omitted using new test data sets, and Phase 3 can be repeated. If an adjustment is completed in the *Evaluation* phase, Phase 1 must be performed. Due to the further enhancement of threats, dangers, and attacks on cyberspace, new applications and services must be continuously integrated into cybersecurity to ensure the highest possible level of protection. Therefore, our process model is cyclical and can be continued anytime.

Key performance indicators

To address Step 6 of our research design, the experts were asked which KPIs they would use to evaluate the performance of the AI for cybersecurity (Fig. 3). The experts have suggested various possible KPIs. The first KPI is *response time* (E3, E4, E7, and E8; Yaseen, 2023). The *response time* measures the resources the AI for cybersecurity needs. It indicated how responsive the infrastructure's performance changes after implementing new AI for cybersecurity. New AI-driven applications and services must not influence business operations to the extent that their performance is negatively affected (Shah, 2021). The second KPI can be derived as an ecological footprint called *energy consumption*. E8 noted that *energy consumption* for AI training and energy used to execute the requests are essential for successful AI implementation. The most mentioned KPI is the *false positive rate* (E1, E2, E4, E5, E6, and E7; Yaseen, 2023). The *false positive rate* measures the false alarms for adversarial attacks within the infrastructure. This means that AI aims to minimize these problems as much as possible. In addition, it is also possible to use the *false positive rate* to determine the actual positive rate, which can be used as a countermeasure for further evaluation. E7 stated that the *false positive rate* must be less than one percent of all the triggered events in a month. Another KPI is described by the *detection accuracy* of threats (E2 and E3; Yaseen, 2023). This indicator is used to assess the correctness of the model. It measures the number of identified threats compared to the overall occasions. E2 compared the solution of AI in network monitoring with the mechanism of virus scanners, which requires that all viruses are detected. E2 also highlighted that the *detection accuracy* of AI must be compared to human assessments. E3 determined that a threat in IT can exist for up to 300 days. Therefore, the AI must be able to recognize them. Additional *Mean-Time-to-Detect (MTTD)* and *Mean-Time-to-Respond (MTTR)* (E3 and E7) measure the solutions' ability to detect and, after successful detection of threats, respond to them. The *MTTD* measures the time the AI for cybersecurity takes from incident to successful detection. If an incident or threat is not detected quickly, no response can be developed, and the potential damage to data integrity increases. *MTTR* is a KPI for AI for cybersecurity that can respond autonomously to detected threats. It measures the time required for the AI to respond to or eliminate a threat after it has been successfully detected. Since detection and response can vary for different threats, the average is measured to provide a more straightforward reference point for monitoring and reporting. Detecting and responding to threats to IT infrastructure quickly is critical to protect corporate data and minimize the potential damage associated with these threats. Finally, *maintenance efforts*, e.g., debugging, indicate the robustness of an AI model and operational cost after implementation (E5; Roshanaei et al., 2024). The *maintenance effort* required indicates the model's weaknesses and assesses the system's ability to respond to long-term changes and unexpected errors. Low maintenance effort indicates the solution's high efficiency and scalability, leading to lower operating cost and higher user satisfaction (E5).

Based on the scientific literature and SSIs, we evaluated the KPIs for their focus on measuring efficiency and reliability. For this purpose, we have evaluated them using a two-dimensional approach. Efficiency evaluates the model's performance regarding resource requirements (e.g., energy and computing power). The reliability of the AI model evaluates the effectiveness of the model in generating and maintaining the required cybersecurity measures and meeting the requirements of our process model. Based on the experts' statements and the references within our iterative five-phase process model, we assigned a weighting to the KPIs in a two-dimensional diagram. We asked the experts to provide a statement for each KPI regarding efficiency and reliability and justify it. Then, independent of the other KPIs, they were asked to create an allocation in our two-dimensional diagram. We individually compared the experts' statements and the allocation, analyzed the justifications, and formed an average of the KPI bubbles based on the allocations. Due to the explanations and discussions with the experts, there was a high level of agreement in the assignment of our two-dimensional diagram, and no outliers occurred. In Fig. 3, the vertical axis indicates efficiency. When a KPI is on the lower end of the scale, its efficiency is low. On the horizontal axis, reliability is indicated. The further to the left a KPI is located, the lower its reliability. For example, KPI *energy consumption* has a high efficiency with low reliability, KPI *false positive rate* has a high reliability with a lower efficiency, and KPI *MTTR* has an equal balance between medium efficiency and medium reliability. The KPI *energy consumption* and *response time* focus on the efficiency of an AI for cybersecurity as they measure the performance of the model and its impact on the business infrastructure. *Response time* also focuses on the model's reliability, as transient responsiveness of the enterprise infrastructure indicates some reliability issues of the solution and service. *MTTR* and *MTTD* center on the efficiency and reliability of the model, as they are time-bound measures of threat response. The maintenance efforts focus on the efficiency and reliability of the model. They measure issues regarding the model's operation and the need for adjustments. Key metrics include the frequency of issues encountered, the necessity for model updates,

and the time spent refining and troubleshooting the solution. *Detection accuracy* and *false positive rate* focus on the reliability of the model. They measure performance regarding the ability to protect the organization's IT infrastructure.

Evaluation and adjustment

Evaluating results and findings is an indispensable component of DSR (Gregor et al., 2020). It ensures our process model's understanding, traceability, and usability (vom Brocke et al., 2020). For an evaluation, we conducted a focus group discussion. The focus group consists of five experts from different cybersecurity sectors. All five experts are primarily responsible for cybersecurity in their companies and organizations and could assess our AI-based process model and KPIs development. In addition, all experts have at least gained experience with AI in their academic or professional careers. The experts were contacted through a networking platform and invited to an online meeting lasting about 90 min. The discussion was recorded and transcribed to ensure that no statements were lost or distorted. Table 5 presents the five experts, including the accompanying group (Group 1, companies and organizations offering AI for cybersecurity; Group 2, companies and organizations implementing AI in cybersecurity; Group 3, consulting companies and organizations) experts' company and organization description and their position.

The experts were introduced to the topic, and our research design was presented. The experts had already received our iterative five-phase process model with an explanation in advance to prepare for the discussion and discuss with other employees in their company and organization. Our process model was discussed and analyzed, and no changes were to be made. However, explanations in the text were specified and already introduced in the corresponding section. In particular, omitting Phases 1 and 2 after a successful adjustment in the *Evaluation* phase was considered sensible and innovative by the experts. E14 said, "Such a step is often forgotten. These cycles are often started from the beginning. A second cyclical process in a cyclical process is innovative and clearly stands out from other frameworks." E12 and E14 emphasized the *Test and Learn* phase, as the experts consider employee training particularly often inadequate. E10 emphasized the *Strategy Analysis* phase because the selection of vendors (Step 9) and the previous analysis (Step 8) are often skipped or poorly implemented. Ambiguities were clarified with experts E11 and E12 based on both feedback loops in Phases 1 and 2, and a corresponding description and demarcation were adopted. Our iterative loops after each phase were considered meaningful, strategic, and practical (E10, E13, and E14).

In addition, our derived KPIs were presented to the focus group. The experts did not receive an assignment of our KPIs in our two-dimensional diagram in advance to avoid influence. KPIs and the two-dimensional diagram were explained. KPIs were discussed and positioned in our two-dimensional diagram within the focus group. In order not to influence the assignment of the KPIs, we only encouraged discussion and, if necessary, provided arguments for the KPIs. We found two minor deviations from our results. The focus group set the KPI *detection accuracy* higher than we did because, although they associate this point with strong reliability, it also can significantly increase efficiency. E13 said: "If we get reliable detection accuracy, then we have a domino effect. Quickly detecting threats, we […] can act efficiently and launch a defense, automatically increasing efficiency." Our KPIs *MTTD* and *MTTR* were heavily discussed. E10 and E12 considered that both KPIs should be given equal weight in the diagram. E11, E13, and E14 argued that the time of the AI for cybersecurity from incident to detection demands more efficiency than reliability. In addition, *MTTD* promotes both efficiency and reliability. The focus group discussion all agreed that *MTTR* and *MTTD* should remain in the positioning while setting *detection accuracy* higher. The changes are included in Fig. 3.

Table 5 Expert (E) profiles for evaluation and adjustment

Expert	Group	Company and organization description	Position
E10	2	Innovation driver for safe, connected, and sustainable mobility	Cybersecurity data consultant
E11	3	Consulting and professional services company focusing on information technology	Cybersecurity business and technology specialist
E12	3	Consulting and services company specializing in IT security, IT security concepts, and the automotive industry	Cybersecurity engineer
E13	2	A leading global provider of transportation and logistics services	Cybersecurity governance specialist
E14	1	Companies for industry-specific products and services in the IT sector	Security specialist and software developer

Comprehensive analysis and future directions

Discussion

We have developed a process model to implement AI successfully in cybersecurity, including different loops in our five phases (Fig. 2). The *Evaluation* phase is an essential iteration in the implementation process because of the performance of the tested AI for cybersecurity. Therefore, known metrics such as false positives can be used. The decision to select AI for cybersecurity depends on specific parameters. Most experts (E1, E3, E4, E5, E7, E9, E10, E11, E12, E13, and E14) agreed that if several providers can fulfill the requirements and the costs are similar, additional aspects must be considered, e.g., certifications regarding a provider's trustworthiness. There are currently no generally applicable certifications, as this area is unregulated (Jawhar et al., 2024; Rampášek et al., 2025). E8 mentions another aspect: "I would also consider the environmental impact, like how much energy one requests to use, because that is also a concern. […] So, that could be a way for your corporate social responsibility solution and service to only use this energy per request." This enables them to distinguish themselves from other providers. Training AI for cybersecurity is resource-intensive and requires graphics processing units that require much energy (E8, E9, E10, and E13).

Based on the SSIs, comparing AI to non-AI-based applications and services in the implementation process is critical. Most experts stated that the installation or deployment would not be different. E9 stated that AI might be more accessible to deploy, "AI-based tools are far easier to deploy because they have been designed to understand autonomously where they are connected […]." Conversely, E3 was critical of complete AI-driven cybersecurity solutions: "AI-based tools are complex. It is not that easy to implement them in a targeted manner and to use them appropriately. This requires experts, and not everyone has them in sufficient numbers." This aspect is controversial, given that a company and organization with a multi-vendor strategy use modular solutions. After the vendor has presented his AI for cybersecurity in the *Test and Learn* phase, a proof-of-concept is agreed upon to verify the AI solution's functionalities. This proof-of-concept can fail due to a high false positive rate, even after several months of training the AI in the corporate environment (E1, E4, E10, E12, E13, and E14). In this case, the decision-makers can abandon the objective of introducing AI for cybersecurity. The advantages identified in the scientific literature and the SSIs overlapped considerably, allowing us to find the decisive advantages. There were fewer similarities regarding the disadvantages. While AI manipulation was highlighted in the literature (Table 2) and found most frequently, data protection was more relevant in the SSIs (E1, E3, E4, E5, E6, E7, E10, E11, E12, and E14). The experts perceive potential manipulation by AI as a concern but not as important as data protection issues. In our focus group discussion, E12 (cybersecurity engineer) and E13 (cybersecurity governance) had different opinions. E12 agreed with the danger of data misuse, it is "[…] an ever-recurring danger. As soon as a system does not work properly, this reduces reliability and data is compromised. Especially with sensitive data […], the protection of data is the top priority. Manipulation […] is much more complex and difficult in an AI system." E13 responded directly to this statement: "Of course, data protection is always a high priority. However, the risk of manipulation is significantly higher than the relevance of data protection. If a system is manipulated, then it can come to complete failure. That would be an economic disaster." It is apparent that the different positions of the experts in their companies require different opinions. In the focus group discussion, E11 (cybersecurity business and technology specialist) said in response to the discussion of the two experts: "Well, it depends on which point of view you look at it from. As a technology specialist, I see the risk of manipulation lower than the relevance of data protection. […] As a business specialist, you have to weigh them up. I am responsible for both, and I can understand both opinions." Adversarial attacks pose more of a theoretical threat, as it is difficult to successfully attack the algorithms behind these AI for cybersecurity (Rosenberg et al., 2021). Attackers can dive more deeply into adversarial attacks as more companies migrate to AI for cybersecurity. E7 stated, that "just a development of own AI model is the best possible solution. But it is rarely probable for a company to develop its own AI model." Therefore, the experts have confirmed that many providers will gain customers. Due to the increase in data, the types of threats and correlations must always be up-to-date. E1 emphasizes that learning the algorithms through reinforcement learning is essential for improving the algorithms and providing more protection. E7 noted that attackers can use AI-based programs like ChatGPT to create malware codes and automated phishing emails. Programs like this raise the risk of a self-learning algorithm finding new methods of successfully attacking companies more frequently. E9 raised the question regarding liability in the event of a model failure. He discussed the desirability of a company holding the supplier responsible for a failure, e.g., shutting down the entire network. A non-AI company has a technological advantage over the actual situation. Such clauses would represent an additional advantage that is not realistic. For this reason, there must be an awareness that these models can fail, as E1 mentioned: "I think these models can fail in unpredictable and unique and weird ways." Supporting a cybersecurity analyst after receiving an alert is seen as a market gap by E4. He mentions that companies

and organizations are trying to write runbooks for the analyst before deciding. The problem is people's limited know-how. Runbooks are a valuable solution for new threats. Regarding the E3 statement to use AI-based programs like ChatGPT, tailored explicitly to the cybersecurity analyst in their applications services, this statement must be discussed because of missing knowledge and research. Currently, companies and organizations tend to automate the first-level analysis (E7), meaning the AI does the "dirty work" (E3) and assists the security analysts with the "low-hanging fruits" (E4). It involves AI for cybersecurity that recognizes threats but does not react independently. In that scenario, analysts will receive an alert that a threat has been detected, and the analysts will decide on further actions. The analyst may be missing knowledge if this threat is new, and the runbook may have no necessary information. AI for cybersecurity can be a possible approach to this barrier, as it can detect and identify threats faster and initiate countermeasures.

Implications and recommendations

Our iterative five-phase process model guides companies and organizations to implement AI in cybersecurity successfully. It serves as a foundation for further research and development and improves efficiency, cost, and quality of implementation. Our critical phases and steps contribute to current and future implementation processes and are grounded on frequently cited scientific literature (Table 4).

Our KPIs enable us to monitor processes and objectives, ensuring transparency. The combination of our process model and KPIs applies to research and practice. Researchers can systematically investigate AI integration, analyze interactions between AI and security processes, and benchmark implementations. Our KPIs allow us to quantify performance, identify weaknesses, and develop new RQs.

For practitioners, our process model and KPIs offer structured guidance to implement AI in a targeted, company-centric way. Monitoring KPIs ensures system efficiency and optimal resource use, improves performance and cost-effectiveness, and thus supports evidence-based decisions on adjustments and investments. Our framework supports early risk identification, strengthens the quality of AI-supported security solutions, increases cyberattack resilience, and minimizes downtime. For regulated companies and organizations, we assist in demonstrating compliance and generating audit-ready documentation.

Our framework supports systematic investigation and practical application, enabling efficient, risk-aware AI integration in cybersecurity. Companies and organizations can specifically adapt and extend our process model, while researchers can generally refine it to explore advanced solutions. The iterative model and KPIs offer an evidence-based approach for adaptive AI implementation, empowering data-driven decisions in IT strategy, resource allocation, and risk management. Integrating KPIs into internal controls ensures continuous monitoring and adaptation to emerging threats and regulations.

Policymakers must establish certification standards, clarify liability frameworks, and ensure transparent governance of AI in cybersecurity-critical sectors. Identified gaps, e.g., the lack of uniform certification procedures and legal uncertainties, highlight the need for political action to foster innovation while maintaining security and transparency. We offer a practical and transferable foundation for company and organization guidelines, funding strategies, and regulatory standards. We support AI's strategic, responsible, and scalable integration into cybersecurity across science, practice, and policy.

Limitations and further research

The experts were familiar with AI in the cybersecurity field, but some had no direct experience with the applications and services or their implementation. The technical requirements for implementing AI in cybersecurity are barely covered. Further research is therefore necessary, and the further RQ (FRQ) arises FRQ1: "Which technical requirements are necessary for the implementation of AI for cybersecurity?" For this purpose, more interviews must be conducted. The selection of experts must expand to include experts from IT security architecture so that our process model can become generalized and expanded. FRQ2: "Which requirements and process steps are mentioned by IT security experts for the architecture of an AI for cybersecurity?" Our process model is designed to be general in order to cover various AI applications and services. Different process model phases are necessary for a different IDS than malware detection. The adaptation of the model varies for different implementations. The phases and steps can have a higher or lower accuracy of suitability for different AI in cybersecurity and must be modified. This leads to FRQ3: "How can our iterative five-phase process model be specifically transferred, and what changes need to be made?" Based on the accuracy rate via Crunchbase, companies and organizations have a rather averse attitude toward designing AI for cybersecurity. Most use security experts to recognize tasks or use AI for cybersecurity as an add-on from providers. The mistrust in AI is a significant barrier to the implementation of AI applications and services and leads to inadequate protection of critical data and company and organization assets; FRQ4: "How can the trustworthiness of AI for cybersecurity be strengthened?" Explainable AI (XAI) can be a solution by addressing the black box problem; XAI enhances understanding of AI decision-making, increasing confidence among cybersecurity experts and management (Gerlach et al., 2022b; Lier

et al., 2023; Meske et al., 2022; Mughal, 2018). It addressed some of the concerns raised in the expert interviews about the trustworthiness of AI. Using XAI, bottlenecks such as management's understanding of the application processes and decisions can be reduced (Lier et al., 2023). A better understanding of AI functions, processes, and decisions will promote business, advancing cybersecurity's in-house development of AI applications and services (Gerlach et al., 2022a, b). This leads to FRQ5: "How can XAI be implemented in AI for cybersecurity, and which advantages arise?".

Conclusions

We have developed a process model for successfully implementing AI in cybersecurity. A systematic and efficient literature review and nine SSIs were conducted to create a unique knowledge base. This combination of academic and practical perspectives enabled the identification of real-world needs and current research gaps. We identified the advantages and disadvantages of AI for cybersecurity and processes for an AI addressing RQ1. Based on different published processes with phases and steps, we identified critical phases and developed a new process model comprising five critical phases and 19 steps to address RQ2. Our five-phase process model is iterative and structured cyclically with interactions due to changing technologies in threats, attacks, and damage. To address RQ3, we identified seven KPIs based on our SSIs and literature review and categorized them in our two-dimensional diagram of efficiency and reliability. The *Evaluation* phase is crucial within the iterative model.

Through our systematic and efficient literature review, which was conducted database-based and AI- and graphic-based, we provided the status quo of AI for cybersecurity. We highlighted AI's advantages (Table 1) and disadvantages (Table 2). Our expert interviews clarified the use of AI-based applications and services. All experts have many years of experience in their fields. The experts for the knowledge base received a semi-structured questionnaire and were asked about possible KPIs and their assignments. The evaluation experts were only sent the process model in advance; the assignment and presentation of the KPIs took place during the focus group discussion. Despite the wide range of opinions, we achieved a high saturation of very similar statements in the fourteen interviews, ensuring the rigor of our results. Through expert interviews and a literature review, we combined the perspectives of practitioners and researchers. We conducted a focus group discussion to evaluate and adjust our findings. As a result, our outcomes are robust and examined, adjustments were added, and we increased usability, understanding, and traceability.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12525-025-00802-x.

Funding Open Access funding enabled and organized by Projekt DEAL. The research project "SiNED – Systemdienstleistungen für sichere Stromnetze in Zeiten fortschreitender Energiewende und digitaler Transformation" also received support from the Lower Saxony Ministry of Science and Culture through the "Niedersächsisches Vorab" grant programme (grant ZN3563) and of the Energy Research Centre of Lower Saxony.

Declarations

Competing interests There are no competing interests.

References

Adams, W. C. (2015). Conducting semi-structured interviews. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (3rd ed., pp. 492–505). Wiley. https://doi.org/10.1002/9781119171386.ch19

Aftergood, S. (2017). Cybersecurity: The cold war online. *Nature, 547*, 30–31. https://doi.org/10.1038/547030a

Aichele, C., & Schönberger, M. (2015). IT-Projektmanagement: Effiziente Einführung in das Management von Projekten. *Springer*. https://doi.org/10.1007/978-3-658-08389-2

Ali, A., Septyanto, A. W., Chaudhary, I., Al Hamadi, H., Alzoubi, H. M., & Khan, Z. F. (2022). Applied artificial intelligence as event horizon of cyber security. *Proceedings of the International Conference on Business Analytics for Technology and Security.* https://doi.org/10.1109/ICBATS54253.2022.9759076

Aloqaily, M., Kanhere, S., Bellavista, P., & Nogueira, M. (2022). Special issue on cybersecurity management in the era of AI. *Journal of Network and Systems Management, 30*(39), 7. https://doi.org/10.1007/s10922-022-09659-3

Alsayed, A., & Bilgrami, A. (2017). E-banking security: Internet hacking, phishing attacks, analysis and prevention of fraudulent activities. *International Journal of Emerging Technology and Advanced Engineering, 7*(1), 109–115.

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, M., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. In *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice,* 291–300. https://doi.org/10.1109/ICSE-SEIP.2019.00042

Ansari, M. F., Dash, B., Sharma, P., & Yathiraju, N. (2022). The impact and limitations of artificial intelligence in cybersecurity: A literature review. *International Journal of Advanced Research in Computer and Communication Engineering, 11*(9), 81–90. https://doi.org/10.17148/IJARCCE.2022.11912

Badhwar, R. (2021). The CISO's next frontier: AI, post-quantum cryptography and advanced security paradigms. *Springer*. https://doi.org/10.1007/978-3-030-75354-2

Berman, D. S., Buczak, A. L., Chavis, J. S., & Corbett, C. L. (2019). A survey of deep learning methods for cyber security. *Information, 10*(4), 35. https://doi.org/10.3390/info10040122

Bérubé, M., Giannelia, T., & Vial, G. (2021). Barriers to the implementation of AI in organizations: Findings from a Delphi study. *Proceedings of the 54th Hawaii International Conference on System Science*. https://doi.org/10.24251/HICSS.2021.805

Brewer, J. L., & Dittman, K. C. (2018). *Methods of IT project management*. Purdue University Press. https://muse.jhu.edu/pub/60/book/83907

Cavusoglu, H., Cavusoglu, H., & Jun, Z. (2008). Security patch management: Share the burden or share the damage? *Management Science, 54*(4), 657–670. https://doi.org/10.1287/mnsc.1070.0794

Charanarur, P., Gundu, S. R., spsampsps Sahu, M. (2025). Cybersecurity-based artificial intelligence healthcare management system. In S. Deb, spsampsps A. K. Sahu (Eds.), *Securing the Digital World*, (1st ed., pp. 128–140). Taylor spsampsps Francis Group. https://doi.org/10.1201/9781032663647

Chen, K. C., Lin, S. C., Hsiao, J. H., Liu, C. H., Molisch, A. F., & Fettweis, G. P. (2020). Wireless networked multirobot systems in smart factories. *Proceedings of the IEEE, 109*(4), 468–494. https://doi.org/10.1109/JPROC.2020.3033753

Chen, X. (2023). Application of artificial intelligence in network security and network Defense. *Proceedings of the International Conference on Computer Simulation and Modeling, Information Security*. https://doi.org/10.1109/CSMIS60634.2023.00047

Craigen, D., Diakun-Thibault, N., & Purse, R. (2014). Defining cybersecurity. *Technology Innovation Management Review*, *4*(10), 13–21. https://doi.org/10.22215/timreview/835

Dangi, A. K., Pant, K., Alanya-Beltran, J., Chakraborty, N., Akram, S. V., & Balakrishna, K. (2023). A review of use of artificial intelligence on cyber security and the fifth-generation cyber-attacks and its analysis. In *Proceedings of the International Conference on Artificial Intelligence and Smart Communication,* 553–557. https://doi.org/10.1109/AISC56616.2023.10085175

Das, R., & Sandhane, R. (2021). Artificial intelligence in cyber security. *Journal of Physics: Conference Series, 1964*(4), 11. https://doi.org/10.1088/1742-6596/1964/4/042072

De Azambuja, A. J. G., Plesker, C., Schützer, K., Anderl, R., Schleich, B., & Almeida, V. R. (2023). Artificial intelligence-based cyber security in the context of industry 4.0 – A survey. *Electronics*, *12*(8), #1920. https://doi.org/10.3390/electronics12081920

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350. https://doi.org/10.1007/s10464-008-9165-0

Eggers, S. (2021). A novel approach for analyzing the nuclear supply chain cyber-attack surface. *Nuclear Engineering and Technology, 53*(3), 879–887. https://doi.org/10.1016/j.net.2020.08.021

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

Fay, R., & Trenholm, W. (2019). The cyber security battlefield AI technology offers both opportunities and threats. Centre for International Governance Innovation. *Governing Cyberspace during a Crisis in Trust*, 45–48. https://www.jstor.org/stable/pdf/resrep26129.11

Gerlach, J., Hoppe, P., Jagels, S., Licker, L., & Breitner, M. H. (2022a). Decision support for efficient XAI services-A morphological analysis, business model archetypes, and a decision tree. *Electronic Markets, 32*(4), 2139–2158. https://doi.org/10.1007/s12525-022-00603-6

Gerlach, J., Werth, O., & Breitner, M. H. (2022b). Artificial intelligence for cybersecurity: Towards taxonomy-based archetypes and decision support. *Proceedings of the 43rd International Conference on Information Systems*. https://aisel.aisnet.org/icis2022/security/security/10

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337–355.

Gregor, S., Chandra Kruse, L., & Seidel, S. (2020). Research perspectives: The anatomy of a design principle. *Journal of the Association for Information Systems, 21(6)*, 1622–1652. https://doi.org/10.17705/1jais.00649.

Gunduz, M. Z., & Das, R. (2020). Cyber-security on smart grid: Threats and potential solutions. *Computer Networks*, *169,* #107094. https://doi.org/10.1016/j.comnet.2019.107094.

Hamm, P., & Klesel, M. (2021). Success factors for the adoption of artificial intelligence in organizations: A literature review. *Proceedings of the 27th Americas Conference on Information Systems*. https://aisel.aisnet.org/amcis2021/art_intel_sem_tech_intelligent_systems/art_intel_sem_tech_intelligent_systems/10/

Hamon, R., Junklewitz, H., & Sanchez, I. (2020). Robustness and explainability of artificial intelligence: From technical to policy solutions. *Publications Office of the European Union*. https://doi.org/10.2760/57493

Handa, A., Sharma, A., & Shukla, S. K. (2019). Machine learning in cybersecurity: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(4), #1306. https://doi.org/10.1002/widm.1306

Harvard Business Review. (2016). *The four phases of project management*. https://hbr.org/2016/11/the-four-phases-of-project-management. Accessed 30 Aug 2024.

Hevner, A., & Chatterjee, S. (2010). Design science research in information systems. *Design Research in Information Systems: Theory and Practice, 22*, 9–22. https://doi.org/10.1007/978-1-4419-5653-8_2

Hu, Y., Kuang, W., Qin, Z., Li, K., Zhang, J., Gao, Y., Li, W., & Li, K. (2021). Artificial intelligence security: Threats and countermeasures. *ACM Computing Surveys, 55*(1), 36. https://doi.org/10.1145/3487890

Hunt, T., Zhu, Z., Xu, Y., Peter, S., & Witchel, E. (2018). Ryoan: A distributed sandbox for untrusted computation on secret data. *ACM Transactions on Computer Systems, 35*(4), 1–32. https://doi.org/10.1145/3231594

IBM. (2024). *Cost of data breach report*. IBM Deutschland GmbH. https://www.ibm.com/reports/data-breach. Accessed 30 Aug 2024.

Irfan, M., Abbas, H., Sun, Y., Sajid, A., & Pasha, M. (2016). A framework for cloud forensics evidence collection and analysis using security information and event management. *Security and Communication Networks, 9*(16), 3790–3807. https://doi.org/10.1002/sec.1538

Jadhav, A. S., & Sonar, R. M. (2011). Framework for evaluation and selection of the software packages: A hybrid knowledge based system approach. *Journal of Systems and Software, 84*(8), 1394–1407. https://doi.org/10.1016/j.jss.2011.03.034

Jang-Jaccard, J., & Nepal, S. (2014). A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences, 80*(5), 973–993. https://doi.org/10.1016/j.jcss.2014.02.005

Jawhar, S., Miller, J., & Bitar, Z. (2024). AI-based cybersecurity policies and procedures. *Proceedings of the 3rd International*

Conference on AI in Cybersecurity. https://doi.org/10.1109/ICAIC60265.2024.10433845

Jayathilaka, H. M. T. N., & Wijayanayake, J. (2025). Systematic literature review on developing an AI framework for SME cybersecurity identification and personalized recommendations. *Journal of Desk Research Review and Analysis, 2*(2), 233–247. https://doi.org/10.4038/jdrra.v2i2.53

Kairu, M., & Shin, S. I. (2022). Evaluating the use of machine learning for cyber security intrusion detection. In *Proceedings of the 24th Southern Association for Information Systems Conference*. https://aisel.aisnet.org/sais2022/27

Kaloudi, N., & Li, J. (2020). The AI-based cyber threat landscape: A survey. *ACM Computing Surveys, 53*(1), 33. https://doi.org/10.1145/3372823

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity, 2*(20), 1–22. https://doi.org/10.1186/s42400-019-0038-7

Krishnasamy, V., & Venkatachalam, S. (2023). An efficient data flow material model based cloud authentication data security and reduce a cloud storage cost using index-level boundary pattern convergent encryption algorithm. *Materials Today: Proceedings, 81*, 931–936. https://doi.org/10.1016/j.matpr.2021.04.303

Kügler, D. (2003). "Man in the middle" attacks on Bluetooth. In *Proceedings of the 7th International Conference on Financial Cryprography,* 149–161. https://doi.org/10.1007/978-3-540-45126-6_11

Kühl, N., Schemmer, M., Goutier, M., & Satzger, G. (2022). Artificial intelligence and machine learning. *Electronic Markets, 32*(4), 2235–2244. https://doi.org/10.1007/s12525-022-00598-0

Le, D. D., Pham, V., Nguyen, H. N., & Dang, T. (2019). Visualization and explainable machine learning for efficient manufacturing and system operations. *Smart and Sustainable Manufacturing Systems, 3*(2), 127–147. https://doi.org/10.1520/SSMS20190029

Lee, A. S., Thomas, M., & Baskerville, R. L. (2015). Going back to basics in design science: From the information technology artifact to the information systems artifact. *Information Systems Journal, 25*(1), 5–21. https://doi.org/10.1111/isj.12054

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

Lier, S. K., Gerlach, J., & Breitner, M. H. (2023). Who needs XAI in the energy sector? A framework to upgrade black box explainability. *Proceedings of the 44th International Conference on Information Systems.* https://aisel.aisnet.org/icis2023/aiinbus/aiinbus/19

Lier, S. K., Gerlach, J., & Breitner, M. H. (2024). What is ethical AI?–Design guidelines and principles in the light of different regions, countries, and cultures. In *Proceedings of the 57th Hawaii International Conference on System Science*, 6848–6858. https://doi.org/10.24251/HICSS.2023.821

Lobb, R., & Colditz, G. (2013). Implementation science and its application to population health. *Annual Review of Public Health, 34*(1), 235–251. https://doi.org/10.1146/annurev-publhealth-031912-114444

Lowry, P. B., Dinev, T., & Willison, R. (2017). Why security and privacy research lies at the centre of the information systems (IS) artefact: Proposing a bold research agenda. *European Journal of Information Systems, 26*(6), 546–563. https://doi.org/10.1057/s41303-017-0066-x

Machado, G. R., Silva, E., & Goldschmidt, R. R. (2021). Adversarial machine learning in image classification: A survey toward the defender's perspective. *ACM Computing Surveys, 55*(1), 38. https://doi.org/10.1145/3485133

Martins, N., Cruz, J. M., Cruz, T., & Abreu, P. H. (2020). Adversarial machine learning applied to intrusion and malware scenarios: A systematic review. *IEEE Access, 8*, 35403–35419. https://doi.org/10.1109/ACCESS.2020.2974752

Mayring, P. (2014). Qualitative content analysis: Theoretical foundation, basic procedures and software solution. *Social Science Open Access Repository*. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173

McIntosh, T., Jang-Jaccard, J., Watters, P., spsampsps Susnjak, T. (2019). The inadequacy of entropy-based ransomware detection. In *Proceedings of the 26th International Conference on Neural Information Processing,* 181–189. https://doi.org/10.1007/978-3-030-36802-9_20

Meske, C., Abedin, B., Klier, M., & Rabhi, F. (2022). Explainable and responsible artificial intelligence. *Electronic Markets, 32*(4), 2103–2106. https://doi.org/10.1007/s12525-022-00607-2

Moghimi, A., Wichelmann, J., Eisenbarth, T., & Sunar, B. (2019). Memjam: A false dependency attack against constant-time crypto implementations. *International Journal of Parallel Programming, 47*(4), 538–570. https://doi.org/10.1007/s10766-018-0611-9

Morgan, S. (2020). *Cybercrime to cost the world $10.5 trillion annually by 2025*. Cybercrime Magazine. https://cybersecurityventures.com/hackerpocalypse-cybercrime-report-2016/. Accessed 30 Aug 2024.

Mughal, A. A. (2018). Artificial intelligence in information security: Exploring the advantages, challenges, and future directions. *Journal of Artificial Intelligence and Machine Learning in Management, 2*(1), 22–34.

Nilsen, P. (2020). Making sense of implementation theories, models, and frameworks. In Albers, B., Shlonsky, A. spsampsps Mildon, R. (Eds.) *Implementation Science 3.0* (pp. 53–79). Springer, Cham. https://doi.org/10.1007/978-3-030-03874-8_3

Nortje, M. A., & Grobbelaar, S. S. (2020). A framework for the implementation of artificial intelligence in business enterprises: A readiness model. *Proceedings of the International Conference on Engineering, Technology and Innovation.* https://doi.org/10.1109/ICE/ITMC49519.2020.9198436

Novikov, I. (2018). *How AI can be applied to cyberattacks*. Forbes. https://www.forbes.com/sites/forbestechcouncil/2018/03/22/how-ai-can-be-applied-to-cyberattacks/. Accessed 30 Aug 2024.

Ogbanufe, O. (2021). Enhancing end-user roles in information security: Exploring the setting, situation, and identity. *Computers & Security*, *108,* #102340. https://doi.org/10.1016/j.cose.2021.102340

Ohm, M., Sykosch, A., & Meier, M. (2020). Towards detection of software supply chain attacks by forensic artifacts. *Proceedings of the 15th International Conference on Availability, Reliability and Security.* https://doi.org/10.1145/3407023.3409183

Ozkan-Okay, M., Akin, E., Aslan, Ö., Kosunalp, S., Iliev, T., Stoyanov, I., & Beloev, I. (2024). A comprehensive survey: Evaluating the efficiency of artificial intelligence and machine learning techniques on cyber security solutions. *IEEE Access, 12*, 12229–12256. https://doi.org/10.1109/ACCESS.2024.3355547

Qi, H., Di, X., & Li, J. (2018). Formal definition and analysis of access control model based on role and attribute. *Journal of Information Security and Applications, 43*, 53–60. https://doi.org/10.1016/j.jisa.2018.09.001

Rampášek, M., Mesarčík, M., & Andraško, J. (2025). Evolving cybersecurity of AI-featured digital products and services: Rise of standardisation and certification? *Computer Law & Security Review, 56*, #106093. https://doi.org/10.1016/j.clsr.2024.106093

Rangrez, U. S., Qadri, S. A., Kumar, C. A., & Kumar, C. J. (2024). Cyber-attack defense system enhanced by artificial intelligence. *Proceedings of the International Conference on Intelligent Systems for Cybersecurity*. https://doi.org/10.1109/ISCS61804.2024.10581124

Reim, W., Åström, J., & Eriksson, O. (2020). Implementation of artificial intelligence (AI): A roadmap for business model innovation.

Artificial Intelligence, 1(2), 180–191. https://doi.org/10.3390/ai1020011

Roohparvar, R. (2023). *How to integrate AI into your cybersecurity strategy – Cyber security solutions, compliance, and con.* Info Guard Security. https://www.infoguardsecurity.com/how-to-integrate-ai-into-your-cybersecurity-strategy/. Accessed 30 Aug 2024.

Rosenberg, I., Shabtai, A., Elovici, Y., & Rokach, L. (2021). Adversarial machine learning attacks and defense methods in the cyber security domain. *ACM Computing Surveys, 54*(5), 1–36. https://doi.org/10.1145/3453158

Roshanaei, M., Khan, M. R., & Sylvester, N. N. (2024). Enhancing cybersecurity through AI and ML: Strategies, challenges, and future directions. *Journal of Information Security, 15*(3), 320–339. https://doi.org/10.4236/jis.2024.153019

Ross, J. (2018). The fundamental flaw in AI implementation. *MIT Sloan Management Review, 59*(2), 10–11.

Saghezchi, F. B., Mantas, G., Violas, M. A., de Oliveira Duarte, A. M., & Rodriguez, J. (2022). Machine learning for DDoS attack detection in industry 4.0 CPPSs. *Electronics*, *11*(4), p. 14. https://doi.org/10.3390/electronics11040602

Salem, A. H., Azzam, S. M., Emam, O. E., & Abohany, A. A. (2024). Advancing cybersecurity: A comprehensive review of AI-driven detection techniques. *Journal of Big Data, 11,* https://doi.org/10.1186/s40537-024-00957-y

Saltz, J. S. (2021). CRISP-DM for data science: Strengths, weaknesses and potential next steps. In *Proceedings of the International Conference on Big Data,* 2337–2344. https://doi.org/10.1109/BigData52589.2021.9671634

Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). AI-driven cybersecurity: An overview, security intelligence modeling and research directions. *Springer Nature Computer Science, 2*(173), 18. https://doi.org/10.1007/s42979-021-00557-0

Schröder, M., Krüger, J., Foroutan, N., Horn, P., Fricke, C., Delikanli, E., Maus, H., spsampsps Dengel., A. (2024). Towards cyber mapping the German financial system with knowledge graphs. *European Semantic Web Conference*, 270–288. https://doi.org/10.1007/978-3-031-60626-7_15

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science, 181*, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Segal, E. (2020). *The impact of AI on cybersecurity*. IEEE Computer Society. https://www.computer.org/publications/tech-news/trends/the-impact-of-ai-on-cybersecurity. Accessed 30 Aug 2024

Sen, R., Heim, G., & Zhu, Q. (2022). Artificial intelligence and machine learning in cybersecurity: Applications, challenges, and opportunities for MIS academics. *Communications of the Association for Information Systems*, *51*(1), 179–209. https://doi.org/10.17705/1CAIS.05109

Serafini, F., & Reid, S. F. (2023). Multimodal content analysis: Expanding analytical approaches to content analysis. *Visual Communication, 22*(4), 623–649. https://doi.org/10.1177/14703572198641 33

Shah, V. (2021). Machine learning algorithms for cybersecurity: Detecting and preventing threats. *Revista Española De Documentación Científica, 15*(4), 42–66. https://doi.org/10.5281/zenodo.10779509

Sharma, S. (2021). Role of artificial intelligence in cyber security and security framework. In N. Bhargava, R. Bhargava, P. S. Rathore, & R. Agrawal (Eds.), *Artificial Intelligence and Data Mining Approaches in Security Frameworks*, (1st ed., pp. 33–63). Wiley. https://doi.org/10.1002/9781119760429.ch3

Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., & Xu., M. (2020). A survey on machine learning techniques for cyber security in the last decade. *IEEE Access, 8*, 222310–222354. https://doi.org/10.1109/ACCESS.2020.3041951

Shin, H., & Lee, J. (1996). A process model of application software package acquisition and implementation. *Journal of Systems and Software, 32*(1), 57–64. https://doi.org/10.1016/0164-1212(95)00045-3

Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE Access, 7*, 53040–53065. https://doi.org/10.1109/ACCESS.2019.2912200

Sikkel, K. (2014). Sourcing lifecycle for software as a service (SAAS). *Proceedings of the International Conference on Advances Science and Contemporary Engineering*. https://doi.org/10.1051/epjconf/20146800010

Siroya, N., & Mandot, M. (2021). Role of AI in cyber security. In N. Bhargava, R. Bhargava, P. S. Rathore, & R. Agrawal (Eds.), *Artificial intelligence and data mining approaches in security frameworks*, (1st ed. pp. 434–444). Wiley. https://doi.org/10.1002/9781119760429.ch1

Templier, M., & Paré, G. (2015). A framework for guiding and evaluating literature reviews. *Communications of the Association for Information Systems*, *37*(1), 112–137. https://doi.org/10.17705/1CAIS.03706

Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets, 31*(2), 447–464. https://doi.org/10.1007/s12525-020-00441-4

Tolido, R., Thieullent, A. L., van der Linden, G., Frank, A., Delabarre, L., Buvat, J., Theisler, J., Cherian, S., & Khemka, Y. (2019). *Reinventing cybersecurity with artificial intelligence - The new frontier in digital security cap*. Gemini Research Institute. https://www.capgemini.com/wp-content/uploads/2019/07/AI-in-Cybersecurity_Report_20190711_V06.pdf. Accessed 30 Aug 2024.

Tong, F., & Yan, Z. (2017). A hybrid approach of mobile malware detection in android. *Journal of Parallel and Distributed Computing, 103*, 22–31. https://doi.org/10.1016/j.jpdc.2016.10.012

Vegesna, V. V. (2023). Enhancing cyber resilience by integrating AI-driven threat detection and mitigation strategies. *Transactions on Latest Trends in Artificial Intelligence, 4*(4), 8.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, 21–54. https://doi.org/10.25300/misq/2013/37.1.02

VERBI Software. (2021). *MAXQDA 2022* [computer software]. Berlin, Germany: VERBI Software. https://www.maxqda.com. Accessed 30 Aug 2024.

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems*, *37(1),* 205–224. https://doi.org/10.17705/1CAIS.03709

vom Brocke, J., Winter, R., Hevner, A., & Maedche, A. (2020). Special issue editorial–Accumulation and evolution of design knowledge in design science research: A journey through time and space. *Journal of the Association for Information Systems, 21(3),* 520–544. https://doi.org/10.17705/1jais.00611

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. Management Information *Systems Quarterly*, *26*(2), xiii-xxiii.

Welte, R., Estler, M., & Lucke, D. (2020). A method for implementation of machine learning solutions for predictive maintenance in small and medium sized enterprises. In *Proceedings of the 53rd CIRP Conference on Manufacturing Systems,* 909–914. https://doi.org/10.1016/j.procir.2020.04.052

Wieczorrek, H. W., & Mertens, P. (2010). Management von IT-projekten: Von der Planung zur Realisierung. *Springer*. https://doi.org/10.1007/978-3-642-16127-8

Xue, Y., Meng, G., Liu, Y., Tan, T. H., Chen, H., Sun, J., & Zhang, J. (2017). Auditing anti-malware tools by evolving android

malware and dynamic loading technique. *IEEE Transactions on Information Forensics and Security, 12*(7), 1529–1544. https://doi.org/10.1109/TIFS.2017.2661723

Yaseen, A. (2023). AI-driven threat detection and response: A paradigm shift in cybersecurity. *International Journal of Information and Cybersecurity, 7*(12), 25–43.

Zhang, J. (2021). Distributed network security framework of energy internet based on internet of things. *Sustainable Energy Technologies and Assessments*, *44*. https://doi.org/10.1016/j.seta.2021.101051

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*, 1029–1053. https://doi.org/10.1007/s10462-021-09976-0