AI and Ethics
https://doi.org/10.1007/s43681-024-00529-z

REVIEW

Securing tomorrow: a comprehensive survey on the synergy of Artificial Intelligence and information security

Ehtesham Hashmi[1] · **Muhammad Mudassar Yamin**[1] · **Sule Yildirim Yayilgan**[1]

Received: 24 February 2024 / Accepted: 17 July 2024

Abstract
This survey paper explores the transformative role of Artificial Intelligence (AI) in information security. Traditional methods, especially rule-based approaches, faced significant challenges in protecting sensitive data from ever-changing cyber threats, particularly with the rapid increase in data volume. This study thoroughly evaluates AI's application in information security, discussing its strengths and weaknesses. It provides a detailed review of AI's impact on information security, examining various AI algorithms used in this field, such as supervised, unsupervised, and reinforcement learning, and highlighting their respective strengths and limitations. The study identifies key areas for future AI research in information security, focusing on improving algorithms, strengthening information security, addressing ethical issues, and exploring safety and security-related concerns. It emphasizes significant security risks, including vulnerability to adversarial attacks, and aims to enhance the robustness and reliability of AI systems in protecting sensitive information by proposing solutions for potential threats. The findings aim to benefit cybersecurity professionals and researchers by offering insights into the intricate relationship between AI, information security, and emerging technologies.

Keywords Artificial Intelligence · Ethical AI · Ethical impact assessment · Information security · Privacy by design

1 Introduction

While technology has many advantages, it may also lead to harassment, violence, and disgrace by encouraging hackers to target computer systems. Concerns regarding cybersecurity and personal security arise as a result of technological innovations' dual nature [122]. Advancements in Artificial Intelligence (AI) are transforming the role of information security, presenting both opportunities and challenges [8]. This manuscript explores the critical intersection of AI with information security, highlighting how AI technologies such as machine learning can enhance security frameworks and address complex cybersecurity threats [114]. Moreover, the deployment of AI in security applications raises important ethical considerations, necessitating a balanced approach to ensure these technologies are used responsibly and fairly. The objectives of this paper are to review the application of AI technologies in enhancing information security measures, analyze the strengths and limitations of these technologies, and discuss the ethical implications of their deployment. We aim to provide a thorough understanding of the potential of AI to revolutionize security practices, along with the associated risks and ethical concerns. In conclusion, our findings reveal that while AI offers significant benefits for security, such as improved threat detection and adaptive defense mechanisms, it also requires careful consideration of ethical issues, including privacy, bias, and accountability. We propose recommendations for integrating ethical AI practices in security applications, aiming to guide future research and implementation in this field.

✉ Muhammad Mudassar Yamin
muhammad.m.yamin@ntnu.no

Ehtesham Hashmi
hashmi.ehtesham@ntnu.no

Sule Yildirim Yayilgan
sule.yildirim@ntnu.no

[1] Department of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology (NTNU), Teknologivegen 22, Gjøvik 2815, Innlandet, Norway

Published online: 30 July 2024

1.1 The evolving role of technology and information security

As the Internet and data volume have significantly expanded over the years, the corresponding increase in cyber risks presents a threat to businesses heavily dependent on data. Information security can be defined as the protection of information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction, aiming to ensure confidentiality, integrity, and availability.[1] Information is critical in governing and sustaining any organization's behavior. Information security strategies traditionally deploy both rule-based and manual methods to protect data and systems from threats. Rule-based methods rely on predefined algorithms or protocols that automatically enforce security measures based on specific conditions. For example, a rule-based intrusion detection system (IDS) might automatically block an IP address after detecting five failed login attempts within a minute, highlighting a system-driven, consistent enforcement approach. Conversely, manual methods require human intervention and decision-making, typically involving security personnel actively monitoring potential threats and making decisions based on real-time data analysis and intuition [114]. An example of a manual method would be a security analyst manually sifting through security logs to identify unusual activity, such as an unusually high volume of data transfer happening at odd hours, which might indicate a data breach or an insider threat [77]. This approach benefits from human expertise but is often slower and less scalable than rule-based methods.

It is critical for both enterprises and individuals, to draw a varied range of stakeholders with the aim of preventing the irreversible impacts of rising security concerns [122]. These days, many corporations seek technology services for faster and more efficient processes [9]. AI technologies have significantly streamlined and enhanced various business processes, offering notable improvements in speed and efficiency. A prime example of this transformation is evident in the banking sector. Traditionally, accessing bank accounts and applying for financial services involved time-consuming procedures and in-person interactions. However, with the integration of AI, these processes have become more user-friendly and efficient. For instance, AI-driven chatbots now facilitate 24/7 customer service, allowing customers to check balances, schedule payments, and even apply for loans without human intervention [49]. However, the limitation lies in the traditional information security-related approaches tend to overlook the human factor, assuming systems operate strictly logically, which may pose security challenges [53]. Although advancements in technology have led to faster and more efficient processes, a significant drawback is the frequent oversight of the human factor. This oversight results in vulnerabilities within information security systems, introducing weaknesses that can lead to security challenges not previously considered [61]. Additionally, challenges like the lack of standardization in handling big data complexities and the ever-evolving diverse nature of cyber threats further contribute to complexities in information security [80]. To ensure optimal performance, these systems require robust protection from threats, emphasizing the need to maintain information security as computer and internet usage has risen, and the significance of information security has grown [111]. Consequently, numerous journals, annual conferences, and workshops now focus on the security aspects of information systems and computing. These forums bring together experts in areas such as cryptology [2], computer science, electrical and computer engineering [43], and information systems, acting as meeting places for professionals to contribute and discuss information security concepts. Traditionally, information security relied on rule-based and manual methods [31].

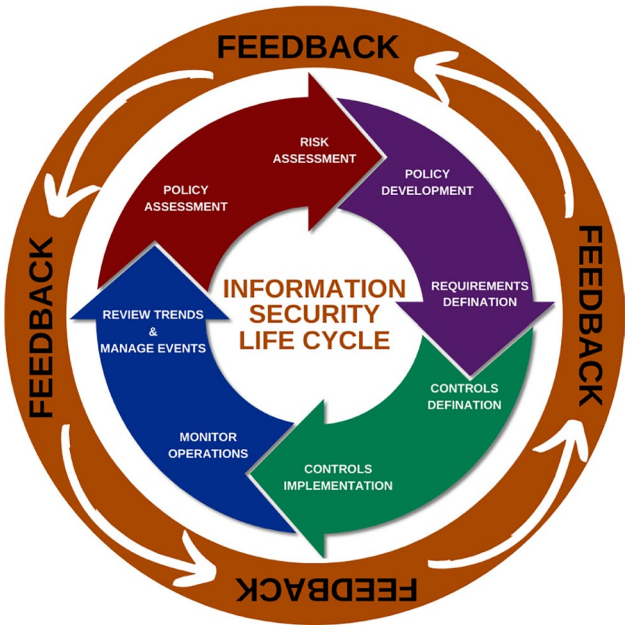

Fig. 1 Information security life cycle

The increase in data volume and complexity of cyber threats exposed the limitations of these approaches, resulting in the disclosure of vulnerabilities to the threat of physical attacks such as breaches and fake identities [4], as well as cyber-attacks such as DDoS [68], phishing [6], and Password Cracking, as well as issues such as Sensor Failure and Budget Failure [22] in the rise of AI. The following Fig. 1 outlines the general life cycle approach to information security, risk assessment identifies potential threats, guiding the

[1] https://csrc.nist.gov/glossary/term/INFOSEC.

development of policies and requirements. Policy development establishes guidelines, followed by control implementation. Continuous monitoring of operations and effective event management ensure a proactive approach to maintaining robust information security measures. The deployment of AI-driven technologies is increasingly crucial in identifying and mitigating information security threats, demonstrating enhanced performance. AI systems, utilizing Machine Learning (ML) algorithms, are adept at examining vast datasets, and identifying anomalies or irregularities indicative of potential security issues. This capability is particularly valuable in addressing complex information security challenges, where traditional security measures might overlook new or sophisticated threats. Additionally, AI significantly improves risk assessment processes. Predictive models, informed by AI, can anticipate potential information security risks based on historical data, thereby aiding in crafting more strategic policies and implementing effective controls. The integration of AI into information security signifies a notable transition from conventional rule-based and manual methods to more proactive, predictive, and automated strategies, effectively meeting the increasing complexity and volume of information security threats.

Consequently, numerous journals, annual conferences, and workshops now focus on the security aspects of information systems and computing. These forums bring together experts in areas such as cryptology, computer science, electrical and computer engineering, and information systems. Notably, contributions from the USENIX Security Symposium [1, 72] and Privacy Enhancing Technologies Symposium (PETS) have been instrumental in shaping our understanding of the ethical implications and security challenges in AI-driven systems [60]. These venues serve as pivotal meeting places for professionals to contribute and discuss the evolving landscape of information security and ethics.

1.2 AI and information security

AI involves developing computer systems that can execute tasks traditionally requiring human intelligence, like learning, problem-solving, and decision-making. The definition of AI has evolved over different decades. According to [92], AI is a concept with fluid boundaries, where the focus is on the essence of the content rather than specific terminology, underscoring the language independence of these definitions and the gradual establishment of its distinct meaning. Chowdhary et al. [29] describe AI as a subset of science and technology aimed at developing intelligent machines to automate manual tasks, significantly influencing various sectors by boosting efficiency and productivity. Wang et al. [133] characterize AI as a fundamental course in computer science, extensively incorporated across various engineering fields. This includes areas like

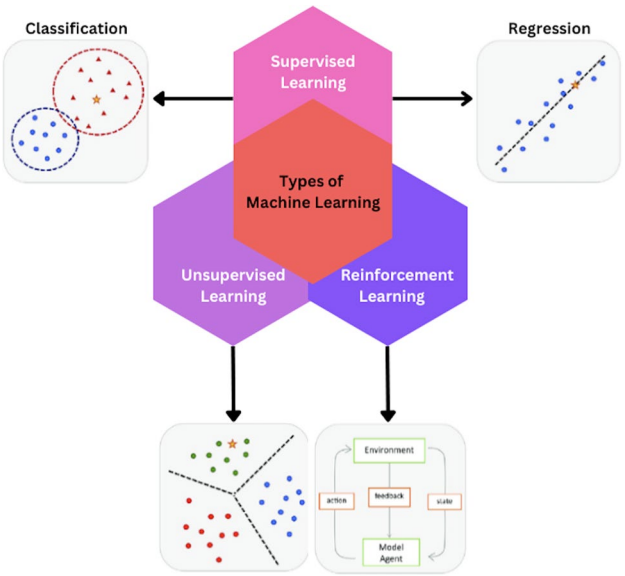

Fig. 2 Types of ML algorithms

automation, language interpretation, robotics, and several expert systems, emphasizing AI's versatility and its broad relevance to multiple disciplines. AI integrates ML techniques, such as supervised, unsupervised, and RL, into the landscape of information security [83]. Supervised learning involves training models on labeled datasets, generating tasks that map inputs to chosen outputs [110]. Unsupervised learning explores patterns within unlabeled data, involving the automated grouping of data into clusters without prior classification or categorization [10]. Reinforcement Learning (RL) utilizes a reward-based system to make decisions in dynamic environments, solving optimization problems by dynamically adapting parameters through interaction with the environment [128]. The following Fig. 2 illustrates the fundamental types of ML,

This survey addresses substantial security risks related to the use of AI in information security, such as vulnerability to adversarial attacks [25]. The research aims to enhance the robustness and reliability of AI systems in protecting sensitive information by investigating and proposing solutions for potential threats. The research also helps to solve privacy concerns associated with the use of massive datasets, ensuring a thorough and secure deployment of AI in information security practices.

1.3 Goals of the paper

1. Review of AI Applications: This survey provides a comprehensive overview of how AI technologies are currently applied in information security.

2. Strengths and Weaknesses: To perform analysis and evaluate the strengths and limitations of using AI in information security.
3. Future Research Direction: A comprehensive survey of recent advancements in AI algorithms for information security, focusing on their effectiveness in cyber threat detection and response, and examining the ethical implications associated with their deployment, including data privacy and bias considerations.
4. Practical Insights for Professionals: This study explores valuable insights and proposed solutions for professionals and researchers in the cybersecurity field to address cybersecurity challenges.
5. Ethical and Societal Impact: In the concluding part, we explore the transformative impact of AI on society and address ethical considerations in AI development.

1.3.1 Understanding rule-based and learning-based systems

The role of AI in information security spans various methodologies, primarily categorized into rule-based and learning-based systems. Rule-based AI systems operate on predefined and explicitly programmed rules. For example, a rule-based intrusion detection system might use rules such as 'block any IP address that attempts failed logins more than five times in 1 min'. In contrast, learning-based AI systems, including machine learning models, learn from data. They adjust their responses based on patterns they detect in the data, without explicit programming of the rules [130]. For instance, a learning-based intrusion detection system might analyze historical traffic data to learn to identify patterns that indicate potential security breaches.

2 AI application in information security

The integration of AI and information security has been extensively researched. This section evaluates existing literature, providing insights into the current state of knowledge and major results linked to the confluence of AI and information security. In order to understand the necessary solution categories to protect against cyberattacks, a well-known cybersecurity framework developed by the National Institute of Standards and Technology (NIST) was implemented in Shen [108]. The framework facilitates a better understanding for cybersecurity professionals and researchers of the various phases: security of information, cybersecurity defense, detection, reaction, and protection [63]. A number of studies have explored the practical implementations of AI algorithms, showcasing their effectiveness in enhancing security measures.

2.1 Enhanced threat detection

Machine Learning (ML) is a core component of AI technologies that significantly advance information security by enabling more sophisticated, adaptive threat detection systems [105]. For example, ML-based algorithms can analyze patterns from vast amounts of data to identify potential threats more rapidly and accurately than traditional methods [14]. This capability is crucial for proactive security measures, adapting to new threats as they emerge.

AI-driven strategies use ML, statistical models, and algorithms for proactive threat identification. They can detect patterns and anomalies that traditional methods might miss. AI-driven threat detection is a proactive strategy that uses ML, statistical models, and algorithms to find and address cybersecurity risks [71]. Lee et al. [71] proposed an approach for threat detection utilizing both traditional ML-based methods such as Support Vector Machine (SVM) [19, 45], Random Forrest (RF) [5], Naive Bayes (NB) [107], and Decision Tree (DT) and Deep Neural Networks (ANNs) including Convolutional Neural Networks (CNNs) [46], Fast CNN (FCNN), Long Short-Term Memory (LSTM) [48]. Their method was evaluated on two real-world datasets, namely NSLKDD[2] and CICIDS2017.[3] The researchers aimed to establish a generalizable security event analysis technique by training on a substantial amount of collected data. Their proposed work involved learning normal and threat patterns while taking into account the frequency of their occurrences. The term 'generalizable security' refers to the development of security models that maintain their effectiveness across different environments and types of data, not just the conditions they were originally trained on. This generalizability is crucial for AI systems in security because threats are constantly evolving and vary significantly across different systems and applications. For example, training AI models on a substantial amount of collected data from diverse sources enables these models to learn and recognize a wide range of threat patterns, thereby improving their ability to generalize and function effectively in different situations that were not part of their initial training set. However, it is important to note that while a large dataset can enhance the potential for generalization, it does not guarantee it. Effective generalization also requires careful selection of training examples, robust model validation methods, and continuous updating of the model to adapt to new threats.

A similar approach for threat detection was used by Le et al. [70] utilizing the CERT[4] dataset. This publicly

[2] https://www.kaggle.com/datasets/hassan06/nslkdd.

[3] https://www.kaggle.com/datasets/cicdataset/cicids2017/code.

[4] https://www.kaggle.com/datasets/mrajaxnp/cert-insider-threat-detection-research.

available dataset contains information related to Traffic Capture, Firewall Logs, Email, and user activities. They employed ML-based methods including Logistic Regression (LR) [5, 52], XGBOOST [62] with different granularity levels. In their study, Sajja et al. [101] introduced a methodology aimed at enhancing the performance of Intrusion Detection Systems (IDS). Intrusion detection and prevention are security measures employed to identify and avert cybersecurity risks to computer systems, networks, infrastructure resources, and more [82]. Their research study utilized both rule-based techniques and learning-based algorithms for the purposes of intrusion detection and classification. Their research work utilized KDD99-DATASET[5] using conventional ML-based methods such as SVM and RF. Fu et al. [38] introduced a Deep Learning (DL) based Network Intrusion Detection (DLNID) approach. Their study utilized the NSL-KDD public benchmark dataset for NID. They applied the Adaptive Synthetic Sampling (ADASYN) method to expand minority class samples, achieving a more balanced dataset. Feature extraction was performed using CNN, and the newly extracted features from an attention mechanism were subsequently fed into a Bi-Directional LSTM (Bi-LSTM) [47], resulting in a notable 0.91 F1 score. In their work, Wu et al. [136] introduced an attention mechanism in DL-based models for intrusion detection, leveraging two publicly available datasets: CICIDS2017 and CIC-DDoS2019.[6] Their proposed model, the Transformer-based Intrusion Detection System (RTIDS), achieved an impressive F1 score of 0.99.

2.2 Anomaly detection

AI excels in identifying unusual activities or patterns in data, which are crucial for spotting potential security threats. Anomaly identification, sometimes referred to as outlier detection or novelty detection in data analysis, is the process of identifying uncommon objects, occurrences, or observations that substantially differ from the majority of the data and fail to fit into a predetermined definition of regular behavior [26, 42]. Generative Adversarial Networks (GANs), serving as unsupervised learning algorithms, have seen widespread application in anomaly detection due to their ability to make abnormal inferences through the adversarial learning of sample representations [137]. Girish et al. [40] introduced a method for detecting anomalies in OpenStack cloud computing. They applied a stacked Bi-LSTM-based model to a dataset collected from OpenStack using collectd.[7] The dataset includes 10 features along with class labels. Their proposed model achieved an accuracy score of 0.94. In their research, Hasan et al. [44] applied traditional ML-based techniques and DL-based algorithms, including SVM, RF, DT, LR, and ANN. Their objective was to identify anomalies in IoT devices using the DS2OS traffic traces dataset.[8] This dataset comprises traces recorded in the IoT environment of DS2OS. Through feature extraction using label encoding [143] and one-hot encoding [136], the researchers achieved remarkable results, attaining a 99% F1 score and accuracy. This highlights the efficacy of their approach in effectively discerning anomalies in IoT network traffic.

Ullah et al. [127] presented a robust and efficient framework that makes use of the capabilities of AI of Things (AIoT) to discover anomalies within Surveillance Big Video Data (BVD). They utilized the dataset created by Sultani et al. [121], which encompasses temporal annotations within videos. This dataset encompasses 13 instances of real-world anomalous activities, such as road accidents, theft, assaults etc. In total, the dataset comprises of 1900 untrimmed surveillance videos, categorized into 950 anomalous and 900 normal videos. In their research, BiLSTM yielded an Area Under Curve (AUC) score of 68%, and the optimization was performed using the Adam optimizer. Hooshmand et al. [54] presented a method for network anomaly detection utilizing a one-dimensional CNN model. Their proposed approach involves segmenting network traffic data into Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and other protocols. Their research study conducted by the authors was founded on the UNSW_NB15 dataset.[9] This dataset comprises a total of two million and 540,044 records. Notably, their achievement included an impressive 97% F1 score specifically for the UDP protocol, demonstrating the effectiveness of their methodology in accurately detecting anomalies in network traffic. Notably, their proposed work achieved an impressive 97% F1 score specifically for the UDP protocol, demonstrating the effectiveness of their methodology in accurately detecting anomalies in network traffic. Xu et al. [139] proposed a data-driven approach for multi-class classification in intrusion and anomaly detection. The dataset employed for their analysis was the KDDcup99[10] dataset. To enhance the quality of the training dataset, they employed the Synthetic Minority Oversampling Technique (SMOTE) algorithm along with mutual information. Various algorithms were utilized to process and filter the data, and ML-based methods such as K-Nearest Neighbors (KNN), SVM, DT, and a bagging classifier. Among all of these

[5] https://www.kaggle.com/code/wailinnoo/intrusion-detection-system-using-kdd99-dataset.

[6] https://www.kaggle.com/code/dhoogla/cic-ddos2019-00-cleaning.

[7] https://collectd.org/features.shtml.

[8] https://www.kaggle.com/datasets/francoisxa/ds2ostraffictraces.

[9] https://www.kaggle.com/datasets/mrwellsdavid/unsw-nb15/data.

[10] https://www.tensorflow.org/datasets/catalog/kddcup99.

techniques, the ensemble method yielded a remarkable accuracy score of 99.7%.

Although high accuracy rates in anomaly detection algorithms are often highlighted, it is crucial to understand that accuracy alone does not guarantee effective security in real-world systems. High accuracy can indicate that the model is proficient at identifying anomalies within the specific dataset it was trained on. However, this does not necessarily mean that the system will perform equally well in practical scenarios where unexpected or novel threats occur. Furthermore, the effectiveness of anomaly detection is highly dependent on the relevance and quality of the features selected during the model training phase. It is essential to incorporate domain knowledge and continual human oversight to ensure that models are not only accurate but also relevant to the evolving nature of security threats. To address this, we advocate for a balanced approach where machine learning assists human analysts by flagging potential anomalies, while humans remain integral in the decision-making process to interpret and validate these findings.

2.3 Malware detection

AI algorithms are effective in identifying and classifying malware, offering significant improvements in protecting against malicious software. Malware, short for malicious software, is code created with the intention of causing harm and is frequently used to infiltrate or exploit a system. The introduction of malware into a computer network environment can yield various effects, contingent on the malware's intended purpose and the configuration of the network [85, 90]. Urooj et al. [129] propose a framework for analyzing reverse-engineered Android applications using ML methods. Their approach focuses on identifying vulnerabilities within smartphone applications. To facilitate their work, they employed various datasets, including MalDroid [76], DefenseDroid,[11] and GD. After pre-processing, they utilized Androguard,[12] an open-source tool, to extract essential features. The research involved training up to six ML algorithms, namely AdaBoost, SVM, DT, KNN, NB, and RF, with the goal of accurately classifying these ML algorithms. Mahdavifar et al. [76] employed a semi-supervised DL-based technique for the classification of Android malware categories. They curated the CICMalDroid2020 dataset,[13] comprising 17,341 of the latest samples across five distinct Android app categories: SMS, Banking, Adware, Benign, and Riskware. Their proposed Pseudo-Label Deep Neural Network (PDNN) algorithm yielded an F1 score of 98%. In addition to achieving a high F1 score, the creation of the CICMalDroid2020 dataset contributes significantly to the field, offering a comprehensive resource for the study and analysis of diverse Android app categories.

In their study, Mohapatra et al. [81] proposed an AI-based approach for malware detection. Their research comprised three primary stages: data processing, decision-making, and detection of malware using a dataset of malware files. To achieve this, they implemented several algorithms, such as RF, LR, DT, KNN, NB, LightGBM [64], and CatBoost [57]. They attained the highest F1 score of 98% in their proposed research study. Vinayakumar et al. [132] introduced ScaleMalNet, a scalable and hybrid DL-based approach designed for real-time deployments, facilitating effective visual malware detection. This method encompasses static, dynamic, and Image Processing (IP) components within a big data framework. The datasets utilized in their research comprised both publicly accessible and private-public datasets [12, 66, 98]. Their research work incorporated a range of traditional ML algorithms, including RF, DT, LR, NB, and KNN. Additionally, DL-based methods such as CNN, GRU, and LSTM were employed. The collective efforts resulted in a noteworthy F1 score of 99%. Notably, their work not only achieves a high F1 score but also emphasizes the significance of combining traditional ML techniques with advanced DL methods for comprehensive malware detection. Yuxin et al. [142] adopted a comparable approach for malware detection, employing the unsupervised Deep Belief Network (DBN) as their proposed method. Their experimentation involved the preparation of four datasets, each comprising 850 malicious files and 850 benign files. In their study, they employed WEKA, KNN, and SVM. Notably, among these methods, DT emerged as the most effective, providing an accuracy score of 97% when utilizing n-gram feature extraction.

The application of ML in reverse-engineering Android applications, as discussed in this section, provides a compelling example of AI as a tool to aid security efforts. While these technologies can significantly streamline the process of identifying vulnerabilities, they also come with trade-offs. One key consideration is the balance between automation and human intervention. While fully automated systems can process vast datasets rapidly, they may lack the nuanced understanding that human experts bring, particularly in complex scenarios involving new or sophisticated attack vectors. An optimal security system often involves some form of human-in-the-loop configuration where machine learning algorithms are used to handle routine analyses and flag anomalies, and security experts step in to provide deeper insights and confirmations. This hybrid approach leverages the speed and efficiency of AI while maintaining the critical judgment and expertise of human analysts.

[11] https://www.kaggle.com/datasets/defensedroid/android-malware-detection.

[12] https://github.com/androguard/androguard.

[13] https://www.kaggle.com/datasets/hasanccr92/cicmaldroid-2020.

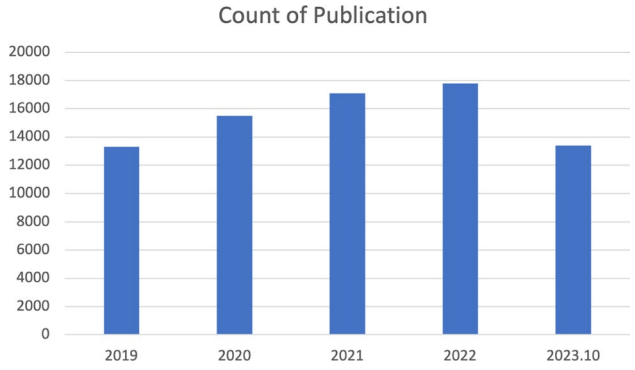

Fig. 3 Threat detection

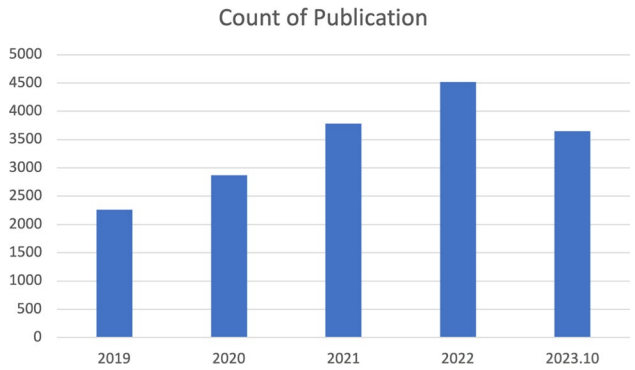

Fig. 4 Malware detection

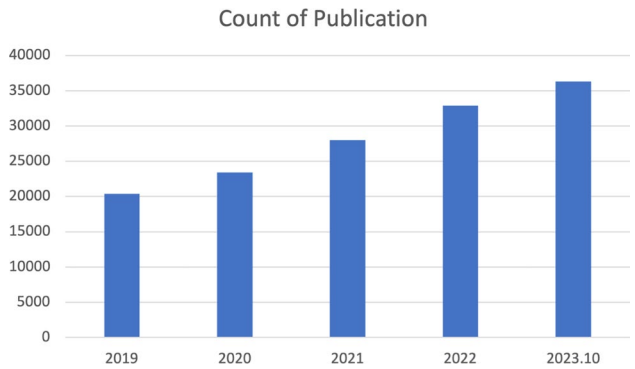

Fig. 5 Anomaly detection

The following Figs. 3, 4 and 5 show the count of publications from Scopus[14] database over the last five years that focus on the synergy of AI and information security across various subjects, including the energy sector, computer science, engineering, agriculture, education, mathematics,

Table 1 Documents by type

Document type	Count of document
Conference Paper	22,431
Article	16,693
Book chapter	1675
Review	1222
Conference review	781
Book	289
Editorial	81
Retracted	40
Short survey	33
Letter	15
Data paper	9
Erratum	9
Note	9

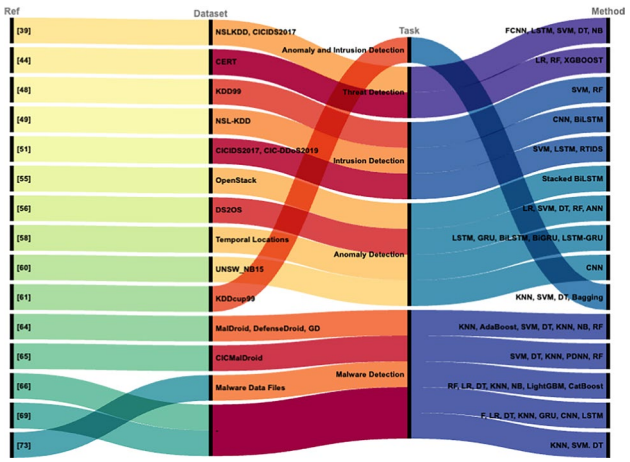

Fig. 6 Related SOTA studies

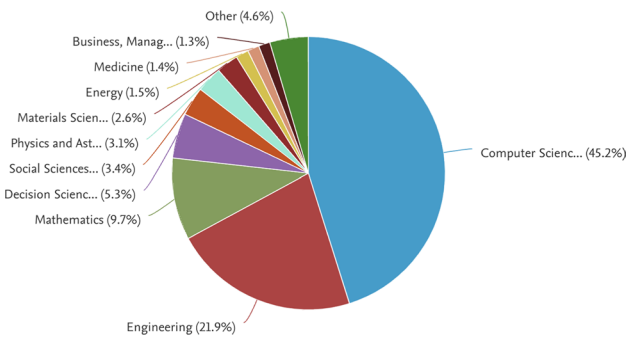

Fig. 7 Documents by subject area

physical science, material science, social science, climate change, and others. Specifically, in the domains of threat

[14] https://www.scopus.com/home.uri.

Table 2 Comparative analysis of selected studies

References	Dataset	Task	Feature set	Results
Lee et al. [71]	NSLKDD, CICIDS2017	Threat Detection	TF-IDF	Accuracy: 94%
Le et al. [70]	CERT	Threat Detection	Granularity Based	UFPR: 32.55
Möller [82]	KDD99	Intrusion Detection	Noise Removal	Accuracy: 94%
Fu et al. [38]	NSL-KDD	Intrusion Detection	One-Hot Encoding, Auto Encoding, Channel Attention	Accuracy: 91%
Wu et al. [136]	CICIDS2017, CIC-DDoS2019	Intrusion Detection	Contextual Embedding	Accuracy: 99%
Girish and Rao [40]	OpenStack	Anomaly Detection	InfluxDB	Accuracy: 94%
Hasan et al. [44]	DS2OS	Anomaly Detection	Label Encoding, One Hot Encoding	Accuracy: 99%
Ullah et al. [127]	Temporal Locations	Anomaly Detection	Contextual Embeddings	AUC: 68%
Hooshmand and Hosahalli [54]	UNSW_NB15	Anomaly Detection	SMOTE	F1 Score: 97%
Xu et al. [139]	KDDcup99	Anomaly and Intrusion Detection	SMOTE	Accuracy: 99.7%
Urooj et al. [129]	MalDroid, DefenseDroid, GD	Malware Detection	Androguard	F1 Score: 96%
Mahdavifar et al. [76]	CICMalDroid	Malware Detection	Androguard	F1 Score: 98%
Mohapatra et al. [81]	–	Malware detection	–	F1 Score: 98%
Vinayakumar et al. [132]	Krčál et al. [66], Anderson and Roth [12], Raff et al. [98]	Malware detection	Sequential, contextual embeddings	F1 Score: 99%
Yuxin and Siyi [142]	Malware data files	Malware detection	N-Grams	F1 Score: 97%

detection, malware detection, and anomaly detection (Table 1, Fig. 6).

The following Table 2 represents the comparative analysis of the current state-of-the-art (SOTA) methods along with Fig. 7, which shows the percentages of produced documents by subject area.

The following Table 2 shows the summarized review of the existing work in the domain of intrusion detection, anomaly detection, malware detection, and threat detection.

Table 2 represents the comprehensive analysis of ML, DL, and advanced AI methods in cybersecurity, particularly for anomaly, threat, and malware detection. The detailed table, covering scientific work, datasets, feature sets, models used, and evaluation measures, serves as a crucial reference for understanding the diverse applications and effectiveness of these methods. This work not only highlights the versatility of AI-based techniques in cybersecurity but also lays the groundwork for future innovations in this rapidly advancing field.

3 AI and society: transformative impact and ethical considerations

The integration of AI into information security practices not only enhances capabilities but also introduces complex ethical issues that warrant thorough evaluation [97]. The ethical concerns arise primarily because AI systems, by their nature, operate with a level of autonomy that can influence decision-making processes directly [118]. This autonomy, if not properly managed, can lead to outcomes that are unintentionally biased, discriminatory, or infringe on privacy. Moreover, AI systems need to follow moral and ethical rules primarily because their decisions can have significant real-world impacts on individuals and communities. While organizations that create and use AI are ultimately responsible for ensuring these systems are ethical, the systems themselves must be designed from the outset to adhere to ethical principles to prevent harm. This is particularly important in sectors like banking and health, where decisions can affect financial stability and well-being. Regarding compliance, while it ensures that systems operate within legal frameworks, ethical AI goes beyond mere compliance. It involves embedding fairness, accountability, transparency, and respect for user privacy into the AI system's design and operation. Compliance ensures legality, but ethics seeks to ensure morality and fairness, which may not always be covered by existing laws. The transformative impact of AI on society spans diverse fields, influencing daily life in areas such as personalized advertising [73], self-driving machinery [36], employment dynamics, and breakthroughs in healthcare [39]. A major challenge in this area is making AI systems that follow moral and ethical rules. To address this, industries need to focus on two things: understanding AI Ethics and finding out how to build Ethical AI. As AI systems become more independent, it is increasingly important to find the right balance between technological growth and what society values as right and wrong. Issues around privacy and respecting human rights and societal norms are important to think about when developing and using new technologies in the field of Information and Communication Technology (ICT) [117, 123].

3.1 Independence of AI systems

When we say that AI systems become more independent, it means that we are referring to the increasing capability of these systems to perform tasks without human intervention, thanks to advancements in AI technologies [115]. This increased automation necessitates robust ethical guidelines and regulatory oversight to ensure that automated decisions are just and fair. While regulations may mandate certain ethical safeguards, the inherent capabilities of AI to learn from vast datasets and adapt over time mean that ongoing monitoring and governance are critical to ensure these systems do not deviate from ethical norms.

3.2 Ethical concerns in AI

The biases in the learning algorithm cause discrimination, the prediction of sensitive personal data such as sexual preferences [120], and the potential of political manipulation through AI highlight a wide range of ethical concerns. Discrimination by algorithms occurs when biased data or biased decision-making criteria are used in AI models, resulting in unfair treatment of certain groups based on race, gender, age, or other characteristics [30]. This form of discrimination is often not a deliberate choice by the company but rather an unintended consequence of using historical data that may reflect past prejudices. Companies are ultimately responsible for the outputs of their AI systems and can be held liable if their systems perpetuate discrimination [65]. It is crucial for organizations to actively monitor, audit, and update their AI systems to mitigate these biases and ensure fairness in automated decisions. Furthermore, it is essential for companies to implement rigorous testing phases to detect and correct biases before deploying AI systems in real-world applications. Implementing ethical AI practices involves not only technical solutions but also governance frameworks that hold organizations accountable for their AI systems' behavior.

One major source of worry is algorithmic bias, which can result in biased consequences. Biases in recruiting algorithms, for example, may perpetuate existing inequities, raising concerns about fairness and equal opportunity in the workplace [23, 51]. Furthermore, the widespread use of AI raises the possibility of political manipulation. The ability of AI algorithms to process huge amounts of data and generate targeted content raises concerns about its possible use for political, societal, economic, and other reasons. This emphasizes the need for robust ethical frameworks to ensure the responsible use of AI, preventing its misuse in manipulating public opinion or democratic processes [18, 94, 100]. With the increasing prevalence of AI, it becomes crucial to confront challenges related to ethics, impact assessment, and broader societal implications. Striking a balance between the advantages and potential drawbacks of AI is imperative in navigating the ethical areas of its increasing prevalence. AI has a wide range of good effects and contributes to societal well-being [32]. Its applications improve living standards, speed up legal processes, generate income, strengthen public safety, and mitigate the environmental and climate implications of human activity [56].

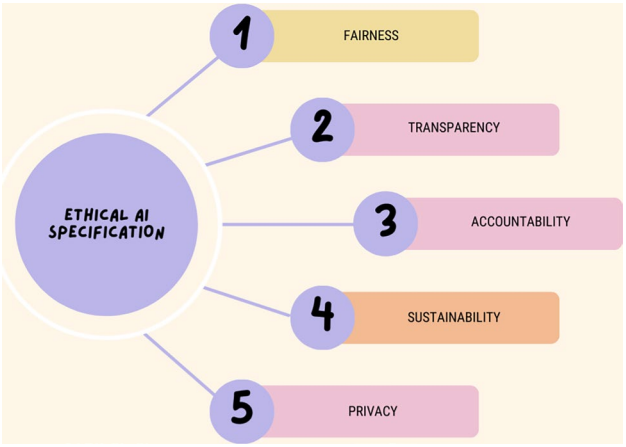

Fig. 8 Ethical AI specifications

The use of AI in security contexts can have broader implications, including geopolitical outcomes. For instance, technologies such as deep fakes and sophisticated hacking tools can be employed to create and spread propaganda, influencing public opinion and potentially disrupting democratic processes [93]. These activities can be linked to larger geopolitical strategies, making it imperative for discussions on AI and security to consider the potential misuse of these technologies in political arenas.

In addition, ensuring transparency, accountability, privacy, and fairness are essential components in building ethical AI systems, emphasizing the need for comprehensive guidelines and practices in these areas [79, 112]. This involves open communication about AI processes, clear accountability mechanisms, protection of user privacy, and the establishment of fair practices to address potential biases. The five fundamental elements of ethical AI are displayed in the following figure 8.

3.3 Ethical AI in information security

Integrating ethical AI into security practices involves ensuring that AI systems operate transparently, accountably, and without bias, particularly when processing personal or sensitive information [35]. Ethical considerations in AI-driven security are vital to maintaining user trust and complying with regulatory standards [88]. For instance, when ML is used to detect fraudulent activities, it must also protect the

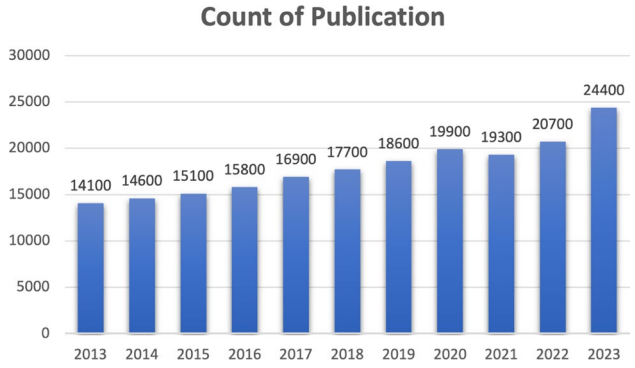

Fig. 9 Impact of AI on educational security practices

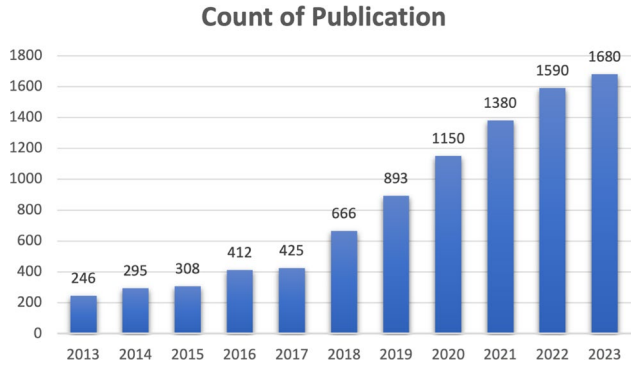

Fig. 10 AI innovations in the financial security landscape

privacy and data rights of individuals, adhering to ethical guidelines to prevent misuse and discrimination.

3.4 Fair AI across diverse domains

In the pursuit of ethical AI, considerations extend to various domains, each demanding fair practices and responsible development. In employment, fair AI strives for unbiased hiring practices and equal opportunities [69, 106], cultivating an inclusive workforce. In healthcare, fair AI contributes to unbiased diagnostics, treatment recommendations, and resource allocation [27, 126], ensuring equitable healthcare access for diverse populations. The financial sector sees fair AI preventing discriminatory practices in lending and decision-making [103], promoting financial inclusivity. Education benefits from fair AI with unbiased assessments and equitable access [75], establishing a level playing field for all learners. Criminal justice systems benefit from fair AI, mitigating biases in risk assessments and sentencing [15], striving for justice without prejudice. Retail and advertising industries benefit from fair AI, ensuring unbiased targeting and recommendations [99], cultivating a diverse marketplace. On social media, fair AI practices mitigate biases in content moderation and information dissemination [41], providing a welcoming online environment.

It is critical to recognize that the financial sector's compliance with fair AI and anti-discriminatory practices is not solely a matter of ethical choice but also a legal requirement [124]. Financial institutions are legally obligated to ensure that their AI systems do not engage in discriminatory practices, as failure to do so can result in significant legal liabilities. However, the commitment to fair AI goes beyond adhering to legal standards. While compliance ensures that financial entities do not violate regulations (such as those pertaining to equal credit opportunities), adopting ethical AI practices involves a proactive approach to fairness that seeks to surpass these regulatory minimums [35, 67]. Ethical AI practices in the financial sector involve designing AI systems

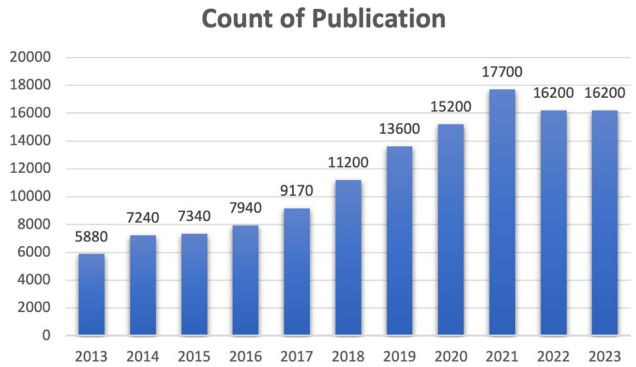

Fig. 11 AI's role in employment security

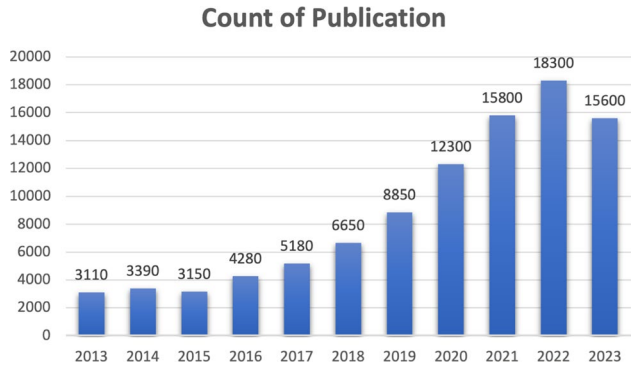

Fig. 12 Advancements of AI in healthcare security

that not only avoid discrimination but also actively promote inclusivity and fairness, regardless of legal compulsion.

The following Figs. 9, 10, 11, 12, 13, and 14 show the distribution of publications across six distinct domains concerning fair and ethical AI practices. This visualization offers insights into the prevalence of research in these areas, reflecting the growing emphasis on responsible AI development and deployment.

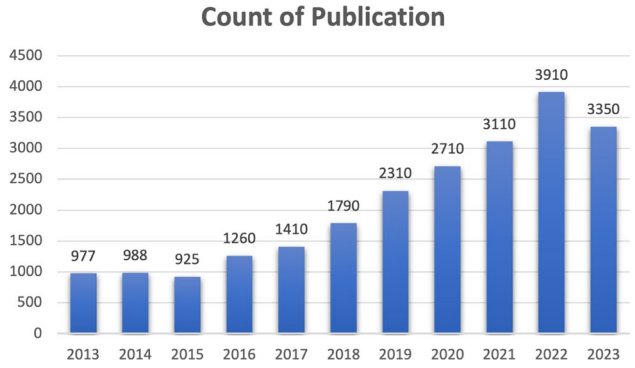

Fig. 13 AI deployment in criminal justice systems

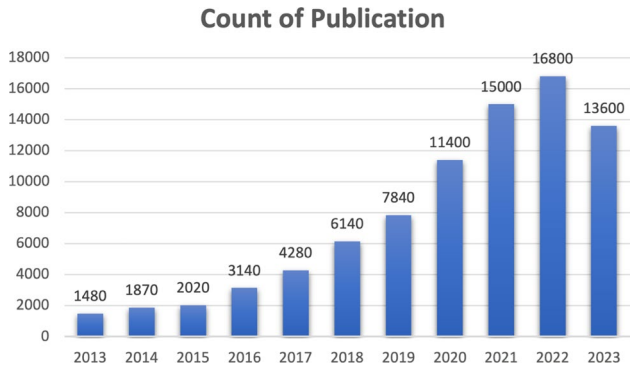

Fig. 14 AI applications in retail and social media security

Table 3 Documents by type

Document type	Count of document
Article	5464
Conference paper	2016
Review	1280
Book	1147
Book chapter	967
Note	98
Editorial	72
Short Survey	24
Letter	19
Conference review	9
Erratum	2
Retracted	1

The following Table 3 shows the number of documents produced from 2000 to 2023 in the fields of computer science and engineering for ethical and fair AI.

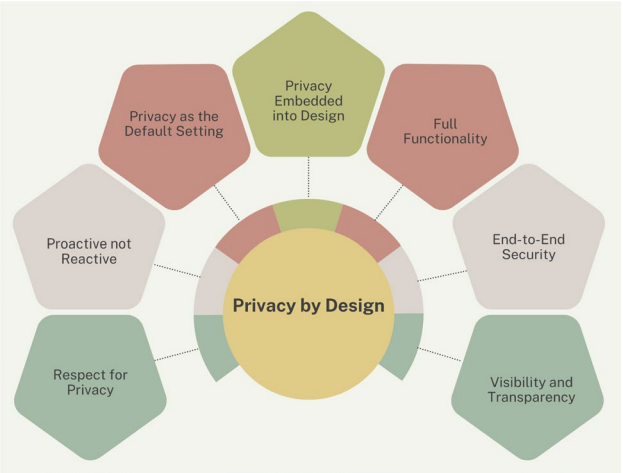

Fig. 15 PbD architecture

4 Ethical governance and privacy protocols for AI development

Maintaining ethical norms and protecting privacy are crucial in the development of AI. This section offers key frameworks such as PbD and the UNESCO Ethical Guidelines, highlighting the significance of EIA in encouraging fairness and accountability in the growth of AI.

4.1 Privacy by design

Privacy by Design (PbD) emerges as a critical ethical paradigm in this era, arguing for the proactive incorporation of privacy protection throughout AI development [21], providing openness, user empowerment, and deep respect for privacy rights [7, 74]. PbD is a concept advocating for the integration of data protection considerations during the system design phase. This approach aims to offer a practical solution that effectively addresses the concerns of data subjects and ensures privacy. The following figure 15 illustrates PbD, highlighting its essential components: making privacy the default setting, being proactive rather than reactive, embedding privacy into the design, maintaining full functionality, ensuring end-to-end security, and promoting visibility and transparency.

Furthermore, PbD emphasizes the significance of including privacy considerations throughout the AI system lifecycle, from original design to deployment and beyond. To address ethical concerns in AI, it's essential to adopt a broad approach that considers the social and ethical aspects of data use, not just the technical side [78]. This strategy helps build trust with users and ensures compliance with global data protection laws, integrating it into the development process from the start [137].

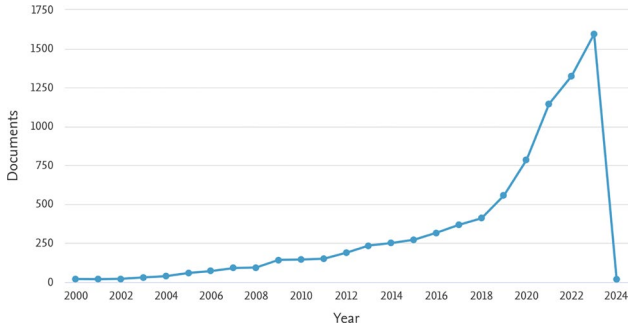

Fig. 16 Studies on PbD and ethical concerns

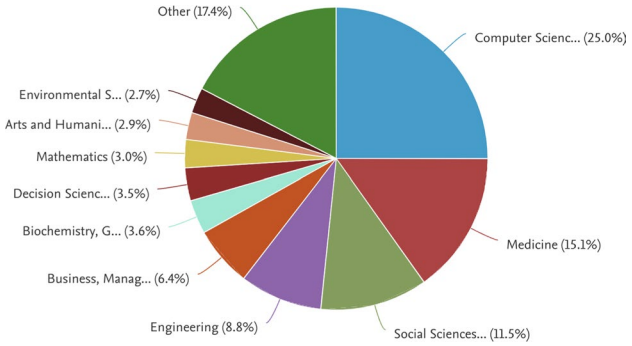

Fig. 17 Documents by subject area

The following Fig. 16 represents the count of documents generated between the years 2000 and 2024 about studies on PbD with ethical considerations.

4.2 UNESCO's ethical framework

UNESCO's recommendation on the Ethics of AI serves as a guiding framework to align AI developments with human rights, dignity, environmental sustainability, fairness, inclusion, and gender equality.[15] Complementing this, UNESCO has introduced two instrumental tools, the Readiness Assessment Methodology [125], and the Ethical Impact Assessments (EIS) in different sectors such as research and education [55, 86], which are designed to promote the incorporation of these moral values into technology breakthroughs from the beginning and ensure responsible and value-driven AI implementation [50].

4.3 Ethical impact assessments

Implementing EIA is a crucial step towards fostering responsible AI development and deployment [59]. Similar to a Privacy Impact Assessment (PIA), EIA may also serve as a method to ensure that stakeholders thoroughly scrutinize ethical implications before deployment [135]. This allows for the implementation of necessary mitigating measures. Ethical Impact Assessments focus on evaluating the potential impacts of AI systems on individuals and society, considering factors such as fairness, safety, privacy, transparency, and accountability [34, 113]. EIA's role in AI extends beyond mere compliance and risk mitigation.

It creates an environment of ethical awareness and proactive responsibility. By systematically evaluating AI systems' ethical implications, EIA ensures that the technology's development aligns with societal values and norms.

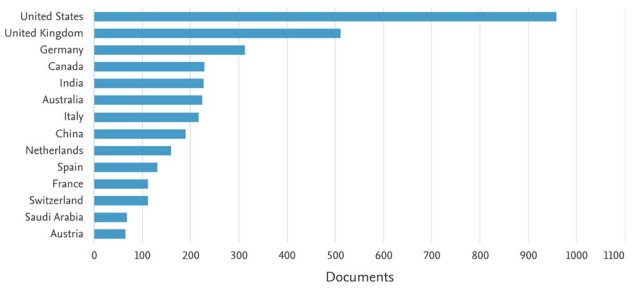

Fig. 18 Global distribution: EIA with AI studies by top 14 countries

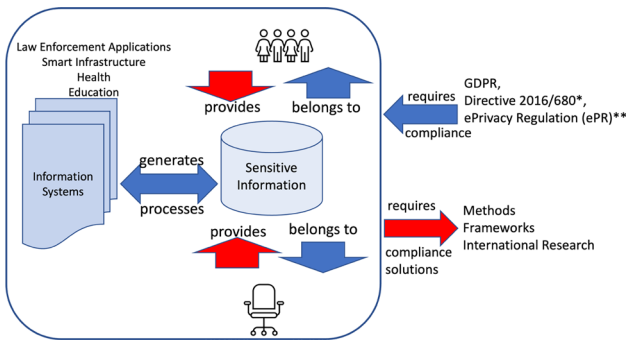

Fig. 19 Data security framework in different sectors

Figure 17 illustrates the percentage distribution of studies conducted using EIA with AI from 2000 to 2023 across various domains. It is evident that a significant proportion of these studies has been implemented in the field of computer science. Additionally, Fig. 18 presents the number of studies conducted by the top 14 countries in the realm of EIA with AI.

[15] https://www.unesco.org/en/articles/ethical-impact-assessment-tool-recommendation-ethics-artificial-intelligence.

5 Proposed data security framework in different sectors

Figure 19 outlines our proposed centralized data security management framework. This framework is adaptable to various sectors including Law Enforcement, Smart Infrastructure, Health, and Education, ensuring tailored security measures that meet the unique needs of each domain. It has been thoughtfully designed and implemented by the Multidisciplinary Research Group on Privacy and Data Protection (MR PET) at the Norwegian University of Science and Technology (NTNU).[16] This initiative reflects the group's commitment to advancing data security technologies while addressing the complex challenges of privacy and protection across diverse fields. The secure repository at the heart of this framework is critical for managing sensitive information across these diverse fields. Law Enforcement applications may involve highly confidential data requiring security protocols, while Smart Infrastructure would necessitate resilient and scalable data protection measures to protect interconnected systems.

In the Health sector, privacy and compliance with regulations like HIPAA [13] are paramount, and for Education, ensuring the confidentiality of student and faculty information is essential [33, 87]. Each sector feeds into and draws from the central repository, with security measures tailored to their unique data sensitivity and regulatory needs. The framework explains the flow of data between these sectors and the repository, highlighting the need for specialized security measures tailored to each sector's requirements. Law Enforcement requires robust protocols to protect classified information, while Smart Infrastructure demands resilient defenses for its networked systems. In Health, privacy and regulatory compliance are crucial, and in Education, safeguarding personal records is key.

6 Challenges and limitations of AI in information security

6.1 Adversarial attacks

The ability of AI systems to withstand and effectively counter adversarial attacks is referred to as adversarial AI resilience. Adversarial attacks represent deliberately changing input data in order to mislead AI models, causing them to make inaccurate or unexpected predictions [104]. Developing AI systems resilient to these attacks is an ongoing challenge, requiring innovative defensive strategies and constant adaptation to emerging attack techniques. Moreover, the evolving nature of adversarial attacks poses a significant limitation, as attackers continuously develop more sophisticated methods to exploit vulnerabilities in AI systems [96]. This arms race between attackers and defenders in AI necessitates not only advanced technical solutions but also a fundamental rethinking of AI model architectures and training methodologies. Additionally, the requirement for extensive datasets to train AI models for adversarial resilience often raises concerns about data privacy and accessibility, further complicating the development of robust AI defenses[91].

6.2 Bias and fairness

AI algorithms may unintentionally reproduce biases existing in training data, resulting in biased results. It is an ethical duty to address bias in AI systems and ensure fairness in decision-making processes. To achieve justice, AI models must be continuously monitored, evaluated, and improved [69]. Justice in AI models refers to the principle of fairness in how AI systems make decisions that affect individuals ensuring that no group or individual is unfairly disadvantaged by automated processes [65]. This concept is closely linked to the broader goal of achieving equity in AI outcomes across diverse demographic groups.

In addition to technical measures, addressing bias in AI necessitates a deep understanding of the socio-cultural contexts from which data originates. This involves identifying and mitigating biases not just in the data, but also in the algorithms' design and implementation processes. Addressing 'bias in the data' involves identifying and correcting skewed data that may lead AI systems to make prejudiced decisions [27]. Mitigation strategies often involve revising the data collection and preparation processes to reflect a more balanced perspective or adjusting the algorithmic model to counteract known biases. However, mitigating bias does not automatically guarantee fairer outcomes. There is a complex trade-off between mitigating bias and maintaining the integrity and usability of the data. Over-correcting for bias, for example, can lead to new forms of biases, sometimes at the expense of other important outcomes such as accuracy or predictive reliability.

Moreover, the subjective nature of what constitutes 'fairness' adds another layer of complexity, as different stakeholders may have varying perspectives on fair outcomes. Achieving consensus on these definitions is crucial but challenging [37, 131]. Furthermore, even with continuous monitoring and updating, the inherent limitations in data representation and the ever-evolving societal values make achieving absolute fairness an elusive goal. This highlights the need for ongoing dialogue and collaboration between technologists, and other stakeholders in society.

[16] https://www.ntnu.no/.

Fig. 20 Top 10 emerging cyber-security threats for 2030 by ENISA

6.3 Resource intensiveness

Creating powerful AI models for real-time threat detection might take time and money. This raises issues about AI system performance, energy utilization, and overall scalability, particularly when faced with the challenges of limited or inadequate datasets. These constraints, which are especially significant for smaller firms with limited resources, may impede the adoption of cutting-edge AI solutions and present issues in keeping up with increasing security threats and technological breakthroughs [3, 11].

The need for high-performance computing resources to process and analyze large amounts of data in real-time further emphasizes the resource intensiveness of advanced AI models [28, 102]. This not only increases operational costs but also contributes to higher energy consumption, raising environmental concerns. Smaller organizations, in particular, may find it challenging to justify the high initial investment and ongoing costs associated with such sophisticated systems [24, 89, 109]. Additionally, the reliance on high-end hardware and software can create dependencies on specific vendors, potentially leading to issues with interoperability and flexibility in integrating with existing security infrastructures. Furthermore, the challenge of ensuring that these resource-intensive AI systems are resilient to disruptions and capable of operating under varying conditions adds another layer of complexity, especially in scenarios where resources are constrained or in fluctuating demand.

It is important to also focus on the specific challenges posed by generative AI and large language models (LLMs). These technologies, which include models like Generative Pre-trained Transformer (GPT) and other similar architectures, are increasingly used in security applications for tasks such as automated threat detection, simulation of cyber attacks, and natural language processing for security protocol compliance [140, 141]. Generative AI and LLMs are particularly resource-intensive, requiring significant computational power not only for initial training but also for ongoing operations [138]. This leads to substantial energy consumption and, consequently, a larger carbon footprint, which is a critical concern in the context of global efforts to reduce greenhouse emissions [58]. The use of these models in security applications can exacerbate environmental impacts, especially as their deployment scales across industries.

7 Top cybersecurity threats by ENISA

The Fig. 20 "Top 10 Emerging Cyber-Security Threats for 2030" from The European Union Agency for Cybersecurity (ENISA)[17] maps out the expected major cybersecurity challenges of the next decade.

It highlights the risk of attackers targeting the supply chain to tamper with software components and the issue of misinformation campaigns disrupting public discourse. Privacy is at stake due to increased digital tracking. Human mistakes and outdated systems pose significant security risks, especially as the cyber and physical worlds converge. The exploitation of smart device data can lead to precise and damaging cyber attacks.

There's a noted concern over the security of space-based assets, such as satellites, essential for global communication. The graphic points to the emergence of complex, multi-layered threats and the shortage of trained cybersecurity professionals. Dependency on international ICT providers could lead to significant systemic failures. Lastly, it flags the potential misuse of AI in cyber attacks, emphasizing the need for vigilant and comprehensive security measures.

8 Future directions

Several intriguing paths for future research and development arise as the fields of AI and information security continue to evolve. Addressing these directions can help to improve the effectiveness, efficiency, and ethical considerations of AI in information security.

8.1 Adversarial AI resilience

The goal of adversarial AI resilience is to create AI systems that can maintain their performance and accuracy even when confronted with well-constructed adversarial inputs [84]. Building AI models that are resistant to adversarial attacks is an important goal for the future. Researchers should focus on

[17] https://www.enisa.europa.eu/news/cybersecurity-threats-fast-forward-2030.

improving the robustness of AI systems, exploring advanced adversarial training methods, and developing creative architectures that can withstand intricate assaults. Understanding adversarial AI's fundamental principles and creating solutions to mitigate weaknesses will be crucial [20].

8.2 Hybrid AI defense strategies

Future efforts should be devoted to developing hybrid AI models that combine the strengths of rule-based systems and ML-based methods. This strategy is intended to provide a comprehensive defense mechanism against constantly changing cyber attacks. Hybrid models can provide increased threat detection capabilities by using the interpretability of rule-based systems and the adaptability of ML [116, 134]. Integrating these hybrid approaches across multiple areas such as education, healthcare, finance, and critical infrastructure has the potential to improve many organizations' overall security situation. Organizations can create robust and context-aware defense systems against emerging cyber threats by adapting hybrid AI models to the specific problems and requirements of each industry [95, 119].

8.3 Ethical and explainable AI practices

As the integration of AI with information security becomes more prevalent, prioritizing ethical considerations and emphasizing the importance of explainability becomes critical. Future research must concentrate on creating AI models that follow ethical norms, ensuring fairness, transparency, and responsibility [16, 46]. Explainable AI (XAI) approaches should be developed to provide explicit insights into AI decision-making processes, increasing user trust and making it easier to identify any ethical concerns [17].

9 Conclusion

This study summarizes how AI technologies like ML and DL have revolutionized threat detection and response mechanisms, offering more efficient, proactive, and adaptive cybersecurity solutions. We briefly mention the challenges, such as data requirements, vulnerability to adversarial attacks, and the need for continuous learning and adaptation in AI models. This survey highlights the importance of addressing ethical issues such as data privacy, bias in AI algorithms, and the need for transparent AI operations in cybersecurity. In conclusion, even though AI has the potential to completely transform information security, its responsible and successful implementation depends on recognizing and resolving these issues and constraints. To overcome these obstacles and guarantee that AI in information security is in line with moral standards, protects privacy, and strengthens cybersecurity overall, researchers, business leaders, and legislators must work together. As technology develops further, utilizing AI's advantages while reducing its inherent drawbacks will require a proactive and flexible strategy.

Author contributions Ehtesham Hashmi: conceptualization, literature analysis, research execution, resources, writing original draft, investigation. Muhammad Mudassar Yamin: visualization, supervision, project management, funding acquisition, research conduct, validation. Sule Yildirim Yayilgan: visualization, supervision, project management, funding acquisition, research conduct, validation.

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

Data availability and access Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical and informed consent for data used Not applicable.

References

1. Adida, B.: Helios: Web-based open-audit voting. In: USENIX security symposium, pp. 335–348 (2008)
2. Agrawal, S., Lin, D.: Advances in Cryptology–ASIACRYPT 2022: 28th International Conference on the Theory and Application of Cryptology and Information Security, Taipei, Taiwan, December 5–9, 2022, Proceedings, Part IV, vol. **13794**. Springer Nature (2023)
3. Ahmad, W., Rasool, A., Javed, A.R., et al.: Cyber security in IoT-based cloud computing: a comprehensive survey. Electronics **11**(1), 16 (2021)
4. Al-Charchafchi, A., Manickam, S., Alqattan, Z. N.: Threats against information privacy and security in social networks: a review. In: Advances in Cyber Security: First International Conference, ACeS 2019, Penang, Malaysia, July 30–August 1, 2019, Revised Selected Papers 1, Springer, pp. 358–372 (2020)
5. Ali, H., Hashmi, E., Yayilgan Yildirim, S., et al.: Analyzing amazon products sentiment: a comparative study of machine and deep learning, and transformer-based techniques. Electronics **13**(7), 1305 (2024)

6. Alkhalil, Z., Hewage, C., Nawaf, L., et al.: Phishing attacks: a recent comprehensive study and a new anatomy. Front. Comput. Sci. **3**, 563060 (2021)
7. Alkhariji, L., De, S., Rana, O., et al.: Semantics-based privacy by design for internet of things applications. Futur. Gener. Comput. Syst. **138**, 280–295 (2023)
8. Al-Khassawneh, Y.A.: A review of artificial intelligence in security and privacy: research advances, applications, opportunities, and challenges. Indonesian J. Sci. Technol. **8**(1), 79–96 (2023)
9. Alkhudhayr, F., Alfarraj, S., Aljameeli, B., et al. Information security: a review of information security issues and techniques. In: 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), IEEE, pp. 1–6 (2019)
10. Alloghani, M., Al-Jumeily, D., Mustafina, J., et al. A systematic review on supervised and unsupervised machine learning algorithms for data science. Supervised and unsupervised learning for data science pp. 3–21 (2020)
11. Ameen, A.H., Mohammed, M.A., Rashid, A.N.: Dimensions of artificial intelligence techniques, blockchain, and cyber security in the internet of medical things: opportunities, challenges, and future directions. J. Intell. Syst. **32**(1), 20220267 (2023)
12. Anderson, H. S., Roth, P.: Ember: an open dataset for training static pe malware machine learning models. (2018). arXiv:1804.04637
13. Anderson, C., Baskerville, R., Kaul, M.: Managing compliance with privacy regulations through translation guardrails: a health information exchange case study. Inf. Organ. **33**(1), 100455 (2023)
14. Azam, Z., Islam, M.M., Huda, M.N.: Comparative analysis of intrusion detection systems and machine learning based model analysis through decision tree. IEEE Access (2023)
15. Bagaric, M., Svilar, J., Bull, M., et al.: The solution to the pervasive bias and discrimination in the criminal justice system: transparent and fair artificial intelligence. Am. Crim. L Rev. **59**, 95 (2022)
16. Balasubramaniam, N., Kauppinen, M., Hiekkanen, K., et al. Transparency and explainability of ai systems: ethical guidelines in practice. In: International Working Conference on Requirements Engineering: Foundation for Software Quality, Springer, pp. 3–18 (2022)
17. Balasubramaniam, N., Kauppinen, M., Rannisto, A., et al.: Transparency and explainability of AI systems: from ethical guidelines to requirements. Inf. Softw. Technol. **159**, 107197 (2023)
18. Bankins, S., Formosa, P.: The ethical implications of artificial intelligence (AI) for meaningful work. J. Bus. Ethics 1–16 (2023)
19. Bansal, M., Goyal, A., Choudhary, A.: A comparative analysis of k-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. Decis. Anal. J. **3**, 100071 (2022)
20. Bartz-Beielstein, T.: Why we need an AI-resilient society. arXiv preprint arXiv:1912.08786 (2019)
21. Bazalytskyi, V.: Artificial intelligence and "privacy by default". Ukr J Int'l L pp. 63 (2023)
22. Bhushan, B., Sahoo, G.: Requirements, protocols, and security challenges in wireless sensor networks: an industrial perspective. Handbook of computer networks and cyber security: principles and paradigms 683–713 (2020)
23. Bornstein, S.: Antidiscriminatory algorithms. Ala L Rev. **70**, 519 (2018)
24. Bravyi, S., Dial, O., Gambetta, J.M., et al.: The future of quantum computing with superconducting qubits. J. Appl. Phys. **132**(16) (2022)
25. Chakraborty, A., Alam, M., Dey, V., et al.: A survey on adversarial attacks and defences. CAAI Trans. Intell. Technol. **6**(1), 25–45 (2021)
26. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. ACM Comput. Surv. (CSUR) **41**(3), 1–58 (2009)
27. Chen, P., Wu, L., Wang, L.: Ai fairness in data management and analytics: a review on challenges, methodologies and applications. Appl. Sci. **13**(18), 10258 (2023)
28. Chidukwani, A., Zander, S., Koutsakis, P.: A survey on the cyber security of small-to-medium businesses: challenges, research focus and recommendations. IEEE Access **10**, 85701–85719 (2022)
29. Chowdhary, K.: Fundamentals of Artificial Intelligence. Springer (2020)
30. Cossette-Lefebvre, H., Maclure, J.: Ai's fairness problem: understanding wrongful discrimination in the context of automated decision-making. AI Ethics **3**(4), 1255–1269 (2023)
31. Croft, R., Newlands, D., Chen, Z., et al. An empirical study of rule-based and learning-based approaches for static application security testing. In: Proceedings of the 15th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 1–12 (2021)
32. Declaration, M.: Montréal declaration for a responsible development of artificial intelligence (2018)
33. Deshmukh, P., Croasdell, D.: Hipaa: Privacy and security in health care networks. In: Information Security and Ethics: Concepts, Methodologies, Tools, and Applications. IGI Global, pp. 2770–2781 (2008)
34. Dhirani, L.L., Mukhtiar, N., Chowdhry, B.S., et al.: Ethical dilemmas and privacy issues in emerging technologies: a review. Sensors **23**(3), 1151 (2023)
35. Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., et al.: Connecting the dots in trustworthy artificial intelligence: from AI principles, ethics, and key requirements to responsible ai systems and regulation. Inf. Fusion **99**, 101896 (2023)
36. Elkholy, H. A., Azar, A. T., Shahin, A. S., et al. Path planning of a self driving vehicle using artificial intelligence techniques and machine vision. In: Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020), Springer, pp. 532–542 (2020)
37. Feng, T., Hebbar, R., Mehlman, N., et al.: A review of speech-centric trustworthy machine learning: privacy, safety, and fairness. APSIPA Trans. Signal Inf. Process. **12**(3), (2023)
38. Fu, Y., Du, Y., Cao, Z., et al.: A deep learning model for network intrusion detection with imbalanced data. Electronics **11**(6), 898 (2022)
39. Gams, M., Kolenik, T.: Relations between electronics, artificial intelligence and information society through information society rules. Electronics **10**(4), 514 (2021)
40. Girish, L., Rao, S.K.: Anomaly detection in cloud environment using artificial intelligence techniques. Computing **105**(3), 675–688 (2023)
41. Gonçalves, J., Weber, I., Masullo, G.M., et al.: Common sense or censorship: how algorithmic moderators and message type influence perceptions of online content deletion. New Media Soc. **25**(10), 2595–2617 (2023)
42. Habeeb, R.A.A., Nasaruddin, F., Gani, A., et al.: Real-time big data processing for anomaly detection: a survey. Int. J. Inf. Manage. **45**, 289–307 (2019)
43. Hariyanto, N., Murjito, E.A., Furqani, J., et al.: Study of static security assessment accuracy results using random forest with various types of training and test datasets. Int. J. Electric. Eng. Inform. **15**(1), 119–133 (2023)
44. Hasan, M., Islam, M.M., Zarif, M.I.I., et al.: Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. Internet Things **7**, 100059 (2019)
45. Hashmi, E., Yamin, M. M., Imran, S., et al. Enhancing misogyny detection in bilingual texts using fasttext and explainable AI.

In: 2024 International Conference on Engineering & Computing Technologies (ICECT), IEEE, pp. 1–6 (2024a)
46. Hashmi, E., Yayilgan, S.Y., Yamin, M.M., et al.: Advancing fake news detection: hybrid deep learning with fasttext and explainable AI. IEEE Access (2024)
47. Hashmi, E., Yayilgan, S.Y.: Multi-class hate speech detection in the norwegian language using fast-rnn and multilingual fine-tuned transformers. Complex Intell. Syst. 1–22 (2024)
48. Hashmi, E., Yayilgan, S.Y., Shaikh, S.: Augmenting sentiment prediction capabilities for code-mixed tweets with multilingual transformers. Soc. Netw. Anal. Min. **14**(1), 86 (2024)
49. Hassan, M., Aziz, L.A.R., Andriansyah, Y.: The role artificial intelligence in modern banking: an exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. Rev. Contempor. Bus. Anal. **6**(1), 110–132 (2023)
50. Havrda, M., Klocek, A.: Well-being impact assessment of artificial intelligence-a search for causality and proposal for an open platform for well-being impact assessment of ai systems. Eval. Program Plann. **99**, 102294 (2023)
51. Hertwig, R., Herzog, S. M., Kozyreva, A.: Blinding to circumvent human biases: deliberate ignorance in humans, institutions, and machines. Perspectives on Psychological Science pp. 17456916231188052 (2022)
52. Hidayat, T.H.J., Ruldeviyani, Y., Aditama, A.R., et al.: Sentiment analysis of twitter data related to Rinca island development using doc2vec and SVM and logistic regression as classifier. Procedia Comput. Sci. **197**, 660–667 (2022)
53. Hitchings, J.: Deficiencies of the traditional approach to information security and the requirements for a new methodology. Comput. Secur. **14**(5), 377–383 (1995)
54. Hooshmand, M.K., Hosahalli, D.: Network anomaly detection using deep learning techniques. CAAI Trans. Intell. Technol. **7**(2), 228–243 (2022)
55. Huang, L.: Ethics of artificial intelligence in education: student privacy and data protection. Sci. Insights Educ. Front. **16**(2), 2577–2587 (2023)
56. Isaak, J., Hanna, M.J.: User data privacy: Facebook, Cambridge analytica, and privacy protection. Computer **51**(8), 56–59 (2018)
57. Jabeur, S.B., Gharib, C., Mefteh-Wali, S., et al.: Catboost model and artificial intelligence techniques for corporate failure prediction. Technol. Forecast. Soc. Chang. **166**, 120658 (2021)
58. Jiang, P., Sonne, C., Li, W., et al.: Preventing the immense increase in the life-cycle energy and carbon footprints of llm-powered intelligent chatbots. Engineering (2024)
59. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**(9), 389–399 (2019)
60. Kaaniche, N., Laurent, M., Belguith, S.: Privacy enhancing technologies for solving the privacy-personalization paradox: taxonomy and survey. J. Netw. Comput. Appl. **171**, 102807 (2020)
61. Kalla, D., Kuraku, S.: Advantages, disadvantages and risks associated with chatgpt and AI on cybersecurity. J. Emerg. Technol. Innov. Res. **10**(10), (2023)
62. Kan, X., Fan, Y., Zheng, J., et al.: Data adjusting strategy and optimized xgboost algorithm for novel insider threat detection model. J. Franklin Inst. **360**(16), 11414–11443 (2023)
63. Kaur, R., Gabrijelčič, D., Klobučar, T.: Artificial intelligence for cybersecurity: literature review and future research directions. Inf. Fusion p 101804 (2023)
64. Ke, G., Meng, Q., Finley, T., et al.: Lightgbm: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. **30**, (2017)
65. Kheya, T. A., Bouadjenek, M. R., Aryal, S.: The pursuit of fairness in artificial intelligence models: a survey. (2024). arXiv preprint arXiv:2403.17333
66. Krčál, M., Švec, O., Bálek, M., et al.: Deep convolutional malware classifiers can learn from raw executables and labels only (2018)
67. Kumar, D., Suthar, N.: Ethical and legal challenges of AI in marketing: an exploration of solutions. J. Inf. Commun. Ethics Soc. (2024)
68. Kumari, P., Jain, A.K.: A comprehensive study of ddos attacks over IoT network and their countermeasures. Comput. Secur. 103096 (2023)
69. Landers, R.N., Behrend, T.S.: Auditing the AI auditors: a framework for evaluating fairness and bias in high stakes AI predictive models. Am. Psychol. **78**(1), 36 (2023)
70. Le, D.C., Zincir-Heywood, N., Heywood, M.I.: Analyzing data granularity levels for insider threat detection using machine learning. IEEE Trans. Netw. Serv. Manage. **17**(1), 30–44 (2020)
71. Lee, J., Kim, J., Kim, I., et al.: Cyber threat detection based on artificial neural networks using event profiles. IEEE Access **7**, 165607–165626 (2019)
72. Lehmann, D., Kinder, J., Pradel, M.: Everything old is new again: binary security of {WebAssembly}. In: 29th USENIX Security Symposium (USENIX Security 20), pp. 217–234 (2020)
73. Li, H.: Special section introduction: artificial intelligence and advertising. J. Advert. **48**(4), 333–337 (2019)
74. Lu, Q., Zhu, L., Xu, X., et al.: Responsible-AI-by-design: a pattern collection for designing responsible AI systems. IEEE Software (2023)
75. Madaio, M., Blodgett, S. L., Mayfield, E., et al. Beyond fairness: structural (in) justice lenses on AI for education. In: The ethics of artificial intelligence in education. Routledge, pp. 203–239 (2022)
76. Mahdavifar, S., Kadir, A.F.A., Fatemi, R., et al.: Dynamic android malware category classification using semi-supervised deep learning. In: 2020 IEEE Intl Conf on Dependable, pp. 515–522. Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), IEEE (2020)
77. Maosa, H., Ouazzane, K., Ghanem, M.C.: A hierarchical security event correlation model for real-time threat detection and response. Network **4**(1), 68–90 (2024)
78. Mashaly, B., Selim, S., Yousef, A. H., et al. Privacy by design: a microservices-based software architecture approach. In: 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), IEEE, pp. 357–364 (2022)
79. Memarian, B., Doleck, T.: Fairness, accountability, transparency, and ethics (fate) in artificial intelligence (AI), and higher education: a systematic review. Comput. Educ.: Artif. Intell. p 100152 (2023)
80. Miloslavskaya, N., Nikiforov, A., Budzko, V.: Standardization of ensuring information security for big data technologies. In: 2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), IEEE, pp. 56–63 (2018)
81. Mohapatra, N., Satapathy, B., Mohapatra, B., et al. Malware detection using artificial intelligence. In: 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, pp. 1–6 (2022)
82. Möller, D.P.: Intrusion detection and prevention. In: Guide to Cybersecurity in Digital Transformation: Trends, Methods, Technologies, Applications and Best Practices, pp. 131–179. Springer (2023)
83. Morales, E. F., Escalante, H. J.: A brief introduction to supervised, unsupervised, and reinforcement learning. In: Biosignal Processing and Classification Using Computational Learning and Intelligence. Elsevier, pp. 111–129 (2022)

84. Moskalenko, V., Kharchenko, V., Moskalenko, A., et al.: Resilience and resilient systems of artificial intelligence: taxonomy, models and methods. Algorithms **16**(3), 165 (2023)
85. Namanya, A. P., Cullen, A., Awan, I., et al. The world of malware: an overview. In: 2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud), IEEE, pp. 420–427 (2018)
86. Nguyen, A., Ngo, H.N., Hong, Y., et al.: Ethical principles for artificial intelligence in education. Educ. Inf. Technol. **28**(4), 4221–4241 (2023)
87. O'herrin, J.K., Fost, N., Kudsk, K.A.: Health insurance portability accountability act (hipaa) regulations: effect on medical record research. Ann. Surg. **239**(6), 772 (2004)
88. Olorunfemi, O.L., Amoo, O.O., Atadoga, A., et al.: Towards a conceptual framework for ethical AI development in it systems. Comput. Sci. IT Res. J. **5**(3), 616–627 (2024)
89. Ometov, A., Molua, O.L., Komarov, M., et al.: A survey of security in cloud, edge, and fog computing. Sensors **22**(3), 927 (2022)
90. Or-Meir, O., Nissim, N., Elovici, Y., et al.: Dynamic malware analysis in the modern era-a state of the art survey. ACM Comput. Surv. (CSUR) **52**(5), 1–48 (2019)
91. Oseni, A., Moustafa, N., Janicke, H., et al. Security and privacy for artificial intelligence: opportunities and challenges. (2021). arXiv preprint arXiv:2102.04661
92. PK, F. A.: What is artificial intelligence? Success is no accident It is hard work, perseverance, learning, studying, sacrifice and most of all, love of what you are doing or learning to do pp. 65 (1984)
93. Pashentsev, E., Bazarkina, D.: Malicious use of artificial intelligence and threats to psychological security in Latin America: common problems, current practice and prospects. In: The Palgrave Handbook of Malicious Use of AI and Psychological Security. Springer, pp. 531–560 (2023)
94. Peters, U.: Algorithmic political bias in artificial intelligence systems. Philos. Technol. **35**(2), 25 (2022)
95. Qin, X., Jiang, F., Cen, M., et al. Hybrid cyber defense strategies using honey-x: a survey. Comput. Netw. 109776 (2023)
96. Qiu, S., Liu, Q., Zhou, S., et al.: Review of artificial intelligence adversarial attack and defense technologies. Appl. Sci. **9**(5), 909 (2019)
97. Radclyffe, C., Ribeiro, M., Wortham, R.H.: The assessment list for trustworthy artificial intelligence: a review and recommendations. Front. Artif. Intell. **6**, 1020592 (2023)
98. Raff, E., Barker, J., Sylvester, J., et al. Malware detection by eating a whole exe. (2017). arXiv:1710.09435
99. Rajeshwari, S., Praveenadevi, D., Revathy, S., et al. 15 utilizing AI technologies to enhance e-commerce business operations. Toward Artificial General Intelligence: Deep Learning, Neural Networks, Generative AI pp. 309 (2023)
100. Rozado, D.: Danger in the machine: the perils of political and demographic biases embedded in AI systems. Manhattan Institute (2023)
101. Sajja, G.S., Mustafa, M., Ponnusamy, R., et al.: Machine learning algorithms in intrusion detection and classification. Ann. Romanian Soc. Cell Biol. **25**(6), 12211–12219 (2021)
102. Samardzic, N., Feldmann, A., Krastev, A., et al. Craterlake: a hardware accelerator for efficient unbounded computation on encrypted data. In: Proceedings of the 49th Annual International Symposium on Computer Architecture, pp. 173–187 (2022)
103. Sargeant, H.: Algorithmic decision-making in financial services: economic and normative outcomes in consumer credit. AI Ethics **3**(4), 1295–1311 (2023)
104. Sarker, I. H.: Multi-aspects AI-based modeling and adversarial learning for cybersecurity intelligence and robustness: a comprehensive overview. Secur. Privacy p e295 (2023)
105. Schmitt, M.: Securing the digital world: protecting smart infrastructures and digital industries with artificial intelligence (AI)-enabled malware and intrusion detection. J. Ind. Inf. Integr. **36**, 100520 (2023)
106. Schwartz, R., Vassilev, A., Greene, K., et al. Towards a standard for identifying and managing bias in artificial intelligence. NIST special publication 1270(10.6028), (2022)
107. Sentuna, A., Alsadoon, A., Prasad, P., et al.: A novel enhanced naïve bayes posterior probability (enbpp) using machine learning: cyber threat analysis. Neural Process. Lett. **53**, 177–209 (2021)
108. Shen, L.: The nist cybersecurity framework: overview and potential impacts. Scitech Lawyer **10**(4), 16 (2014)
109. Shen, Z., Deifalla, A.F., Kamiński, P., et al.: Compressive strength evaluation of ultra-high-strength concrete by machine learning. Materials **15**(10), 3523 (2022)
110. Shetty, S. H., Shetty, S., Singh, C., et al. Supervised machine learning: algorithms and applications. Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools and Applications pp. 1–16 (2022)
111. Shiau, W.L., Wang, X., Zheng, F.: What are the trend and core knowledge of information security? A citation and co-citation analysis. Inf. Manage. **60**(3), 103774 (2023)
112. Shin, D.: User perceptions of algorithmic decisions in the personalized AI system: perceptual evaluation of fairness, accountability, transparency, and explainability. J. Broadcast. Electron. Media **64**(4), 541–565 (2020)
113. Siau, K., Wang, W.: Artificial intelligence (AI) ethics: ethics of AI and ethical AI. J. Database Manage. (JDM) **31**(2), 74–87 (2020)
114. Sontan, A.D., Samuel, S.V.: The intersection of artificial intelligence and cybersecurity: challenges and opportunities. World J. Adv. Res. Rev. **21**(2), 1720–1736 (2024)
115. Soori, M., Arezoo, B., Dastres, R.: Artificial intelligence, machine learning and deep learning in advanced robotics, a review. Cogn. Robot. (2023)
116. Srivastava, K., Shekokar, N.: Design of machine learning and rule based access control system with respect to adaptability and genuineness of the requester. EAI Endorsed Trans. Pervasive Health Technol. **6**(24), e1–e1 (2020)
117. Stahl, B. C., Eden, G., Jirotka, M.: Responsible research and innovation in information and communication technology: identifying and engaging with the ethical implications of icts. Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society pp. 199–218 (2013)
118. Stahl, B.C., Brooks, L., Hatzakis, T., et al.: Exploring ethics and human rights in artificial intelligence-a Delphi study. Technol. Forecast. Soc. Chang. **191**, 122502 (2023)
119. Steingartner, W., Galinec, D., Kozina, A.: Threat defense: cyber deception approach and education for resilience in hybrid threats model. Symmetry **13**(4), 597 (2021)
120. Strizzi, J.M., Di Nucci, E.: Ethical and human rights concerns of sexual orientation change efforts: commentary on Sullins (2022). Arch. Sex. Behav. **52**(3), 865–867 (2023)
121. Sultani, W., Chen, C., Shah, M.: Real-world anomaly detection in surveillance videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6479–6488 (2018)
122. Taherdoost, H.: Cybersecurity vs. information security. Procedia Comput. Sci. **215**, 483–487 (2022)
123. Tavani, H. T.: Expanding the standard ict-ethics framework in an era of AI. J. Inf. Ethics **29**(2), (2020)

124. Tóth, Z., Blut, M.: Ethical compass: the need for corporate digital responsibility in the use of artificial intelligence in financial services. Organ. Dyn. p 101041, (2024)
125. Tripathi, S., Gupta, M.: A holistic model for global industry 4.0 readiness assessment. Benchmarking: Int. J. **28**(10), 3006–3039 (2021)
126. Ueda, D., Kakinuma, T., Fujita, S., et al.: Fairness of artificial intelligence in healthcare: review and recommendations. Jpn. J. Radiol. 1–13 (2023)
127. Ullah, W., Ullah, A., Hussain, T., et al.: Artificial intelligence of things-assisted two-stream neural network for anomaly detection in surveillance big video data. Futur. Gener. Comput. Syst. **129**, 286–297 (2022)
128. Uprety, A., Rawat, D.B.: Reinforcement learning for IoT security: a comprehensive survey. IEEE Internet Things J. **8**(11), 8693–8706 (2020)
129. Urooj, B., Shah, M.A., Maple, C., et al.: Malware detection: a framework for reverse engineered android applications through machine learning algorithms. IEEE Access **10**, 89031–89050 (2022)
130. Uszko, K., Kasprzyk, M., Natkaniec, M., et al.: Rule-based system with machine learning support for detecting anomalies in 5g wlans. Electronics **12**(11), 2355 (2023)
131. Varona, D., Suárez, J.L.: Discrimination, bias, fairness, and trustworthy AI. Appl. Sci. **12**(12), 5826 (2022)
132. Vinayakumar, R., Alazab, M., Soman, K., et al.: Robust intelligent malware detection using deep learning. IEEE Access **7**, 46717–46738 (2019)
133. Wang, P.: On defining artificial intelligence. J. Artif. General Intell. **10**(2), 1–37 (2019)
134. Wang, S., Pei, Q., Zhang, Y., et al.: A hybrid cyber defense mechanism to mitigate the persistent scan and foothold attack. Secur. Commun. Netw. **2020**, 1–15 (2020)
135. Wright, D., Mordini, E.: Privacy and ethical impact assessment. In: Privacy Impact Assessment. Springer, pp. 397–418 (2012)
136. Wu, Z., Zhang, H., Wang, P., et al.: Rtids: a robust transformer-based approach for intrusion detection system. IEEE Access **10**, 64375–64387 (2022)
137. Xia, X., Pan, X., Li, N., et al.: Gan-based anomaly detection: a review. Neurocomputing **493**, 497–535 (2022)
138. Xu, M., Yin, W., Cai, D., et al. A survey of resource-efficient llm and multimodal foundation models. (2024) arXiv preprint arXiv: 2401.08092
139. Xu, H., Sun, Z., Cao, Y., et al.: A data-driven approach for intrusion and anomaly detection using automated machine learning for the internet of things. Soft. Comput. **27**(19), 14469–14481 (2023)
140. Yamin, M. M., Hashmi, E., Ullah, M., et al. Applications of llms for generating cyber security exercise scenarios (2024)
141. Yenduri, G., Ramalingam, M., Selvi, G. C., et al. Gpt (generative pre-trained transformer)–a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. IEEE Access (2024)
142. Yuxin, D., Siyi, Z.: Malware detection based on deep learning algorithm. Neural Comput. Appl. **31**, 461–472 (2019)
143. Zhang, P., Kang, Z., Yang, T., et al. Lgd: label-guided self-distillation for object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 3309–3317 (2022)