International Journal of Information Security (2025) 24:43
https://doi.org/10.1007/s10207-024-00959-0

REGULAR CONTRIBUTION

Artificial intelligence for system security assurance: A systematic literature review

Shao-Fang Wen[1] · **Ankur Shukla**[2] · **Basel Katt**[1]

Accepted: 7 December 2024 / Published online: 14 December 2024
© The Author(s) 2024

Abstract
System Security Assurance (SSA) has emerged as a critical methodology for organizations to verify the trustworthiness of their systems by evaluating security measures against industry standards, legal requirements, and best practices to identify any weakness and demonstrate compliance. In recent years, the role of Artificial Intelligence (AI) in enhancing cybersecurity has received increased attention, with an increasing number of literature reviews highlighting its diverse applications. However, there remains a significant gap in comprehensive reviews that specifically address the integration of AI within SSA frameworks. This systematic literature review seeks to fill this research gap by assessing the current state of AI in SSA, identifying key areas where AI contributes to improve SSA processes, highlighting the limitations of current methodologies, and providing the guidance for future advancements in the field of AI-driven SSA.

Keywords Cybersecurity · Security assurance · Artificial intelligence · Systematic literature review

Abbreviations

ACO	Ant Colony Optimization
AI	Artificial Intelligence
AMOE	Assessment and Management of Organizational Evidence
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
ANP	Analytic Network Process
BAGS	Bayesian Attack Graph for Smart Grid
BDI	Belief-Desire-Intention
CART	Classification And Regression Tree
CAV	Connected and Autonomous Vehicles
CBR	Case-Based Reasoning
CC	Common Criteria
charCNN	Character-level Convolutional Neural Networks
CNN	Convolutional Neural Networks
CPS	Cyber-Physical System
CSA	Clonal Selection Algorithm
CSA	Clonal Selection Algorithm
CVS	Common Vulnerabilities and Exposure
CWE	Common Weakness Enumeration
DDPG	Deep Deterministic Policy Gradient
DL	Deep Learning
DNN	Deep Neural Network
DQN	Deep Q-Learning Network
DRL	Deep Reinforcement Learning
DSA	Dynamic Security Assessment
DSS	Decision Support Systems
ENBPP	Enhanced Naïve Bayes Posterior Probability
ESASCF	Expert-System Automated Security Compliance Framework
FIS	Fuzzy Inference System
FNN	Feed-Forward Networks
GA	Genetic Algorithms
GAN	Generative Adversarial Network
GRU	Gated Recurrent Unit
HFL	Horizontal Federated Learning
HIS	Health information Systems
HMM	Hidden Markov Models
ICS	Industrial Control System
ICT	Information and Communication Technology
IIoT	Industrial Internet of Things
IMoT	Internet of Medical Things

✉ Shao-Fang Wen
shao-fang.wen@ntnu.no

[1] Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Gjøvik, Norway

[2] Institute for Energy Technology, Halden, Norway

IoT	Internet of Things
ISGFs	Information Security Governance Frameworks
IT2FLS	Interval Type-2 Fuzzy Logic System
LM	Levenberg-Marquardt
LSTM	Long-Short Term Memory
LTE	Long Term Evolution
MCP	Misuse Case Programming
MDP	Markov Decision Processes
ML	Machine Learning
MLP	Multi-Layer Perceptron
MONA	Mission-Oriented Network Analysis
MPIC	Massive Personal Information Clustering
MSN	Mobile Social Networks
NER	Named Entity Recognition
NLP	Natural Language Processing
OPMDP	Partially Observable Markov Decision Processes
OSSTM	Open-Source Security Testing Methodology Manual
OT	Operational Technology
PAA	Probabilistic Arithmetic Automata
PDCA	Plan-Do-Check-Action
PII	Personally Identifiable Information
PLM	Pre-trained Language Model
PSO	Particle Swarm Optimization
QML	Quantum ML
QRA	Quantitative Risk Assessment
RL	Reinforcement Learning
RMCM	Restricted Misuse Case Modeling
RNN	Recurrent Neural Network
SARSA	State–action–reward–state–action
SCADA	Supervisory Control and Data Acquisition
SPORF	Sparse Projection Oblique Randomer Forests
SSA	System Security Assurance
SSG-AFL	Static Seed Selector- American Fuzz Lop
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
TOM	Threat Operating Model
TOPSIS	Order Preference by Similarities to Ideal Solution
VANET	Vehicular Ad Hoc Network
WGAN	Wasserstein Generative Adversarial Network
W-HMM	Weighted Hidden Markov Model
XAI	Explainable AI
XGBoost	Extreme Gradient Boosting

1 Introduction

Information and Communication Technology (ICT) has become the backbone of businesses across diverse industries in today's interconnected world. It facilitates data-driven decision-making, enhances operational efficiency, and simplifies communication processes. Given the widespread reliance on ICT systems, it is paramount for businesses to prioritize comprehensive security measures to safeguard these systems and maintain a robust security posture. These measures are crucial for mitigating risks such as unauthorized access, data breaches, service interruptions, and other cyber threats. However, merely implementing security measures is not enough. Organizations must also provide verifiable evidence of the sufficiency and efficacy of their security implementations [81, 183]. This evidence not only demonstrates the organization's commitment to security but also reassures stakeholders about the safeguards in place to protect critical information and assets.

System Security Assurance (SSA) has emerged as an effective method for organizations to assess and ensure the trustworthiness and dependability of their systems [83]. By performing SSA, organizations can evaluate a system's correctness and security, thereby enhancing its functionality and mitigating potential risks [20]. SSA specifically involves assessing, recording, and monitoring the security posture of systems to determine whether the implemented security features, practices, procedures, and architecture align effectively with the security objectives [147]. One of SSA's primary benefits is its contribution to regulatory compliance. Organizations are often required to adhere to industry standards, legal requirements, and best practices to demonstrate compliance. This necessitates evaluating their security measures against these benchmarks and identifying any gaps or areas of non-compliance that require attention [53].

Parallel to the evolution of SSA, there have been significant efforts to develop Artificial Intelligence (AI)-based solutions for a wide range of cybersecurity challenges [84]. AI's ability to rapidly process millions of data points to predict and prevent cyberattacks makes it an essential tool for enhancing SSA. By integrating AI into security frameworks, organizations can automate complex security tasks and support security teams more effectively, thereby improving the overall efficiency and responsiveness of cybersecurity operations [34, 197]. The increasing collaboration between AI and cybersecurity researchers is fostering a rich body of research aimed at developing advanced solutions for proactive threat identification, automated assurance evidence collection, and continuous compliance assessment. This synergy is crucial in addressing the dynamic challenges posed by modern cyber threats, ensuring that security systems are both resilient and adaptable [34].

Table 1 Previous literature reviews in AI and cybersecurity

Year	Reference	Focus of the study
2023	[151]	This study reviews literature on Internet of Things (IoT) security intelligence, including the IoT paradigm, IoT-based smart environments, related security concerns with machine learning solutions
2023	[35]	This study examines the research on AI-driven cyberattacks and how it relates to Industry 4.0 cybersecurity
2023	[84]	This study examines the practical applications of AI in cybersecurity, focusing on the five cybersecurity functions (Identify, Protect, Detect, Respond, and Recover)
2023	[123]	This study examines the use of AI techniques, specifically machine learning (ML) and deep learning (DL), to detect attacks and malicious software
2022	[1]	This study reviews the use of AI methods, particularly ML and DL, to detect cybersecurity threats on the IoT environment
2021	[197]	This paper summarizes AI-based approaches for cybersecurity applications, including user access authentication, network situation awareness, dangerous behavior monitoring, and abnormal traffic identification
2020	[181]	This study investigates the application of AI in cybersecurity, particularly in intrusion detection systems. It presents AI's role in reducing complexity, training times, and false alarms in combating cybercrimes

In recent years, researchers have highlighted the diverse applications of AI in cybersecurity through literature reviews, as shown in Table 1. Despite the increasing volume of such reviews, there remains a significant gap in those that specifically address the integration of AI within SSA. Historically, syntheses of AI applications in SSA have been limited, with no comprehensive reviews detailing how AI techniques are implemented throughout the SSA spectrum. This gap underscores the need for a systematic exploration to enhance the security frameworks of ICT systems. In this study, we conduct a Systematic Literature Review (SLR), aiming to fill this research gap by assessing the current state of AI in SSA, identifying emerging trends, and outlining the limitations of current methodologies. This comprehensive overview intends not only to illuminate existing AI-driven security solutions but also to guide future advancements in the field.

The rest of the paper is structured as follows. Section 2 establishes the fundamental concepts of SSA and introduces the SSA framework. In Sect. 3, we outline the research methodology used to conduct this literature review and explains the data extraction process. Section 4 delves into data synthesis, applying the SSA framework to categorize and analyze the findings. Section 5 delves deeply into the various AI techniques used in SSA, while Sect. 6 examines the application of AI techniques in enhancing SSA across several application domains. Section 7 identifies research gaps, setting the stage for future research directions. Section 8 highlights the limitations of our study. Finally, Sect. 9 presents the main conclusions.

2 Background

This section outlines the fundamental concepts of SSA and introduces the SSA framework. This is the first step towards establishing reliable communication in this field.

2.1 The concept of system security assurance

According to Anderson [13], *Assurance* is the process of assessing the probability of a system failure in a particular way. Simply put, security assurance in the context of security is our estimate of how long the system will remain secure. Spears, Barki, and Barton [165] define security assurance as the level of confidence in meeting security requirements from a substantive standpoint. It entails making sure that the system is safeguarded against potential threats, weaknesses, and attacks by having the appropriate security measures in place. Security assurance is defined as "the measure of confidence that the security features, practices, procedures, and architecture of an information system accurately mediate and enforce the security policy" in the NIST Special Publication [147] (Page 26). This paper follows Katt and Prashe's [83] definition of security assurance, which views it as an attribute associated with the confidence that security requirements are met, while vulnerabilities and weaknesses are either tolerated or addressed. With this definition, the term "system security assurance" refers to a systematic process of assessing a system's security posture in order to determine its overall security status and conditions.

Figure 1 illustrates the concept of SSA, a methodology used to assess and ensure the trustworthiness and efficacy of ICT systems in meeting specified security requirements. At the core of SSA is the evaluation of an ICT system's security capabilities, which are attributes or functionalities designed to defend against security threats. These capabilities stem

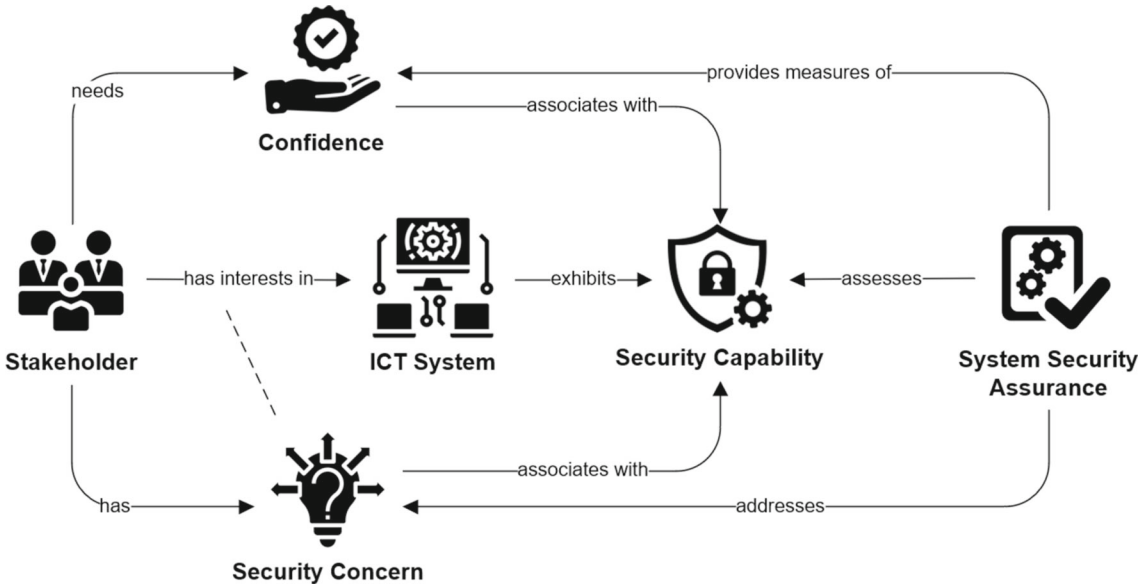

Fig. 1 Illustration of the concepts of system security assurance

from the security concerns raised by stakeholders—individuals or entities with interests in the ICT system. Stakeholders' concerns directly influence the security capabilities that need to be developed and integrated into the IT system [180]. Once implemented, these capabilities undergo a thorough SSA to assess their effectiveness and sufficiency in addressing the identified security needs. This evaluation process, in turn, provides a measure of confidence to the stakeholders, validating that the system meets the necessary security standards and behaves as expected under potential threats. The cycle of assessing capabilities and gaining confidence highlights the dynamic nature of SSA, emphasizing continuous improvement and adaptation to evolving security threats and stakeholder needs.

2.2 System security assurance framework

In the context of SSA, no additional security controls are implemented; instead, the focus is solely on assessing the existence and effectiveness of the existing controls. This approach ensures that the current security measures are thoroughly evaluated to verify their adequacy and effectiveness. Typically, SSA employs a systematic approach that includes a number of activities, starting from the identification of relevant security policies and standards [58, 73]. The ultimate objective of SSA is to drive improvements in the system's security posture [136]. Figure 2 depicts our proposed SSA framework, a strategic methodology for systematically assessing, testing, and analyzing the system's security controls and practices. Anchored in the Plan-Do-Check-Act (PDCA) cycle, this framework is characterized by specific activities and goals that significantly enhance the effectiveness of the security assurance process. The following sections detail how each component of the framework integrates into the broader context of SSA, emphasizing the importance of evaluating the existing security infrastructure to identify areas of improvement and ensure ongoing compliance with security policies and standards.

2.2.1 Plan (P)—security requirement and criteria development

In the context of our SSA framework, the "Plan" phase is fundamentally about setting a robust foundation for security management. This phase involves developing or refining the security policies that outline how the organization approaches SSA. It includes standards, guidelines, and procedures for maintaining security. The initial step in this phase is identifying and understanding relevant security policies, standards, and regulations that apply to the system or organization. This ensures alignment with legal mandates and best practices in cybersecurity. With these elements in place, the next phase (Do) involves the active execution and testing of assurance techniques to gather evidence in accordance with the established security requirements and criteria.

2.2.2 Do (D)—Assurance Technique and Evidence Collection

The "Do" phase primarily revolves around the active execution of security testing and assurance techniques to gather crucial evidence about a system's security posture. This phase is vital for operationalizing the security plans and strategies

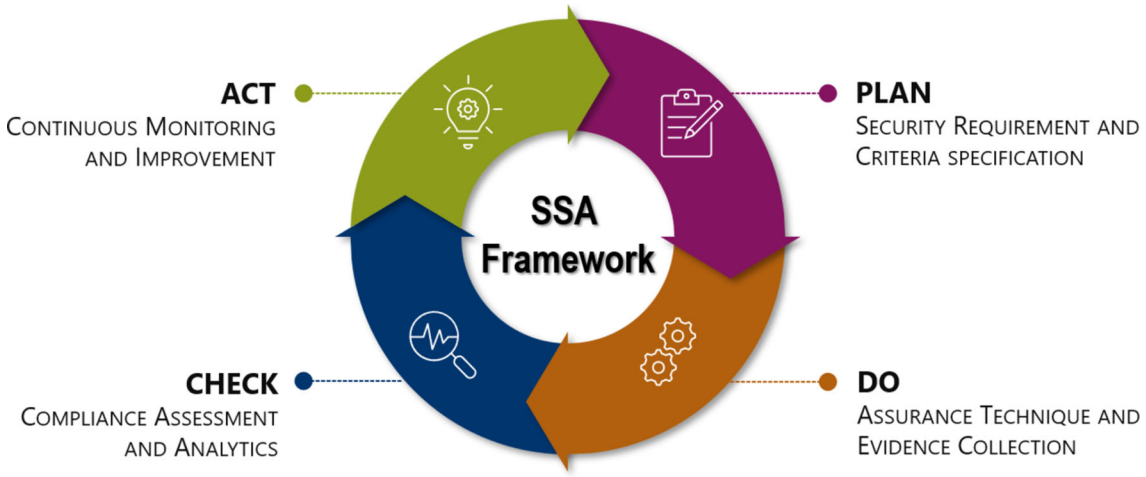

Fig. 2 The proposed system security assurance framework, anchored in the PDCA cycle

developed during the "Plan" phase, serving as the practical implementation stage where these plans are put to the test. The primary goal of the "Do" phase is the comprehensive collection of data and evidence about the fulfillment of the predefined security requirements and the existence of system vulnerabilities. This phase includes extensive security testing through simulated attacks that mimic potential real-world breaches, stress tests to evaluate the system under extreme conditions, and other specific tests designed to pinpoint weaknesses. The data and evidence gathered during the "Do" phase serve as the foundation for the "Check" phase.

2.2.3 Check (C)—compliance assessment and analytics

During the "Check" phase, a comprehensive evaluation of the collected data and implemented security measures takes place. It entails conducting a thorough review of the data and evidence gathered during the "Do" phase to determine how well the current security measures meet predefined security criteria and standards. This phase uses detailed vulnerability analysis and risk assessment to identify and prioritize current security gaps, as well as forecast potential future vulnerabilities. Such analyses aid in understanding the severity and potential impact of each identified risk, allowing the organization to better align security initiatives with business priorities and compliance requirements. Furthermore, the security evaluation and analysis segment evaluate the adequacy and functionality of existing security controls, ensuring that they effectively mitigate identified risks and meet predefined security standards. The outcome of the "Check" phase is a comprehensive understanding of the system's security strengths and weaknesses. It highlights areas where the security strategy has succeeded and identifies where improvements are necessary. This phase ultimately sets the stage for the "Act" phase, where the insights gained are used to refine and enhance the security strategies and controls.

2.2.4 Act (A)—continuous monitoring and improvement

The "Act" phase captures the essence of proactive and continuous improvement in a system's security posture. This phase is characterized by a commitment to refining and optimizing security measures, drawing on insights and feedback gained during the previous "Check" phase. This continuous improvement process entails more than just making incremental changes; it also includes reevaluating and improving existing security measures to better protect against emerging threats and vulnerabilities.

This dynamic cycle of implementation, evaluation, and adjustment not only strengthens the system's defenses but also feeds insights back into the "Plan" phase, effectively closing the PDCA cycle and preparing for the next round of security evaluation. This iterative process ensures that system security management is a continuous and evolving practice that effectively adapts to new challenges and environments.

3 Literature review methodology

In this paper, we conduct a SLR to catalog the application of AI techniques within the SSA domain and identify areas needing further research. Our SLR methodology follows the guidelines set forth by Kitchenham [90, 91], ensuring a rigorous and structured approach to our review process. This methodology allows us to systematically identify, evaluate, and synthesize the existing research, providing a comprehensive overview of the current state of AI applications in SSA and highlighting potential directions for future investigation.

4 Research questions

In order to comprehensively explore the integration of AI techniques within the field of SSA, this study seeks to address several key areas of inquiry. The following research questions (RQs) have been formulated to guide the systematic review and analysis of the current state of AI research, its application within SSA frameworks, and its impact across various domains. Additionally, these questions aim to identify existing research gaps and propose directions for future research, ensuring a holistic understanding of this evolving field:

RQ1: What AI research has been conducted that is relevant to SSA, and how can these studies be systematically positioned within the SSA framework?

RQ2: How are specific AI techniques being utilized to support and improve different aspects and SSA?

RQ3: How do AI techniques support SSA across different application domains?

RQ4: What are the existing research gaps in the integration of AI techniques within the domain of SSA, and what are the potential directions for future research in this field?

4.1 Search strategy

This section outlines how the literature search was conducted to ensure comprehensiveness and relevance.

4.1.1 Database and sources

For this SLR work, Scopus was chosen as the sole bibliometric database to identify and gather relevant research studies. Scopus was chosen for this study due to its extensive coverage and high-quality indexing. Research shows that nearly all journals indexed in Web of Science are also included in Scopus and Dimensions [164]. Moreover, Scopus indexes 66.07% more unique journals compared to Web of Science, providing broader access to relevant literature. This extensive coverage ensures that our review captures a comprehensive range of high-quality publications in the fields of AI techniques and SSA. By utilizing Scopus, we ensure access to an extensive and diverse collection of relevant publications, ensuring the reliability and validity of our review.

4.1.2 Keywords and search terms

The search string for this SLR was carefully crafted to ensure comprehensive coverage and precision in retrieving relevant studies. The design of the search string incorporates two primary components: SSA keywords and security-enhanced keywords, combined with AI-related terms.

The SSA keywords encompass a broad range of terms directly related to various aspects of security assurance, including assessment, evaluation, verification, and vulnerability analysis. These keywords are intended to capture studies focusing on the core activities and metrics used in security assurance frameworks. The security-enhanced keywords are designed to specifically include the term "security" in conjunction with related techniques and methods. This ensures that the search results are highly relevant to security-focused research. Examples of such keywords include "security risk assessment," "security audit," and "security code review." This approach ensures that only those studies that explicitly address security aspects within these techniques are retrieved. Lastly, the AI keywords segment includes terms related to various AI technologies and methodologies such as 'machine learning', 'deep learning', 'natural language processing', and 'data mining'. By integrating these AI terms, the search string is aimed at capturing studies that explore the application of AI techniques in enhancing security assurance.

Overall, the search string is constructed as follows:

((SSA Keywords) OR (Security − Enhanced Keywords))
AND (AI Keywords)

Table 2 provides a detailed breakdown of each segment, with keywords combined using logical operators to ensure comprehensive coverage.

4.1.3 Search period

The search period was set from 2016 to 2023. This period was chosen to capture the most recent and relevant advancements in the integration of AI techniques within SSA. The field of AI has witnessed rapid development and significant breakthroughs in recent years, especially since 2016. This period marks the rise of more sophisticated ML algorithms and neural networks, which have been increasingly applied across various security domains [32, 134]. These advancements have enabled AI to handle more complex tasks, improve accuracy, and provide real-time solutions for cybersecurity challenges. By focusing on this period of time, we make sure that the review includes the most recent findings, innovations, and approaches, providing a current and thorough understanding of the trends and prospects in this field.

4.2 Inclusion and exclusion criteria

4.2.1 Inclusion criteria

The inclusion criteria are designed to ensure that the selected articles are highly relevant, technically robust, and contribute valuable insights into the integration of AI within SSA. Each criterion has been carefully formulated to encompass key aspects of the research scope, ensuring a comprehensive and focused review of the literature. The criteria are as follows:

Table 2 The segments of the search string

SSA keywords	Security-enhanced keywords	AI keywords
“security assurance" OR "security requirement" OR "security assessment" OR "security evaluation" OR "security verification" OR "security validation" OR "security analysis" OR "security metric" OR "security score" OR "security profile" OR "security target" OR "vulnerability scan" OR "vulnerability scanner" OR "vulnerability scanning" OR "vulnerability assessment" OR "vulnerability analysis" OR "vulnerability test" OR "vulnerability testing" OR "vulnerability detection" OR "threat assessment" OR "threat analysis" OR "threat model" OR "threat modeling" OR "threat evaluation" OR "security test" OR "security testing" OR "penetration test" OR "fuzzy test" OR "fuzz testing" OR "ethical hacking"	"security" AND ("risk assessment" OR "risk analysis" OR "risk evaluation" OR "audit" OR "certification" OR "assurance technique" OR "assurance method" OR "compliance check" OR "formal verification" OR "code review" OR "static analysis" OR "code analysis" OR "dynamic testing" OR "dynamic verification")	"artificial intelligence" OR "ML" OR "DL" OR "artificial neural network" OR "generated AI" OR "large language model" OR "natural language processing" OR "text mining" OR "feature extraction" OR "data mining" OR "sentiment analysis" OR "computer vision" OR "reinforcement learning" OR "predictive analytics" OR "unsupervised learning" OR "supervised learning" OR "semi-supervised learning" OR "transfer learning" OR "algorithmic modeling" OR "knowledge representation" OR "expert systems" OR "pattern recognition" OR "anomaly detection" OR "automated reasoning" OR "fuzzy logic" OR "genetic algorithms" OR "intelligent agents" OR "robotics"

- **Topical Alignment:** The core subject matter of the paper must pertain to the system security, explicitly addressing applications or implications for the use of AI in this domain.
- **Relevance to SSA**: The article should discuss methodologies, techniques, or practices that support the SSA process by providing evidence to measure the confidence that the system meets its security requirements. This includes risk assessments, vulnerability detection, security analysis, security audits, compliance checks, and formal verification processes.
- **Relevance to AI**: The article should prominently feature AI techniques, either as the primary focus or as a significant part of the methodology. For example, studies that incorporate ML as a core component of their research are considered relevant.
- **Within the Defined Types of Systems**: Papers must fit within specific system types, including IT systems (software and hardware) and Operational Technology (OT) systems such as power/smart grid systems, 5G, Industrial Control Systems (ICS), automotive control systems, Health information Systems (HIS), Cyber-Physical Systems (CPS), and business and financial systems.
- **Technical Depth**: Articles must provide detailed technical information regarding the AI methodologies discussed or utilized. This includes comprehensive descriptions of algorithms, data handling techniques, and the experimental setup, ensuring the inclusion of robust and replicable studies.
- **Empirical Evidence**: The studies must present solid empirical evidence supporting their claims, conclusions, or hypotheses related to AI and system security assurance. This includes data, case studies, experimental results, or other forms of empirical validation.
- **Language**: The article must be written in English to ensure accessibility and comprehensibility for the review team.
- **Full Research and Peer-Reviewed Paper**: The article must be a full research paper, including comprehensive details about the study’s methodology, results, and conclusions. Editorials, books, book chapters, and summaries of conference and symposium proceedings will be excluded.

4.2.2 Exclusion criteria

To ensure the integrity and relevance of this systematic literature review, certain types of publications will be excluded based on the following criteria. These criteria have been formulated to complement the inclusion criteria and ensure that

only high-quality, original research is considered. The exclusion criteria are as follows:

- **Secondary Literature**: Reviews, summaries, editorials, commentaries, and meta-analyses will be excluded to ensure that the review focuses solely on original research studies.
- **Incomplete or Preliminary Reports**: To ensure that only fully developed and finalized studies are considered, papers that are marked as preliminary, incomplete, or that present unfinished research will be excluded.
- **Duplicated Studies**: Multiple publications reporting the same research or data (e.g., a conference paper followed by a journal article) will be excluded to prevent duplication of findings.
- **General IT/Security Overviews**: General IT or system security papers that do not specifically address SSA or the integration of AI within this field will be excluded.
- **Unrelated AI Applications**: Studies that discuss AI applications irrelevant to system security, such as AI in marketing, AI for entertainment, etc., will be excluded.
- **Source Code-Based Studies** This review excludes papers focused on source code-based security techniques such as code review, code scanning, static analysis, and other code-level methods. The rationale is that the SSA targets the entire system, emphasizing system-level security assurance over detailed internal code analysis. Many systems are proprietary or do not readily provide source code access, making these methods less applicable across a broader spectrum of systems. Therefore, this exclusion allows the review to focus on AI-driven methodologies that enhance security assurance more universally, without relying on access to source code.
- **Malware Studies:** In this SLR, we have opted to exclude papers primarily centered on malware detection. This exclusion is strategic, emphasizing a proactive rather than reactive security posture. Malware detection typically responds to threats after they manifest, which contrasts with our goal of bolstering preemptive defenses through vulnerability detection. By focusing on identifying and addressing potential vulnerabilities early, our approach aligns with the preventive objectives of SSA, aiming to enhance system resilience by fortifying defenses against potential threats before they result in breaches.
- **Systems-Specific Exclusions**: Papers that focus exclusively on certain specialized systems will be excluded to maintain the review's focus on traditional IT and OT systems. These exclusions include AI-based systems and blockchain-based systems, allowing for a more in-depth exploration of AI's role in these specific contexts.

4.3 Data extraction

The data extraction process started from a broad database search to a precise selection of relevant studies. Figure 3 shows in detail the data extraction process for this SLR study. Initially, a comprehensive search was conducted using Scopus, yielding 3,251 potential papers. A quick scan of the titles and abstracts of this extensive collection was conducted in order to eliminate studies that did not appear to be relevant to the themes of SSA and AI techniques. This filtering reduced the pool to 1,669 papers. Following that, a thorough full-text review was conducted on these studies, with an emphasis on the originality, methodological soundness, and significance of their contributions to SSA. This process helped to reduce the list of studies to 330. The final stage involved a meticulous synthesis and evaluation process, which narrowed the list to 149 high-quality articles. In, Sects 4–6 we provide an in-depth discussion of how the selected articles integrate AI techniques within the SSA framework, exploring their application cross several application domains. In this context, when we refer to 'Article', we are specifically discussing the studies from our SLR. This designation aids in clearly distinguishing these studies throughout our discussion, making it easier and more beneficial for readers to follow the specific applications and impacts of AI techniques in SSA.

5 Data synthesis based on the SSA framework

In this chapter, we delve into the synthesis of data extracted from the selected articles, focusing on addressing RQ1: *What AI research has been conducted that is relevant to SSA, and how can these studies be systematically positioned within the SSA framework?* Fig. 4 depicts the data synthesis structure based on the SSA framework. This analysis aims to map out and evaluate the extensive body of AI research across the various phases of the SSA framework—Plan, Do, Check, and Act. By systematically categorizing the findings of each study within this structured framework, we can better understand the roles and impacts of AI technologies in enhancing security assurance. Through this synthesis, we hope to provide a comprehensive overview that aligns current AI research with the practical demands of SSA, shedding light on both accomplishments and areas for future exploration.

5.1 Security requirements and criteria development

Table 3 presents a summary of the main characteristics of each selected study within the phase of security requirements and criteria development. Detailed descriptions of the categories in this function are provided below.

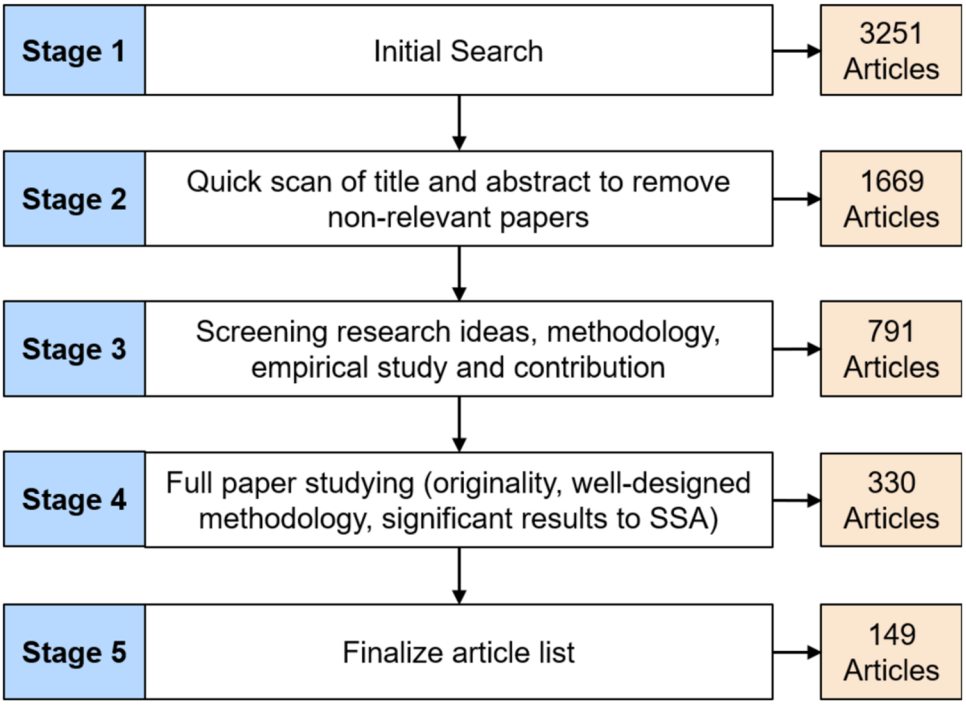

Fig. 3 Data extraction process in SLR

5.1.1 Security requirement elicitation

The evolution of *Security Requirements Elicitation* has transitioned from labor-intensive, manual processes to sophisticated, technology-driven approaches, largely due to advancements in AI, such as NLP and ML. Traditionally, this process involved extensive stakeholder interviews and manual document analysis, which were time-consuming and prone to human error, often resulting in inconsistent security documentation. Now, AI-powered tools are capable of interpreting complex technical language, extracting pertinent information from large datasets, and predicting security needs based on learned patterns. These tools analyze unstructured data from various sources to generate structured, actionable security specifications that align closely with organizational policies. We further categorize security requirement elicitation into three sub-

categories: automated requirement generation, requirement analysis and requirement elicitation through user engagement.

5.1.1.1 Automated requirements generation *Automated Requirements Generation* utilizes advanced AI technologies, including ML and advance algorithms, to streamline the creation and refinement of security requirements. This approach efficiently analyzes data to predict and customize security needs, significantly improving the accuracy and speed of developing security strategies. It enables rapid adaptation to emerging threats, ensures comprehensive and current security measures, and minimizes oversights.

Articles [182] and [79] provide examples of how automated systems can be used to categorize and identify security issues for software development projects. In Article [182], an ontology-based framework is examined that makes use of semantic rules to automatically identify security issues in use case models, guaranteeing that security verification standards are met from the start of the project. Similar to this, the article [79] employs text mining and decision tree algorithms to effectively categorize security requirements. This approach helps software engineers promptly address security requirements from software requirement specification documents and enhances the security of the finished software product.

Conversely, Articles [125] and [26] explore the role of ML in the efficient synthesis and documentation of security requirements. Article [125] introduces a classifier that streamlines the extraction and categorization of security-related requirements from extensive regulatory documents, significantly reducing the manual effort required for compliance in regulated sectors such as the automotive industry. Article [26] presents a hybrid approach combining ML and a rule-based expert system to predict security functional requirements and evaluation assurance levels for ICT products based on the Common Criteria (CC) framework. This method facilitates the generation of detailed security documentation, simplifying the certification process and ensuring that products comply with essential security standards.

5.1.1.2 Requirements analysis Once security requirements are established, AI in *Automated Requirement Analysis* plays

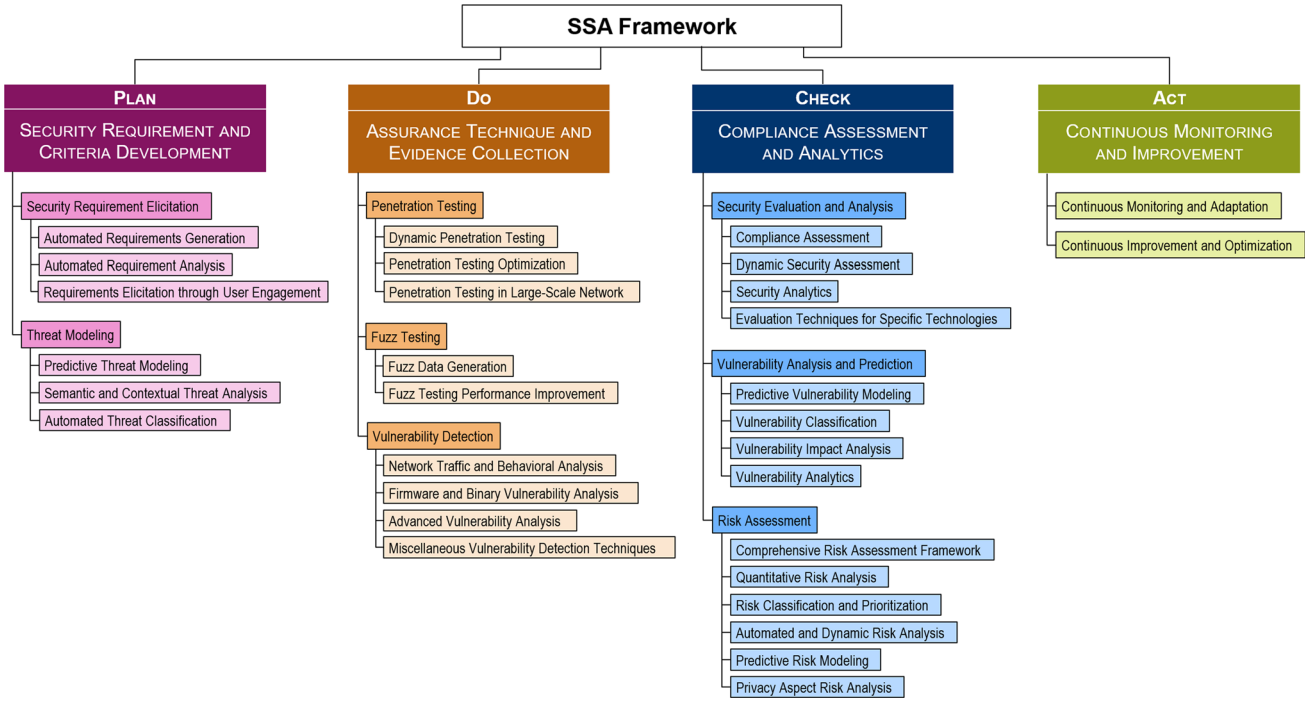

Fig. 4 SLR Data synthesis structure based on the SSA framework, organized according to the PDCA methodology

Table 3 Summary of the selected articles in security requirements and criteria development

Category	Sub-category	Selected article
Security requirement elicitation	Automated requirements generation	[26, 79, 125, 182]
	Automated requirement analysis	[6, 18, 60, 97, 115, 116]
	Requirements elicitation through user engagement	[61, 82, 114]
Threat modeling	Predictive threat modeling	[10, 12, 19, 67, 152]
	Semantic and contextual threat analysis	[68, 161]
	Automated threat classification	[88, 154, 193]

a pivotal role in evaluating and optimizing these specifications. It employs sophisticated computational tools to systematically analyze and interpret security needs from large datasets. This approach utilizes advanced algorithms, including ML and NLP, to automatically extract, categorize, and evaluate security requirements from various sources such as system documentation, user inputs, and regulatory guidelines. This method significantly reduces human error and bias, leading to more precise and reliable SSA planning.

The Misuse Case Programming (MCP) methodology, proposed in Articles [116] and [115], exemplifies the use of NLP to enhance security requirement analysis. This approach automatically generates security test cases from misuse case specifications, addressing both positive and negative security requirements. By translating natural language descriptions into executable test cases, MCP enables thorough and automated security evaluations. This methodology supports the identification and testing of security vulnerabilities, improving the overall robustness and compliance of software systems.

DL and ML algorithms play a crucial role in transforming security requirement analysis. Article [6] proposes a scoring system for information security governance frameworks in the banking sector, which uses DL to convert expert survey responses into predictive models. This system enables organizations to proactively evaluate and select the most effective security frameworks based on expert input, minimizing risk and resource expenditure. Additionally, Decision Support Systems (DSS) for security-control identification (Article [17]) and automated selection of security controls (Article [18]) leverage historical assessment data to recommend relevant controls, enhancing the efficiency and accuracy of security control implementation and aligning them with compliance requirements.

Fuzzy logic systems and ontology-based approaches also significantly contribute to security requirement analysis. Article [60] proposes the Interval Type-2 Fuzzy Logic System (IT2FLS) to automate cyber security assessments by capturing intra- and inter-expert uncertainty and using fuzzy logic rules to enhance the adequacy of security requirement assessments. This method ensures a nuanced evaluation of security requirements, improving the precision of security analyses. Furthermore, the AI methodology for comparing cloud service security certifications (Article [97]) utilizes NLP to map certification standards, laws, and user requirements into a structured ontology. This system allows for rapid and precise comparisons across different certification schemes, simplifying the task of aligning cloud service offerings with diverse security standards and legal requirements.

5.1.1.3 Requirements elicitation through user engagement

Requirements Elicitation through User Engagement emphasizes direct interaction with stakeholders to gather comprehensive security needs. Utilizing AI-driven tools such as NLP, structured scenarios, and case-based reasoning, this approach ensures precise and contextual security requirements are captured directly from user interactions.

Article [61] explores a structured approach to eliciting security requirements through natural language scenarios, utilizing user stories to articulate detailed, context-rich security scenarios that translate stakeholder interactions into actionable requirements. Similarly, Article [114] employs Restricted Misuse Case Modeling (RMCM) to integrate security and privacy requirements into the development process of multi-device software ecosystems, facilitating comprehensive engagement with users to capture a wide array of security concerns and potential threats. Article [82] enhances this interactive process through Case-Based Reasoning (CBR), which uses past cases and scenarios to refine the elicitation of security requirements, ensuring that the lessons learned from previous engagements are effectively applied to new situations.

5.1.2 Threat modeling

Threat Modeling plays an essential role in SSA by systematically identifying and analyzing potential security threats. This proactive approach helps define and prioritize security requirements, setting the stage for more robust security requirement elicitation [159]. Threat modeling can be significantly improved with AI. AI technologies such as ML and NLP automate and refine the analysis of large datasets in order to identify potential vulnerabilities. This integration ensures that the security requirements derived from threat modeling are both comprehensive and precise, allowing organizations to develop targeted strategies to effectively address identified risks.

5.1.2.1 Predictive threat modeling

Predictive Threat Modeling leverages AI to foresee and mitigate potential security threats, significantly enhancing the security posture of various systems. By analyzing historical data and current trends, these methodologies enable proactive identification and management of vulnerabilities.

Several studies demonstrate the integration of ML and AI techniques to advance predictive threat modeling. For instance, multiple approaches utilize ML algorithms to analyze data and predict cyber threats. Articles [152] and [10] both highlight the use of ML to evaluate risk factors and predict future threats. These models continuously learn from new and historical data, dynamically updating their predictions to maintain accurate threat assessments. This ongoing learning process enhances the resilience of security measures in cloud environments and critical infrastructure by identifying potential attack vectors based on past incidents and user behaviors.

Incorporating reinforcement learning, Article [67] showcases the use of the State–Action–Reward–State–Action (SARSA) algorithm to identify and analyze security vulnerabilities in CPS. By simulating attack scenarios, this approach helps determine the most critical vulnerabilities, allowing for targeted and robust security measures. Similarly, Article [12] employs ML within the software development lifecycle to predict potential vulnerabilities, ensuring that security measures are continually updated and refined. Further enhancing predictive threat modeling, the Threat Operating Model (TOM) presented in Article [19] utilizes AI to analyze extensive cyber threat intelligence. By clustering and summarizing threat intelligence data, TOM provides early warnings and identifies potential threats with high accuracy.

5.1.2.2 Semantic and contextual threat analysis

Semantic and Contextual Threat Analysis in cybersecurity utilizes NLP to perform an in-depth exploration of the semantics embedded in textual data, facilitating a nuanced understanding of cyber threats specific to various contexts. By parsing and interpreting the subtle meanings and contextual cues within large volumes of text—from technical documents to social media posts—this approach allows security analysts to detect and decode complex threat patterns and potential vulnerabilities that might otherwise go unnoticed. This strategy is best shown in Article [68], which carefully analyzes and categorizes cyber threats from unstructured sources, including news articles and social media, especially within healthcare systems, using NLP and Named Entity Recognition (NER) technologies. By using this approach, security threats can be proactively monitored and identified, enabling responses to be customized to the unique requirements of the healthcare industry. Similarly, Article [161] uses novel NLP methods to carry out an extensive threat assessment and management procedure that is specific to healthcare environments. It scans

multiple data sources to detect potential threats, using semantic analysis to evaluate their impact and likelihood, thereby providing actionable insights that support effective decision-making and risk management.

5.1.2.3 Automated threat classification *Threat Classification* is an essential process in SSA that involves the systematic identification and categorization of cyber threats, which utilizes predefined rules and databases of known threats to filter and classify incoming data, identifying potential security risks based on recognized patterns. The integration of AI into automated threat classification enables analysis and learning from the data, adapting to new and evolving threats dynamically, going beyond static databases and rule-based algorithms.

Article [88] exemplifies this approach by using ML algorithms to analyze network traffic data within a cloud computing framework, effectively detecting and classifying network threats based on their characteristics. This system is designed to be scalable and efficient, capitalizing on cloud resources for real-time data processing and proactive threat mitigation. Similarly, Article [193] focuses on refining cybersecurity audits through systematic profiling of cyber threats using feature extraction and selection techniques, which aid auditors in identifying potential vulnerabilities and preparing effective audit strategies. Additionally, Article [154] utilizes an Enhanced Naïve Bayes Posterior Probability (ENBPP) model to improve threat detection and classification, integrating a risk assessment function that allows for more accurate predictions by addressing feature interdependencies.

5.2 Assurance techniques and evidence collection

A summary of the selected studied in the category of assurance techniques and evidence collection is shown in Table 4. Below are thorough explanations of each of the categories included in this section.

5.2.1 Penetration testing

Penetration Testing serves as a proactive measure to identify and address potential vulnerabilities within a system or network before they can be exploited by malicious actors. This form of testing involves simulating cyberattacks under controlled conditions to assess the effectiveness of existing security measures and pinpoint weaknesses in an organization's cybersecurity defenses [155]. AI-driven penetration testing tools, which can adjust their strategies based on real-time data, can simulate complex cyber-attack scenarios more dynamically and realistically than traditional methods [120].

5.2.1.1 Dynamic penetration testing *Dynamic Penetration Testing*, integral to security assurance, employs AI to continuously refine and adapt evaluation strategies in real-time. This approach is critical for responding effectively to evolving network conditions and emerging threats, enhancing the robustness of security assessments.

Given the dynamic nature of network environments, agile and adaptive penetration testing methodologies are essential. Articles [44] and [45] exemplify this by employing Reinforcement Learning (RL) to model penetration testing tasks as a partially observed Markov decision process, continuously improving testing strategies based on network interactions. This approach dynamically adapts to new threats and configurations, reducing dependence on human intervention and enhancing systematic and autonomous testing processes. Additionally, Article [43] applies similar dynamic testing methodologies within vehicular ad-hoc networks, using a RL model to simulate network scenarios for connected and autonomous vehicles. This allows for automated decision-making processes, improving test efficiency and efficacy by adjusting strategies based on real-time network states and potential attack paths.

Lastly, Article [140] introduces AgentPen, a system that conducts autonomous penetration testing using a Deep Neural Network (DNN) and Q-learning. This agent learns from each interaction and adjusts its strategy in real time to adapt to new and changing environments, eliminating the need for predefined models. The neural network component improves the learning process, allowing the agent to handle more complex scenarios and outperforming traditional, rigid models.

5.2.1.2 Penetration testing optimization In order to optimize penetration testing processes, several studies have investigated novel techniques for refining security testing methodologies, with a particular emphasis on the incorporation of advanced technologies and algorithms. Article [57], for example, discusses the use of optimization algorithms, specifically Ant Colony Optimization (ACO), to improve ethical hacking procedures in healthcare information systems. This method simulates natural foraging behaviors to identify the most effective attack paths, demonstrating how bio-inspired algorithms can significantly improve the systematic discovery of vulnerabilities.

Articles [64] and [30] advance DL by utilizing Deep Q-Learning Networks (DQNs). Article [64] applies this technology to automate the detection and evaluation of optimal attack paths in network environments. Similarly, Article [30] uses a Deep-Q Network to create strategic and autonomous attack graphs. These efforts represent a significant shift toward automating complex decision-making processes, allowing for more efficient test execution while reducing human oversight. Adding to this technological

Table 4 Summary of the selected article in assurance techniques and evidence collection

Category	Sub-category	Selected article
Penetration testing	Dynamic penetration testing	[43–45, 140]
	Penetration testing optimization	[30, 57, 64, 71, 78, 103, 142, 178]
	Penetration testing in large-scale network	[47, 100, 101, 190]
Fuzz testing	Fuzz data generation	[104, 127, 148, 198]
	Fuzz testing performance improvement	[50, 96, 124, 138]
Vulnerability detection	Network traffic and behavioral analysis	[25, 28, 31, 40, 45, 72, 141]
	Firmware and binary vulnerability analysis	[167, 184]
	Advanced vulnerability analysis	[42, 113, 118, 133, 137, 191]
	Miscellaneous vulnerability detection techniques	[33, 51, 70, 111, 174, 185]

focus, Article [71]evaluates the GyoiThon penetration testing tool, which enhances the detection of vulnerabilities in web applications through its ML feature. By employing a Naïve Bayes algorithm, GyoiThon systematically analyzes website features to identify potential security flaws, showcasing a more automated and effective approach to vulnerability assessment.

Further advances are demonstrated by creative models and procedures. Article [103] provides a more thorough assessment of network vulnerabilities by incorporating social engineering elements into penetration test simulations. In order to improve policy performance and learning efficiency, Article [178] improves DL networks using methods such as Dueling network architectures and the Epsilon Greedy-UCB algorithm. In the meantime, Article [142] optimizes attack planning in intricate network environments by fusing RL with an ontology-based framework. Lastly, Article [78] utilizes a CNN to automate attack path generation and optimization, transforming attack graphs into executable codes for interactive testing.

5.2.1.3 Penetration testing in large-scale network Recent research on penetration testing for large-scale networks has centered on developing methodologies that address the complexities and scale inherent in modern network architectures. Notably, Articles [101] and [47] introduced hierarchical and distributed testing models, which are useful in maximizing the scope and efficiency of penetration tests. These papers advocate for a structured approach in which tasks are distributed among multiple specialized agents (Article [101]) or across different network clusters (Article [47]). This division provides strategic oversight and detailed local analysis, making the penetration testing process more manageable and focused.

Articles [190] and [100], which focus on improving the decision-making procedures used in penetration testing, highlight additional developments in the field. Article [190] improves the planning and execution of penetration tests, increasing the speed and accuracy of vulnerability identification by combining attack graph tools and sophisticated decision-making algorithms. However, Article [100] streamlines the decision-making process in penetration testing by minimizing ineffective explorations and concentrating efforts on likely successful strategies by utilizing a Markov decision process framework.

5.2.2 Fuzz testing

Fuzz Testing involves inputting vast amounts of random data ("fuzz") into software systems to trigger errors, crashes, or security loopholes [92]. Fuzz testing uses AI techniques like deep reinforcement learning, neural networks, and DQNs to improve test input generation, intelligently optimize input mutations, and fine-tune mutation strategies. This greatly increases the efficacy, coverage, and efficiency of vulnerability detection.

5.2.2.1 Fuzz data generation *Fuzz Data Generation* is a pivotal aspect of software testing that involves creating random data inputs to test software systems for vulnerabilities, such as crashes or bugs. Recent innovations in this domain have integrated advanced ML techniques to automate and enhance the fuzz testing process. Notably, the methodologies used in Articles [104] and [127] both make use of Wasserstein Generative Adversarial Networks (WGANs). By creating test data for ICS on its own without requiring comprehensive protocol specifications, Article [104] creates an intelligent fuzzing technique that significantly increases testing effectiveness and vulnerability detection. Similar to this, Article [127]'s

GANFuzzer greatly improves the detection of potential security breaches by using WGAN enhanced with a self-attention mechanism to generate realistic and varied test cases for industrial control protocols.

Conversely, Article [148] uses Recurrent Neural Networks (RNNs) to generate HTML tags for browser fuzz testing. With this method, creating test cases is streamlined and syntactic rules are followed while introducing novel inputs to cause unexpected behaviors in browsers. This results in broader coverage and more accurate identification of new bugs. Furthermore, Article [198] presents the SeqFuzzer framework, which effectively identifies security flaws in intricate systems like the EtherCAT protocol by using a sequence-to-sequence DL model to comprehend and replicate the structures of stateful industrial protocols.

5.2.2.2 Fuzz testing performance improvement The use of AI in fuzz testing has greatly improved the performance and efficiency of vulnerability detection in SSA. Researchers have developed various AI techniques to create methods and frameworks that optimize the fuzz testing process, resulting in more precise and effective results.

Articles [96, 124], and [50] provide examples of how ML can be applied to improve the fuzz testing process. Article [96] reduces computational demands and improves testing efficacy by using a DQN to intelligently select input mutations that are more likely to expose vulnerabilities. Similarly, Article [124] increases the rate of vulnerability discovery and code path coverage by utilizing ML to enhance test input generation and analysis. In addition, Article [50] presents the "DRLFCfuzzer," which optimizes fuzz testing under particular format constraints by leveraging deep reinforcement learning. By modifying the mutation strategy to adhere to the software's input format constraints, this tool improves the identification of software vulnerabilities.

Furthermore, AI-enhanced fuzz testing has advanced with the help of specialized frameworks covered in Articles [138] and [49]. In order to improve vulnerability detection in complex 5G systems, Article [138] describes the DEFT framework, which analyzes log files from NextG network fuzz tests using neural networks and word embedding techniques. Article [49], on the other hand, presents the Static Seed Selector- American Fuzz Lop (SSG-AFL) framework, which greatly enhances program coverage and bug detection efficiency by optimizing seed selection in fuzzing reactive systems through the use of a static seed generator.

5.2.3 Vulnerability detection

Vulnerability Detection is a critical component of SSA, encompassing a wide range of techniques for identifying security weaknesses within various systems and applications. The incorporation of AI into this process has transformed the field, providing powerful tools to improve the accuracy, efficiency, and speed of detecting vulnerabilities. AI-powered techniques, such as ML and neural networks, enable real-time monitoring and analysis, revealing hidden vulnerabilities and providing more insight into potential security breaches.

5.2.3.1 Network traffic and behavioral analysis *Network Traffic and Behavioral Analysis* involves the monitoring and evaluation of data flowing through a network to detect irregular activities and potential security threats. This process is improved by advanced AI techniques, which analyze data flows intelligently and look for odd patterns that could indicate security breaches. By using AI in this way, network behavior can be assessed more precisely and dynamically, which helps identify threats early and greatly enhances an organization's overall security posture.

Research papers like Articles [21, 72], and [31] show how AI can help advance vulnerability detection by using behavioral analytics and network traffic analysis. In Article [21], the MANDRAKE methodology (a method for vulnerabilities detection based on the IoT network packet)—which focuses on smart home scenarios—is presented. It uses ML and network traffic entropy to identify vulnerabilities in Internet of Things (IoT) devices without requiring physical device access. Similar to this, Article [72] describes AppMine, a framework that analyzes temporal dependencies in application behavior to detect anomalies in web applications running in Docker containers. AppMine uses unsupervised learning models like PCA, one-class Support Vector Machine (SVM), and Long-Short Term Memory (LSTM) neural networks. A ML-based framework for identifying, categorizing, and mitigating botnet vulnerabilities is covered in Article [31], which significantly improves network security against such attacks.

Articles [28, 40], and [141] highlight the variety of AI applications by concentrating on specific network contexts. Article [28] automatically analyzes Long Term Evolution (LTE) documentation to find possible security risks in cellular networks by utilizing ML and NLP. Gradient Boosting is a high-efficiency substitute for conventional methods in vulnerability scanning and penetration testing procedures. It is applied in Article [40] to analyze port responses. Lastly, Article [141] uses sequence models to identify security flaws in wireless networks, demonstrating how AI can analyze communication patterns and flag possible security risks in advance.

5.2.3.2 Firmware and binary analysis *Firmware and Binary Analysis* examine the low-level code in software programs and devices to identify vulnerabilities and ensure compliance with security standards. This type of analysis is critical for detecting security flaws in the closed-source binaries commonly found in mobile and IoT devices, where traditional source code auditing is not possible. Integrating AI

into this field allows for the detection of complex patterns and anomalies that traditional methods may miss, significantly increasing the accuracy and efficiency of vulnerability detection.

The integration of AI in vulnerability analysis of firmware and binary programs is advancing rapidly, as evidenced by a vulnerability and patch presence detection framework for executable binaries, named PATCHECKO, described in Article [167] and the research presented in Article [184]. The PATCHECKO framework employs a hybrid approach that combines dynamic binary analysis and static analysis with DL to scrutinize mobile and IoT firmware binaries. This method significantly enhances precision and reduces false positives by using ML to compare binary function pairs and identify similarities linked to known vulnerabilities. Similarly, Article [184] utilizes DL models, including CNN, LSTM, and a combined CNN-LSTM model, to analyze function call sequences during program execution, identifying potential vulnerabilities. This approach surpasses traditional methods by capturing both the structural and temporal dynamics of function calls, highlighting the profound impact of AI on improving firmware and binary analysis techniques.

5.2.3.3 Advanced vulnerability analysis In order to increase the efficiency and accuracy of vulnerability detection, a number of studies have proposed an advanced approach that uses real-time data processing to recognize and address security threats as they materialize. By utilizing AI techniques, this method analyzes system calls, network traffic, and user interactions within applications to assess how the system behaves while it is being used. For instance, Articles [42] and [118] focus on Android systems, employing ML to examine system calls and application behaviors to efficiently identify malware and security flaws. Both studies highlight the use of AI to manage and categorize the vast amounts of data from system calls, detecting unusual patterns that indicate potential security risks.

Articles [133] and [113] examine the application of AI in the field of industrial and network systems for the purpose of monitoring and safeguarding complex environments. Article [133] focuses on real-time data processing and feature extraction capabilities as it assesses various ML algorithms for their efficacy in identifying vulnerabilities in the Industrial Internet of Things (IIoT). A big data security analysis platform that uses Classification And Regression Tree (CART) decision trees and Spark-streaming to analyze network data streams in real-time and improve the detection of network vulnerabilities and exploits is covered in Article [113].

Furthermore, advanced analysis in web applications and firmware vulnerability detection are covered in Articles [137] and [191], respectively. In Article [137], anomalies and possible security breaches in heterogeneous web applications are identified through real-time analysis of HTTP and SOAP transactions using Probabilistic Arithmetic Automata (PAA). In order to improve the accuracy of firmware vulnerability detection in IoT devices, Article [191] presents an AI-driven method that makes use of the Clonal Selection Algorithm (CSA). This approach eliminates the need for substantial training datasets.

5.2.3.4 Miscellaneous vulnerability detection techniques Several innovative studies stand out for their novel approaches to vulnerability detection, which fall into a more diverse category due to their disparate methodologies and applications. These papers demonstrate how AI can be used to address specific and diverse security challenges across multiple software platforms.

One notable approach is presented in Article [33], which introduces a depth-wise separable Convolutional Neural Network (CNN) for cybersecurity vulnerability assessment through CAPTCHA breaking. This method enhances the detection of vulnerabilities in CAPTCHA systems, which are widely used to differentiate human users from automated bots. By efficiently breaking CAPTCHAs, the model identifies weaknesses that could be exploited by automated attacks, thereby improving the robustness of web security measures. Simultaneously, Article [70] leverages Multi-Layer Perceptron (MLP) to refine software vulnerability detection, improving predictive accuracy by processing results from various base models and handling nonlinear data.

Moreover, Article [51] develops a tailored methodology to optimize the selection and configuration of vulnerability detection tools for specific software, ensuring the deployment of the most effective tools to uphold robust security standards. In the realm of IoT devices, Article [111] employs NLP and a multi-layer matching algorithm to enhance the accuracy of vulnerability assessments by analyzing unstructured Common Vulnerabilities and Exposure (CVE) descriptions and correlating them with device information, thus reducing false positives and negatives.

Additionally, the use of black box testing methodologies is evolving. Article [174] applies state machine inference to detect vulnerabilities in mobile applications, specifically targeting Android platforms, by effectively modeling app behaviors. Meanwhile, Article [185] examines the security of financial systems against adversarial attacks, evaluating the vulnerabilities in credit card fraud detection models and showcasing the challenges in ML-based systems.

5.3 Compliance assessment and analytics

Table 5 displays a summary of the selected studies in the compliance assessment and analytics category. Detailed explanations of each sub-category in this section can be found below.

Table 5 Summary of the selected article in compliance assessment and analytics

Category	Sub-category	Selected article
Security evaluation and analysis	Compliance assessment	[36, 46, 56]
	Dynamic security assessment	[74, 106, 108, 163, 170]
	Security analytics	[38, 132, 171]
	Evaluation techniques for specific technologies	[3, 87, 105, 144, 149, 166]
Vulnerability analysis and prediction	Predictive vulnerability modeling	[85, 102, 126, 150, 158]
	Vulnerability classification	[76, 112, 117, 131]
	Vulnerability impact analysis	[66, 99, 199]
	Vulnerability analytics	[9, 41, 77, 128, 139]
Risk assessment	Comprehensive risk assessment framework	[5, 22, 23, 48, 62, 75, 95, 122, 175, 179]
	Quantitative risk analysis	[7, 65, 89, 176, 189, 192, 195, 196]
	Risk classification and prioritization	[8, 143, 194]
	Automated and dynamic risk analysis	[4, 16, 109, 119, 162]
	Predictive risk modeling	[11, 24, 63, 93, 94, 157, 187]
	Privacy aspect risk analysis	[39, 54, 80]

5.3.1 Security evaluation and analysis

This section provides an in-depth exploration of methodologies and tools designed to evaluate and analyze system security posture. It examines how advanced AI techniques are applied to assess system vulnerabilities, ensure compliance with security standards, and gauge the effectiveness of implemented security measures.

5.3.1.1 Compliance assessment In the context of SSA, *Compliance Assessment* refers to systematically reviewing and verifying that a system' security posture complies with established security standards, regulations, and best practices [180]. AI techniques have been increasingly applied to streamline and enhance the accuracy of these assessments, making them more efficient and reliable. We present papers in this category that introduce sophisticated modeling methods for dynamically assessing and managing compliance, utilizing sophisticated computational techniques to adjust to modifications and guarantee continuous adherence to security protocols.

The tool Assessment and Management of Organizational Evidence (AMOE), introduced in Article [36], automates the extraction and evaluation of organizational evidence from cloud services. To increase the transparency of security and privacy measures, AMOE assesses data from textual policy documents using NLP and Question-Answering techniques. With the help of this tool, continuous audit-based certification that complies with European cybersecurity standards can be supported, greatly minimizing the amount of manual labor needed for ongoing audits of cloud service providers. Similarly, Article [56] centers on the semi-automatic assessment of security features in software requirements. This method compares software requirements documents to recognized security standards like ISO and OWASP by using NLP and ML. By assessing semantic relationships through textual entailment and neural networks, the method determines whether the security requirements are adequately specified and compliant with the standards. Meanwhile, Article [46] presents the Expert-System Automated Security Compliance Framework (ESASCF), which employs rule-based expert systems to automate network security compliance tasks such as vulnerability assessments and penetration testing. By capturing and generalizing expert knowledge, this system reduces the time and resources required for security audits and enhances the consistency and accuracy of compliance measures.

5.3.1.2 Dynamic security assessment *Dynamic Security Assessment* (DSA) enables real-time evaluation of system vulnerabilities and stability, supporting proactive management of security risks. AI techniques have transformed how DSA is conducted, allowing for continuous monitoring and assessment of changing conditions. Several studies highlight different AI-driven approaches to enhance the effectiveness and efficiency of DSA across various systems.

In power systems, data-driven and hybrid ML techniques have shown significant promise. Article [106] introduces a DSA framework using Sparse Projection Oblique Randomer Forests (SPORF) for real-time stability analysis. By integrating phasor measurement units, this method continuously monitors system conditions, addressing class imbalance and feature selection to ensure accurate transient stability assessments. Article [163] enhances DSA

with a hybrid-extreme learning machine (ELM) approach, integrating Particle Swarm Optimization (PSO) and Levenberg–Marquardt (LM) algorithms. This ensemble method uses transient energy function terms and fault location data to predict stability margins, optimizing performance in real-time security assessments for power systems.

Adaptive learning and continuous monitoring are pivotal in dynamic assessment frameworks. Article [108] employs semi-supervised learning, utilizing both labeled and unlabeled data to dynamically adapt to new conditions, reducing the need for extensive simulations and improving assessment efficiency. Article [74] presents a three-stage framework for ICSs using a Weighted Hidden Markov Model (W-HMM). This model updates its parameters to reflect changes in system behavior and emerging threats, ensuring accurate real-time security assessments. Finally, AI techniques are also applied to software vulnerabilities. Article [170] discusses Mission-Oriented Network Analysis (MONA), an AI-driven approach that uses neuro-fuzzy systems, regression analysis, and neural networks for dynamic vulnerability impact analysis. By handling ambiguity in vulnerability data and identifying underlying patterns, this model supports real-time, data-driven assessments critical for addressing security threats.

5.3.1.3 Security analytics *Security Analytics* leverages data analytics and advanced algorithms to enhance the detection, analysis, and response to security threats. This process involves collecting and analyzing vast amounts of data from various sources, such as network traffic, user behaviors, and system logs, to uncover patterns and anomalies that may indicate potential security risks [156]. Leveraging AI and ML, these methodologies enhance the accuracy and efficiency of security evaluations across various domains.

Article [171] introduces a DSS that uses Genetic Algorithms (GAs) to optimize security configurations in critical infrastructures. By integrating the Open-Source Security Testing Methodology Manual (OSSTM) with additional security indicators and vulnerability data, the DSS provides a robust framework for evaluating and optimizing security measures dynamically. Similarly, Article [38] focuses on IoT networks, developing an automated security assessment framework that combines ML and NLP to predict vulnerability severity and map out potential attack paths. This approach enhances the efficiency and accuracy of IoT security analysis, providing actionable insights for cybersecurity professionals. Additionally, Article [132] proposes a novel methodology to calculate software security scores by semantically comparing Common Weakness Enumeration (CWE) and CVE using NLP. This enhanced semantic analysis results in more accurate security scores, helping developers and security professionals prioritize and mitigate vulnerabilities effectively.

5.3.1.4 Evaluation techniques for specific technologies Security evaluation techniques designed for specific technologies ensure that the unique challenges and vulnerabilities of each technological domain are effectively addressed. These techniques use advanced AI methodologies to improve the accuracy and efficiency of security assessments, thereby providing strong frameworks for protecting critical systems. This section examines various AI-driven security evaluation techniques used in power systems, cyber-physical systems, 5G networks, the Internet of Medical Things (IoMT), smart grid databases, and healthcare devices.

Several studies focus on the application of DL for security assessment in different technological domains. Article [166] details a DL-based feature extraction framework using deep autoencoders to optimize system security assessments in power systems. This framework translates high-dimensional data into a lower-dimensional, informative feature space, enhancing the ability of classifiers to detect safe and unsafe operational states. Article [105] proposes a CNN methodology for automating security audits of smart grid databases. The CNN efficiently extracts and analyzes data, identifying potential security risks and improving the speed and accuracy of security audits compared to traditional methods.

Other studies emphasize the use of AI-driven models to assess vulnerabilities in more specialized systems. Article [87] focuses on CPS, employing DL models to generate stealthy attacks that disrupt system operations without detection. By simulating real-time attacks using Feed-Forward Networks (FNNs) and RNNs, this approach evaluates the resilience of CPS under adversarial conditions. In the realm of 5G networks, Article [149] presents a framework utilizing ML to construct attack graphs that identify vulnerabilities in the 5G core network, particularly within its software-defined networking and network function virtualization components. This method highlights how vulnerabilities in protocols like the 5G authentication and key agreement can be exploited to compromise network security.

Additionally, studies focus on emerging technologies such as IoMT and healthcare devices. Article [144] explores Quantum ML (QML) for security assessments in the IoMT, leveraging quantum computing principles to handle the complexity and velocity of IoMT data more efficiently. This approach offers a robust and faster analysis of potential security threats, enhancing the accuracy of security assessments. Article [3] details a method using the Fuzzy Analytic Network Process (ANP) integrated with Order Preference by Similarities to Ideal Solution (TOPSIS) to evaluate the efficacy of security techniques for healthcare devices. By systematically collecting and analyzing quantitative and qualitative data, this methodology ranks security mechanisms based on their performance, supporting the optimization of security implementations in healthcare environments.

5.3.2 Vulnerability analysis and prediction

This section explores advanced methodologies and techniques for identifying, evaluating, and forecasting potential security vulnerabilities within systems and applications. By leveraging advanced AI technologies, this section highlights how these approaches enhance the accuracy, efficiency, and effectiveness of vulnerability assessments.

5.3.2.1 Predictive vulnerability modeling *Predictive Vulnerability Modeling* is a proactive cybersecurity approach that uses advanced analytics and AI techniques to forecast potential security weaknesses. By analyzing historical data, system configurations, and known vulnerabilities, this method identifies patterns and correlations that indicate future security risks. In this category, we include articles that develop predictive models to estimate the likelihood and severity of vulnerabilities using advanced AI computational techniques.

Several studies highlight the application of AI in predictive vulnerability modeling. Article [85] presents a multiclass hybrid approach that utilizes ML models to estimate software vulnerability vectors and severity scores. This model's predictive accuracy is assessed through detailed error analysis and validation techniques, optimizing its sensitivity to different vulnerability severity levels. Similarly, Article [158] proposes an XLNet-based model for predicting CVSS metric values from vulnerability descriptions, automating the traditionally manual and labor-intensive process. Both studies emphasize the importance of precise and reliable predictions in enhancing the efficiency of vulnerability assessments.

Another group of studies explores the use of GAs and DL for vulnerability prediction. Article [150] introduces a method combining deep symbiotic GAs with LSTM and Gated Recurrent Unit (GRU) models to enhance feature selection and predict software vulnerabilities based on phenotypic patterns. Additionally, Article [102] leverages Pre-trained Language Models (PLMs) and prompt learning to predict vulnerability characteristics from descriptions, demonstrating improved performance with fewer training data. These methodologies illustrate how integrating DL and GAs can enhance the predictive capabilities of vulnerability models. Lastly, addressing the complexity of heterogeneous web environments, Article [126] proposes a model using PAA to predict and evaluate security risks. This approach provides a dynamic and adaptable framework for assessing and mitigating vulnerabilities in diverse web applications.

5.3.2.2 Vulnerability classification *Vulnerability Classification* involves categorizing security weaknesses based on various attributes, such as severity, exploitability, and potential impact. Advanced AI techniques, including ML and DL, have significantly enhanced the accuracy and efficiency of vulnerability classification, providing detailed and actionable insights into cybersecurity threats. Several studies highlight the application of DL and ML models in this field. For instance, Article [112] introduces a character-level Convolutional Neural Network (charCNN) to predict the exploitability of software vulnerabilities by analyzing detailed textual patterns in vulnerability descriptions. This method outperforms traditional models in precision and robustness. Similarly, Article [131] presents an AI-assisted framework for active directory environments using graph-based techniques and Random Forest classifiers. This framework models network structures to identify and classify potential attack paths, continuously learning and adapting to new data to ensure up-to-date vulnerability assessments.

Other studies focus on integrating ML models to improve the accuracy and speed of vulnerability classification. Article [76] utilizes text-mining techniques and a majority voting system to harmonize vulnerability severity scores, providing a robust foundation for ML models to classify vulnerabilities in cyber-physical systems. This approach automates the severity assessment, aiding in efficient resource allocation and remediation prioritization. Similarly, Article [117] proposes the Cyber-threats and Vulnerability Information Analyzer, which employs both unsupervised and supervised ML models to analyze data from databases like NVD and MITRE, reducing human intervention and enhancing the speed and accuracy of vulnerability classification.

5.3.2.3 Vulnerability impact analysis *Vulnerability Impact Analysis* involves evaluating the potential consequences of security vulnerabilities within a system. This process is essential for understanding the severity and potential damage of identified vulnerabilities, enabling organizations to prioritize mitigation efforts effectively. AI techniques, particularly ML, have significantly enhanced the accuracy and efficiency of vulnerability impact analysis, providing detailed insights that inform proactive security measures networks and devices.

A novel methodology for device-type fingerprinting in IIoT environments is introduced in Article [199], utilizing ML techniques to analyze TCP/IP protocol headers and open ports. This approach identifies device types and their characteristics, aiding in vulnerability impact analysis and enabling targeted security measures based on precise device configurations. For the impact analysis of network vulnerabilities, Article [99] employs ML algorithms to identify network service dependencies and assess the potential effects of vulnerabilities on ongoing network activities. This methodology pinpoints critical network components that could be exploited, allowing operators to concentrate security efforts on the most vulnerable areas, thereby enhancing the network's overall security posture. Additionally, Article [66]

explores a graph-based approach to analyze how vulnerabilities propagate through interconnected systems. By mapping interdependencies and potential attack paths, this method provides a comprehensive impact analysis, helping prioritize security measures where they are most needed and supporting informed, strategic decisions in vulnerability management.

5.3.2.4 Vulnerability analytics *Vulnerability Analytics* include the ability to predict potential security flaws and prioritize them for remediation. This field often relies on historical data to forecast future risks and utilizes advanced techniques to provide accurate, actionable insights into security vulnerabilities. By incorporating AI techniques, such as ML, neural networks, and advanced visualization methods, vulnerability assessments can be made more accurate, interpretable, and efficient.

Several studies utilize ML to enhance vulnerability analytics. Article [139] introduces a mechanism for assessing network security through hyper-heuristic ML techniques. This approach combines real-world data with genetic programming to evolve graph-based heuristics, predicting attack paths and identifying potential breaches. Similarly, Article [41] describes a methodology for vulnerability analysis using multiclass perceptron's to classify asset vulnerabilities based on metrics like severity and the age of the last scan. These classifications help prioritize assets for remediation, enhancing overall security management.

Other studies focus on improving the interpretability and usability of vulnerability assessments. Article [9] presents an approach that integrates Explainable AI (XAI) and advanced visualization techniques to make vulnerability scoring systems more understandable for security analysts. By clustering vulnerabilities using natural language features and visual analytics, this study provides a comprehensive view of relationships and severities within a dataset, improving decision-making processes. Additionally, Articles [77] and [128] explore automated vulnerability discovery and assessment methods. Article [77] integrates ML algorithms with data from vulnerability databases to evaluate the impact and severity of vulnerabilities in cyber-physical systems, facilitating prioritized responses. Article [128] utilizes structural and social influence metrics to develop a vulnerability assessment scheme for Mobile Social Networks (MSNs), providing a proactive tool for identifying and mitigating potential security threats based on node mobility and social interactions.

5.3.3 Risk assessment

Risk Assessment serves as a foundational element in evaluating the overall security posture of a system, providing a comprehensive understanding of the risks that could impact its integrity, confidentiality, and availability. This section delves into sophisticated AI techniques for identifying, evaluating, and prioritizing potential risks within systems.

5.3.3.1 Comprehensive risk assessment framework Risk assessment frameworks are methodical techniques that are employed to recognize and assess possible threats to the security posture of a system. These frameworks help organizations to understand their risk exposure, develop mitigation strategies, and enhance their overall system security [186]. Traditional risk assessment frameworks often rely on qualitative and quantitative methods, but the integration of AI has revolutionized these processes, enabling more accurate, scalable, and context-aware assessments. AI technologies such as ML, fuzzy logic, and neural networks significantly enhance risk assessment frameworks by automating complex data analyses and providing dynamic, data-driven insights.

Several studies highlight the application of AI in developing comprehensive risk assessment frameworks. Article [23] introduces a model for selecting cloud service providers using ML to automate risk analysis. By converting qualitative questionnaire data into quantitative risk values, this model aids customers in making informed decisions based on security, privacy, and service delivery metrics. Similarly, Article [48] focuses on autonomous vehicle perception systems, integrating ISO/SAE 21434 standards [69] with AI methodologies to address cyber-physical threats. The framework employs ML algorithms to continuously evaluate risks, enhancing the precision of threat identification and the robustness of security measures.

In the realm of critical infrastructure, Article [95] proposes an integrated cyber security risk management framework that leverages ML, fuzzy set theory, and cyber threat intelligence. This framework effectively predicts various cyber risks and assesses control effectiveness, supporting proactive risk management. Additionally, Article [62] presents a risk assessment framework utilizing GAs to automate and optimize risk management processes. This approach incorporates variables such as asset value, impact, and likelihood, allowing for customized and efficient identification of critical security risks.

Fuzzy logic and ML also play a pivotal role in enhancing risk assessments. Article [22] employs fuzzy rule sets to better quantify and manage uncertainties in network security environments. By correlating and assessing security events, this method improves the accuracy of risk assessments over traditional approaches. Meanwhile, Article [75] focuses on network information security, utilizing ML algorithms to analyze large datasets, identify patterns, and detect anomalies. This framework enhances the understanding of the network's security state and informs targeted reinforcement strategies.

In Article [122], a ML-based risk analysis is applied to Android applications, utilizing permission analysis to create

a risk index that flags possible security risks. By enhancing the ability to predict app behavior, this dynamic assessment method supports preventive security measures. Finally, a CNN-based framework for network information security is presented in Article [179], which can adjust to real-time data in order to deliver precise and timely risk assessments. This strategy guarantees thorough and ongoing security assessments, greatly enhancing the decision-making procedures involved in network security management.

5.3.3.2 Quantitative risk analysis *Quantitative Risk Assessment* (QRA) focuses on the measurement and quantification of security risks, specifically analyzing the potential impacts and vulnerabilities within systems to provide numerical or categorical risk levels. Various studies have proposed different methodologies to perform QRA, each leveraging advanced AI technologies to enhance the accuracy and effectiveness of the risk assessment process, providing more dynamic, adaptable, and precise risk assessments.

AI technologies have been increasingly applied in QRA to handle the complexity and dynamism of modern security threats. For instance, Bayesian networks, as discussed in Article [176] and Article [65], model complex probabilistic relationships and manage uncertainties in cyber-physical systems and Supervisory Control and Data Acquisition (SCADA) networks. These networks integrate real-time and historical data to continuously update and refine the risk model, improving the detection and analysis of new and evolving threats.

Fuzzy logic-based approaches also offer nuanced assessments of security risks by handling uncertainties and partial truths. Article [192] and Article [7] demonstrate the application of fuzzy logic in different contexts. The fuzzy logic-based risk scoring system for mobile applications (Article [192]) evaluates app permissions and features against predefined security criteria, incorporating results from various antivirus tools to enhance accuracy. Additionally, the Fuzzy Inference System (FIS) used in the security assessment of information assets (Article [7]) generates nuanced security ratings based on the assets' sensitivity and criticality, providing a more comprehensive evaluation compared to traditional binary classifications.

AI-driven methodologies also significantly enhance the precision and efficiency of QRA by leveraging big data analytics. Article [195] introduces an AI methodology using the fuzzy C-means clustering algorithm to evaluate network risks by clustering statistical data on network intrusions. This approach detects subtle and complex patterns indicative of potential threats, and the integration of big data technologies facilitates handling large data volumes, crucial for comprehensive risk assessments. Similarly, advanced ML techniques like tree boosting are employed for the security evaluation of mobile applications (Article [89]), analyzing API behavior patterns and comparing them against known malicious APIs to assign risk scores, thus facilitating proactive risk management.

Additionally, Article [196] illustrates the application of Artificial Neural Networks (ANNs) to process large datasets and identify patterns, enhancing the accuracy of security evaluations by continuously adapting to new threats. The ANN model quantifies potential security risks by detecting patterns and anomalies, ensuring that the risk analysis remains dynamic and data-driven, thereby providing a comprehensive and precise evaluation of information security risks.

5.3.3.3 Risk classification and prioritization *Risk Classification and Prioritization* focuses on identifying, evaluating, and ranking vulnerabilities within systems to effectively allocate resources and mitigate potential threats. These processes aim to provide accurate and scalable assessments by incorporating AI, enabling organizations to achieve context-aware evaluations. As a result, AI enhances decision-making and improves the overall security posture by ensuring that resources are directed towards the most critical vulnerabilities.

Recent studies demonstrate the application of AI in improving risk classification and prioritization. Article [143] introduces a ML-based model using SVM to classify organizations into cyber risk categories based on their cyber posture. By incorporating features such as maturity, complexity, and attractiveness indexes derived from public data, the model provides a predictive classification of organizations into low or high cyber risk classes. Further enhancing vulnerability risk prioritization, Article [194] presents a vulnerability risk prioritization system, called LICALITY, that integrates logical reasoning and DL. Unlike traditional methods like the CVSS, LICALITY captures attackers' preferences using a threat modeling method and a neuro-symbolic model blending neural networks and probabilistic logic programming. This allows for a more dynamic and context-aware risk prioritization by analyzing both the criticality and likelihood of exploitation. Additionally, Article [8] introduces an AI-driven approach combining NLP and ML to dynamically associate vulnerabilities with corresponding exploits. By processing textual descriptions and incorporating site-specific threat intelligence, this method tailors risk assessments to the actual tactics and preferences of potential adversaries.

5.3.3.4 Automated and dynamic risk analysis *Automated and Dynamic Risk Analysis* leverages AI techniques to monitor and evaluate security threats in real-time, reducing the need for manual intervention. By analyzing data from various sources such as system logs, network traffic, and user

behaviors, these techniques can identify vulnerabilities, predict potential threats, and assess their impact.

Articles [162] and [16] both utilize NLP and ML techniques for cybersecurity analysis. Article [162] applies AI models like BERT and Extreme Gradient Boosting (XGBoost) to process extensive natural language texts from the web, extracting and categorizing cybersecurity-related information within the healthcare ecosystem. Similarly, Article [16] employs a decision support system to analyze risks based on historical data and current security threats. This system also incorporates adversarial AI techniques to simulate attacks, continuously refining the risk assessment process through dynamic data interaction.

Another significant AI application involves Bayesian networks and ANNs. Article [4] presents a method for automating information security risk assessment in distributed computing networks, using data from information security means sensors to detect and assess threats in real-time. Bayesian networks model interdependencies between security variables, while ANNs predict potential security breaches by analyzing data patterns. Article [109] combines Hidden Markov Models (HMM) with attack graph models for network security risk assessment, capturing the probabilistic nature of network states and modeling uncertainties. The Viterbi algorithm calculates the most probable state transitions, identifying potential attack paths and intentions, thereby offering a quantitative mechanism for assessing risks by analyzing interdependencies and causal relations between vulnerabilities.

Lastly, AI and cloud technologies have been integrated into Industry 4.0 cybersecurity risk assessments. Article [119] outlines a method using an ICS testbed to conduct penetration tests in a simulated environment, leveraging AI to analyze data and detect anomalies. The cloud infrastructure scales the testing environment, enhancing analytical capabilities and enabling comprehensive risk evaluations.

5.3.3.5 Predictive risk modeling *Predictive Risk Modeling* focuses on developing models that learn from historical data to assess risks, allowing proactive management and mitigation. Several studies utilize ML, fuzzy logic, and neural networks to enhance the accuracy and interpretability of risk assessments. For instance, Articles [94]and [63] highlight the use of fuzzy logic combined with ML classifiers to predict risk types and assess asset criticality in cyber-physical systems and develop interpretable fuzzy scoring systems for dynamic risk evaluation, respectively.

Neural networks, particularly in advanced configurations, play a pivotal role in predictive risk modeling for dynamic environments. Article [157] presents a FNN-based model, named FNN-CRAM for managing cyber risks in online gaming firms, specifically targeting DDoS attacks. Article [93] uses a three-layer perceptron to assess risks in large-scale dynamic networks like smart cities, optimizing predictive accuracy through backpropagation. Both models analyze complex patterns in data, providing detailed risk assessments and mitigation strategies tailored to their respective domains.

Ensemble learning and Bayesian-based models further enhance predictive capabilities by integrating diverse data sources and leveraging probabilistic modeling. Article [187] introduces non-intrusive network security risk assessment prediction model, named NiNSRAPM, which is an ensemble learning-based non-intrusive network security risk assessment model using techniques like Random Forest, XGBoost, and LightGBM. Article [11] applies Bayesian neural networks for predicting cybersecurity risk severity, improving accuracy by incorporating historical data and expert insights. Additionally, Article [24] utilizes ML and big data analytics to model security risks in smart grid infrastructures, identifying patterns that signal potential threats and enabling proactive risk mitigation.

5.3.3.6 Privacy aspect risk analysis *Privacy Aspect Risk Analysis* in SSA focuses on identifying and mitigating risks related to personal data privacy, particularly in the context of modern technologies such as the Internet of Things (IoT) and mobile applications. Article [80] introduces the Massive Personal Information Clustering (MPIC) model, which assesses the risks associated with the aggregation of Personal Identifiable Information (PII) by IoT device manufacturers. By analyzing how personal data from multiple devices can be clustered to increase privacy risks, this model highlights the importance of understanding aggregated risks to protect user privacy in the IoT context. Similarly, Article [54] proposed Enterprise Smartphone Apps Risk Assessment (ESARA) framework, which uses the App behavior analyzer and App perception analyzer to monitor and analyze the security and privacy perceptions of enterprise smartphone apps. This dual analysis ensures that app activities align with security policies and that user perceptions are considered, providing a comprehensive view of privacy risks. Article [39] addresses privacy risks in AI-powered mobile cloud applications by introducing the Ability-based Scale for Intelligent Applications (ASIA) to measure app intelligence and a semi-quantitative threat modeling method. This approach systematically identifies potential security and privacy threats, offering risk scores that assess the effectiveness of existing security mechanisms against AI-driven challenges.

5.4 Continuous monitoring and improvement

This section delves into the AI methodologies and technologies that facilitate real-time monitoring and continuous enhancement of a system's security posture. Table 6 summarize the Selected Article in the category of continuous

Table 6 Summary of the selected articles in continuous monitoring and improvement

Category	Selected Article
Continuous monitoring and adaptation	[14, 55, 65, 74, 110, 160, 169]
Continuous improvement and optimization	[98, 106, 107, 139, 146, 171, 177]

monitoring and improvement. Detailed explanations of each category in this section are given below.

5.4.1 Continuous monitoring and adaptation

Continuous Monitoring and Adaptation entail the real-time evaluation and dynamic adjustment of security measures in response to immediate threats and changes within a system. This process necessitates continuous monitoring of security metrics and system performance in order to detect anomalies and vulnerabilities as they emerge. AI helps with this by using advanced techniques like ML and neural networks to analyze massive amounts of data in real time, identifying patterns that indicate potential threats. These AI models can then prompt immediate security protocol updates, ensuring that defense mechanisms are responsive and evolving.

Building on its earlier discussion in the context of dynamic security assessment, Article [74] also plays a significant role in the theme of continuous monitoring and improvement. Using the W-HMM, the three-stage framework for industrial control systems updates its parameters in real-time, allowing it to adjust to system changes and effectively counter new threats. Similarly, in its contribution in the QRA, Article [65] uses Bayesian networks for SCADA systems, combining offline batch learning with online incremental learning to adapt dynamically to new data. This approach ensures that risk assessments remain accurate and relevant, even as network conditions and threats evolve.

Adaptive modeling and certification also play a vital role in maintaining security across various systems. In network environments, real-time monitoring significantly boosts security. Article [110] presents a security assessment framework for swarm networks, integrating a MLP and an XGBoost classifier to continuously analyze network traffic, facilitating immediate threat detection and classification. Similarly, Article [14] addresses continuous certification of non-functional properties in service-based systems by employing ML to automate re-certification as operational environments evolve. This approach ensures that certifications remain valid and up to date by identifying and assessing behavioral changes.

Specialized domains like IoT, cloud environments, and supply chains benefit significantly from continuous monitoring. Article [160] details a method for monitoring IoT systems using audit hooks and ML to detect anomalies in real-time. Article [169] explores a neuro-fuzzy model for dynamic security evaluation in cloud environments, addressing various QoS parameters to provide adaptive assessments. Article [55] describes a approach (under the EU funded project FISHY) for continuous security assurance in supply chain ecosystems, particularly in autonomous driving. This FISHY approach integrates ongoing monitoring and adaptive security controls to maintain high security standards across the supply chain.

5.4.2 Continuous improvement and optimization

Continuous Optimization and Improvement aims to strengthen security measures over time and make them more resilient to changing threats. This methodology entails a methodical examination of performance data, pinpointing opportunities for improvement, and executing modifications to gradually optimize the overall security posture. AI methods that drive the iterative improvement of security protocols, like data analytics, DL, and genetic algorithms, are vital to this process. For instance, SPORF are used in Article [106] to assess dynamic security in power systems. This technique ensures long-term stability and improves the system's responsiveness to new threats by continuously optimizing security protocols based on real-time data. Similarly, Article [171] describes a DSS that analyzes and optimizes security configurations in critical infrastructures using genetic algorithms.

In the context of smart grids and supply chains, continuous improvement and optimization are vital for maintaining robust security. Article [146] proposes a distributed DSA method based on Horizontal Federated Learning (HFL) and differential privacy, namely EFedDS, to dynamically assess and optimize grid stability. This approach ensures continuous adaptation to changes while preserving data privacy. Article [177] details RL-BAGS tool: Reinforcement Learning-Bayesian Attack Graph for Smart Grid System, which uses RL algorithms to refine and update security strategies based on ongoing risk assessments. This tool enhances the resilience of smart grids by ensuring that security measures are progressively aligned with the latest cybersecurity developments.

Further advancements are seen in automated tools and dynamic metrics for proactive security measurement adjustment. Article [139] discusses the development of automated network security metrics that allow network administrators to proactively adjust and strengthen defenses based on evolving threats. These metrics ensure continuous improvement by

dynamically refining security strategies without the need for manual intervention.

6 The application of AI techniques in SSA

This section investigates the application of various AI methodologies in SSA, highlighting their significance and effectiveness in supporting and improving SSA processes, thereby answering RQ2. Tables 7 summarize the AI techniques applied in the selected articles, detailing their specific implementations and contributions to SSA. Due to the wide range of articles, only the most representative ones have been selected for discussion. The following sections delve into specific AI techniques, including supervised learning, unsupervised learning, probabilistic methods, reinforcement learning, DL, and NLP, illustrating how each contributes to a robust and dynamic security framework.

6.1 Unsupervised learning

Unsupervised learning techniques allows for the discovery of hidden patterns and structures in large datasets without the need for labeled data. These techniques are widely used in fields such as market segmentation, image recognition, and anomaly detection in various industries [37]. In the context of cybersecurity, unsupervised learning plays a crucial role in identifying patterns and detecting anomalies within vast amounts of security data [172]. By leveraging these techniques, security systems can improve vulnerability detection across multiple domains by automating the analysis of complex data. For instance, in IoT environments, techniques like entropy analysis of network traffic effectively uncover vulnerabilities, ensuring robust security measures without needing direct device access (Article [21]). Clustering techniques, such as those used to classify botnet vulnerabilities based on their characteristics and behaviors, enable precise identification and targeted countermeasures (Article [31]). Anomaly detection focused on system call patterns in Android applications identifies deviations indicative of malware, enhancing the detection of sophisticated threats that evade traditional methods (Article [118]). In the IIoT, unsupervised learning models leverage datasets like WUSTL-IIoT-2018, ICS-SCADA, and CICDDoS2019 to detect vulnerabilities and mitigate Distributed Denial of Service (DDoS) attacks with high accuracy (Article [31]).

6.2 Supervised learning

Supervised learning techniques enables models to learn from labeled datasets and make accurate predictions. These techniques are widely used in applications such as image classification, speech recognition, and medical diagnosis [129], where the goal is to map input data to known output labels. In the domain of cybersecurity, supervised learning leverages labeled data to train models that can accurately predict and classify security threats, vulnerabilities, and risks.

At the foundation of these techniques are Naïve Bayes, Decision Trees, and Random Forests, which are extensively utilized in SSA (Articles [14, 85, 154]). Naïve Bayes classifiers are applied to classify software vulnerabilities and estimate security vectors by analyzing text data from security reports and logs. Decision Trees, used for classifying security risks and threats, facilitate hierarchical decision-making based on feature values. Complementing these, Random Forests aggregate the predictions of multiple decision trees to enhance the accuracy of threat detection and risk assessment, offering a robust solution for complex security scenarios.

Building on these foundational algorithms, SVMs (Article [117]) and MLPs (Article [42]) provide advanced capabilities for malware detection and the identification of zero-day vulnerabilities. SVMs classify vulnerability by finding the optimal hyperplane that separates benign and malicious behaviors, ensuring high precision in detection systems. MLPs, as neural networks, detect complex patterns in security data, offering the capability to identify sophisticated threats that might evade simpler models.

Furthermore, XGBoost plays a crucial role in analyzing risk factors related to virtual machine configurations and user behavior to predict potential cyber threats in cloud environments (Articles [23, 152]). This method significantly improves the accuracy of risk predictions by efficiently handling large datasets and complex feature interactions. In addition to these classification techniques, Text Analysis Methods such as Term Frequency-Inverse Document Frequency (TF-IDF) (Article [85]) are employed to convert security-related text data into numerical features. These methods are essential for processing and analyzing large volumes of security documents, enabling more accurate threat identification and risk assessments.

Complementing text analysis, Ontology-Based Frameworks leverage structured knowledge representations to automate the detection of security issues from use case models, thereby enhancing the precision and reliability of security analysis (Article [182]). To optimize threat modeling and risk assessment, ensemble methods combine multiple ML models, improving overall performance by integrating various algorithms (Articles [12, 187]). These methods provide comprehensive security solutions by enhancing the accuracy and robustness of the assessments.

Table 7 Summary of AI techniques in SSA

AI technique	AI model and algorithm	Description	Selected article
Unsupervised learning	Entropy analysis	Analyzes network traffic entropy in IoT environments to detect vulnerabilities, identifying deviations in expected network behavior without needing direct device access	[21]
	Clustering techniques	Classifies botnet vulnerabilities and privacy risks in IoT by grouping similar characteristics and behaviors, aiding in understanding and addressing specific types of threats	[31, 80]
	Anomaly detection	Monitors system call patterns in Android applications and IIoT systems to identify deviations indicative of malware, enhancing the detection of sophisticated threats	[118, 133]
Supervised learning	Naïve bayes	Used to classify software vulnerabilities, it helps in estimating security vectors and severity scores by analyzing text data such as security reports and logs	[85]
	Decision trees	Applied for the classification of security risks and threats, decision trees help in making decisions based on the hierarchical structure of feature values	[154]
	Random forests	Employed to enhance the accuracy of threat detection and risk assessment, random forests aggregate the predictions of multiple decision trees	[14]
	Support vector machine	Utilized in vulnerability detection systems, SVMs classify vulnerabilities by finding the optimal hyperplane that separates benign and malicious software behaviors	[42]
	Multi-layer perceptron	Used for detecting complex patterns in security data, MLPs are neural networks that can identify zero-day vulnerabilities and other sophisticated threats	[25]
	Extreme gradient boosting	Applied to predict potential cyber threats in cloud environments, XGB analyzes risk factors related to VM configurations and user behavior	[23, 152]
	Text analysis methods	Used to process and analyze large volumes of security-related text data, methods like TF-IDF convert text into numerical features that highlight important terms	[85]
	Ontology-based frameworks	Leveraged to automate the detection of security issues from use case models, ontology-based frameworks use structured knowledge	[182]
	Ensemble Methods	Utilized to optimize threat modeling and risk assessment, ensemble methods combine multiple ML models to improve overall performance	[12, 187]

Table 7 (continued)

AI technique	AI model and algorithm	Description	Selected article
Probabilistic methods	Bayesian network	Used for dynamic risk assessment by modeling interdependencies and updating risk assessments with real-time and historical data	[4, 11, 65, 143, 176]
	Graph-based and ML techniques	Applied to assess security risks by classifying network vulnerabilities based on complex interdependencies and misconfigurations	[131]
	Enhanced posterior probability	Refines threat detection and classification by integrating a risk assessment function to address feature interdependencies	[154]
	Weighted hidden markov model	Dynamically models security states and adapts to changes triggered by known and unknown cyberattacks using the Baum-Welch algorithm	[74]
	Massive personal information clustering	Addresses privacy risks in the IoT domain by predicting and mitigating the aggregation of PII from IoT devices	[80]
Fuzzy logic	Fuzzy set theory integration	Integrates fuzzy set theory with ML classifiers to predict risks, assess control effectiveness, and model uncertainty in assessments	[60, 94, 95]
	Neuro-fuzzy systems	Employs ANFIS to evaluate quality parameters, providing accurate and objective assessments	[169]
	Fuzzy rule-based systems	Uses fuzzy rule sets, association rule mining algorithms and fuzzy-based classification to handle uncertainties and improve the quantification of dynamic environments	[7, 22, 195]
	Fuzzy logic for decision support	Develops a fuzzy logic-based inference system to analyze data and infer risk levels, supporting preemptive risk mitigation	[3, 192]
Optimization techniques	GAs and variants	Utilize GAs to optimize security configurations and risk parameters dynamically, enhancing efficiency and accuracy in vulnerability predictions	[62, 150, 171]
	Hybrid optimization methods	Combine methods like hybrid-extreme learning machine with PSO and CSA to enhance real-time security assessments and detection precision	[163, 191]
	Ant colony optimization	Enhances ethical hacking effectiveness in HIS by systematically analyzing vulnerabilities and potential breach points	[57]
	Evolutionary algorithms	Employs an interpretable evolutionary fuzzy scoring system to optimize scoring accuracy and reduce rule complexity in cyber security risk assessments	[63]

Table 7 (continued)

AI technique	AI model and algorithm	Description	Selected article
	Adaptive heuristics and evolved metrics	Provides proactive and dynamic adjustments to network defenses, automatically refining security measures as conditions change or new threats emerge	[139]
Reinforcement Learning	Hierarchical and deep reinforcement learning	Utilizes hierarchical DRL and DQN to decompose and automate penetration testing tasks, optimizing decision-making in large-scale network environments	[30, 47, 100, 101, 140, 190]
	Social engineering and network graphs	Incorporates social engineering factors into penetration testing frameworks using RL within modified network graph models, enhancing effectiveness in cybersecurity testing	[103]
	Reinforcement learning in specialized environments	Applies Q-Learning and SARSA to automate penetration testing in CAVs and smart grids, assessing vulnerabilities and optimizing security measures in these specialized contexts	[43, 67, 177]
	Deep reinforcement learning for fuzzing	Leverages DRL models like DDPG to optimize software fuzzing processes under specific format constraints, improving vulnerability discovery	[50]
	Hybrid frameworks combining rl and ontology-based agents	Combines ontology-based cognitive BDI-agent with RL to optimize attack planning and execution in dynamic and uncertain environments, enhancing penetration testing efficiency	[142]
Deep Learning	Convolutional neural networks	Utilize CNNs for cybersecurity vulnerability assessment, automated security audits, and penetration testing by generating test paths and code from real attack scenarios	[33, 78, 105]
	Recurrent neural networks and LSTMs	Apply RNNs and LSTMs for fuzz testing web browsers, detecting software vulnerabilities through sequence analysis, and monitoring vulnerabilities in wireless networks	[107, 148]
	Hybrid CNN-LSTM models	Combine CNNs and LSTMs to detect software vulnerabilities by analyzing binary programs	[184]
	Deep reinforcement learning	Enhance fuzzing processes and penetration testing strategies using DRL models like DQN Networks and WGANs	[96, 104]
	Neural networks for risk assessment	Apply feedforward neural networks and MLPs for risk assessment and management in various domains, processing extensive data to identify and quantify potential security risks	[70, 157]

Table 7 (continued)

AI technique	AI model and algorithm	Description	Selected article
Natural Language Processing	Character-level CNN	Predicts the exploitability of software vulnerabilities by extracting fine-grained character-level features from unstructured descriptions	[112]
	BERT and XGBoost	Analyzes cyber threats and vulnerabilities in the healthcare sector by extracting information from natural language documents for real-time threat identification and risk management	[68, 161, 162]
	Pre-trained language models	Utilizes prompt learning to predict vulnerability severity and exploitability characteristics from vulnerability descriptions, improving performance with fewer training data	[102]
	Misuse case programming	Translates misuse case specifications into executable test cases to identify potential vulnerabilities and ensure completeness and consistency in security requirements	[114–116]
	Text analysis for security requirements	Automates the classification of security requirements in Software Requirement Specification documents, aiding developers in early implementation of security measures	[79]
	Automated organizational evidence assessment	Uses NLP to automate the assessment of organizational evidence within cloud services, enhancing transparency and efficiency in security and privacy measures	[8, 36]

6.3 Probabilistic method

Probabilistic methods provide a framework for modeling uncertainty and making predictions using probability distributions. These techniques are used in a variety of applications, including speech recognition, NLP, and financial forecasting, to help deal with data's inherent uncertainty and variability [173].

Bayesian networks, a type of probabilistic method, utilize graphical models to represent the probabilistic relationships among variables, making them essential for complex decision-making and uncertainty management [59]. For instance, the Bayesian Attack Graph for Smart Grid (BAGS) tool evaluates the likelihood of cyber component compromise within smart grid systems, facilitating effective risk analysis and resource allocation to bolster situational awareness and resilience against cyber-physical attacks (Article [176]). Similarly, in SCADA systems, Bayesian networks integrate real-time data from intrusion detection systems with historical data to dynamically update and detect potential threats (Article [65]), employing Leaky Noisy-OR gates to manage uncertainties and estimate risk values despite incomplete data.

In healthcare organizations, Bayesian-based ML models classify cyber risks by analyzing features such as maturity and complexity indices, thus offering a robust framework for real-world cyber risk assessment (Article [143]). In distributed computing networks, the combination of Bayesian networks with ANN automates the information security risk assessment process. These models analyze data from various sensors to dynamically detect and assess threats, modeling interdependencies among security variables and predicting potential breaches by recognizing patterns in the data (Article [4]).

Enhanced probabilistic techniques, such as the ENBPP model, advance threat detection and classification by incorporating a risk assessment function. This model addresses feature interdependencies often overlooked by standard Naïve Bayes classifiers, thereby reducing false positives and expediting threat processing, which is critical for timely and accurate security evaluations (Article [154]). In ICS security, the W-HMM dynamically models security states and adapts

to changes triggered by both known and unknown cyberattacks. By applying the Baum-Welch algorithm, this model improves the accuracy and timeliness of security assessments, reflecting real-time security states and managing the impact of unforeseen threats (Article [74]).

Graph-based techniques and clustering also play a significant role in security assessments. In active directory environments, a methodological framework integrating graph-based and ML techniques effectively assesses security risks by classifying network vulnerabilities based on complex interdependencies and misconfigurations (Article [131]). In the IoT domain, the MPIC model addresses privacy risks by predicting and mitigating the aggregation of PII across IoT devices (Article [80]).

Dynamic assessment frameworks employing probabilistic methods further enhance security evaluations. The dynamic assessment framework for ICS security uses a W-HMM to provide both qualitative and quantitative evaluations, adapting to security state changes and managing unknown threats more effectively (Article [74]). Bayesian-based ML models are also employed to predict the severity of cybersecurity risks, integrating the expertise of risk assessors to offer scalable and accurate solutions for quantitative risk assessments (Article [11]).

6.4 Fuzzy logic

Fuzzy logic techniques manage uncertainty and providing flexible risk assessments. These methods enhance traditional risk management practices and offer robust decision-support mechanisms across various applications [52]. Fuzzy Set Theory Integration combines fuzzy set theory with ML classifiers and comprehensive assessment models to predict risks and assess control effectiveness. This approach leverages fuzzy set theory and cyber threat intelligence to identify critical assets and evaluate risks in real-world scenarios (Article [95]). Similarly, fuzzy set theory enhances risk prediction and control assessment in cyber-physical systems by modeling uncertainty through IT2FLS (Articles [60, 94]).

Building on the foundation provided by fuzzy set theory, neuro-fuzzy systems further refine the evaluation process by employing Adaptive Neuro-Fuzzy Inference Systems (ANFIS) (Article [169]). These systems evaluate quality parameters in cloud environments, providing accurate and objective assessments that improve resource allocation and address significant security vulnerabilities.

Complementing these neuro-fuzzy approaches, Fuzzy Rule-Based Systems use fuzzy rule sets, association rule mining, and fuzzy clustering to handle uncertainties and improve risk quantification in dynamic environments (Articles [7, 22, 195]). Techniques like fuzzy C-means clustering enhance the precision of security assessments by analyzing statistical data and identifying complex patterns in network behavior.

Fuzzy logic for decision support creates inference systems that evaluate data and infer risk levels to enable proactive risk mitigation as a way to complement these techniques. To improve protection and security measures, techniques such as the Fuzzy Analytic Network Process (FANP) and TOPSIS are employed to rank security techniques according to their efficacy (Articles [3, 192]).

6.5 Optimization techniques

Optimization techniques leverage advanced algorithms to optimize configurations, predictions, and assessments, ensuring robust security across different domains [168]. Several studies highlight the application of GAs and their variants for optimizing security parameters and configurations. In critical infrastructure, a decision support system utilizes GAs to quantitatively assess and optimize security configurations by exploring various solutions quickly and efficiently. This approach ensures comprehensive and dynamically adaptable assessments (Article [171]). Similarly, in information security risk management, GAs optimize risk parameters dynamically, making the process more efficient compared to traditional methods (Article [62]). The integration of DL with symbiotic GAs further enhances predictive modeling for detecting software vulnerabilities, significantly improving the accuracy of vulnerability predictions (Article [150]).

Hybrid optimization techniques are integral to the real-time evaluation of security measures and the identification of vulnerabilities. For power systems, a hybrid-extreme learning machine combined with PSO and the LM algorithm enhances transient stability assessment under fault conditions (Article [163]), demonstrating superior performance over traditional methods. In the field of firmware vulnerability detection for IoT devices, the CSA from artificial immune systems is employed (Article [163]). With the use of principal component analysis and the Relief algorithm to fine-tune feature selection and weight calculation, this approach increases detection precision without requiring large training datasets.

ACO and evolutionary algorithms further illustrate the application of optimization techniques in SSA. For HIS, ACO enhances ethical hacking effectiveness by systematically analyzing vulnerabilities and potential breach points, outperforming traditional methods in experimental simulations (Article [57]). An interpretable evolutionary fuzzy scoring system employs evolutionary algorithms to optimize scoring accuracy and reduce rule complexity in cyber security risk assessments, enhancing both interpretability and practical applicability (Article [63]). Additionally, adaptive heuristics and evolved security metrics provide proactive and dynamic adjustments to network defenses (Article [139]). These adaptive systems refine security measures automatically as network conditions change or new threats emerge,

ensuring continuous improvement and preemptive adjustments to mitigate potential vulnerabilities.

6.6 Reinforcement learning

RL focuses on training agents to make a sequence of decisions by rewarding desired behaviors and penalizing undesired ones, thereby optimizing their performance over time through trial and error [27]. Several studies focus on the application of RL in penetration testing by modeling penetration testing tasks as Markov Decision Processes (MDPs) or Partially Observable Markov Decision Processes (POMDPs). For instance, hierarchical DRL models decompose penetration testing tasks into manageable subtasks, integrating expert prior knowledge to improve learning efficiency and reduce unnecessary explorations (Article [101]). This method is practical for large-scale network scenarios, making penetration testing automation more applicable. Similarly, RL algorithms, such as Q-Learning and DQN, construct attack graphs to map security threats and feasible attack paths, optimizing decision-making in identifying and exploiting network vulnerabilities (Articles [30, 47, 100, 140, 190]).

Incorporating social engineering elements into penetration testing frameworks significantly enhances their realism and overall effectiveness. By leveraging RL algorithms within modified network graph models, these approaches train and evaluate penetration testing strategies that include tactics like phishing and pretexting, providing a dynamic and realistic setting for cybersecurity testing (Article [103]).

RL also proves valuable in specialized environments such as Connected and Autonomous Vehicles (CAVs) and smart grids. For CAVs, Q-Learning automates penetration tests in Vehicular Ad Hoc Networks (VANETs), addressing the limitations of traditional testing methods by simulating cyber-attacks to assess vulnerabilities (Article [43]). In smart grids, RL techniques such as SARSA and Q-Learning applied to Bayesian Network models help system engineers compute optimal policies for scanning or patching components, highlighting critical vulnerabilities and suggesting effective security measures (Articles [67, 177]).

Advanced methodologies like DRLFCfuzzer utilize DRL to enhance software fuzzing techniques under specific format constraints. This methodology employs the Deep Deterministic Policy Gradient (DDPG) model to enhance the fuzzing process, thereby optimizing the identification of software vulnerabilities while adhering efficiently to format constraints (Article [50]). Hybrid frameworks that combine ontology-based cognitive Belief-Desire-Intention (BDI)-agent with RL techniques optimize attack planning and execution, handling dynamic and uncertain environments efficiently (Article [142]). These frameworks utilize a knowledge base derived from expert penetration testers to systematically learn and adapt, improving the overall efficacy of penetration tests.

6.7 Deep learning

DL, a subset of ML, involves neural networks with many layers that can learn to represent data with multiple levels of abstraction, making it highly effective for complex tasks such as image and speech recognition, NLP, and autonomous driving [86]. DL techniques leverage various neural network architectures to process large datasets, uncover complex patterns, and provide actionable insights, significantly improving the effectiveness and efficiency of security assessment.

Several studies focus on using CNNs and their variants for vulnerability assessment and risk management. For instance, CNNs are employed to evaluate the robustness of security measures by breaking CAPTCHAs (Article [33]) and in automating security audits of smart grid databases to identify potential security risks rapidly and accurately (Article 133). Additionally, CNN-based models are applied to automate penetration testing by generating test paths and code from real attack scenarios, improving the efficiency and effectiveness of security testing processes (Article [78]).

Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks are particularly effective in handling sequential data and temporal dependencies, which are crucial for detecting complex vulnerabilities. These models are used in fuzz testing web browsers (Article [148]), detecting vulnerabilities in software through sequence analysis of function calls (Article [184]), and monitoring vulnerabilities in wireless communication networks (Article [107]). Hybrid models combining CNNs and LSTMs further improve vulnerability detection. For example, CNN-LSTM models are used to detect software vulnerabilities by analyzing binary programs, significantly outperforming traditional methods (Article [184]). Such combined approaches leverage the strengths of both architectures to provide more comprehensive security assessments.

Deep reinforcement learning (DRL) techniques optimizes the fuzzing processes and penetration testing strategies. Models like DQN enhance software vulnerability testing by predicting promising input mutations, reducing computational loads and improving testing effectiveness (Article [96]). Furthermore, WGANs are used to generate fuzzing data for ICSs, significantly improving test efficiency and vulnerability detection (Article [104]).

Neural networks are also utilized for risk assessment and management in various domains. FNNs are applied to assess and mitigate cyber risks in massively multiplayer online gaming firms (Article [157]), while MLPs are integrated into software vulnerability discovery models to enhance predictive accuracy (Article [70]). These models process extensive

data to identify and quantify potential security risks, providing precise and timely risk assessments.

6.8 Natural language processing

NLP enables computers to understand, interpret, and generate human language, facilitating applications such as machine translation, sentiment analysis, and chatbots. NLP techniques leverage sophisticated AI models to extract, classify, and analyze unstructured text data (or in other languages) in order to derive meaningful insights [15].

Several studies have utilized NLP for automating the identification and assessment of vulnerabilities from textual descriptions. For example, charCNN and PLMs are used to predict the exploitability of software vulnerabilities from unstructured vulnerability descriptions, highlighting features such as confidentiality impact and attack vector (Article [102, 112]). This approach helps in prioritizing vulnerabilities for timely remediation, providing a more nuanced understanding of potential threats.

In the healthcare sector, NLP techniques like BERT and XGBoost are employed to analyze cyber threats and vulnerabilities by extracting information from natural language documents, such as news reports and online databases (Articles [68, 161, 162]). These methods enhance real-time threat identification and risk management, ensuring up-to-date threat intelligence in dynamic environments. Similarly, NLP-driven methodologies are used to automate the assessment of organizational evidence within cloud services, improving the transparency of security and privacy measures through continuous audit-based certification (Article [36]).

NLP also plays a critical role in generating and analyzing security requirements from natural language documents. Techniques such as MCP and RMCM translate misuse case specifications and security requirements into executable test cases and security models (Articles [114–116]). This process aids in identifying potential vulnerabilities and ensuring the completeness and consistency of security requirements. Additionally, NLP-based frameworks automate the classification of security requirements found in software requirement specification documents (Article [79]), aiding developers in recognizing and implementing necessary security measures early in the software development lifecycle.

Furthermore, NLP techniques are used to improve the accuracy and relevance of security threat assessments. For instance, tools like AMOE and systems for vulnerability risk prioritization utilize NLP to extract and analyze security-related information, enhancing the contextual understanding of vulnerabilities and the efficiency of security assessments (Articles [8, 36]). These approaches leverage NLP to convert free-form text into structured formats, facilitating easier and more consistent security evaluations.

7 Application domains analysis

In this section, we focus on the application of AI techniques in enhancing SSA across several key system types, addressing RQ3. While the literature review covers a broad spectrum of systems, this section highlights only the most critical ones: web applications, mobile applications, IoT systems, cyber-physical systems, power and smart grid, ICSs, 5G network systems, connected and autonomous vehicles, and healthcare system. Table 8 outlines the selected articles for each application domain in our literature review. The following sections examine how AI-driven solutions are being implemented to address security challenges these systems face.

7.1 Web applications

AI-driven approaches in web application SSA encompass a variety of techniques aimed at automating and enhancing penetration testing, vulnerability detection, and risk analysis. The integration of AI not only streamlines the detection processes but also improves the accuracy and efficiency of security assessments, making it a crucial tool for developers and security analysts in protecting web applications. For instance, ML frameworks like AppMine utilize unsupervised learning models to detect anomalies in application behaviors, particularly in containerized environments such as Docker. These models are adept at recognizing complex patterns and temporal dependencies that indicate potential security vulnerabilities, significantly improving the scope and accuracy of detections (Article [72]).

Techniques such as RL and ML have been applied to automate and optimize Web application security tests. For example, systems like GyoiThon leverage algorithms like Naïve Bayes to enhance the detection of vulnerabilities in web applications with content management systems, outperforming traditional methods in identifying a broader range of security issues (Articles [71] and [45]). In additional, DL models have been developed to automate the generation of web penetration testing code, addressing the challenge of manual test path generation in dynamic network environments and significantly improving the speed and effectiveness of security testing (Article [78]). Furthermore, models utilizing PAA predict and evaluate security risks, allowing for a dynamic and adaptive security posture that can quickly respond to new and evolving threats in heterogeneous web environments (Article [126]).

7.2 Mobile applications

AI's impact is particularly significant in the realm of vulnerability detection and risk assessment for mobile applications, providing strong protection for both users and systems. Leveraging ML algorithms, several studies have developed

Table 8 Summary of the selected article based on application domains

Application domain	selected article
Web application	[41, 45, 71, 72, 78, 126]
Mobile applications	[8, 39, 42, 54, 89, 118, 122, 137, 174, 177, 192]
IoT system	[21, 38, 80, 111, 133, 148, 160, 167, 191, 199]
Cyber-physical system	[76, 77, 87, 94, 95]
Power and smart grid system	[24, 67, 98, 105, 106, 108, 113, 146, 163, 166, 169, 176]
5G network System	[128, 138, 149]
Connected and autonomous vehicles	[43, 48, 55, 141]
Industrial control system	[10, 65, 74, 104, 119, 127]
Healthcare systems	[3, 57, 68, 143, 144, 162]

methods that improve the detection of security vulnerabilities by analyzing patterns in system calls, app permissions, and user behaviors (Articles [42, 118]). Risk assessment frameworks also benefit from AI, where NLP and ML techniques assess the privacy and security risks of mobile apps installed on devices within enterprise environments (Articles [54, 192]). These frameworks evaluate app behaviors and public perceptions, analyzing vulnerabilities and facilitating enterprises in mitigating potential security threats effectively. Additionally, novel ML approaches like tree boosting have been adapted to evaluate security threats in mobile apps more objectively, moving away from subjective user reviews to API-based evaluations (Article [89]). Furthermore, the implementation of AI in mobile application security extends to dynamic risk modeling. Innovations such as using ML to construct detailed risk indexes based on app permissions enable a nuanced assessment of how permissions are requested and exploited by apps, enhancing the identification of overprivileged apps that may pose security risks (Article [122]).

7.3 IoT systems

The inherent complexities and the vast scale of interconnected devices in IoT systems present unique security challenges that require advanced solutions. AI methodologies provide the capability to analyze large volumes of data from diverse IoT devices and networks, predict potential vulnerabilities, and automate the detection and mitigation processes. For instance, methods that automate the process of detecting vulnerabilities by analyzing network traffic or firmware characteristics are vital for maintaining the security integrity of IoT systems (Articles [21, 191]). Furthermore, AI has been instrumental in predicting and assessing security risks within IoT networks. Several studies have focused on utilizing AI to model and predict vulnerabilities based on device behavior and network traffic patterns (Articles [80, 133, 167, 199]). These predictive models help in pre-empting potential attacks and facilitating timely responses to emerging threats, thereby enhancing the overall resilience of IoT systems.

7.4 Cyber-physical systems

Cyber-Physical Systems (CPS) involve the seamless interaction of physical and software components, with embedded computers and networks monitoring and controlling physical processes, often with feedback loops where physical processes affect computations and vice versa. AI-driven methodologies are adept at uncovering vulnerabilities, predicting potential security threats, and providing decision support for optimal security configurations. Automated methods for assessing vulnerability severity, as seen in papers (Articles [76, 77]), utilize AI to parse and harmonize data from multiple security databases, leading to more accurate and consistent vulnerability scores. Risk prediction and management are significantly enhanced by AI, employing advanced algorithms that identify critical assets and analyze and predict potential security breaches based on historical and real-time data (Articles [94, 95]).

7.5 Power and samrt grid systems

Various studies have developed innovative AI methodologies to address both the cyber and physical security challenges faced by modern power and smart grids. AI is not only transforming how security risks are assessed and managed but also how these systems adapt to new challenges, such as those posed by the integration of large volumes of renewable energy sources.

Probabilistic models, as seen in Article [176], use Bayesian Attack Graphs to enhance resilience against cyber-physical attacks on power grids, aiding in risk analysis and optimal resource allocation. DL techniques, illustrated in Articles [105, 106, 166], and [24], employ neural networks such as SPORF and CNN to improve accuracy and efficiency in real-time security assessments and feature extraction for power system state evaluations.

Optimization techniques, highlighted in Article [163], integrate methods with extreme learning machines to enhance dynamic security assessments of power systems under fault conditions. Supervised learning methods utilize semi-supervised learning, transfer learning, and dense neural networks to handle fast-changing operating conditions and improve the stability and security of power grids (Articles [108] and [98]). Additionally, fuzzy logic (Article [169]) and RL (Article [67]) techniques are employed to evaluate quality of service parameters and simulate attack scenarios in power grids, respectively, providing accurate assessments and robust security measures. Other innovative approaches, such as HFL (Article [146]) and big data analytics (Article [113]), demonstrate the diverse application of AI in enhancing power and smart grid security.

7.6 Industrial control systems

ICS encompass a variety of systems and technologies used to monitor, control, and automate industrial processes. These systems are integral to critical infrastructure sectors such as energy, water treatment, manufacturing, and transportation. Integrating AI techniques into ICS significantly enhances SSA, addressing vulnerabilities, optimizing risk assessments, and improving predictive threat analysis. For instance, the use of Bayesian networks enables dynamic detection of cybersecurity risks in SCADA systems (Article [65]), continuously updating threat detection by integrating real-time and historical data.

DL techniques, such as those employing Generative Adversarial Networks (GANs), automate the generation of test data for vulnerability detection, eliminating the need for detailed protocol specifications and improving accuracy (Article [104]). Another method uses a GAN-based framework to streamline fuzz testing for industrial control protocols, enhancing the efficiency of vulnerability detection processes (Article [127]). Additionally, optimization techniques dynamically adjust to security state changes caused by cyberattacks, providing real-time qualitative and quantitative evaluations to maintain ICS operational integrity (Article [74]). Additionally, supervised learning techniques are essential for anticipating cyberattacks by analyzing the actions and intentions of attackers (Article [10]), strengthening preventive security protocols, and guaranteeing that ICS can anticipate and neutralize threats.

7.7 5G network systems

5G Network Systems represent the latest evolution in mobile network technology, designed to deliver faster speeds, lower latency, and more reliable connections than previous generations. The increased complexity and interconnectivity of 5G networks also introduce significant cybersecurity challenges [2]. The broader attack surface, combined with the critical nature of the services supported by 5G, necessitates advanced security measures to protect against potential threats and vulnerabilities. One innovative application of AI in 5G security is found in the DEFT framework, designed specifically for fuzz testing in NextG networks, including 5G systems (Article [138]). Moreover, the integration of ML in the security analysis of 5G core network infrastructures is exemplified by research focusing on software-defined networking and network function virtualization components (Article [149]). In addition, AI is used to assess and classify vulnerabilities in MSNs within the 5G context (Article [128]). The research utilizes ML classifiers based on user interactions and movement patterns.

7.8 Connected and autonomous vehicles

Connected and Autonomous Vehicles (CAVs) represent a revolutionary advancement in transportation technology, integrating connectivity features with autonomous driving capabilities. These vehicles are equipped with a myriad of sensors, communication systems, and sophisticated algorithms that enable them to communicate with each other and with infrastructure [2]. An integrated approach combining traditional threat modeling with systems theory-based analysis (STPA-Sec) addresses cyber-physical threats specific to autonomous vehicle operations (Article [48]). In autonomous driving technologies, supply-chain security is crucial due to the complex ecosystem of hardware and software components. AI techniques provide continuous security assessment, adapting to new threats and ensuring compliance with standards (Article [55]). For the Internet of Vehicles (IoV), the dynamic nature of data necessitates novel security assessment methods like SAMCT, which uses Microcontroller Unit (MCU) chip temperature to predict device security states.

AI-driven RL automates penetration testing in Vehicular Ad Hoc Networks (VANETs), improving efficiency and comprehensiveness by simulating cyber-attacks and assessing vulnerabilities (Article [43]). Moreover, AI techniques, such as sequence models, identify vulnerabilities in wireless communication networks, predicting patterns that indicate security risks and enhancing proactive security measures (Article [141]).

7.9 Healthcare systems

In the healthcare sector, AI-driven techniques are crucial for enhancing SSA. As healthcare systems integrate more digital technologies and connect to the IoMT, they face significant cybersecurity challenges. The vast amount of sensitive health data transmitted and stored by these devices makes them prime targets for cyber-attacks. AI tools enable real-time

threat detection, risk assessment, and proactive security management, ensuring healthcare ecosystems remain resilient against cyber vulnerabilities.

Within HIS, AI-driven ethical hacking approaches such as ACO are employed to systematically analyze vulnerabilities and potential breach points. This method surpasses traditional techniques, offering improved security measures to protect sensitive health data (Article [57]). For the IoMT, QML addresses the critical challenge of securing devices with limited computing resources. By applying advanced semi-supervised learning models, healthcare systems can better assess and mitigate vulnerabilities in devices that handle sensitive patient information (Article [144]).

NLP techniques are extensively used to analyze and manage cyber threats in healthcare. By extracting data from reports, news, and other unstructured sources, NLP enhances threat intelligence and situational awareness, thereby improving the overall security posture of healthcare information infrastructures (Articles [68, 161, 162]). ML models, combined with probabilistic methods, classify healthcare organizations into various risk categories based on parameters such as maturity and complexity. This classification aids in assessing cyber risks and implementing appropriate security measures, thereby enhancing organizational resilience (Article [143]). Additionally, a mathematical model using the Fuzzy ANP integrated with the Technique for TOPSIS prioritizes security techniques for healthcare devices. This model helps developers and manufacturers in implementing the most effective security measures to protect against unauthorized access and potential breaches (Article [3]).

8 Research gaps and future research directions

This section seeks to address RQ4 by identifying the current limitations from our literature review and proposing strategic directions for future research to bridge these gaps.

8.1 Contextual understanding and customization of security requirements

One significant research gap in AI-supported security requirement elicitation and criteria development is the lack of contextual understanding and customization capabilities.

Research such as in Articles [116] and [61] highlights the difficulties of translating natural language security requirements into actionable security measures, indicating a gap in AI's ability to grasp context and ambiguity inherent in natural language. These limitations, including potential misinterpretations and oversights of subtle security demands, underscore the necessity for models that can better comprehend the intricacies and nuances of contextual information. Innovations like the MCP methodology (Article [116]) use NLP to bridge this gap by generating security test cases from misuse case specifications. However, the variability in data quality and evolving security threats call for future advancements in NLP and ML technologies. These improvements should focus on enhancing the ability of AI systems to detect and interpret the contextual subtleties of security requirements, leading to more customized and precise security assessments.

Moreover, the integration of structured approaches to eliciting security requirements, as discussed in Article [61], shows a promising avenue for customization. Future research should explore the development of adaptive AI frameworks that dynamically update and refine their understanding based on new data and emerging security scenarios. Such frameworks should be capable of seamlessly integrating diverse data sources and maintaining a high level of customization to accurately reflect the specific security needs of different environments.

8.2 Scalability and adaptability of vulnerability detection

There are large research gaps in the area of vulnerability identification when it comes to scalability and adaptation. When used in large-scale, diverse systems, current AI models frequently struggle to generalize across different network topologies and encrypted traffic (Articles [21] and [42]), emphasizing the need for models that can dynamically update and handle large datasets efficiently. Addressing these limitations could involve the integration of adaptive learning techniques like RL, which allows AI models to learn and adapt in response to new threats continuously, reducing the dependency on large labeled datasets.

Moreover, the effectiveness of current models against sophisticated and less conventional attack vectors, as noted in Articles [33] and [72], highlights a gap in handling complex scenarios. Federated learning could offer a robust solution here, allowing for the decentralized training of models across various nodes to learn from diverse and realistic data sources. This approach not only enhances the robustness of AI systems but also preserves data privacy, providing a strategic advantage in detecting anomalies across heterogeneous network environments.

Furthermore, the need for real-time processing capabilities and the reduction of false positives are pressing concerns (Article [21, 118]). Advanced anomaly detection techniques, such as those based on DL architectures like GANs or Autoencoders, could significantly enhance the capability to identify subtle anomalies and deviations in data patterns. These techniques offer a promising avenue for future research, potentially improving both the accuracy

and efficiency of vulnerability detection systems under the demanding conditions of operational security environments. This combined approach of leveraging advanced ML strategies and novel architectural models promises to overcome current limitations and pave the way for more resilient AI-driven SSA solutions.

8.3 Automatic assurance evidence synthesis and interpretation

The synthesis and interpretation of assurance evidence through automated systems are essential components of SSA, particularly as organizations strive to meet complex regulatory and security standards efficiently. Despite advancements in automation technologies, significant limitations in current approaches have been highlighted in studies like Article [36], which need to be addressed through future research.

A primary limitation is the ability of existing systems to automatically synthesize and interpret heterogeneous data sources (Article [38]). Current systems often struggle to integrate and make sense of the diverse formats and structures of evidence, ranging from logs and real-time monitoring data to audit reports and previous compliance assessments. This challenge is exacerbated by the varying levels of detail and context provided within these data sources, which can lead to inconsistencies in interpretation and potential oversights in compliance and risk assessment.

Future research should focus on developing more sophisticated data integration and analytics frameworks, capable of processing and synthesizing information from a broad array of sources and formats without losing the context or integrity of the data. Techniques such as advanced data fusion and ML models trained on domain-specific datasets, as suggested by the approaches Article [170], could offer significant improvements in how systems handle the complexity of assurance evidence.

Furthermore, automated systems must not only aggregate data but also interpret it in a manner that aligns with the regulatory and security frameworks applicable to the organization. To enhance this capability, future work could explore the application of semantic analysis technologies and the development of context-aware ML models that adapt their processing and interpretation strategies based on the regulatory environment, similar to innovations described in Article [56].

Additionally, the interaction between automated systems and human auditors is an area ripe for development. Enhancing the interfaces through which these systems present synthesized evidence can help ensure that human auditors can easily understand and verify the conclusions drawn by the system. Research into user-centered design and interactive visual analytics, as discussed in Article [132], could play a crucial role in improving how assurance evidence is presented and utilized.

8.4 Integration of compliance metrics with real-time data streams

A notable research gap in "Compliance Assessment and Analytics" is the integration of compliance metrics with real-time data streams. Typically, such issues are inferred from broader discussions about the limitations of current AI-driven compliance models, addressed in Articles [36] and [56], which discuss automated compliance using static policy documents. The future direction in this area should focus on the development of systems that can seamlessly ingest and analyze real-time data from multiple sources to provide instantaneous compliance assessments. This involves leveraging technologies like stream processing and real-time analytics to monitor compliance continuously, rather than relying on periodic reviews that may not capture the most current data. For example, employing advanced event-processing engines that can detect patterns and anomalies in real-time can help organizations immediately identify potential compliance breaches and react swiftly to mitigate risks. Enhancing the capability to perform real-time compliance checks will ensure that security measures and regulations are being followed precisely at the moment they are needed, thereby enhancing the overall security framework's responsiveness and effectiveness.

8.5 Quantitative measures of compliance impact

In the domain of "Compliance Assessment and Analytics," one research gap lies in the development of quantitative measures that directly link compliance actions with their impact on security effectiveness. A significant limitation, as discussed in Article [176] is the inability of current models to simulate real-world scenarios accurately, which can lead to a misrepresentation of the compliance impact. This gap underscores the need for a more metric-driven approach to evaluate how specific compliance measures influence the overall resilience and security of systems. Future research should focus on creating robust quantitative frameworks that can measure the direct impact of compliance on security posture. This involves developing new metrics that not only assess the implementation and adherence to security policies but also quantify their effectiveness in mitigating risks and preventing breaches. For instance, statistical models could be designed to correlate specific compliance metrics with incident reduction rates, providing a clear, data-driven insight into the effectiveness of compliance measures [153]. Additionally, the introduction of ML techniques to predict the future state of system security based on compliance data could revolutionize how organizations measure the return on investment in compliance activities. By establishing these

quantitative links, organizations can make more informed decisions about where to allocate resources and how to prioritize compliance efforts to maximize security outcomes.

8.6 Adaptation to regulatory changes

Adapting to regulatory changes remains a significant challenge in the field of SSA, as highlighted in the ongoing research and literature. For example, as described in Articles [106] and [171], even advanced AI-driven security systems often require manual updates to align with new regulations, which can be both time-consuming and prone to human error. Additionally, the inherent limitations of predictive models (Article [171]) mean that there can be discrepancies between predicted and actual security needs, which are exacerbated by evolving regulatory landscapes. A crucial research gap lies in the development of security frameworks that can automatically interpret and integrate regulatory changes into their operational protocols.

Future research should focus on developing adaptive frameworks that can seamlessly integrate regulatory updates with minimal manual intervention. This includes the development of NLP techniques that can understand the context and implications of new regulations and ML models that can dynamically adjust security settings without manual intervention. Moreover, incorporating ML models that are capable of adaptive and incremental learning will allow security assurance frameworks to continuously update and evolve in line with regulatory changes. Additionally, exploring the use of advanced analytics and artificial intelligence to predict regulatory trends and proactively adjust security policies can further enhance the agility of compliance frameworks.

8.7 Continuous system security assurance

Continuous system security assurance faces significant limitations, primarily due to the dependency on high-quality, diverse training data and the challenges of scaling to complex, large-scale environments. Articles [61] highlights how variability and data quality issues can compromise ML models, while Articles [65] and [139] emphasize the computational overhead associated with real-time data processing and continuous learning. These limitations are particularly noticeable in dynamic environments like ICS and smart grids, where the security landscape is continuously evolving, necessitating frequent updates and refinements to maintain accuracy and reliability.

Future research should focus on enhancing data quality and diversity through techniques like data augmentation, synthetic data generation, and federated learning, which can create more robust datasets [145]. Additionally, designing scalable algorithms for efficient real-time processing, leveraging edge computing and distributed processing frameworks, is crucial [130]. Finally, exploring adaptive and incremental learning models [188] that continuously update with new data will ensure that security evaluations remain effective against emerging threats, providing a proactive defense mechanism.

8.8 AI Transparency and insight for security assurance

In the domain of security assurance and evidence collection, the transparency and insight of AI models present significant research gaps that hinder broader adoption and effectiveness. Many AI models, especially those utilizing DL, often function as black boxes, offering limited insight into their decision-making processes. This opacity poses a challenge for security professionals who need to understand and trust the AI's outputs to take appropriate actions (Article [70, 184]). Additionally, regulatory requirements and industry standards increasingly demand transparent AI solutions to ensure accountability and compliance. Addressing these gaps necessitates the development of methods that can clarify the internal workings of AI models, making their conclusions more transparent and comprehensible. For instance, the integration of XAI frameworks into security systems can significantly enhance their usability (Article [57, 68]). Future research should focus on creating frameworks for transparent AI that integrate seamlessly with existing security systems. Techniques such as attention mechanisms, model-agnostic explanations, and advanced visualization tools should be employed to demystify AI decisions (Article [72, 191]). Furthermore, interdisciplinary approaches combining AI with human expertise can enhance understandability, ensuring AI-driven insights are actionable and reliable. Developing standardized metrics and benchmarks for evaluating AI transparency will support this effort, fostering trust and facilitating the integration of advanced AI techniques in security assurance and evidence collection practices (Article [51, 113]).

8.9 Ethical considerations in security assurance

While the provided papers focus on technical challenges, limitations, and advancements in SSA, none explicitly address the ethical considerations integral to this field. Ethical issues such as privacy, data protection, algorithmic fairness, and transparency are critical and should be incorporated into the design and implementation of security assurance frameworks (Article [135]). In particular, personal data privacy is a key concern when AI models are used to analyze large-scale security data. AI-driven SSA frameworks often require vast amounts of data for accurate threat detection, which can infringe on individual privacy if not managed carefully.

Future research should prioritize developing robust privacy-preserving techniques, such as homomorphic encryption, differential privacy, and federated learning [29], to ensure that sensitive data is anonymized or protected without compromising the efficacy of security assessments.

Additionally, algorithmic accountability is essential in ensuring that AI models used for security are transparent, fair, and auditable. XAI techniques must be integrated into security assurance frameworks to enable stakeholders to understand how decisions are made, especially in high-stakes contexts such as access control or incident response. This is vital to mitigate risks of bias and to ensure that security decisions do not unfairly target specific user groups or contexts (Article [121]). A deeper investigation into how AI can be made both fair and transparent is crucial, as biases in AI models can lead to unequal treatment across demographics or different organizational structures.

Moreover, incorporating ethical risk assessments into the development and deployment of security technologies is critical to identifying and mitigating potential ethical issues early on. This involves addressing legal and social implications, such as ensuring compliance with data protection laws (e.g., GDPR) and aligning security measures with broader societal values. By focusing on these concrete steps, future research can help ensure that AI-driven SSA systems are both ethically sound and socially acceptable, ultimately enhancing trust in their deployment.

8.9.1 Scalability and practical implementation challenges

One notable gap identified in the literature is the limited exploration of the scalability of AI techniques when applied to large, real-world systems such as cloud environments, 5G networks, and other large-scale infrastructures. While many studies focus on proof-of-concept implementations or domain-specific applications (e.g., IoT, IIoT), few explicitly address how these AI techniques perform when scaled to handle the vast amounts of data and high-speed processing demands characteristic of such environments. Certain AI methods, particularly distributed learning frameworks (e.g., federated learning) and reinforcement learning techniques, have been noted for their potential scalability, as evidenced by studies applying these models in large-scale cybersecurity environments, such as dynamic penetration testing in large-scale networks (Article [101]) and security testing for cloud configurations (Article [23]). However, more comprehensive evaluations of scalability across different industries and infrastructures are lacking. Future research should focus on investigating the real-world scalability of these AI techniques, particularly in dynamic and complex environments such as cloud computing and 5G networks. By addressing this gap, the applicability and robustness of AI-driven SSA solutions can be better understood and enhanced for practical, large-scale implementations.

9 Limitations

In conducting this SLR on the application of AI in SSA, we carefully designed our methodology to focus on the most relevant studies and manage the scope of our research. However, it is essential to recognize that certain inherent limitations arise from our methodological choices. These limitations are discussed below to provide context for the findings and to outline the boundaries within which they should be interpreted.

Framework Specificity: Our SLR is specifically tailored to a distinct SSA framework, which may not align with frameworks utilized by other researchers. While this approach enhances the relevance and depth of our analysis within this framework, it limits the generalizability of our findings across different SSA conceptualizations. As a result, our review may not encompass the full diversity of perspectives and methodologies present in the broader field of system security assurance, potentially omitting valuable insights from alternate frameworks.

Search Strategy: Our decision to use Scopus as the primary database for this review, combined with carefully chosen keywords, was designed to strike a balance between comprehensiveness and specificity. However, this strategy comes with its own set of limitations. While Scopus provides extensive coverage, it does not encapsulate all scholarly databases, possibly omitting relevant studies published elsewhere. Furthermore, the specificity of our keywords, aimed at enhancing search precision, may fail to retrieve pertinent studies employing different terminologies. These methodological decisions, although pragmatic, could restrict the scope of literature captured, subtly influencing the breadth of our review.

Exclusion Criteria: The exclusion criteria for our review were carefully designed to maintain focus and manageability but also limit the scope of our study. By deliberately excluding areas such as source code-based studies, malware detection, and research on AI-based or blockchain systems, we narrow our focus to traditional IT and OT systems and more recent studies from 2016 onwards. This selective approach helps streamline the review process but may lead to less comprehensive coverage of the full spectrum of AI applications in SSA. These exclusions are intentional to align the review with our specific research objectives and ensure clarity and depth in the areas we cover.

These limitations reflect deliberate decisions to tailor the review to specific research needs. However, they emphasize the importance of taking these factors into account when

interpreting the findings and propose avenues for future research to expand on the areas not covered in this study.

10 Conclusion

This SLR delves deeply into the integration of AI techniques within the field of SSA, providing an extensive exploration across a range of IT and OT systems. By following a rigorous SLR methodology, our study reviewed a total of 3,251 articles from the Scopus database, ultimately focusing on 149 high-quality papers that illustrate AI's significant impact on enhancing SSA processes such as requirement elicitation, evidence collection, compliance assessment, and continuous monitoring.

Our analysis aligns with the defined stages of SSA — Plan, Do, Check, Act — by showcasing how AI techniques like unsupervised and supervised learning, probabilistic methods, fuzzy logic, deep learning, and natural language processing not only enhance the detection and response capabilities of SSA but also support the continuous adaptation and improvement of security measures. This is particularly evident in diverse application domains such as web applications, mobile platforms, IoT systems, cyber-physical systems, smart grids, industrial controls, 5G networks, connected vehicles, and healthcare systems.

Despite the advancements, our review identifies critical gaps in the systematic integration of AI within SSA frameworks, especially in enhancing data quality and developing adaptive learning models that can effectively adjust to new threats. These gaps underscore the necessity for future research to focus on the scalability and adaptability of AI-driven SSA solutions, the integration of real-time compliance metrics, and the development of frameworks that can dynamically adjust to regulatory changes. Additionally, ensuring AI transparency and addressing ethical considerations in security assurance are pivotal for gaining stakeholder trust and enhancing the effectiveness of security measures.

In conclusion, while AI significantly advances SSA by automating complex tasks and enhancing threat detection and compliance, there is a continuous need for developing methods that address the evolving nature of cyber threats and the dynamic digital landscape. Future research should prioritize closing the identified gaps, thereby ensuring that AI-driven SSA methods remain robust, scalable, and capable of upholding high security standards. This review not only highlights the current state of AI applications in SSA, but also provides a roadmap for future advancements, with the goal of strengthening security frameworks as integral components of an increasingly interconnected world.

Acknowledgements This work has received funding from the Research Council of Norway through the SFI Norwegian Centre for Cybersecurity in Critical Sectors (NORCICS) project no. 310105

Funding Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital).

Declarations

Conflict of interest The authors declare that they do not have conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Abdullahi, M., et al.: Detecting cybersecurity attacks in internet of things using artificial intelligence methods: a systematic literature review. Electronics **11**(2), 198 (2022)
2. Ahmed, H.U., et al.: Technology developments and impacts of connected and autonomous vehicles: an overview. Smart Cities **5**(1), 382–404 (2022)
3. Ahmed, S., Alhumam, A.: Unified computational modelling for healthcare device security assessment. Comput. Syst. Sci. Eng. **37**(1), 1–18 (2021)
4. Akhmetov, B., et al.: Automation of information security risk assessment. Int. J. Electr. Telecommun. **68**, 549–555 (2022)
5. Al-Turkistani, H.F. and A. AlFaadhel. 2021. "Cyber resiliency in the context of cloud computing through cyber risk assessment". In 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA). IEEE.
6. Al Batayneh, A.A., Qasaimeh, M., Al-Qassas, R.S.: A scoring system for information security governance framework using deep learning algorithms: a case study on the banking sector. ACM J. Data Inform. Quality (JDIQ) **13**(2), 1–34 (2021)
7. Alonge, C.Y., et al. 2020. "Information asset classification and labelling model using fuzzy approach for effective security risk assessment". In 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS). IEEE.
8. Alperin, K., et al. 2019. "Risk prioritization by leveraging latent vulnerability features in a contested environment". In Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security.
9. Alperin, K.B., A.B. Wollaber, and S.R. Gomez. 2020. "Improving interpretability for cyber vulnerability assessment using focus and context visualizations". In 2020 IEEE Symposium on Visualization for Cyber Security (VizSec). IEEE.

10. Alqudhaibi, A., et al.: Predicting cybersecurity threats in critical infrastructure for industry 4.0: a proactive approach based on attacker motivations. Sensors **23**(9), 4539 (2023)
11. Alshammari, F.H.: Design of capability maturity model integration with cybersecurity risk severity complex prediction using bayesian-based machine learning models. SOCA **17**(1), 59–72 (2023)
12. Althar, R.R., et al.: Automated risk management based software security vulnerabilities management. IEEE Access **10**, 90597–90608 (2022)
13. Anderson, R.: Security engineering: a guide to building dependable distributed systems. Wiley, Haboken (2020)
14. Anisetti, M., C.A. Ardagna, and N. Bena. 2023. "Continuous Certification of Non-functional Properties Across System Changes". In International Conference on Service-Oriented Computing. Springer.
15. Bahja, M. 2020. "Natural language processing applications in business". E-Business-higher education and intelligence applications.
16. Basile, C., et al.: Design, implementation, and automation of a risk management approach for man-at-the-end software protection. Comput. Security **132**, 103321 (2023)
17. Bettaieb, S., et al. 2019. "Decision support for security-control identification using machine learning". In International Working Conference on Requirements Engineering: Foundation for Software Quality. Springer.
18. Bettaieb, S., et al.: Using machine learning to assist with the selection of security controls during security assessment. Empirical Soft. Eng. **25**, 2550–2582 (2020)
19. Bo, T., et al. 2019. "Tom: A threat operating model for early warning of cyber security threats". In Advanced Data Mining and Applications: 15th International Conference, ADMA 2019, Dalian, China, November 21–23, 2019, Proceedings 15. Springer.
20. Boyce, J. and D. Jennings. 2002. "Information assurance: Managing organizational IT security risks". volume: Butterworth-Heinemann.
21. Brezolin, U., Vergütz, A., Nogueira, M.: A method for vulnerability detection by IoT network traffic analytics. Ad Hoc Netw. **149**, 103247 (2023)
22. Cai, W., Yao, H.: Research on Information Security Risk Assessment Method Based on Fuzzy Rule Set. Wirel. Commun. Mob. Comput. **2021**, 1–12 (2021)
23. Cayirci, E., et al.: A risk assessment model for selecting cloud service providers. J. Cloud Comput. **5**(1), 14 (2016)
24. Chehri, A., Fofana, I., Yang, X.: Security risk modeling in smart grid critical infrastructures in the era of big data and artificial intelligence. Sustainability **13**(6), 3196 (2021)
25. Chen, G., Wang, H., Zhang, C.: Mobile cellular network security vulnerability detection using machine learning. Int. J. Inf. Commun. Technol. **22**(3), 327–341 (2023)
26. Chen, J.-L., et al. 2023. "Security Document Generation for Common Criteria Using Machine Learning and Rule-based Expert System". In 2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT). IEEE.
27. Chen, L., et al.: Decision transformer: reinforcement learning via sequence modeling. Adv. Neural Inform. Process. Syst. **34**, 15084–15097 (2021)
28. Chen, Y., et al. 2021. "Bookworm game: Automatic discovery of lte vulnerabilities through documentation analysis". In 2021 IEEE Symposium on Security and Privacy (SP). IEEE.
29. Cheng, L., Liu, F., Yao, D.: Enterprise data breach: causes, challenges, prevention, and future directions. Wiley Interdisciplin. Rev. Data Mining Know. Discovery **7**(5), e1211 (2017)
30. Chowdhary, A., et al. 2020. "Autonomous security analysis and penetration testing". in 2020 16th International Conference on Mobility, Sensing and Networking (MSN). IEEE.
31. Chu, Z., Han, Y., Zhao, K.: Botnet vulnerability intelligence clustering classification mining and countermeasure algorithm based on machine learning. IEEE Access **7**, 182309–182319 (2019)
32. Chui, M., et al. 2023. "The state of AI in 2023: Generative AI's breakout year".
33. Dankwa, S., Yang, L.: An efficient and accurate depth-wise separable convolutional neural network for cybersecurity vulnerability assessment based on CAPTCHA breaking. Electronics **10**(4), 480 (2021)
34. Das, R. and R. Sandhane. 2021. "Artificial intelligence in cyber security". in Journal of Physics: Conference Series. IOP Publishing.
35. De Azambuja, A.J.G., et al.: Artificial intelligence-based cyber security in the context of industry 4.0—a survey. Electronics **12**(8), 1920 (2023)
36. Deimling, F. and M. Fazzolari. 2023. "AMOE: A Tool to Automatically Extract and Assess Organizational Evidence for Continuous Cloud Audit". In IFIP Annual Conference on Data and Applications Security and Privacy. Springer.
37. Dike, H.U., et al. 2018. "Unsupervised learning based on artificial neural network: A review". In 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS). IEEE.
38. Duan, X., et al. 2021. "Automated security assessment for the internet of things". In 2021 IEEE 26th Pacific Rim International Symposium on Dependable Computing (PRDC). IEEE.
39. Elahi, H., et al.: On the characterization and risk assessment of ai-powered mobile cloud applications. Comput Standards Interfaces **78**, 103538 (2021)
40. Ferdinand, M.R., S. Mandala, and D. Oktaria. 2021. "Host Vulnerability Analysis Using Supervised Learning Based on Port Response". In 2021 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA). IEEE.
41. Flanagan, K., et al. 2016. "SAVIO R: security analytics on asset vulnerability for information abstraction and risk analysis". In 2016 UKSim-AMSS 18th International Conference on Computer Modelling and Simulation (UKSim). IEEE.
42. Garg, S., Baliyan, N.: A novel parallel classifier scheme for vulnerability detection in android. Comput. Electr. Eng. **77**, 12–26 (2019)
43. Garrad, P., Unnikrishnan, S.: Reinforcement learning in VANET penetration testing. Results Eng. **17**, 100970 (2023)
44. Ghanem, M.C. and T.M. Chen. 2018. "Reinforcement learning for intelligent penetration testing". In 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4). IEEE.
45. Ghanem, M.C., Chen, T.M.: Reinforcement learning for efficient network penetration testing. Information **11**(1), 6 (2019)
46. Ghanem, M.C., et al. 2023. "ESASCF: expertise extraction, generalization and reply framework for optimized automation of network security compliance". IEEE Access.
47. Ghanem, M.C., Chen, T.M., Nepomuceno, E.G.: Hierarchical reinforcement learning for efficient and effective automated penetration testing of large networks. J. Intel. Inform. Syst. **60**(2), 281–303 (2023)
48. Ghosh, S., et al.: An integrated approach of threat analysis for autonomous vehicles perception system. IEEE Access **11**, 14752–14777 (2023)
49. Godboley, S., et al. 2022. "Ssg-afl: Vulnerability detection for reactive systems using static seed generator based afl". In 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC). IEEE.
50. Gong, K., et al. 2022. "DRLFCfuzzer: fuzzing with Deep-Reinforcement-Learning under Format Constraints". In 2022 2nd

International Conference on Electronic Information Engineering and Computer Technology (EIECT). IEEE.
51. Grieco, G. and A. Dinaburg. 2018. "Toward smarter vulnerability discovery using machine learning". In Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security.
52. Gupta, P.: Applications of fuzzy logic in daily life. Int. J. Adv. Res. Comput. Sci. **8**(5), 1795 (2017)
53. Hale, M.L., Gamble, R.F.: Semantic hierarchies for extracting, modeling, and connecting compliance requirements in information security control standards. Requirements Eng. **24**, 365–402 (2019)
54. Hatamian, M., S. Pape, and K. Rannenberg. 2019. "ESARA: a framework for enterprise smartphone apps risk assessment". In ICT Systems Security and Privacy Protection: 34th IFIP TC 11 International Conference, SEC 2019, Lisbon, Portugal, June 25–27, 2019, Proceedings 34. Springer.
55. Hatzivasilis, G., et al. 2023. "Continuous Security Assurance of Modern Supply-Chain Ecosystems with Application in Autonomous Driving: The FISHY approach for the secure autonomous driving domain". In 2023 IEEE International Conference on Cyber Security and Resilience (CSR). IEEE.
56. Hayrapetian, A. and R. Raje. 2018. "Empirically analyzing and evaluating security features in software requirements". In Proceedings of the 11th Innovations in Software Engineering Conference.
57. He, Y., et al.: Artificial intelligence-based ethical hacking for health information systems: simulation study. J. Med. Int. Res. **25**, e41748 (2023)
58. Hecker, A. and M. Riguidel. 2009. "On the operational security assurance evaluation of networked IT systems". In Smart Spaces and Next Generation Wired/Wireless Networking: 9th International Conference, NEW2AN 2009 and Second Conference on Smart Spaces, ruSMART 2009, St. Petersburg, Russia, September 15–18, 2009. Proceedings. Springer.
59. Heckerman, D.: Bayesian networks for data mining. Data Min. Knowl. Disc. **1**, 79–119 (1997)
60. Hibshi, H., T.D. Breaux, and C. Wagner. 2016. "Improving security requirements adequacy". In 2016 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE.
61. Hibshi, H., Jones, S.T., Breaux, T.D.: A systemic approach for natural language scenario elicitation of security requirements. IEEE Trans. Dependable Secure Comput. **19**(6), 3579–3591 (2021)
62. Hosam, O. 2022. "Intelligent risk management using artificial intelligence". In 2022 Advances in Science and Engineering Technology International Conferences (ASET). IEEE.
63. Hsieh, C.-H., et al. 2015. "Cyber security risk assessment using an interpretable evolutionary fuzzy scoring system". In 2015 International Carnahan Conference on Security Technology (ICCST). IEEE.
64. Hu, Z., R. Beuran, and Y. Tan. 2020. "Automated penetration testing using deep reinforcement learning". In 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). IEEE.
65. Huang, K., et al. 2017. "Application of Bayesian network to data-driven cyber-security risk assessment in SCADA networks". In 2017 27th International Telecommunication Networks and Applications Conference (ITNAC). IEEE.
66. Huff, P. and Q. Li. 2021. "Towards automated assessment of vulnerability exposures in security operations". In Security and Privacy in Communication Networks: 17th EAI International Conference, SecureComm 2021, Virtual Event, September 6–9, 2021, Proceedings, Part I 17. Springer.
67. Ibrahim, M., Elhafiz, R.: Security analysis of cyber-physical systems using reinforcement learning. Sensors **23**(3), 1634 (2023)
68. Islam, S., S. Papastergiou, and S. Silvestri. 2022. "Cyber threat analysis using natural language processing for a secure healthcare system". In 2022 IEEE Symposium on Computers and Communications (ISCC). IEEE.
69. ISO. 2021. "ISO/SAE 21434:2021 road vehicles—cybersecurity engineering". volume.
70. Jabeen, G., et al. 2019. "An integrated software vulnerability discovery model based on artificial neural network". In SEKE.
71. Jagamogan, R.S., et al. 2022. "Penetration Testing Procedure using Machine Learning". In 2022 4th International Conference on Smart Sensors and Application (ICSSA). IEEE.
72. Jana, I. and A. Oprea. 2019. "AppMine: Behavioral analytics for web application vulnerability detection". In Proceedings of the 2019 ACM SIGSAC Conference on Cloud Computing Security Workshop.
73. Jaskolka, J. 2020. "Recommendations for effective security assurance of software-dependent systems". In Intelligent Computing: Proceedings of the 2020 Computing Conference, Volume 3. Springer.
74. Ji, X., et al.: A three-stage dynamic assessment framework for industrial control system security based on a method of W-HMM. Sensors **22**(7), 2593 (2022)
75. Jiang, R. and L. Wan. 2022. "Network Information Security Risk Assessment Method Based on Machine Learning Algorithm". In International Conference on Advanced Hybrid Information Processing. Springer.
76. Jiang, Y. and Y. Atif. 2020. "An approach to discover and assess vulnerability severity automatically in cyber-physical systems". In 13th international conference on security of information and networks.
77. Jiang, Y., Atif, Y.: Towards automatic discovery and assessment of vulnerability severity in cyber–physical systems. Array **15**, 100209 (2022)
78. Jiao, J., H. Zhao, and H. Cao. 2021. "Using Deep Learning to Construct Auto Web Penetration Test". In Proceedings of the 2021 13th International Conference on Machine Learning and Computing.
79. Jindal, R., R. Malhotra, and A. Jain. 2016. "Automated classification of security requirements". In 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE.
80. Jinhong, Y., K. Chul-Soo, and M.M.H. Onik. 2019. "Aggregated risk modelling of personal data privacy in internet of things". In 2019 21st International Conference on Advanced Communication Technology (ICACT). IEEE.
81. Johnson, E.C.: Security awareness: switch to a better programme. Netw. Secur. **2006**(2), 15–18 (2006)
82. Jung, J.-W., Lee, S.-W.: Security requirement recommendation method using case-based reasoning to prevent advanced persistent threats. Appl. Sci. **13**(3), 1505 (2023)
83. Katt, B., Prasher, N.: "Quantitative security assurance", in exploring security in software architecture and design. IGI Global (2019). https://doi.org/10.4018/978-1-5225-6313-6.ch002
84. Kaur, R., Gabrijelčič, D., Klobučar, T.: Artificial intelligence for cybersecurity: literature review and future research directions. Inform. Fusion **97**, 101804 (2023)
85. Kekül, H., Ergen, B., Arslan, H.: A multiclass hybrid approach to estimating software vulnerability vectors and severity score. J. Inform. Security Appl. **63**, 103028 (2021)
86. Khan, M., et al.: Deep learning methods and applications. Deep Learn. Convergence Big Data Analytics (2019). https://doi.org/10.1007/978-981-13-3459-7_3
87. Khazraei, A., et al. 2022. "Learning-based vulnerability analysis of cyber-physical systems". In 2022 ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS). IEEE.
88. Kim, H., et al.: Design of network threat detection and classification based on machine learning on cloud computing. Cluster Comput. **22**, 2341–2350 (2019)

89. Kim, K., et al.: Risk assessment scheme for mobile applications based on tree boosting. IEEE Access **8**, 48503–48514 (2020)
90. Kitchenham, B.: Procedures for performing systematic reviews. Keele, UK, Keele University **33**(2004), 1–26 (2004)
91. Kitchenham, B. 2007. "Guidelines for performing systematic literature reviews in software engineering". Technical report, Ver. 2.3 EBSE Technical Report. EBSE.
92. Klees, G., et al. 2018. "Evaluating fuzz testing". In Proceedings of the 2018 ACM SIGSAC conference on computer and communications security.
93. Krundyshev, V. 2020. "Neural network approach to assessing cybersecurity risks in large-scale dynamic networks". In 13th International Conference on Security of Information and Networks.
94. Kure, H.I., et al.: Asset criticality and risk prediction for an effective cybersecurity risk management of cyber-physical system. Neural Comput. Appl. **34**(1), 493–514 (2022)
95. Kure, H.I., Islam, S., Mouratidis, H.: An integrated cyber security risk management framework and risk predication for the critical infrastructure protection. Neural Comput. Appl. **34**(18), 15241–15271 (2022)
96. Kuznetsov, A., et al. 2019. "Automated Software Vulnerability Testing Using In-Depth Training Methods". in CMIS.
97. Labaj, M., K. Rástočný, and D. Chudá. 2019. "Towards Automatic Comparison of Cloud Service Security Certifications". In International Conference on Current Trends in Theory and Practice of Informatics. Springer.
98. Lam, H.A., Dong, Z.Y.: Transfer learning based dynamic security assessment. IET Gener. Transm. Distrib. **15**(16), 2333–2343 (2021)
99. Lange, M., F. Kuhr, and R. Möller. 2016. "Using a deep understanding of network activities for network vulnerability assessment". In Proceedings of the 1st International Workshop on AI for Privacy and Security.
100. Li, Q., et al.: INNES: an intelligent network penetration testing model based on deep reinforcement learning. Appl. Intell. **53**(22), 27110–27127 (2023)
101. Li, Q., et al.: A hierarchical deep reinforcement learning model with expert prior knowledge for intelligent penetration testing. Comput. Security **132**, 103358 (2023)
102. Li, X., et al. 2023. "Prediction of vulnerability characteristics based on vulnerability description and prompt learning". In 2023 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER). IEEE.
103. Li, Y., et al.: An intelligent penetration test simulation environment construction method incorporating social engineering factors. Appl. Sci. **12**, 6186 (2022)
104. Li, Z., et al.: An intelligent fuzzing data generation method based on deep adversarial learning. IEEE Access **7**, 49327–49340 (2019)
105. Lin, X., et al. 2023. "Research on Security Audit Technology of Smart Grid Database Based on Neural Networks". In 2023 8th International Conference on Computer and Communication Systems (ICCCS). IEEE.
106. Lin, Y., Wang, X.: A data-driven scheme based on sparse projection oblique randomer forests for real-time dynamic security assessment. IEEE Access **10**, 79469–79479 (2022)
107. Liu, H., N. Wang, and S. Liang. 2021. "Wireless communication network security intelligent monitoring system based on machine learning". In Journal of Physics: Conference Series. IOP Publishing.
108. Liu, R., Verbič, G., Ma, J.: A new dynamic security assessment framework based on semi-supervised learning and data editing. Electr. Power Syst. Res. **172**, 221–229 (2019)
109. Liu, S.-c. and Y. Liu. 2016. "Network security risk assessment method based on HMM and attack graph model". In 2016 17th IEEE/ACIS international conference on software engineering, artificial intelligence, networking and parallel/distributed computing (SNPD). IEEE.
110. Liu, Z., et al.: A novel deep learning based security assessment framework for enhanced security in swarm network environment. Int. J. Critical Infrastruct. Protect. **38**, 100540 (2022)
111. Luo, J. and J. Wang. 2021. "Vulnerability assessment of iot devices through multi-layer keyword matching". In 2021 International Conference on Computer, Internet of Things and Control Engineering (CITCE). IEEE.
112. Lyu, J., et al. 2021. "A character-level convolutional neural network for predicting exploitability of vulnerability". In 2021 International Symposium on Theoretical Aspects of Software Engineering (TASE). IEEE.
113. Ma, L. 2021. "Research on Vulnerability Exploitation and Detection Technology Based on Big Data Analysis". In 2021 IEEE International Conference on Industrial Application of Artificial Intelligence (IAAI). IEEE.
114. Mai, P.X., et al.: Modeling security and privacy requirements: a use case-driven approach. Inf. Softw. Technol. **100**, 165–182 (2018)
115. Mai, P.X., et al. 2019. "MCP: A security testing tool driven by requirements". In 2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion). IEEE.
116. Mai, X.P., et al. 2018. "A natural language programming approach for requirements-based security testing". In 29th IEEE International Symposium on Software Reliability Engineering (ISSRE 2018). IEEE.
117. Malik, A.A. and D.K. Tosh. 2023. "Dynamic Vulnerability Classification for Enhanced Cyber Situational Awareness". In 2023 IEEE International Systems Conference (SysCon). IEEE.
118. Malik, Y., C.R.S. Campos, and F. Jaafar. 2019. "Detecting android security vulnerabilities using machine learning and system calls analysis". In 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C). IEEE.
119. Matsuda, W., et al. 2019. "Cyber security risk assessment on industry 4.0 using ics testbed with ai and cloud". In 2019 IEEE conference on application, information and network security (AINS). IEEE.
120. McKinnel, D.R., et al.: A systematic literature review and meta-analysis on artificial intelligence in penetration testing and vulnerability assessment. Comput. Electr. Eng. **75**, 175–188 (2019)
121. Mehrabi, N., et al.: A survey on bias and fairness in machine learning. ACM Comput. Surv. (CSUR) **54**(6), 1–35 (2021)
122. Merlo, A. and G.C. Georgiu. 2017. "Riskindroid: Machine learning-based risk analysis on android". In ICT Systems Security and Privacy Protection: 32nd IFIP TC 11 International Conference, SEC 2017, Rome, Italy, May 29–31, 2017, Proceedings 32. Springer.
123. Mijwil, M., Salem, I.E., Ismaeel, M.M.: The significance of machine learning and deep learning techniques in cybersecurity: a comprehensive review. Iraqi J. Comput. Sci. Math. **4**(1), 87–101 (2023)
124. Mishin, I. and O. Saltykova. 2021. "Methods for improving Fuzzing-Testing Using Machine Learning and visualisation of results". In 2021 International Conference on Information Technology and Nanotechnology (ITNT). IEEE.
125. Mohamad, M., et al. 2022. "Identifying security-related requirements in regulatory documents based on cross-project classification". In Proceedings of the 18th International Conference on Predictive Models and Data Analytics in Software Engineering.
126. Moshika, A., et al.: Vulnerability assessment in heterogeneous web environment using probabilistic arithmetic automata. IEEE Access **9**, 74659–74673 (2021)

127. Mukhopadhyay, S. 2021. "Industrial Control Protocol Fuzzing using Deep Adversarial Networks". In 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3). IEEE.
128. Nagaraj, K., et al. 2019. "Vulnerability assessment and classification based on influence metrics in mobile social networks". In Proceedings of the 17th ACM International Symposium on Mobility Management and Wireless Access.
129. Nasteski, V.: An overview of the supervised machine learning methods. Horizons. b **4**, 51–62 (2017)
130. Nastic, S., et al.: A serverless real-time data analytics platform for edge computing. IEEE Internet Comput. **21**(4), 64–71 (2017)
131. Nebbione, G., Calzarossa, M.C.: A methodological framework for AI-assisted security assessments of active directory environments. IEEE Access **11**, 15119–15130 (2023)
132. Nourin, S.M., G. Karabatis, and F.C. Argiropoulos. 2021. "Measuring Software Security Using Improved CWE Base Scores".
133. Nwakanma, C.I., et al. 2022. "Effective Industrial Internet of Things Vulnerability Detection Using Machine Learning". In 2022 5th Information Technology For Education And Development (ITED). IEEE.
134. O'Malley, J. 2018. "The 10 Most Important Breakthroughs in Artificial Intelligence". Tech Radar.
135. Olorunfemi, O.L., et al.: Towards a conceptual framework for ethical AI development in IT systems. Computer Science & IT Research Journal **5**(3), 616–627 (2024)
136. Ouedraogo, M., et al.: Appraisal and reporting of security assurance at operational systems level. J. Syst. Softw. **85**(1), 193–208 (2012)
137. Padmanaban, R., et al. 2019. "Security analytics for heterogeneous Web". In 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). IEEE.
138. Peng, Y., et al. 2023. "DEFT: A Novel Deep Framework for Fuzz Testing Performance Evaluation in NextG Vulnerability Detection". IEEE Access.
139. Pope, A.S., et al. 2018. "Automated design of network security metrics". In Proceedings of the Genetic and Evolutionary Computation Conference Companion.
140. Pozdniakov, K., et al. 2020. "Smart security audit: Reinforcement learning with a deep neural network approximator". In 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA). IEEE.
141. Qian, H.Z. and W. Yong. 2021. "Research on Detection Method of Wireless Communication Network Security Vulnerability Based on Sequence Model". In 2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA). IEEE.
142. Qian, K., et al. 2021. "Ontology and reinforcement learning based intelligent agent automatic penetration test". In 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). IEEE.
143. Rafaiani, G., et al. 2023. "A Machine Learning-based Method for Cyber Risk Assessment". In 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS). IEEE.
144. Rajawat, A.S., et al.: Quantum machine learning for security assessment in the internet of medical things (IoMT). Future Internet **15**(8), 271 (2023)
145. Ramzan, F., et al.: Generative adversarial networks for synthetic data generation in finance evaluating statistical similarities and quality assessment. AI **5**(2), 667–685 (2024)
146. Ren, C., et al. 2023. "EFedDSA: An efficient differential privacy-based horizontal federated learning approach for smart grid dynamic security assessment". IEEE Journal on Emerging and Selected Topics in Circuits and Systems.
147. Ross, R.S. 2011. "Managing information security risk: Organization, mission, and information system view".
148. Sablotny, M., B.S. Jensen, and C.W. Johnson. 2019. "Recurrent neural networks for fuzz testing web browsers". In Information Security and Cryptology–ICISC 2018: 21st International Conference, Seoul, South Korea, November 28–30, 2018, Revised Selected Papers 21. Springer.
149. Saha, T., Aaraj, N., Jha, N.K.: Machine learning assisted security analysis of 5G-network-connected systems. IEEE Trans. Emerg. Top. Comput. **10**(4), 2006–2024 (2022)
150. Şahin, C.B., Dinler, Ö.B., Abualigah, L.: Prediction of software vulnerability based deep symbiotic genetic algorithms: phenotyping of dominant-features. Appl. Intell. **51**(11), 8271–8287 (2021)
151. Sarker, I.H., et al.: Internet of things (iot) security intelligence: a comprehensive overview, machine learning solutions and research directions. Mobile Netw. Appl. **28**(1), 296–312 (2023)
152. Saxena, D., et al. 2023. "An AI-driven VM threat prediction model for multi-risks analysis-based cloud cybersecurity". IEEE Transactions on Systems, Man, and Cybernetics: Systems.
153. Sengupta, A. 2021. "A Stakeholder-Centric Approach for Defining Metrics for Information Security Management Systems". In International Conference on Risks and Security of Internet and Systems. Springer.
154. Sentuna, A., et al.: A novel enhanced naïve bayes posterior probability (ENBPP) using machine learning: cyber threat analysis. Neural. Process. Lett. **53**, 177–209 (2021)
155. Shah, S., Mehtre, B.M.: An overview of vulnerability assessment and penetration testing techniques. J. Comput. Virology Hack. Tech. **11**, 27–49 (2015)
156. Shah, V.: Machine learning algorithms for cybersecurity: detecting and preventing threats. Revista Espanola de Documentacion Cientifica **15**(4), 42–66 (2021)
157. Sharma, K., Mukhopadhyay, A.: Cyber-risk management framework for online gaming firms: an artificial neural network approach. Inf. Syst. Front. **25**(5), 1757–1778 (2023)
158. Shi, F., et al.: XLNet-based prediction model for CVSS metric values. Appl. Sci. **12**(18), 8983 (2022)
159. Shostack, A.: Threat modeling: Designing for security. Wiley, Haboken (2014)
160. Shrestha, I. and M. Hale. 2019. "Detecting dynamic security threats in multi-component IoT systems".
161. Silvestri, S., et al.: Cyber threat assessment and management for securing healthcare ecosystems using natural language processing. Int. J. Inf. Secur. **23**(1), 31–50 (2024)
162. Silvestri, S., et al.: A machine learning approach for the NLP-based analysis of cyber threats and vulnerabilities of the healthcare ecosystem. Sensors **23**(2), 651 (2023)
163. Singh, M., Chauhan, S.: A hybrid-extreme learning machine based ensemble method for online dynamic security assessment of power systems. Electr. Power Syst. Res. **214**, 108923 (2023)
164. Singh, V.K., et al.: The journal coverage of web of science, scopus and dimensions: a comparative analysis. Scientometrics **126**, 5113–5142 (2021)
165. Spears, J.L., Barki, H., Barton, R.R.: Theorizing the concept and role of assurance in information systems security. Inform. Manag. **50**(7), 598–605 (2013)
166. Sun, M., Konstantelos, I., Strbac, G.: A deep learning-based feature extraction framework for system security assessment. IEEE Trans. Smart Grid **10**(5), 5007–5020 (2018)
167. Sun, P., et al. 2020. "Hybrid firmware analysis for known mobile and iot security vulnerabilities". In 2020 50th annual IEEE/IFIP international conference on dependable systems and networks (DSN). IEEE.
168. Sun, S., et al.: A survey of optimization methods from a machine learning perspective. IEEE Trans. Cyber. **50**(8), 3668–3681 (2019)

169. Tabassum, N., et al.: Qos based cloud security evaluation using neuro fuzzy model. Comput. Mater. Continua **70**(1), 1127–1140 (2022)
170. Tatarinova, Y., Sinelnikova, O.: Constructing a model for the dynamic evaluation of vulnerability in software based on public sources. Eastern-European J. Enterprise Technol. **6**(2), 114 (2021)
171. Tortorelli, A., et al.: A decision support tool for optimal configuration of critical infrastructures. Int. J. Crit. Infrastruct. **18**(2), 105–127 (2022)
172. Usmani, U.A., A. Happonen, and J. Watada. 2022. "A review of unsupervised machine learning frameworks for anomaly detection in industrial applications". In Science and Information Conference. Springer.
173. Uusitalo, L., et al.: An overview of methods to evaluate uncertainty of deterministic models in decision support. Environ. Model. Software **63**, 24–31 (2015)
174. van der Lee, W. and S. Verwer. 2018. "Vulnerability detection on mobile applications using state machine inference". In 2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). IEEE.
175. Viktoriia, H., H. Hnatienko, and T. Babenko. 2021. "An intelligent model to assess information systems security level". In 2021 Fifth World Conference on Smart Trends in Systems Security and Sustainability (WorldS4). IEEE.
176. Wadhawan, Y., AlMajali, A., Neuman, C.: A comprehensive analysis of smart grid systems against cyber-physical attacks. Electronics **7**(10), 249 (2018)
177. Wadhawan, Y. and C. Neuman. 2018. "Rl-bags: A tool for smart grid risk assessment". In 2018 International Conference on Smart Grid and Clean Energy Technologies (ICSGCE). IEEE.
178. Wang, P., et al. 2022. "DUSC-DQN: An Improved Deep Q-Network for Intelligent Penetration Testing Path Design". In 2022 7th International Conference on Computer and Communication Systems (ICCCS). IEEE.
179. Wang, Y.-f. and W.-n. He. 2021. "Research on Network Information Security Risk Assessment Based on Artificial Intelligence". In Multimedia Technology and Enhanced Learning: Third EAI International Conference, ICMTEL 2021, Virtual Event, April 8–9, 2021, Proceedings, Part I 3. Springer.
180. Wen, S.-F., Katt, B.: Exploring the role of assurance context in system security assurance evaluation: a conceptual model. Inform. Comput. Security **32**(2), 159–178 (2024)
181. Wiafe, I., et al.: Artificial intelligence for cybersecurity: a systematic mapping of literature. IEEE Access **8**, 146598–146612 (2020)
182. Williams, I., et al.: An automated security concerns recommender based on use case specification ontology. Auto. Softw. Eng. **29**(2), 42 (2022)
183. Williams, P.: Information security governance. Inf. Secur. Tech. Rep. **6**(3), 60–70 (2001)
184. Wu, F., et al. 2017. "Vulnerability detection with deep learning". In 2017 3rd IEEE international conference on computer and communications (ICCC). IEEE.
185. Xiao, J., et al.: Black-box attack-based security evaluation framework for credit card fraud detection models. INFORMS J. Comput. **35**(5), 986–1001 (2023)
186. Yadav, S.B., Dong, T.: A comprehensive method to assess work system security risk. Commun. Assoc. Inform. Syst. **34**(1), 8 (2014)
187. Yang, J.-Z., et al. 2022. "NiNSRAPM: An Ensemble Learning Based Non-intrusive Network Security Risk Assessment Prediction Model". In 2022 7th IEEE International Conference on Data Science in Cyberspace (DSC). IEEE.
188. Yang, Y., et al. 2019. "Adaptive deep models for incremental learning: Considering capacity scalability and sustainability". In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.
189. Ye, Z., Y. Guo, and A. Ju. 2019. "Zero-day vulnerability risk assessment and attack path analysis using security metric". In Artificial Intelligence and Security: 5th International Conference, ICAIS 2019, New York, NY, USA, July 26–28, 2019, Proceedings, Part IV 5. Springer.
190. Yi, J., Liu, X.: Deep reinforcement learning for intelligent penetration testing path design. Appl. Sci. **13**(16), 9467 (2023)
191. Yu, M., et al. 2019. "Vulnerability Detection in Firmware Based on Clonal Selection Algorithm". In 2019 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE.
192. Yüksel, A.S., et al.: Implementation of a web-based service for mobile application risk assessment. Turk. J. Electr. Eng. Comput. Sci. **25**(2), 976–994 (2017)
193. Zakaria, K.N., et al. 2019. "Feature extraction and selection method of cyber-attack and threat profiling in cybersecurity audit". In 2019 International Conference on Cybersecurity (ICoCSec). IEEE.
194. Zeng, Z., et al.: Locality—likelihood and criticality: Vulnerability risk prioritization through logical reasoning and deep learning. IEEE Trans. Netw. Serv. Manage. **19**(2), 1746–1760 (2021)
195. Zhang, Q. 2019. "Research on quantitative analysis of security of network risk based on big data". In 2019 International Conference on Robots & Intelligent System (ICRIS). IEEE.
196. Zhang, Y. and Z. Rao. 2020. "Research on information security evaluation based on artificial neural network". In 2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE). IEEE.
197. Zhang, Z., et al. 2022. "Artificial intelligence in cyber security: research advances, challenges, and opportunities". Artificial Intelligence Review, pages 1–25.
198. Zhao, H., et al. 2019. "SeqFuzzer: An industrial protocol fuzzing framework from a deep learning perspective". In 2019 12th IEEE Conference on software testing, validation and verification (ICST). IEEE.
199. Zhou, F., et al.: Fingerprinting IIoT devices through machine learning techniques. J. Signal Proc. Syst. **93**, 779–794 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.