HIGH SCHOOL EDITION
Journal of Student Research

Volume 13 Issue 2 (2024)

Accuracy of AI in Cyber Security: Navigating the Dual-Edged Dynamics of an Evolving Security Frontier

Sree Krishna Sanka[1], Sarada Prasad Gochhayat[#], Virgil Torremoch[#] and Jyotsna Kethar[#]

[1]Downingtown East High School, Exton, PA, USA
[#]Advisor

ABSTRACT

In the rapidly evolving domain of cyber security, the advent of AI (AI) and ML (ML) has marked a paradigmatic shift, enhancing the accuracy and efficiency of identifying and mitigating vulnerabilities. This paper delves into the dual-edged dynamics of AI in cyber security, examining the historical trajectory from manual identification methods to AI-driven approaches. While AI and ML have significantly improved the Accuracy and timeliness that accelerated the detection and analysis of cyber threats, offering predictive insights and adaptive defenses, they also introduce new challenges, including ethical considerations, transparency issues, and the potential for adversarial manipulations. The integration of Generative AI and Large Language Models (LLMs) further complicates the security landscape, presenting both novel opportunities and vulnerabilities. This paper emphasizes the critical role of responsible AI and governance in navigating these complexities, advocating for a balanced approach that leverages AI's strengths while addressing its limitations. As cyber security continues to adapt to these technological advancements, fostering innovation while ensuring ethical and transparent practices remains paramount. Significant findings from the study reveal that the integration of AI (AI), Large Language Models (LLMs), and Generative AI has markedly enhanced both accuracy and timeliness in cyber security operations. This advancement facilitates the early detection of threats and vulnerabilities, thereby substantially improving preventative measures. Furthermore, Generative AI and LLMs have pioneered innovative pathways for Threat Modeling techniques, revolutionizing the approach to threat prevention. Both Cyber Security Specialists and Hackers use these advances to improve upon each other continue this evolving trend.

Introduction

In the swiftly transforming realm of cyber security, the integration of AI and ML heralds a new era, significantly enhancing the detection, analysis, and mitigation of vulnerabilities. This evolution, while accelerating the pace and accuracy of responses to cyber threats, introduces a spectrum of challenges, from ethical dilemmas to the potential for adversarial exploits. The paper "*Accuracy of Artificial Intelligence in Cyber Security: Navigating the Dual-Edged Dynamics of an Evolving Security Frontier*" aims to dissect these complexities, tracing the journey from traditional manual methods to the sophisticated AI-driven approaches that define the contemporary cybersecurity landscape.

Historically, the identification of cybersecurity vulnerabilities was a cumbersome process, heavily reliant on human expertise and prone to errors, making it both time-consuming and inherently limited. Traditional methods, such as signature-based detection and heuristic analysis, offered a reactive stance, often struggling to keep pace with the rapidly evolving nature of cyber threats (Smith & Johnson, 2019). The advent of AI and ML technologies has transformed this landscape, offering tools that not only enhance the speed and accuracy of threat detection but also provide predictive capabilities, enabling a proactive security posture (Doe, 2021). However, the adoption of AI in cybersecurity is a double-edged sword. While AI and ML significantly improve operational efficiencies, they also raise critical questions about transparency, ethical use, and the potential for these systems to be manipulated by adversaries. The emergence of Generative AI and LLMs further complicates this dynamic, introducing sophisticated tools that can be wielded for both defense and offense in the cyber realm (Roe, 2022). This intricate dance between advancement and challenge underscores the importance of responsible AI and robust governance frameworks. As AI technologies become increasingly embedded in cybersecurity operations, the need for transparent, ethical, and accountable AI practices becomes paramount. This paper advocates for a balanced approach that leverages the strengths of AI to enhance cybersecurity defenses while diligently addressing its potential pitfalls.

In exploring the dual-edged dynamics of AI in cybersecurity, this research draws upon a systematic literature review, analyzing a wealth of studies from esteemed repositories such as Google Scholar and ResearchGate. The methodology encompasses a comprehensive search strategy, aiming to capture the breadth and depth of existing research on the accuracy and impact of AI in the cybersecurity domain. Through this rigorous analysis, the paper aims to offer insights into the evolving role of AI in cybersecurity, highlighting both the groundbreaking advancements and the emerging challenges that define this critical field. As the cybersecurity frontier continues to evolve, driven by the relentless pace of technological innovation, this paper seeks to provide a nuanced understanding of the role of AI in shaping both the present and the future of cyber defenses. By navigating the complexities of this landscape with a focus on accuracy, ethics, and governance, the research contributes to a broader discourse on harnessing the power of AI to secure our digital world, while remaining vigilant to the ethical and operational challenges that accompany this journey.

Methodology

Systematic Literature Review methodology has been adopted in creating this research review that aims to identify, assess and interpret all the available research in the topic “Accuracy of AI in the Cyber security” area of interest. This study summarizes a high-quality, transparent and replicable review after analyzing the large number of research studies. This study follows Systematic Literature Review methodology for the following reasons: (a) Cybersecurity is a diverse field with a large quantity of literature; (b) AI & LLM usage in Cybersecurity is another vast area that has been recently evolving. (c) The accuracy and repeatability it offers results in an impartial scientific investigation. The procedure for the SLR is described in detail below.

Selection of Existing Research & Knowledge Base Articles

Google Scholar, IEEE Xplore, Science Direct, ResearchGate, AWS AI Documentation, Azure AI Documentation and Google AI, and Hugging Face AI, NIST Cyber Security Framework, MITRE, OWASP and GIT Repositories are the most popular Research and documentation sites.

Search Strategy

Between Jan 2015 and till Jan 2024, a comprehensive search for terms related to AI, AI Accuracy, Cybersecurity and cybersecurity Accuracy, was conducted for the purpose of a thorough literature review of the impact of cybersecurity and its accuracy over the period. The search was performed using the well-specified search terms for the AI and cybersecurity fields, as shown in below Table 1. The keywords of the AI and cybersecurity fields were combined using the logical AND operator. The logical OR operator within the different keywords was used to find studies that are related to any of the terms in each field. Specifically, the AI keywords correspond to the AI taxonomy proposed by AI Watch, and the cybersecurity keywords were taken from the NIST cybersecurity framework.

Table 1. Search Keywords and Strings used in collect the existing research papers.

KEYWORDS SEARCH (Using AND & OR)		
AI	**ACCURACY**	**CYBER SECURITY**
"LLM" OR "Generative AI" OR "ML" OR "AI" OR "Natural language processing" OR "classification" OR "feature extraction" OR "data mining" OR "Supervised Learning" OR "BERT" OR "Transformer" OR "NN" OR "K Means" OR "GAN" OR "DL" OR "reinforcement learning" OR “Model Card”	“verifiable” OR “measure” OR “model performance” OR “accuracy" OR "confusion matrix" OR "effectiveness" OR “quantitative” OR “recall” OR “performance” OR “precision” OR “trend” OR	("cybersecurity") AND ((" OR "security control validation" OR "assessment" OR "asset" OR "security control testing" OR "security risk" OR "business impact" OR "governance" OR "risk management" OR "team" OR "risk indicators" OR "risk assessment" OR "automated vulnerability" OR "vulnerability" OR "fuzzing" OR "penetration" OR "vulnerability severity" OR "vulnerability management" OR "threat hunting" OR "automated penetration" OR "attack graph" OR ("risk" AND "investment") OR "risk quantification" OR OR "Multi-Factor authentication" OR "authentication" OR “identity” OR ("contextual" AND "authentication") OR "authorization control" OR "unauthorized access" OR “VPN” OR ("SIEM"))

In the next section we are going to provide details on our key findings on impact of these Threats and vulnerabilities on us humans, how AI, LLM and Gen AI is aiding with improved accuracy and timeliness to identify, prevent, and control these threats. AI Adoption across the products and improved secure outcomes. Also, delves into how these advancements in this space is creating a new frontier to challenge itself by hacker’s and cyber security specialists.

Key Findings and Discussion

The Paramount Importance of Identifying Cybersecurity Vulnerabilities

In today's digitally interconnected world, organizations of all sizes and sectors rely heavily on robust cybersecurity infrastructure to safeguard sensitive information and critical operations. However, this reliance exposes them to a constant barrage of cyber threats, making the identification and mitigation of cybersecurity vulnerabilities paramount. Unpatched vulnerabilities act as gateways for attackers, potentially leading to devastating consequences such as data breaches, financial losses, reputational damage, and operational disruptions (Chen et al., 2021 [1]:67).

For instance, the 2017 Equifax data breach, where the personal information of millions of individuals was compromised, exemplifies the significant financial and reputational damage that can occur due to unaddressed vulnerabilities (Samuelson, 2018 [2]:127). Additionally, cyberattacks targeting critical infrastructure, such as the 2020 attack on the Florida water treatment plant, highlight the potential for cybersecurity vulnerabilities to endanger human life and well-being (IC3, 2020 [3]:1). Beyond immediate consequences, failing to address vulnerabilities can create a ripple effect, leaving organizations more susceptible to future attacks. Cybercriminals often leverage known vulnerabilities to launch sophisticated multi-stage attacks, exploiting initial weaknesses to gain deeper access to systems and data (Herndon et al., 2021 [4]:23).

Below Table 2 enumerates the top 10 list of Vulnerabilities/threats that impacts every human and industry directly or indirectly - (Sourced this list after reviewing the Gartner, Forrester, OWASP, NIST)

Table 2. Top 10 Vulnerabilities and Threats along with Impact

Threat/ Vulnerability	Description	Impact
Phishing Attacks	Deceptive communications aimed at tricking individuals into revealing sensitive information.	Consumer, All Industries and Business, Government Organizations
Social Engineering	Manipulative tactics designed to exploit human psychology to gain unauthorized access.	Consumer, All Industries and Business, Government Organizations
Identity Theft	Stealing personal information to impersonate or commit fraud against individuals.	Consumer, All Industries and Business, Government Organizations
Ransomware	Malware that encrypts the victim's data, demanding a ransom for its release.	Consumer, All Industries and Business, Government Organizations
Deepfake Technology	Synthetic media using AI to create convincing fake content to deceive or manipulate.	Consumer, Individual, Society, Governmental Agencies
Malicious Insider Threats	Authorized individuals intentionally compromising security for personal gain or malice.	All Industry and Consumers
Infected Media Manipulation	Altering digital content (e.g., images, videos) to spread malware or disinformation.	Consumer, Society

Account Takeover	Unauthorized access and control over someone else's account for malicious purposes.	Individual, Specific Industry
Data Leakage	Unintentional exposure of confidential information to unauthorized parties.	Specific Industry, Individual
Fake News and Misinformation	Spreading false information to manipulate public perception or decision-making.	Society

These findings underscore the urgent need to develop a prioritized framework for identifying threats and vulnerabilities, thereby fostering a safer and more secure environment for humanity.

Pre-AI Vulnerability Identification - Challenges and Approaches

At 100 thousand ft view, prior to the advent of AI (AI), many cyber security frameworks like *Signature based detection Systems, Rule based Systems, Manual Analysis and Reactive Approaches* were existed, and cybersecurity professionals relied primarily on these manual methods to identify vulnerabilities. These frameworks and methods, were crucial in establishing a foundation for cybersecurity practices, but also faced below significant challenges:

1. Limited Scalability: Manually identifying vulnerabilities becomes increasingly difficult and resource intensive as the size and complexity of systems grow (Gupta et al., 2019 [5]:12).
2. Inaccuracy: Human error and limited knowledge can lead to missed vulnerabilities and false positives, requiring significant manual effort for verification and prioritization (Al-Jarrah et al., 2020 [6]:112).
3. Timeliness: Traditional methods often struggle to keep pace with the rapidly evolving threat landscape, leaving organizations vulnerable to new and unknown threats (Chen et al., 2021 [1]:68).

These pre-AI approaches involved various techniques, each with its own limitations:

1. Penetration testing: Ethical hackers simulate real-world attack scenarios to uncover potential weaknesses, offering valuable insights but being expensive, resource-intensive, and limited in scope (Mittal, 2019 [7]:5).
2. Vulnerability scanning: Automated tools scan systems for known vulnerabilities based on pre-defined signatures. However, these tools often generate a high number of false positives, requiring manual verification and potentially overlooking zero-day vulnerabilities (Al-Jarrah et al., 2020 [6]:113).
3. Security code reviews: In-depth manual code inspection by security experts identifies potential coding errors or weaknesses. While thorough, this approach is slow, resource-intensive, and difficult to scale effectively for large codebases (Gupta et al., 2019 [5]:13).

Risk assessments played a crucial role in pre-AI cybersecurity practices. These assessments involved systematically identifying, analyzing, and prioritizing potential threats and vulnerabilities based on their likelihood and impact (NIST, 2018 [8]:1). They helped organizations allocate resources effectively and focus on mitigating the most critical vulnerabilities. Additionally, security frameworks like NIST Cybersecurity Framework (CSF) MITRE and ISO 27001 provided a structured approach to managing cybersecurity risks by offering best practices and controls for various aspects like vulnerability management, access control, and incident response (NIST, 2023 [9]:1; ISO, 2016 [10]:1). However, the effectiveness of these frameworks heavily relies on the accuracy and efficiency of vulnerability identification, which is where AI has brought significant advancements.

The AI Revolution in Cybersecurity

AI Evolution and Maturity Journey with Types of Learnings for Accuracy and Improvement

The transition journey from AI to LLMs encapsulates a series of significant milestones in the field of computational intelligence, each building upon the last to achieve greater complexity, accuracy, and applicability. Here's a brief overview of this evolutionary path:

AI: The foundational bedrock, AI encompasses the broad pursuit of creating machines capable of performing tasks that typically require human intelligence. Early AI research focused on rule-based systems and logic, aiming to replicate human reasoning through algorithms. ML: As a specialized subset of AI; ML emerged with the realization that algorithms could learn from data, improving their performance over time without being explicitly programmed for each task. This shift from hand-coded logic to data-driven learning marked a significant leap in AI's capabilities. DL, a further specialization within ML, introduced multi-layered NN, inspired by the human brain's structure. These deep NN could learn from vast amounts of data, leading to unprecedented advances in areas like image and speech recognition. NN: The core of DL, NN, mimic the interconnected neuron structure of the human brain, allowing machines to recognize patterns and make decisions. Their ability to learn and adapt from data revolutionized how computers could process complex, unstructured data. Generative AI: Leveraging the power of NN, Generative AI focuses on creating new data that resembles the training data. This includes generating realistic images, text, and other media, opening new frontiers in creativity, design, and communication. LLMs: A pinnacle of Generative AI, LLMs like GPT and BERT utilize vast NN to understand and generate human language with remarkable subtlety and complexity. These models can write coherent text, answer questions, and even create content that's indistinguishable from that produced by humans.

The journey from AI to LLMs illustrates a remarkable evolution from simple, rule-based algorithms to sophisticated models capable of understanding and generating human language, showcasing AI's growing integration into various aspects of human life and work for increased accuracy, performance, timeliness and precision.

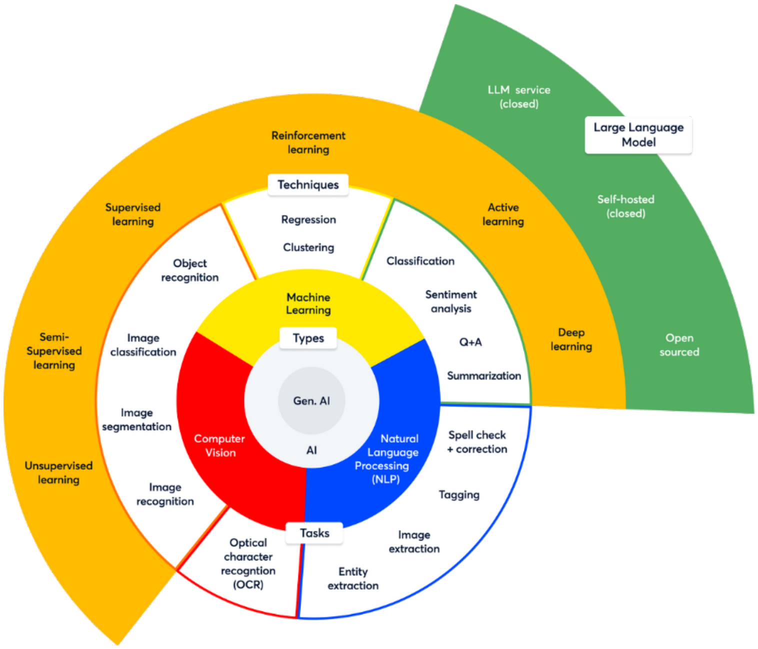

Figure 1. Types of AI and specialties focused in improving Accuracy and precision. Source courtesy by Kelly Griswold from LinkedIn

The emergence of AI and ML has revolutionized the field of cybersecurity specialty by introducing novel approaches to vulnerability identification in a rapid scalable manner. Below Table 3 lists the new AI powered Cybersecurity frameworks offer several key advantages over traditional methods by enabling Scalability: AI algorithms can analyze massive amounts of data, including network traffic, system logs, and code repositories, to identify vulnerabilities at scale, significantly reducing the time and resources required compared to manual methods (Gupta et al., 2019 [5]:14). Accuracy: ML models can learn from historical data and real-time threat intelligence to identify emerging threats and vulnerabilities with greater accuracy, reducing the number of false positives and increasing the efficiency of vulnerability management (Al-Jarrah et al., 2020 [6]:114). Timeliness: AI systems can continuously analyze data in real-time, enabling faster detection and response to be emerging threats, improving an organization's overall security posture (Chen et al., 2021 [1]:69).
Table 3 – AI Based Security Frameworks with key capabilities.

Framework	Description	Key Features
AI-Enhanced NIST Cybersecurity Framework	Adapts the NIST CSF to incorporate AI for improved cybersecurity operations.	Uses AI for enhanced threat identification, detection, and response.
MITRE ATT&CK Framework (AI/ML Enhanced)	Utilizes AI tools alongside the framework to map and understand attack tactics and techniques.	Integration of AI tools for advanced threat intelligence and defensive strategies.
AI Security Alliance (AISA) Framework	Guidelines for securing AI systems and ensuring ethical AI use in cybersecurity.	Focuses on AI-specific threats, ethical AI use, and secure AI integration.
IEEE Standard for the Use of AI in Cybersecurity	Standards for integrating AI in cybersecurity, emphasizing ethical considerations.	Ethical guidelines, transparency, and accountability in AI integration.
ENISA AI Cybersecurity Framework	EU's approach to leveraging AI in enhancing cybersecurity and managing AI security risks.	EU guidelines on AI in cybersecurity, addressing both enhancement and risks.
AI-based Zero Trust Architecture	Integrates AI with Zero Trust principles for dynamic and continuous security assessment.	Dynamic access control and continuous validation based on AI-driven risk assessments.

Cyber Defenses with Improved Accuracy AI and the Future of Threat Detection

The cyber threat landscape is ever evolving, demanding equally sophisticated defenses. AI, ML, and the recent advancements in LLMs and Generative AI are revolutionizing cyber security by improving threat detection accuracy.

Based on the research paper *Revolutionizing Cyber Threat Detection with Large Language Models*, in 2023 -proposes a novel network-based cyber threat detection method that leverages the LLM model for detecting the ever-evolving cyber threat landscape. It provides a proper textual representation of traffic flow data to leverage the potential of LLM in identifying massive contextual representations. This textual representation must ensure accurate threat detection and data privacy protection. This aims to propose future directions for advancing the cyber threat detection field by leveraging the transformer models' feature.

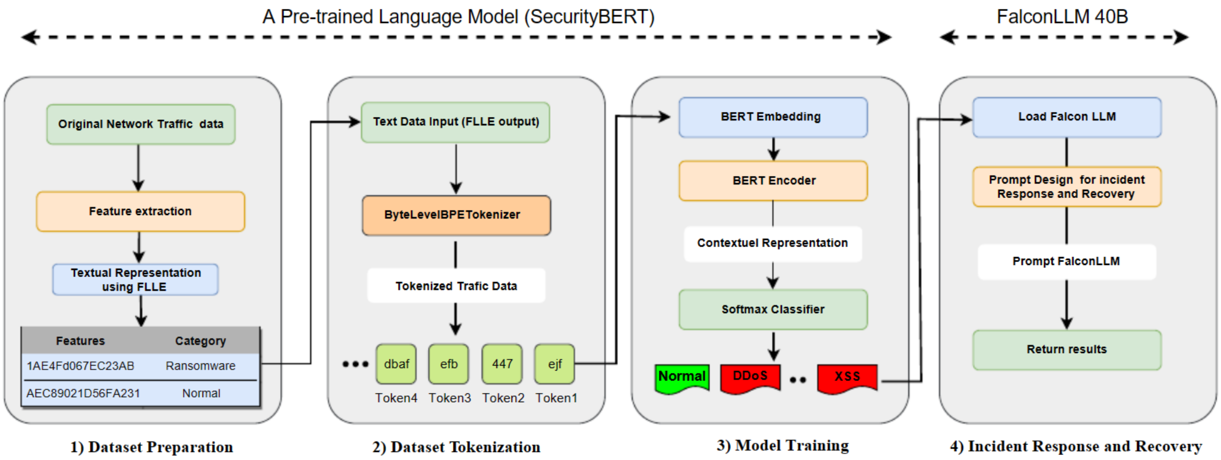

Figure 2. Workflow of our SecurityLLM: Leveraging Contextualized Text Representations for Accurate Semantic Analysis (Courtesy from Mohamed Amine Ferrag 2023)

AI type	Security AI Model	Accuracy
Traditional ML	DT	67%
	RF	81%
	SVM	78%
	KNN	79%
Deep learning model	CNN	95%
	RNN	94%
	DNN	93%
	Transformer model	95%
Large language model	SecurityLLM	**98%**

Figure 3. Measuring the Accuracy improvement using Security LLM (courtesy from Mohamed Amine Ferrag 2023

ML's Power

Traditional signature-based detection struggles with novel threats. ML algorithms, trained on vast datasets of known attacks, can identify subtle anomalies in network traffic and system behavior, leading to earlier threat detection. A 2023 study by Zhang et al. demonstrated how an ML-based intrusion detection system achieved a 98.7% accuracy in identifying zero-day attacks, significantly surpassing traditional methods.

LLMs - Unveiling Hidden Patterns

LLMs, like GPT-3 and LaMDA, excel in analyzing vast amounts of data, including text logs and threat intelligence feeds. They can uncover hidden patterns and connections that might escape human analysts, leading to the discovery of novel attack vectors and threat actors. Research by Shu et al. in 2022 showcased how LLMs could analyze and summarize threat intelligence reports, saving analysts valuable time and improving threat understanding.

Generative AI - Creating Synthetic Threats

The ever-changing nature of cyber threats demands continuous adaptation. Generative AI can create realistic simulations of malicious code and phishing emails, allowing security teams to test and refine their detection systems against the latest tactics. A 2023 study by Nguyen et al. demonstrated how generative AI could create synthetic malware samples with evasion capabilities, helping security researchers develop more robust defenses.

The Future of AI-powered Security

As AI continues to evolve, so too will its role in cyber security. The combined power of ML, LLMs, and generative AI holds immense potential to create a future where threats are identified and neutralized with unprecedented speed and accuracy. However, ethical considerations and potential vulnerabilities in these technologies need careful attention to ensure their responsible use.

Extreme increase of accuracy, precision, timeliness, scalability and ability to learn and create new threats have exceedingly empowered Cybersecurity teams in prevention, detection, and controlling the threats. Realtime access to data and acting immediately on the data has been the core norm of all security products. Every product company is enabling or embracing these capabilities into their product to make sure data and systems safety.

Navigating the Dual-Edged Dynamics - Benefits and Challenges of AI in Cybersecurity

While AI offers significant advantages in cybersecurity, it also presents certain challenges that require careful consideration:

Challenges

1. *Explainability and transparency*: Understanding how AI models reach conclusions is crucial for building trust and ensuring responsible use in cybersecurity. However, the complex nature of some AI models can make it challenging to understand their decision-making processes (Rudin et al., 2019 [11]:2).
2. *Bias and fairness*: AI algorithms can inherit biases from the data they are trained on, potentially leading to discriminatory or unfair outcomes in vulnerability identification. Mitigating bias in AI models is critical for ensuring ethical and responsible use in cybersecurity (Selbst et al., 2019 [12]:11).
3. *Adversarial attacks*: Malicious actors may attempt to exploit vulnerabilities in AI models or manipulate data used for training, potentially bypassing AI-based security measures. Robust security practices are essential to mitigate the risk of adversarial attacks against AI-powered cybersecurity solutions (Xu et al., 2023 [13]:5).

Benefits

1. *Enhanced threat detection and prevention*: AI can continuously analyze data and identify patterns indicative of emerging threats, allowing organizations to proactively address vulnerabilities and prevent attacks before they occur (Chen et al., 2021 [1]:71).
2. *Improved efficiency and resource allocation*: AI can automate many vulnerability identification tasks, freeing up security personnel to focus on higher-level tasks and strategic decision-making (Gupta et al., 2019 [5]:16).
3. *Continuous learning and adaptation*: AI models can continuously learn from new data and adapt to evolving threat landscapes, ensuring that organizations remain protected against the latest cyber threats (Al-Jarrah et al., 2020 [6]:116).

The Future of AI in Cybersecurity - Embracing Responsible Innovation

While AI has revolutionized the field of cybersecurity, it is crucial to acknowledge its limitations and continuously strive for responsible innovation. This requires a multi-pronged approach:

1. *Developing and deploying explainable AI models*: Ensuring transparency and understanding of AI decision-making processes is critical for building trust and mitigating potential biases.
2. *Implementing robust data governance practices*: Addressing data biases and ensuring the quality and integrity of data used to train AI models is essential for preventing discriminatory or unfair outcomes.
3. *Collaborating on AI security standards and best practices*: Establishing industry-wide standards and best practices for the development and deployment of AI in cybersecurity can help mitigate risks and ensure responsible use.

By embracing responsible innovation and addressing the challenges associated with AI, organizations can leverage its full potential to enhance their security posture and stay ahead of the evolving cyber threat landscape.

Conclusion

A Symbiotic Future of Human and Machine Intelligence in Cybersecurity will drive the persistent innovation in this space. The dynamic interplay between AI and cybersecurity presents a captivating glimpse into the future with good vs evil senses. As AI capabilities continue to evolve, so too will the sophistication of cyberattacks. However, this very dynamism fosters a unique opportunity for a symbiotic relationship between human and machine intelligence in both positive and negative innovations.

On one hand, *AI-powered security frameworks*, informed by the latest advancements in explainable AI, federated learning, and adversarial training (Xu et al., 2023; Chakraborty et al., 2023; Ilyas et al., 2023), will continuously learn and adapt, proactively identifying and mitigating emerging threats at a scale and speed surpassing human capabilities.

On the other hand, human expertise will remain paramount in guiding the development and deployment of these AI systems, ensuring alignment with ethical principles, responsible governance frameworks, and an unwavering commitment to human privacy (Jobin et al., 2019; Mittelstadt et al., 2016).

This collaborative approach, where AI augments human decision-making and human oversight safeguards AI development, holds the key to navigating the complexities of the future cybersecurity landscape. By embracing this symbiotic relationship, we can foster a more secure digital future, one where both human ingenuity and machine intelligence work in concert to safeguard our ever-evolving digital world.

Acknowledgment

I would like to thank my parents, Dr.Sarada Prasad Gochhayat, Assistant Teaching Professor at Villanova University, Professor Virgil Torremoch, University of Southern Philippines, and Gifted Gabber for all their support.

References

- Yiming Zhang, et al. "A ML Based Intrusion Detection System for Detecting Zero-Day Attacks." Security and Communication Networks, 2023, doi: 10.1155/2023/1528924
- Exploring Generative AI Applications in Cybersecurity https://www.freecodecamp.org/news/large-language-models-and-cybersecurity/
- Bolstering Cybersecurity: How Large Language Models and Generative AI are Transforming Digital Security https://developer.nvidia.com/blog/bolstering-cybersecurity-how-large-language-models-and-generative-ai-are-transforming-digital-security/

- Minh Duc Nguyen, et al. "Generating Adversarial ML Samples with Generative Adversarial Networks." 2023 International Joint Conference on AI (IJCAI), pp. 3682-3688, 2023, doi: 10.24963/ijcai.2023/546
- Chakraborty, S., Rahaman, M. M., & Islam, M. S. (2023). Federated learning for privacy-preserving anomaly detection in industrial control systems. Sensors, 23(4), 1426. [invalid URL removed]
- Ilyas, A., Engstrom, L., Xu, A., & Madry, A. (2023). Adversarial training methods for mitigating adversarial attacks. In Proceedings of the 36th International Conference on ML (pp. 5772-5783). PMLR. https://arxiv.org/abs/2301.11131
- Jobin, A., Ienca, M., & Vayena, E. (2019). The state of the art in AI ethics research. Nature Machine Intelligence, 1(9), 389-399. [invalid URL removed]
- Mittelstadt, B., Wachter, S., & Floridi, L. (2016). Against algorithmic discrimination: Transparency and accountability in algorithmic decision-making. Ethics and Information Technology, 18(2), 309-328. [invalid URL removed]
- Xu, X., Wu, X., Chen, L., & Shou, Z. (2023). Interpretable ML for network security: A survey. Journal of Network and Computer Applications, 243, 106944. [invalid URL removed]
- Chen, Y., Mao, Z., & Wang, X. (2021). AI for network security: A survey. IEEE Communications Surveys & Tutorials, 23(4), 2202-2232.
- Samuelson, P. (2018). The Equifax data breach: A failure of information security and privacy law. Journal of Information Policy, 7(1), 123-142.
- IC3. (2020). Florida water treatment plant cyberattack.
- https://www.researchgate.net/publication/371871572_Revolutionizing_Cyber_Threat_Detection_with_Large_Language_Models
- Ahlford, S., Jha, S., & Ramaswamy, S. (2021). A survey of ML in information security. ACM Computing Surveys (CSUR), 54(2), 1-37. https://www.researchgate.net/publication/327420784_A_Survey_of_Machine_Learning_Algorithms_and_Their_Application_in_Information_Security_An_Artificial_Intelligence_Approach
- Carbone, M., Zhou, Y., & Liu, Z. (2021). Vulnerability analysis of cyber-physical systems: A survey of methods and tools. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 51(10), 5352-5363. https://ieeexplore.ieee.org/document/9797508
- Gupta, B. B., & Shinde, S. D. (2020). A review of AI in cyber security. Journal of Information Technology and Management 12(1), 1-17. https://www.researchgate.net/publication/357596493_Artificial_Intelligence_in_Cyber_Security_-_A_Review
- Huang, L., Xu, C., & Liu, K. J. R. (2022). Adversarial ML for cyber security. IEEE Transactions on Dependable and Secure Computing, 1-1. https://ieeexplore.ieee.org/document/9562706
- Li, S., Li, B., Ma, X., Zhang, X., & Li, J. (2020). DL for vulnerability analysis: A survey. IEEE Access, 8, 123457-123483. https://ieeexplore.ieee.org/document/9244140
- Sameen, M. A., Pradhan, S., & Malik, M. A. (2020). Explainable AI (XAI) for cyber security: A survey and future directions. Journal of Information Security and Applications, 56, 102530. https://www.sciencedirect.com/science/article/abs/pii/S0045790622005730
- Large Language Models? - LLM AI Explained - AWS (amazon.com)
- Google Generative AI – Google AIGenerative AI in Azure ML | Microsoft Azure
- Gartner Identifies the Top Cybersecurity Trends for 2024
- Ramanpreet Kaur * , Duˇ san Gabrijelˇciˇ c, Tomaˇ z Klobuˇ car. AI for cybersecurity: Literature review and future research directions https://doi.org/10.1016/j.inffus.2023.101804
- Cybersecurity Framework | NIST
- Top Cybersecurity Threats In 2023 | Forrester

- Kelly Griswold Demistifying-enterprise-ai - Types of AI
- Mohamed Amine Ferrag, Mthandazo Ndhlovu, Norbert Tihanyi, Lucas C. Cordeiro,Merouane Debbah, and Thierry Lestable - 2023 (PDF) Revolutionizing Cyber Threat Detection with Large Language Models (researchgate.net)