International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| 9th ICAILLMBA (ALLAM) 2025-Conference Proceedings
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.ICAILLMBA-120

Original Article

Impact of Artificial Intelligence in various phases of Cyber Security: A Comprehensive Survey

Sadhana kodali[1],Pradeepini Gera[2],RaviSankar Malladi[3]
[1,2,3]Department of CSE, Koneru Lakshmaiah Education Foundation,Vaddeswaram,Guntur Dt.

Abstract - *Artificial Intelligence which is a powerful technology has a great impact in many of the cybersecurity tasks like detection of new attacks, predictive intelligence, threat detection, AI enabled cyber defense etc. AI enhanced detections are quite useful to have high accuracy in threat detection and response which strengthens the cybersecurity teams to automate the security issues & attacks and to accelerate the process of detection. Identification of threats accurately is crucial especially if the threat is AI generated, where the attackers can automate and create sophisticated malware, it can be a complicated task. This study offers a comprehensive review of various AI methods useful in different phases of Cyber Security and an insight into AI enabled Cyber-attacks. The AI methods and their functionality is compared to learn how well they perform in mitigating the risks of cyber security at various phases. The study provides clear understanding of the impact of AI on cybersecurity in automation, reinforced cyber defense and intelligence in threat detection. This study emphasizes the use of AI tools for employee training used in threat detection that mitigate the risks of cyber-attacks. In all the phases of Cyber Security we need continuous monitoring and frequent vigilance for breaches and threats which require the involvement of AI tools. A thorough understanding of the life cycle of Cyber Security and various AI tools involved in each phase is mentioned in this work as a comprehensive study.*

Keywords - *Cyber Security, Artificial Intelligence, Detection, Response, Recovery, Cyber-Attacks.*

1. Introduction

Today's trend for Cyber Security involves the usage of AI and ML algorithms. Most of the trends in threat detection and prevention are focusing on the leveraging AI techniques. The importance of AI security trends increases when coming to IOT based devices security and prominence of the Zero Trust model. The analysis of large datasets and various malicious patterns would be complex with the traditional and manual techniques, when compared to the usage of AI based threat detection. Due to the growing number of devices on the internet the traditional security measures to protect the devices from these vulnerabilities is highly hectic and requires sophisticated AI algorithms to protect the devices from cyber-attacks. The traditional approaches for Cyber Security, which includes the static rule based, signature based detection for identifying and mitigating threats may have certain limitations to detect sophisticated threats and attacks.

The National Institute of Standards and Technology (NIST) in the United States introduced the Cyber security Framework (CSF) in 2014. This framework was designed to guide organizations in identifying, preventing, and responding to cyber security threats and vulnerabilities. Since its initial release, the CSF has undergone several revisions. In 2018 Version 1.1 was released and was originally named Framework for Improving Critical Infrastructure Cyber security. The most recent update, Version 2.0, was published in 2024 and rebranded simply as the Cyber security Framework. The CSF serves as a foundational reference for organizations aiming to implement cyber security best practices and align with international standards.

This survey is conducted through a structured and systematic review of existing literature, standards, and industrial practices related to the application of Artificial Intelligence across various phases of the cybersecurity lifecycle. The study primarily aligns its analysis with the NIST Cybersecurity Framework (CSF) 2.0, using its core functions Govern, Identify, Protect, Detect, Respond, and Recover as the foundational structure for categorizing AI-driven cybersecurity techniques. Peer-reviewed journal articles, conference papers, white papers, and authoritative reports published by organizations such as NIST, CISA, MITRE, and Gartner were reviewed to understand both defensive and offensive uses of AI. Particular attention was given to recent developments in machine learning, deep learning, and generative AI that influence threat detection, asset management, risk analysis, and cyber-attack automation.

The survey further incorporates a comparative and analytical approach by mapping AI techniques to specific cybersecurity phases and evaluating their effectiveness, advantages, and limitations. Real-world use cases, tools, and frameworks such as AI-driven asset discovery systems, SIEM platforms, OSINT reconnaissance tools, and adversarial AI detection mechanisms are analyzed to understand practical deployments. Additionally, emerging risks introduced by Generative AI are examined using the NIST AI Risk Management Framework (AI RMF) and ARIA evaluation models to highlight governance, ethical, and operational challenges. By synthesizing insights from academic research, industry solutions, and government frameworks, this survey provides a holistic view of how AI strengthens cybersecurity defenses while simultaneously introducing new threat

vectors that require continuous monitoring and adaptive risk management. Section 2 focuses on the Overview, and other sections cover various phases.

2. Overview of Cyber Security Framework

The Cybersecurity Framework has the following components[2]:

- CSF Core: Comprises a hierarchy of functions, along with the categories and sub categories which detail high-level cybersecurity outcomes as taxonomy.
- CSF Organizational Profile: This profile describes an organizations target cybersecurity posture with respect to CSF Core functions.
- CSF Tiers:CSF Tiers are used to characterise the accuracy of cyber security risk governance and management practices. This Chapter focuses on the impact of AI in various phases of cybersecurity especially corresponding to CSF Core. The Cybersecurity Framework Core is a set of cybersecurity outcomes arranged as functions, followed by category and sub category .NIST CSF is built with respect to the following core functions: Govern, Identify, Protect, Detect,Respond,Recover.

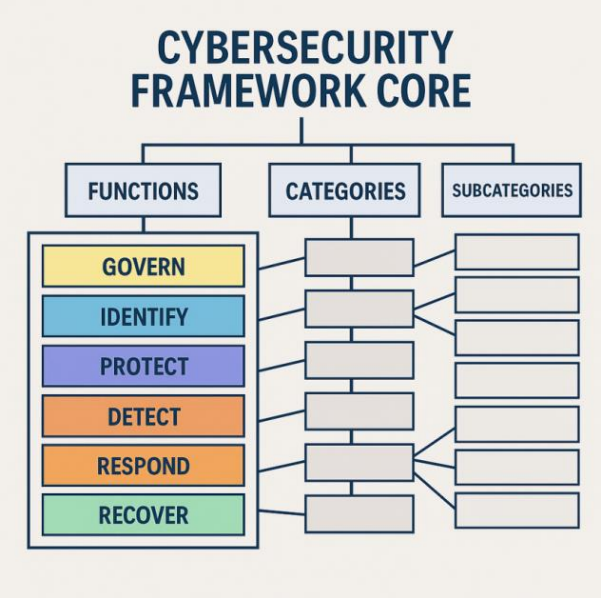


Fig 1: Cybersecurity Core and Functions

The CSF Core and the functions are represented in Figure 1 where CSF Core is categorised as Functions, Categories and Subcategories. The Functions of Cybersecurity can be observed as the backbone from which the other cybersecurity components are built around[3].Any cybersecurity risk management decisions can be built around these pillars which are the functions of CSF and these functions also correspond to the phases cybersecurity life cycle which are represented in Figure 2.

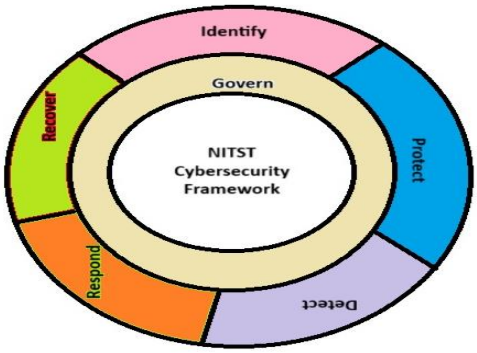


Fig 2: CSF Functions Which Also Correspond the Phases of Cybersecurity Life Cyle.0

2.1. Impact of AI on various phases of cyber security life cycle

The Govern phase which is a part of the NIST CSF which is responsible for developing cyber security strategies that align with the business goals. More elaborately the organizational structures and roles are clearly defined, risk management, continuous improvement and monitoring cyber security policies.

Fig 3: Core Phases of Cyber Security Life Cycle

3. Govern phase

3.1. AI for risk management and analysis

AI based tools can effectively detect risks as sophisticated algorithms are used and can be unbiased. While the human based risk management can be biased and depend on personal beliefs. Using AI tools for risk management is advantageous for processing large amounts of data for detecting, evaluating and minimising risks.

NIST AI Risk management framework(RMF) has emerged due to the increasing complexities and the framework has the following functions: Govern,Map, Measure and Manage. This framework was initiated in 2021 and released in January 2023 [4].NIST is integrating AI into cyber security which reflects the significance and necessity of AI in all phases of the cyber security lifecycle. The NIST AI RMF mainly focuses on risks involved with AI development and AI deployment.

NIST devises an evaluation program for Assessing Risks and Impacts of AI (ARIA) [5] for providing trustworthy and safe AI approaches. ARIA addresses the gap and evaluates between the AI functionality and the real world, provides essential information if AI systems are valid, secure, safe, reliable once they are deployed. ARIA also helps the individuals and society to understand the impact of AI .ARIA has initial evaluation known as 0.1 Pilot evaluation plans, and three kinds of evaluations[6]:model testing, red-teaming and field testing.

- Model Testing: Used to confirm claimed model capabilities.
- Red-teaming: This approach is used to understand the occurrence of risks, under what conditions do violative outcomes encounter. It is kind of stress testing which are used to explore risk boundaries.
- Field testing :Used to potential positive and negative impacts of applications.

The ARIA evaluation environment has three layers which provides the possible AI risks which may have both positive and negative impacts , for whom does the given risk create impact and why does the risk created. The evaluation environment has three layers namely testing layer, annotation layer and measurement layer as represented in Figure 4.The testing layer has test scenarios and hosts user interactions with the submitted applications. There are three kinds of test scenarios namely Model testing, Red-Teaming and Field testing. The ARIA experimentation environment is represented in Figure 5 where the new risk measurements, methods, metrics are obtained by simulating the real world conditions[7].ARIA builds a systematic as well as repeatable risk management methods.

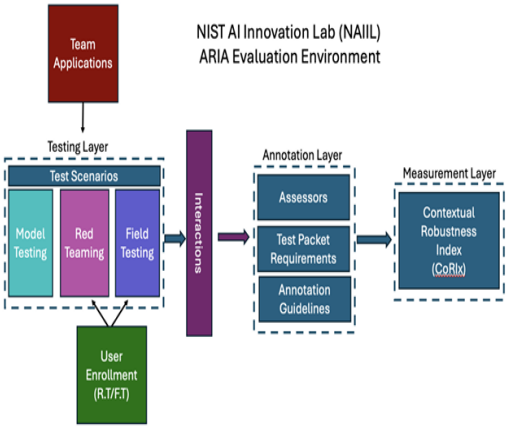


Fig 4: NIST ARIA Evaluation Environment

Source:https://aichallenges.nist.gov/aria/docs/evaluationplan.pdf

AI risks may vary at various stages of AI lifecycle while the Generative AI (GAI) risks can aggravate AI risks [8].The stages of AI life cycle include design, development, deployment, operation and /or decommissioning. Further risks are categorised in [9] as Technical / Model risks , Misue by humans , ecosystem / societal risks which are tabulated including the specific risks under each category in Table 1.While few GAI risks are complex, unknown and are exacerbated due to lack of visibility GAI training data.

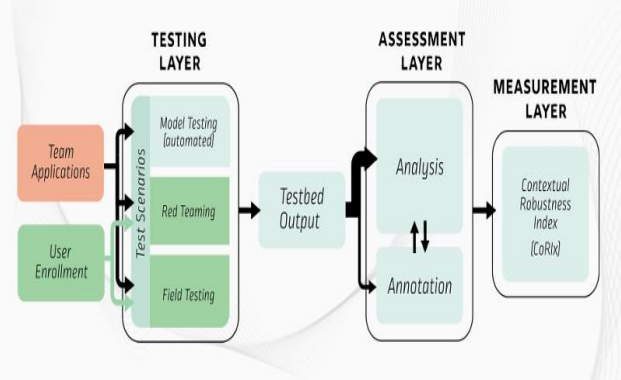


Fig 5: Depicts Aria Experimentation Environment by Simulating Real World Conditions and Risk Assessment
Source:https://aichallenges.nist.gov/aria/docs/ARIA_Program_Overview.pdf

AI risks may vary at various stages of AI lifecycle while the Generative AI (GAI) risks can aggravate AI risks [8].The stages of AI life cycle include design, development, deployment,operation and /or decommissioning. Further risks are categorised in [9] as Technical / Model risks , Misue by humans , ecosystem / societal risks which are tabulated including the specific risks under each category in Table 1.While few GAI risks are complex, unknown and are exacerbated due to lack of visibility GAI training data. AI risks may vary at various stages of AI lifecycle while the Generative AI (GAI) risks can aggravate AI risks [8].The stages of AI life cycle include design, development, deployment, operation and /or decommissioning. Further risks are categorised in [10] as Technical / Model risks , Misue by humans , ecosystem / societal risks which are tabulated including the specific risks under each category in Table 1.While few GAI risks are complex, unknown and are exacerbated due to lack of visibility GAI training data.

- Confabulation: Generating a false or erroneous content which can mislead the users.
- Dangerous recommendations: Generating violent, dangerous or hateful content or recommendations.
- Data Privacy: Leakage or disclosure of sensitive information.
- Value chain and Component Integration: Integration of data from third party components which may be improperly processed ,cleaned or other issues that reduce data transparency for downstream users.
- Harmful Bias or Homogenization: Amplification of improper or incorrect presumptions related to history, society, or between languages and sub groups which leads to undesired homogeneity which skews the system
- Chemical, Biological, Radiological, Nuclear (CBRN) information :Ease of access to sensitive data like design capabilities, dangerous weapons, materials or agents.
- Human-AI Configuration: Interactions of humans with AI systems which may lead to anthropomorphising or developing algorithm aversion, emotional tanglement or over relying GAI systems
- Abusive , obscene ,degrading Content :Ease of production of obscene or abusive images which may cause harm.
- Information Integrity: Low-level barrier for generating the content which may not distinguish un-certainties and misinformation.
- Information Security: Low-grade barriers in lieu of offensive cyber capabilities, which may compromise the availability, confidentiality and availability training data, code etc.
- Environmental Impacts: Impacts of high resource utilization of GAI models which impact ecosystems.
- Intellectual Property: Eased production of licensed, trademarked and copyrighted content ,un-authorised exposure of trade secrets. Controversial to risks, the trustworthy AI characteristics for every risk is tabulated in Table 2.These characteristics of AI are to be well noted when AI applications are developed especially for the use of cyber security.

3.2. Asset Management

Asset management tools are used for the classifying, managing, monitoring and identifying various IT assets like hardware, data, software and systems for ensuring compliance and security. The traditional approaches of asset management focus on the inventory and cost but the AI based asset management tools focus on the cyber security risk. A novel method for automatically identifying cyber assets on a network using behavioral data analysis rather than relying on static inventory methods is proposed in [11]. It introduces a framework that captures network traffic, applies machine learning techniques (like clustering and classification), and identifies unknown or rogue devices in real-time. A network behavior-based asset detection method is presented in this framework. Both the known and unknown assets can fit this framework. Using AI can reduce

human effort in asset discovery and inventorying. However modern cyber security asset management leverages behavioral analytics to detect unmanaged devices or shadow IT, improving network hygiene.AI driven system with continuous learning is proved to be more efficient by adjusting the inventory and risk management [12].The continuous learning approach makes it adaptive and proactive in response to real time threat detection.

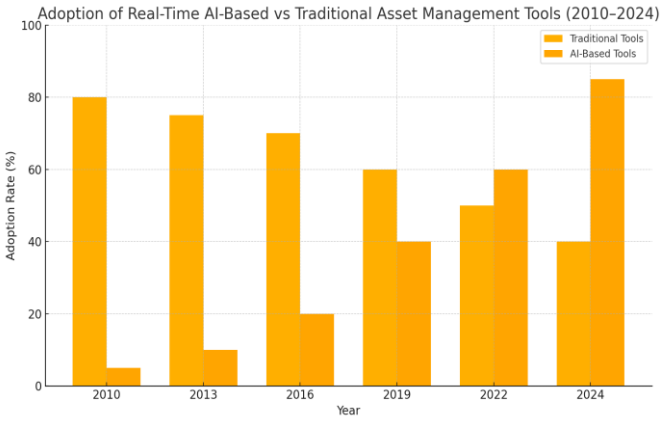


Fig 6: Comparison of Traditional Asset Management Tools vs. AI Based Asset Management Tools

The bar graph in Figure 6 illustrates the adoption trend of traditional asset management tools versus AI-based real-time systems over the past 15 years. It is evident that AI-based asset management systems have seen a steep increase in adoption, particularly after 2015, while the use of traditional tools has declined steadily. Some widely used AI based Asset Management tools include Axonius, Tanium, Qualys Global IT Asset Inventory, Service Now ITOM with Security Operations,Rapid7 InsightVM, Cisco Cyber Vision.

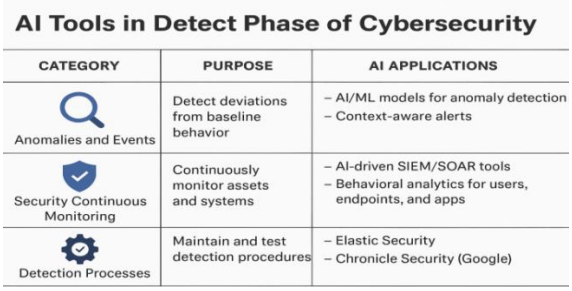
4. Protect Phase

The protect phase in cyber security lifecycle is very essential step which establishing security controls to minimise and protect any damage from cyber threats. The categories under this vital function include Access Control, Data Security, Awareness and Training, information Protection processes and procedures and Protective technology.[13] The use of AI tools for access control can be used to enhance security, user experience and efficiency. Embedding Machine Learning algorithms and AI tools will improve accuracy and provide a better analysis for the user behaviour and potential threats. Facial recognition is a way for physical access control and many AI driven automation tasks are used for access management[14]t.AI has the potential to analyse historic data, compromised systems and predict the attack patterns using proactive hunt for threats. Improved access control techniques like the facial recognition systems, biometric authentication, personalised access controls use AI algorithms for a better performance.

Data Security AI is used for managing information protection with policies and procedures. Many AI applications are used for real-time content inspection and classification.AI leverages intelligent encryption management. Awareness and Training Using AI personalised training modules are evolving with the use of NLP. AI phishing simulations can be used to identify Phishing attacks.

5. Detect Phase

AI provides real time threat intelligence in this phase by anomaly detection and behavioural analytics with improved accuracy. During Anomaly detection Contextual AI helps reduce false positives by combining threat intelligence with system behaviour[15]. In Security Continuous monitoring category AI improves the process by improved automating log analysis and even correlation. Additionally Machine Learning algorithms are employed to trace threats across huge datasets in real time.[16]



AI Tools in Detect Phase of Cybersecurity

CATEGORY	PURPOSE	AI APPLICATIONS
Anomalies and Events	Detect deviations from baseline behavior	– AI/ML models for anomaly detection – Context-aware alerts
Security Continuous Monitoring	Continuously monitor assets and systems	– AI-driven SIEM/SOAR tools – Behavioral analytics for users, endpoints, and apps
Detection Processes	Maintain and test detection procedures	– Elastic Security – Chronicle Security (Google)

Fig 7: Categories in Detect Phase and AI Tools Used

6. Recover phase

Using in AI enhaces the effectiveness by improving the speed by automating the recovery actions which include restoring backups, re-enabling services. Improves precision in AI models which prioritize recovery tasks based on asset criticality and past impact analysis. Using AI can increase the **insight generation by** highlighting recurring weaknesses and suggests improvements to business continuity plans. Using Natural Language Processing (NLP) transforms technical recovery steps into actionable insights for non-technical stakeholders.

7. Impact of AI in Cyberattack life cycle

The cyber-attack life-cycle also has various phases like reconnaissance, weaponization, Delivery, exploitation, installation, command and control, these phases are also known as cyber kill chain which are used by the attacker to make a progress at various stages is represented in Figure 8. Using AI driven Open Source Intelligence(OSINT) can transform the cyber intelligence and threat detection by the ethical hackers or at the time of penetration testing and help the cyber security professionals to thoroughly understand the data vulnerable to attacks, leaked credentials, identify the threat actors more efficiently. Summarizing the AI based OSINT and reconnaissance tools in [17] gives the importance of AI algorithms to protect data and resources from cyber-attacks.

Fig 8: Cyber-Attack life Cycle or Cyber Kill Chain Represented with Various Phases

8. AI in Reconnaisance phase

Reconnaisance gathers information about the target which happens before attacking a target or even beneficial while performing a penetration test.AI powered reconnaissance tools can benefit an ethical hacker or a cyber-security expert to have open source intelligence on attack surfaces more efficiently.AI based OSINT helps is gathering and analysis of vast amounts of data from various real world sources like websites, social media platforms, discussion forums ,Government databases, publications ,Public IOT devices and sensor data.

Reconnaisance can be named Passive Reconnaisance if data gathered is OSINT based and data gathering is done without direct interaction of the target. Whereas Active reconnaissance is probing the target network, domain or infrastructure. The benefit of using AI is that patterns can be detected, data collection can be automated all with high accuracy making the tasks suitable for actionable insights.

Reconnaisance trends kept increasing the past 5 years and are accelerating. The year 2023 shows a significant increase. Botnets are helping the attackers which leaves the attack undetected. Using the National alerts US CISA ,some statistics and use cases the graph is represented in Figure 9.

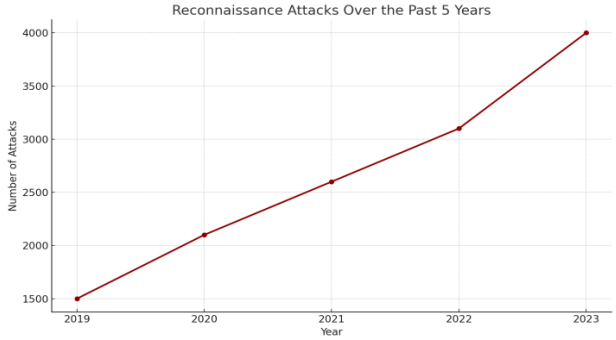


Fig 9: Reconnaisance Attacks over the Past Five Years (2019-2023)

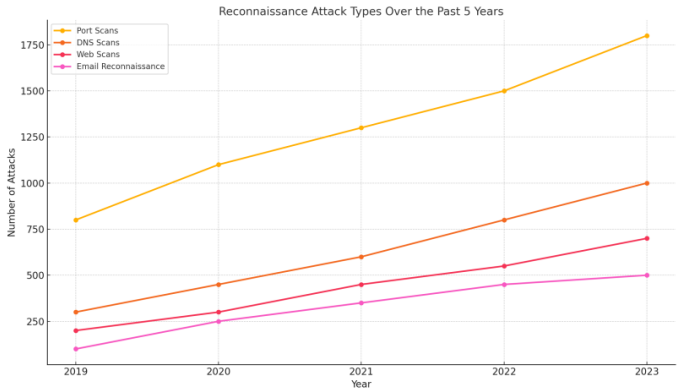


Fig 10: Reconnaisance Attack Types Based On Port Scans, Dns Scans, Web Scans and E-Mail Reconnaissance

Benefits of AI enabled reconnaissance and OSINT tools are listed as follows:

- AI enables automated data from scraping the web from real time resources in huge amounts.
- AI algorithms can be used to analyse anomalies, patterns and threats.
- AI algorithms can be used for sentiment analysis and social media analysis.
- AI algorithms can scan Dark web forums and market places for identifying any leaked credentials and can track malicious activities.

Table 1: Integrating Reconnaissance Detection across the Cybersecurity Lifecycle

Phase	Impact
Identify	Organizations need to maintain an up-to-date list of inventories of exposed assets and misconfigurations.
Protect	Organizations need to Implemen recon deception (honeypots, tarpits) which is very vital.
Detect	Need to do a Focus shift to detecting scan signatures, abnormal behavior in DNS or HTTP logs.
Respond	Need to correlate and investigate recon events with exploit attempts
Recover	Use recon detection to improve incident postmortems and threat intelligence sharing.

AI enhanced image recognition, video recognition and improvised facial recognition can be useful for deep fake detection. Some well-known AI based OSINT tools and reconnaissance tools and strategies are listed in Table 10 these tools are considered to be most powerful tools for gathering information and securing the environment.

9. Weaponization

Using AI in a malicious or destructive purposes is known as weaponization. This may include using drones or robots which are called autonomous weapons, AI based cyber-attacks, misinformation, Deepfakes, Data poisoning, model manipulation, social engineering automation all these are well known examples of weaponization.

Defensive AI techniques can be implemented to detect and prevent weaponization. Tools used for defensive AI especially for Cyber and Network Defense include Darktrace, Cloudstrike Falcon,Vectra AI. In order to verify Content Integrity, audio/video synthesis and Deepfake identification , Reality defender , Sensiti AI ,Microsoft Video authenticator are used.To identify adversarial AI attacks developers use IBM Adversarial Toolbox is used ,Google's Magica detects potential AI attacks which include manipulation of Files. Recorded Future is a threat intelligence and monitoring tool to identify threat actors, Threat Connect is used for correlation mapping and mapping real-time attacks.

10. Conclusion

The study finds that embedding AI to techniques for threat detection, vulnerability management has a faster and more accurate way of predicting threats when compared to traditional approaches.AI models can detect zero day attacks and can indicate exploitation attacks with high accuracy. The Machine Learning approaches used in Malware analysis can classify malware depending on code instead of known signatures.AI can be used to parse the dark-web and forums by employing Natural Language Processing to perform the required scans.AI can detect and prevent malicious download and filter phishing mails. Using AI can accelerate the detection of threats, can handle large volumes of data, prioritizes various threats, evolves and adapts with threat intelligence. However there are major challenges which need to be addressed when ineffective models are used, poor training data used to learn, increased cost and complexity. To overcome these limitations, robust training of the models and regular audits on the model can address these issues.

References

[1] NIST CSF https://en.wikipedia.org/wiki/NIST_Cybersecurity_Framework

[2] The NIST Cybersecurity Framework (CSF) 2.0. National Institute of Standards and Technology,Gaithersburg,MD,NIST CSWP 29. https://doi.org/10.6028/NIST.CSWP.29
[3] Buczak, A. L., & Guven, E. (2016).*A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection.* IEEE Communications Surveys & Tutorials, 18(2), 1153–1176.
[4] NIST AI RMF https://doi.org/10.6028/NIST.AI.600-1
[5] NIST ARIA https://ai-challenges.nist.gov/aria
[6] NIST ARIA 0.1 Pilot evaluation https://ai-challenges.nist.gov/aria/docs/evaluation_plan.pdf
[7] Sommer, R., & Paxson, V. (2010). *Outside the Closed World: On Using Machine Learning for Network Intrusion Detection.*IEEE Symposium on Security and Privacy.
[8] AI RMF GAI profile https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
[9] AI Risks www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai
[10] Ibrahim, A., Valli, C., McAteer, I. et al. A security review of local government using NIST CSF: a case study., J Supercomput 74, 5171–5186 (2018).
[11] Y. Fan, H. Li, S. Li, and K. Huang, Cyber Asset Identification Based on Network Behavior Analysis,IEEE
[12] J. Liu, L. Yang, and Q. Zhao ,Dynamic Cyber Asset Management Using AI for Threat Detection ,IEEE Transactions on Network and Service Management, 2021
[13] https://www.gartner.com/en/articles/how-ai-is-redefining-cybersecurity
[14] https://www.technologyreview.com/2021/04/22/1023623/ai-cybersecurity/
[15] https://attack.mitre.org
[16] Security Information and Event Management (SIEM) – https://www.gartner.com
[17] OSINT and Reconnaisance https://www.webasha.com/blog/ai-driven-osint-reconnaissance-how-artificial-intelligence-is-transforming-cyber-intelligence-and-threat-detection