Open Access License Notice

Cybersecurity Threats and Mitigation Strategies in AI Applications

Sajjad Bhuiyan
School of Information Studies
Syracuse University
Syracuse, USA
mhbhuiya@syr.edu
0009-0004-7408-3416

Joon S. Park
School of Information Studies
Syracuse University
Syracuse, USA
jspark@syr.edu
0000-0003-1925-2155

Abstract—The integration of artificial intelligence (AI) into daily life and critical infrastructure has elevated the importance of addressing cybersecurity concerns within AI applications. While AI systems offer numerous benefits, such as enhanced efficiency, automation, and decision-making, they also introduce novel vulnerabilities and threats. Ensuring the security and reliability of these systems is crucial. This paper investigates key cybersecurity challenges associated with AI, including data privacy, integrity, adversarial attacks, and the ethical implications of AI in security. Additionally, it examines the role of Shapley Additive explainable AI in promoting transparency, allowing for greater interpretability of AI models and insights into decision-making processes.

Keywords—*AI security, cybersecurity, cyber threats, generative AI, explainable AI, data privacy*

I. INTRODUCTION

In the digital age, the integration of Artificial Intelligence (AI) across various sectors has led to transformative changes, offering unprecedented opportunities for innovation and efficiency. However, these advancements introduce complex cybersecurity challenges that impact individuals, organizations, and society. As AI becomes increasingly embedded in daily life and critical infrastructure, securing these systems against malicious attacks, unauthorized access, and unintended consequences is critical. We explore key cybersecurity issues in AI applications, such as data privacy breaches, adversarial attacks, ethical considerations, and AI-driven cyber threats. Understanding these challenges is essential for developing robust security measures, ensuring ethical AI use, and maintaining trust in digital systems. A collaborative approach among technologists, policymakers, and stakeholders is necessary to balance AI's potential with effective risk mitigation.

This paper provides a comprehensive analysis of AI-related cybersecurity concerns, identifying AI's limitations, assessing specific cyber threats, and evaluating strategies for mitigating these risks. It also addresses the ethical and legal implications of AI security and examines the future direction of explainable AI. By making these topics accessible, this paper serves as a resource for developers, business managers, and government agencies, offering insights into both technical and strategic considerations necessary for managing and securing AI systems. Protecting privacy and remaining vigilant are crucial as AI continues to shape the digital landscape.

II. CYBERSECURITY THREATS WITH AI APPLICATIONS

Protecting data privacy and integrity is a significant concern in AI applications, as these systems often need large amounts of sensitive data for training and operation.

A. Data Poisoning

Attackers can manipulate AI by introducing malicious data into its training set, leading to flawed decision-making. Data integrity is crucial for the accuracy and reliability of AI systems. Malicious actors can manipulate training data through data poisoning attacks, deliberately introducing inaccuracies to compromise the system's performance. In Data Poisoning, attackers can manipulate machine learning systems according to their goals [1].

AI models are trained using data sets. In a data poisoning attack, as depicted in Figure 1, malicious data is introduced into the training dataset.

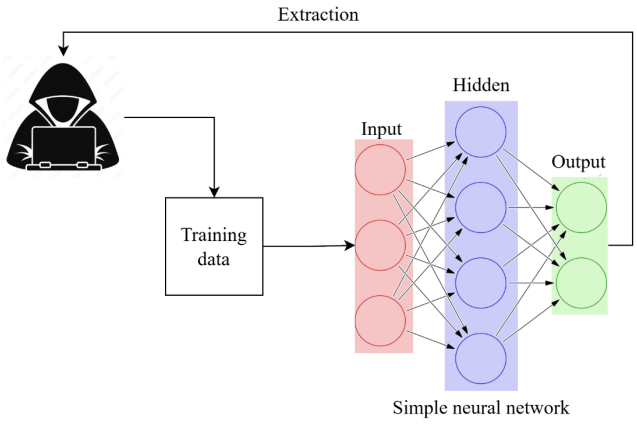

Fig. 1. Data poisoning in action

Figures 2 and 3 show the area under the curve (AUC) score with the training dataset with no data poisoning and 25% data poisoning [1]. From the figure, we can observe a significant difference between false positive rates. A false positive is a result that is incorrectly identified as positive. We can also observe significant differences in True Positive results from the AUC curve.

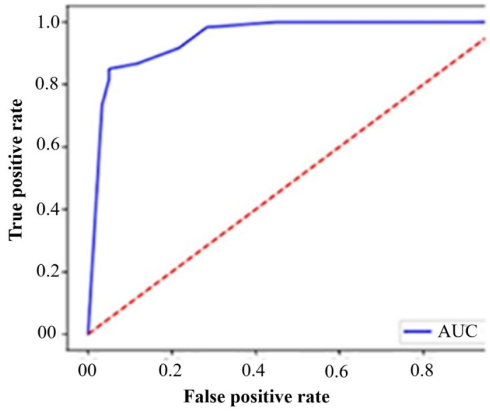

Fig. 2. A Training Data Set with no Data Poisoning [1]

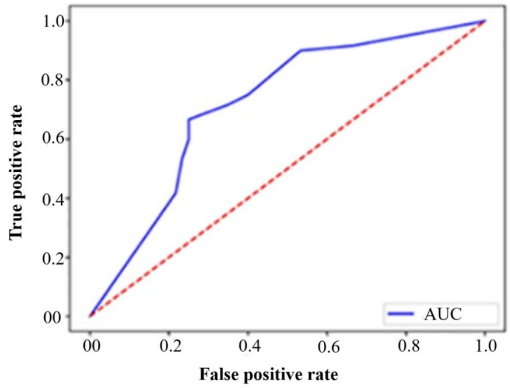

Fig. 3. A training dataset with 25% data poisoning [1]

To prevent data poisoning, we should ensure data integrity through rigorous validation, cleaning, and anomaly detection to spot malicious inputs. We should also strengthen model robustness with adversarial training and data augmentation. Additionally, we need to regularly audit and update data sources while implementing strong access controls and continuous monitoring to protect against unauthorized interference.

B. Model Theft

Unauthorized access to AI models, known as model theft or extraction, allows attackers to replicate and misuse the AI application. This is a significant concern in AI cybersecurity, especially given AI's limitations. Figure 4 illustrates the six steps of machine learning, the targeted ML model, and the training data.

AI model theft in cybersecurity can have severe consequences, including compromised security, as stolen models may be used to bypass defenses, leading to heightened vulnerability. Intellectual property loss results in financial damages and competitive disadvantages [27]. Privacy breaches occur when models trained on sensitive data are exposed, risking the confidentiality of personal information. Organizations may face regulatory penalties for failing to safeguard their models and data. Additionally, dealing with model theft can strain resources, diverting attention from other crucial cybersecurity tasks. Lastly, such theft undermines trust in AI-based security systems, casting doubt on their effectiveness and reliability.

Addressing model theft requires robust security measures, ongoing monitoring, and breach mitigation strategies. Techniques like model watermarking, differential privacy, and secure multiparty computation can enhance AI model security and reduce theft risks.

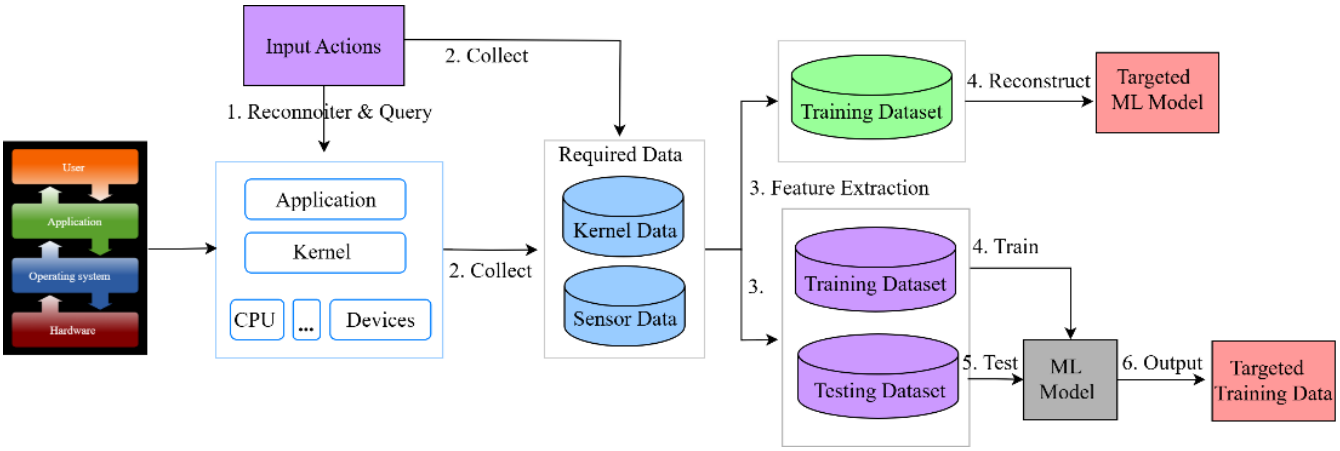

Fig. 4. Model theft uses machine learning techniques to acquire information about the ML model illicitly.

C. *Adversarial Attacks*

Adversarial attacks significantly limit AI in cybersecurity by creating inputs that mislead AI models into making incorrect decisions, thereby reducing the effectiveness of AI-powered solutions. Figure 5 illustrates adversarial attacks in the MIMO (Multiple-input multiple-output) system. A MIMO system is a technology used in wireless communications where multiple antennas are employed at both the transmitter and receiver end to improve communication performance by increasing data throughput and link reliability [9]. Figure 5 is the result of the cumulative distribution function (CDF) of per-user spectral efficiencies (SEs) in scenarios with and without an adversarial attack (specifically, the Basic Iterative Method or BIM), where AI solutions are implemented in both scenarios. It is evident from the data that the performance of SE drastically deteriorates under adversarial attacks [4].

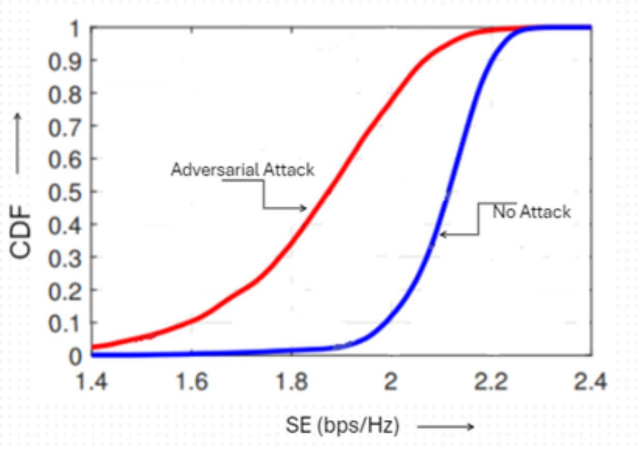

Fig. 5. Effect of adversarial attack on MIMO [4]

In cybersecurity, adversarial attacks manipulate inputs to deceive AI models, undermining their reliability and effectiveness and potentially allowing malicious activities to go undetected. These attacks increase vulnerability to cyber threats, erode trust in AI systems, and raise ethical and legal concerns, highlighting the need for ongoing advancements in defensive techniques to ensure resilience and compliance.

Researchers are exploring strategies to mitigate adversarial attacks, such as adversarial training, robustness checks, and new model architectures. Collaboration within the cybersecurity community is crucial for sharing knowledge on emerging threats and defenses.

D. *Reverse Engineering*

In AI and cybersecurity, reverse engineering involves attackers analyzing AI models to understand their function, identify vulnerabilities, or extract proprietary information. Figure 6 illustrates a sample model reverse-engineering attack.

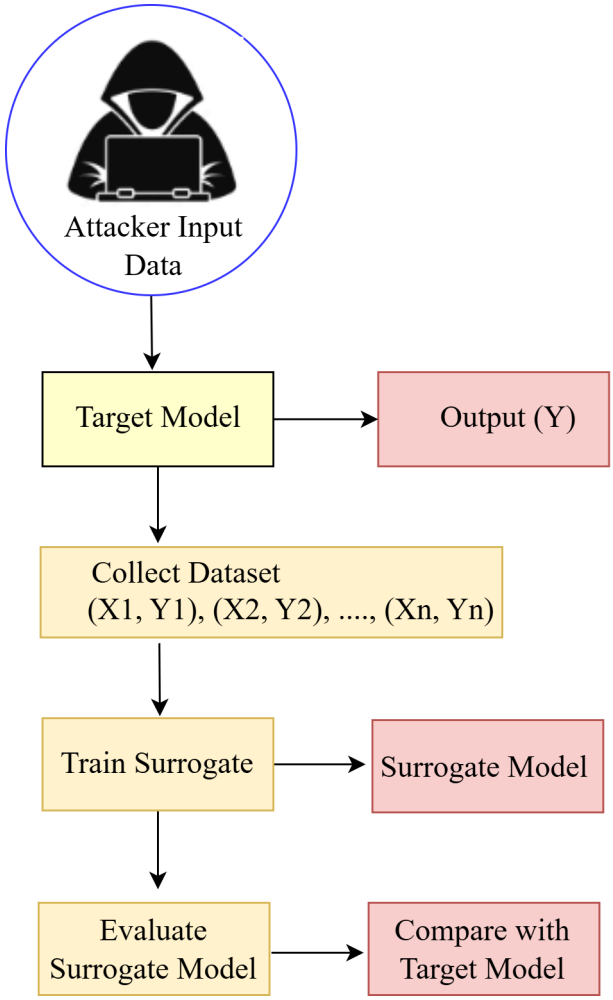

Fig. 6. Model reverse-engineering attack

Reverse engineering presents significant challenges for AI in cybersecurity by exposing vulnerabilities that attackers can exploit to bypass detection or trigger false positives [10]. This process also risks intellectual property theft, as proprietary algorithms and data may be stolen, allowing competitors or malicious actors to replicate or undermine the model's functionality. By understanding an AI system's decision-making process, attackers can craft inputs to evade detection, thereby compromising the system's effectiveness and granting unauthorized access to sensitive data. Additionally, reverse engineering can lead to the malicious replication of AI models, enabling the creation of convincing spam or phishing messages. As AI models increase the attack surface, additional protections such as obfuscation and tamper-detection become necessary, though they complicate deployment and maintenance. These risks raise both legal and ethical concerns, emphasizing the need to protect intellectual property while ensuring the security and integrity of AI systems.

Addressing reverse engineering requires a multi-layered approach that combines technical safeguards with legal protections. Techniques like model obfuscation, secure enclaves, and legal measures (e.g., copyright and patents) can protect AI models. Ongoing monitoring and updates are also essential to identify and address emerging vulnerabilities, ensuring robust cybersecurity defenses.

E. Privacy Leaks

Privacy leaks in the context of Artificial Intelligence (AI) in cybersecurity refer to unintended disclosures of sensitive or personal information through AI models. For instance, model inversion attacks can reveal sensitive details about training data from AI outputs, such as personal or proprietary information. Membership inference attacks might disclose whether specific data was used in training, exposing individual data or past security details. Data extraction via prediction APIs can uncover sensitive information about models or their training data, while transfer learning risks privacy by potentially leaking sensitive data in new contexts. Additionally, insufficient data anonymization and model overfitting can lead to privacy breaches, exposing personal or sensitive information.

Privacy leaks in AI-driven cybersecurity can harm individuals, damage reputations, and lead to legal issues under regulations like GDPR or CCPA. Mitigating these risks involves data protection strategies such as minimization, anonymization, and access controls. Advanced methods like differential privacy, federated learning, and secure multi-party computation can further protect privacy while leveraging AI.

III. IMPACT OF GENERATIVE AI IN CYBERSECURITY

Generative AI, which includes technologies capable of producing data, content, and simulations that resemble human-generated output, has significant implications for cybersecurity and privacy. Its impact is multifaceted, offering both innovative solutions to enhance security and new challenges that need careful management. Here are some of the key aspects of generative AI's impact on cybersecurity and privacy:

A. Positive Impacts

Generative AI enhances threat detection by simulating cyber-attacks, identifying vulnerabilities, and strengthening security measures. It also automates security tasks by generating configurations and policies, adapting to evolving threats, and reducing manual workload for dynamic security management. In addition, generative AI can create realistic phishing simulations to train users to recognize and respond to threats. It also supports data privacy by using differential privacy to produce anonymized datasets, protecting individual privacy while maintaining data utility for AI training.

B. Negative Impacts

Generative AI can produce sophisticated phishing content that is difficult to distinguish from legitimate communications, posing challenges for individuals and security systems [29]. It can also create deepfakes, including realistic images, videos, and audio, leading to impersonation, fraud, and challenges in identity verification. Additionally, generative AI can create or modify malware, making it more adaptable and challenging to detect, thus accelerating the cyber arms race. If trained on personal data, it may also unintentionally generate outputs with sensitive information and amplify biases in training data, leading to unfair or discriminatory outcomes with privacy and ethical implications.

C. Continuous Risk Mitigation

The dual-edged nature of generative AI's impact on cybersecurity and privacy necessitates a balanced approach to its deployment and regulation.

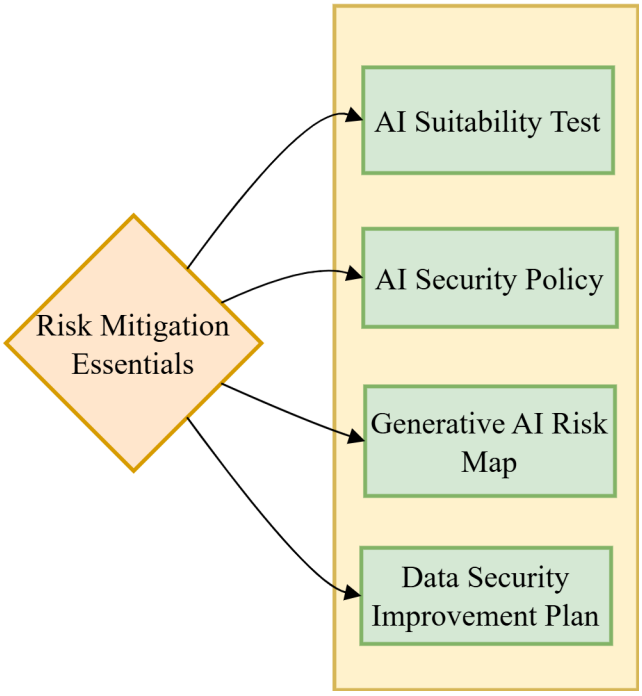

Fig. 7. Generation AI risk mitigation essentials

Figure 7 highlights generative AI risk mitigation essentials. Effective risk mitigation requires continuous research into secure and ethical AI practices, along with the development of robust defenses against AI-generated threats [23]. It is also essential to create frameworks that ensure transparency and accountability in AI systems. Furthermore, international collaboration is crucial to establish norms and guidelines that govern the use of generative AI, protecting against its misuse while leveraging its potential for positive contributions to cybersecurity and privacy.

IV. PRIVACY AND ETHICAL CONCERNS WITH AI APPLICATIONS

Privacy and ethical considerations in AI-driven cybersecurity involve technology, human values, and societal norms.

A. *Privacy, Bias, Security, and Fairness in AI*

AI systems rely on vast amounts of data, raising critical privacy and consent concerns. Transparent data collection processes and precise consent mechanisms are essential to ensure the ethical handling of personal information. Moreover, AI algorithms can amplify biases in training data, resulting in discriminatory outcomes. To ensure fairness, AI development must prioritize bias mitigation, using fairness-aware machine learning techniques and diverse datasets. Proper accountability frameworks and compliance with regulations such as GDPR are crucial for protecting privacy and maintaining ethical standards.

B. *Transparency and Human Control in AI*

AI's opaque decision-making processes challenge transparency, making it essential to develop explainable AI (XAI) systems that non-experts can understand and assess. Algorithmic transparency is crucial for ensuring fairness, allowing users to trace decision-making methods and data sources. As AI takes on more decision-making roles, there are growing ethical concerns about eroding human control. AI should assist, not replace, human decision-making, ensuring that responsibility remains with humans rather than machines. Clear audit trails and ethical oversight are necessary to maintain accountability in these systems.

C. *Security and Ethical AI Development*

The rapid adoption of AI introduces new security risks, making robust security measures vital to prevent vulnerabilities and misuse. Access controls, regular security assessments, and incident response planning are key to addressing potential breaches. In addition, the use of AI in surveillance by governments or corporations raises ethical concerns about privacy and individual freedoms. Safeguards must be implemented to protect human rights and avoid oppressive practices. Ethical AI development also requires educating developers on cybersecurity, privacy, and ethical risks during system design, ensuring compliance with regulations and ethical principles.

V. AI-AUGMENTED CYBER ATTACK SCENARIOS

In this section, we analyze potential attack scenarios, each based on real-life attacks that have previously occurred. By incorporating AI capabilities, these scenarios reveal how AI can be used to enhance the scale and effectiveness of such attacks, especially in critical sectors.

A. *AI-augmented Attacks in Healthcare*

In 2017, the WannaCry ransomware attack highlighted the healthcare sector's vulnerability to cyber threats, particularly as it disrupted the UK's NHS [22]. Though not AI-specific, this incident demonstrated the potential risks to AI-powered healthcare tools like diagnostic systems and patient monitoring. Adversarial attacks on AI in medical imaging are another threat, where slight input modifications can mislead AI systems into misclassifying benign conditions as malignant, jeopardizing patient safety and diagnostic accuracy. Similarly, AI-powered wearable health trackers are vulnerable to attacks that could intercept or manipulate patient data, resulting in incorrect health assessments. Additionally, AI applications that process electronic health records (EHR) are at risk of privacy breaches, where inadequate anonymization or weak access controls can lead to unauthorized data access. To mitigate these risks, healthcare providers must strengthen security measures, including encryption, stringent access controls, and frequent security audits, to protect patient data and uphold the reliability of AI-driven healthcare solutions.

B. *AI-augmented Attacks in the Financial Industry*

The 2010 "Flash Crash" exposed the susceptibility of financial markets to manipulation via algorithmic trading. Although these systems aren't fully AI-based, they incorporate AI in high-frequency trading, raising concerns about the potential for AI-driven market disruptions and emphasizing the importance of cybersecurity. Financial institutions depend on AI-powered fraud detection, yet adversarial attacks can exploit these systems, enabling fraudsters to evade detection, resulting in significant financial and reputational harm. Fintech startups, heavily reliant on AI and big data, are especially vulnerable to cyberattacks due to the sensitive financial information they handle. A breach could lead to identity theft, fraud, and regulatory fines, eroding consumer trust in AI-based fintech. Additionally, AI trading bots used by investors can be misused for market manipulation, where coordinated bot activity inflates stock prices, and regulatory bodies struggle to mitigate these risks. These scenarios demonstrate the complex cybersecurity landscape AI introduces in the financial sector. To address these risks, financial institutions must focus on encryption, access controls, and regular threat detection, while regulatory oversight and collaboration are key to securing an AI-driven financial ecosystem.

C. *AI-augmented Attacks on Autonomous Vehicles*

In 2015, researchers successfully hacked a Jeep Cherokee's infotainment system, gaining remote control over key functions such as steering and acceleration, underscoring the cybersecurity risks faced by connected vehicles. As autonomous vehicles increasingly rely on AI and network connectivity, these vulnerabilities could be exploited to jeopardize safety and control systems [26]. AI-powered vision systems, critical for navigation, are vulnerable to adversarial attacks where minor modifications to road signs or pedestrian images can mislead the system, posing significant safety risks. A malware infection in a fleet of autonomous vehicles could disrupt essential functions or compromise sensor data, resulting in accidents or unauthorized access to personal information. Furthermore, while telemetry data generated by autonomous vehicles improves AI performance and safety, breaches of this sensitive data can lead to privacy violations and identity theft, undermining public trust. These scenarios highlight the pressing need for robust cybersecurity measures in autonomous vehicles, requiring collaboration between

automakers, tech companies, regulators, and experts to establish security standards, detect threats, and respond to potential breaches. Continuous research and innovation are vital to ensuring the resilience of AI-driven vehicle systems.

VI. ENHANCING AI SECURITY WITH EXPLAINABLE AI

In discussing the future of AI security, we examine the role of Explainable AI (XAI) in improving the security of Internet of Things (IoT) networks through Intrusion Detection Systems (IDS). Given the complexity and opacity of many machine-learning (ML) models, transparent and interpretable predictions are essential for fostering trust and reliability in decision-making processes [6].

A prominent XAI method is Shapley Additive Explanations (SHAP), which interprets ML models by using Shapley values from game theory to fairly allocate contributions among features. XAI techniques clarify the decision-making processes of ML models, providing transparency that is critical for cybersecurity applications where understanding alerts is crucial for trust and accurate decision-making. SHAP can be applied to various machine learning models—whether tree-based, neural networks, or linear models—to explain individual predictions, while aggregated explanations offer insights into the model's overall behavior and feature importance. By fairly distributing feature importance and providing consistent explanations, SHAP enhances transparency and builds trust in AI models.

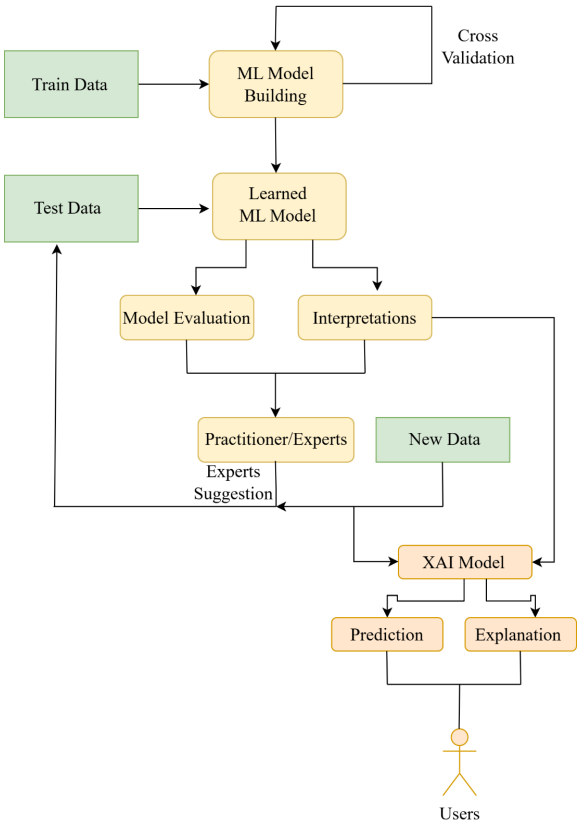

Fig. 8. Framework of XAI model

As illustrated in Figure 8, the XAI framework consists of three phases: Training, Prediction, and Interpretation / Explanation, where the latter phase focuses on improving interpretability by analyzing the factors influencing predictions. This is especially valuable in clinical decision-making and strengthens cybersecurity by improving attack detection and promoting effective communication between AI systems and human analysts in IoT networks.

VII. THE FUTURE OF AI CYBERSECURITY: EMERGING TRENDS

A. *AI-Powered Cyber Defense*

The future of AI cybersecurity will be shaped by advancements in AI technologies, innovative solutions, and collaboration among stakeholders to address emerging threats as follows.

B. *Adversarial Machine Learning*

As AI-based defenses evolve, adversaries will leverage AI to develop increasingly sophisticated cyberattacks. Adversarial machine learning will present significant challenges as attackers exploit AI vulnerabilities to evade detection and manipulate data. Addressing these threats demands ongoing research into adversarial robustness and the creation of advanced countermeasures.

C. *Explainable AI in Cybersecurity*

The rising demand for transparency in AI systems will drive the implementation of explainable AI (XAI) in cybersecurity. XAI will enable analysts to understand AI decisions, identify vulnerabilities, and interpret recommendations, fostering greater trust, accountability, and collaboration in securing AI systems.

D. *Privacy-Preserving AI Security*

With growing concerns over data privacy and regulatory compliance, privacy-preserving AI techniques will become vital. Technologies such as federated learning, homomorphic encryption, and differential privacy will allow for secure data sharing and analysis without compromising sensitive information.

E. *Cybersecurity for AI Applications*

As AI technologies continue to evolve, securing AI systems will be essential. This will involve protecting models, algorithms, and training data from tampering or exploitation. Best practices, including secure development, model verification, and runtime defenses, will be critical to maintaining the integrity and reliability of AI systems.

F. *Regulatory and Ethical Frameworks*

Governments and regulators will play a key role in shaping AI cybersecurity by developing frameworks, standards, and guidelines. These frameworks will incorporate ethical principles such as fairness, accountability, and transparency to mitigate risks and promote responsible AI use.

Overall, the future of AI cybersecurity will be characterized by constant innovation, adaptation, and collaboration to address evolving cyber threats, safeguarding digital assets, infrastructure, and individuals in an increasingly AI-driven world.

VIII. CONCLUSION

Cybersecurity concerns in AI applications are complex and multifaceted, encompassing technical, ethical, and regulatory challenges. Protecting AI systems from these threats demands a comprehensive strategy, including strong data protection, resilience against adversarial attacks, and careful consideration of the ethical dimensions of AI security. To prevent AI exploitation, robust security measures, continuous monitoring, and adversarial training are essential in addressing vulnerabilities. Shapley Explainable AI plays a key role by offering transparent and interpretable insights into model decisions, fostering trust and accountability. Collaboration among researchers, industry experts, and policymakers is crucial to developing standards, guidelines, and best practices for securing AI applications and ensuring their safe, beneficial use in society.

REFERENCES

[1] F. A. Yerlikaya and S. Bahtiyar, "Data poisoning attacks against machine learning algorithms," Expert Systems with Applications, vol. 208, p. 118101, Dec. 2022.

[2] L. Cui et al., "A Covert Electricity-Theft Cyberattack Against Machine Learning-Based Detection Models," IEEE Transactions on Industrial Informatics, vol. 18, no. 11, pp. 7824-7833, Nov. 2022, doi: 10.1109/TII.2021.3089976.

[3] Y. Miao, C. Chen, L. Pan, Q. Han, J. Zhang, and Y. Xiang, "Machine Learning Based Cyber Attacks Targeting on Controlled Information: A Survey."

[4] Ö. F. Tuna and F. E. Kadan, "A Novel Method to Mitigate Adversarial Attacks on AI-Driven Power Allocation in D-MIMO," in Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom), Istanbul, Turkiye, 2023, pp. 336-341, doi: 10.1109/BlackSeaCom58138.2023.10299750.

[5] K. Yoshida, T. Kubota, S. Okura, M. Shiozaki, and T. Fujino, "Model Reverse-Engineering Attack using Correlation Power Analysis against Systolic Array Based Neural Network Accelerator," in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), Seville, Spain, 2020, pp. 1-5, doi: 10.1109/ISCAS45731.2020.9180580.

[6] J. V. Rani, H. A. Saeed Ali, and A. Jakka, "IoT Network Intrusion Detection: An Explainable AI Approach in Cybersecurity," in Proc. 4th Int. Conf. Commun., Comput. Ind. 6.0 (C216), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/C2I659362.2023.10430601.

[7] N. Papernot et al., "The Limitations of Deep Learning in Adversarial Settings," in Proc. IEEE Eur. Symp. Security Privacy (EuroS&P), Saarbruecken, Germany, 2016, pp. 372-387, doi: 10.1109/EuroSP.2016.36.

[8] A. Jakka and J. Vakula Rani, "An Explainable AI Approach for Diabetes Prediction," in Innovations in Computer Science and Engineering, H. S. Saini, R. Sayal, A. Govardhan, and R. Buyya, Eds., Singapore: Springer, 2023, vol. 565, pp. 15-27, doi: 10.1007/978-981-19-7455-7_2.

[9] B. R. Manoj, M. Sadeghi, and E. G. Larsson, "Adversarial Attacks on Deep Learning Based Power Allocation in a Massive MIMO Network," in Proc. IEEE Int. Conf. Commun. (ICC), Montreal, QC, Canada, 2021, pp. 1-6, doi: 10.1109/ICC42927.2021.9500424.

[10] K. Yoshida et al., "Model Reverse-Engineering Attack using Correlation Power Analysis against Systolic Array Based Neural Network Accelerator," in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), Seville, Spain, 2020, pp. 1-5, doi: 10.1109/ISCAS45731.2020.9180580.

[11] C. Basile, D. Canavese, L. Regano, and P. Falcarin, "A meta-model for software protections and reverse engineering attacks," J. Syst. Softw., vol. 150, pp. 3-21, 2019.

[12] F. A. Yerlikaya, "Data poisoning attacks against machine learning algorithms," Expert Syst. Appl., vol. 208, p. 118101, Dec. 2022.

[13] N. Papernot et al., "The Limitations of Deep Learning in Adversarial Settings," in Proc. IEEE Eur. Symp. Security Privacy (EuroS&P), Saarbruecken, Germany, 2016, pp. 372-387, doi: 10.1109/EuroSP.2016.36.

[14] G. D'Angelo, M. Ficco, and F. Palmieri, "Malware detection in mobile environments based on autoencoders and api-images," J. Parallel Distrib. Comput., vol. 137, pp. ---, 2019.

[15] H. Hang, N. Cheng, Y. Zhang, and Z. Li, "Label flipping attacks against naive bayes on spam filtering systems," Appl. Intell., vol. ---, 2021.

[16] D. He, S. Zeadally, N. Kumar, and J. Lee, "Anonymous authentication for wireless body area networks with provable security," IEEE Syst. J., vol. 11, no. 4, pp. 2590-2601, 2017.

[17] P. M. Santos, B. R. Manoj, M. Sadeghi, and E. G. Larsson, "Universal adversarial attacks on neural networks for power allocation in a massive MIMO system," IEEE Wireless Commun. Lett., vol. 11, no. 1, pp. 67-71, 2022.

[18] B. Kim et al., "Adversarial attacks against deep learning based power control in wireless communications," in Proc. IEEE Globecom Workshops, 2021, pp. 1-6.

[19] I. Alsmadi and R. Saraireh, "IoT intrusion detection system based on machine learning techniques with explainability," Comput. Mater. Continua, vol. 67, no. 1, pp. 1205-1223, 2021.

[20] P. Blanchard et al., "Machine learning with adversaries: Byzantine tolerant gradient descent," in Adv. Neural Inf. Process. Syst. 30: Annu. Conf. Neural Inf. Process. Syst., 2017, pp. 119-129.

[21] M. Jagielski et al., "Subpopulation data poisoning attacks," CoRR, abs/2006.14026, 2020.

[22] S. -C. Hsiao and D. -Y. Kao, "The static analysis of WannaCry ransomware," 2018 20th International Conference on Advanced Communication Technology (ICACT), Chuncheon, Korea (South), 2018, pp. 153-158, doi: 10.23919/ICACT.2018.8323680.

[23] S. Li et al., "Hidden backdoors in human-centric language models," CoRR, abs/2105.00164, 2021.

[24] L. Muñoz-González et al., "Poisoning attacks with generative adversarial nets," CoRR, abs/1906.07773, 2019.

[25] A. Roque and S. K. Damodaran, "Explainable AI for Security of Human-Interactive Robots," Int. J. Human-Comput. Interact., vol. 38, no. 18-20, pp. 1789-1807, 2022, doi: 10.1080/10447318.2022.2066246.

[26] A. O. A. Zaabi, C. Y. Yeun and E. Damiani, "Autonomous Vehicle Security: Conceptual Model," 2019 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific), Seogwipo, Korea (South), 2019, pp. 1-5, doi: 10.1109/ITEC-AP.2019.8903691.

[27] Z. Zhang, H. A. Hamadi, E. Damiani, C. Y. Yeun and F. Taher, "Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research," in IEEE Access, vol. 10, pp. 93104-93139, 2022, doi: 10.1109/ACCESS.2022.3204051.

[28] S. Neupane et al., "Security Considerations in AI-Robotics: A Survey of Current Methods, Challenges, and Opportunities," IEEE Access, vol. 12, pp. 22072-22097, 2024, doi: 10.1109/ACCESS.2024.3363657.

[29] Simmons, T., & Park, J. S. (2024, June). "Cybersecurity Challenges and Opportunities with Generative AI. Advanced Education for Cyber Security Track," the 23rd European Conference on Cyber Warfare and Security. Published.