Discussion-3 The Role of AI in Cybersecurity

Discussion Topic:

The use of AI in cybersecurity is growing rapidly. Use the Internet to research the latest developments in cybersecurity AI. Provide an example of cybersecurity AI, including how it works and what platforms are using it. How can adversarial AI attacks be defended against?

My Post:

Hello Class,

I appreciated this discussion topic as I am fascinated by the immense advances in AI tech and development that have happened in the last couple of years. AI research has been around for quite a while. As early as 1950, the year that Alan Turing published “Computing Machinery and Intelligence” […], which proposed the "Turing Test" as a way to measure a machine's ability to mimic intelligent behavior. However, the AI field was formally defined/launched/founded by the 1956 Dartmouth workshop.

On that side note, today, AI is starting to be integrated in almost every field of computer science and information technology. One of these fields is cybersecurity.
In cybersecurity, AI is revolutionizing the way organizations approach security (McCabe, 2025). AI is now being used:

- To proactively detect threats in real time by identifying zero-day exploits through behavioral analytics, by preventing advanced persistent threats (APTs) before they infiltrate critical systems, and by stopping lateral movement within enterprise networks by dynamically segmenting users and devices.
- To combat phishing by analyzing email structures, language nuances, voice, video and embedded links.
- To support zero trust architecture (ZTA) by continuously verifying users and devices based on real-time behavioral analysis, by blocking unauthorized access and lateral movement, and by dynamically adjusting access controls based on evolving risk scores

(McCabe, 2025)

An example of AI being used to protect and secure a system is Microsoft Sentinel. Microsoft Sentinel is cloud-native security information and event management (SIEM) and Security Orchestration, Automation, and Response (SOAR) solution (Microsoft, 2024). Sentinel uses AI to Machine Learning (ML) to collect entire organization data, including Microsoft 365, Azure, and other third-party sources. Then it uses AI/ML to analyze this data to filter/detect suspicious activities and previously unseen threats that traditional methods might miss. If a potential threat is detected, Sentinel uses AI to correlate related alerts to generate the picture/structure of the attack chain. The AI can also take a more proactive role by triggering automated responses to contain the threat; for example, it can trigger an automated response that will isolate an infected device or block a malicious IP address.

On the other side of the coin, AI has enabled cybercriminals to exploit vulnerabilities in a more effective way, by helping them to avoid detection, execute more sophisticated attacks, and scale their operations

(Ross, 2025). Additionally, the integration of AI/LM within cybersecurity and systems brings new vulnerabilities, such as adversarial AI or adversarial machine learning (ML). Adversarial AI “seeks to inhibit the performance of AI/ML systems by manipulating or misleading them” (Stanham, 2023). These attacks can be primarily categorized as:

- Evasion attacks, which are subtle and malicious inputs that can bypass detection by an AI model.
- Poisoning attacks, these attacks are meant to poison/contaminate the data used to train AI/ML models.
- Model extraction/theft, these attacks aim to reverse engineer a model by repeatedly querying it, and using the outputs to recreate the model, or by illegally accessing the model’s parameters.

Here is a list of the top 10 Machine Learning security risks by OWASP (n.d.):

- **ML01:2023 Input Manipulation Attack**
- **ML02:2023 Data Poisoning Attack**
- **ML03:2023 Model Inversion Attack**
- **ML04:2023 Membership Inference Attack**
- **ML05:2023 Model Theft**
- **ML06:2023 AI Supply Chain Attacks**
- **ML07:2023 Transfer Learning Attack**
- **ML08:2023 Model Skewing**
- **ML09:2023 Output Integrity Attack**
- **ML10:2023 Model Poisoning**

Defending against adversarial AI attacks is possible, but “it requires a comprehensive approach to security that considers vulnerabilities across different systems” (Paloalto, n.d.). This comprehensive approach to security involves training AI models on a dataset that includes adversarial examples; filtering processes that are capable of removing or flagging potentially malicious inputs and malicious data within training sets; and implementing a "never trust, always verify" protocol within the AI implementation framework.

-Alex

References:

McCabe, M. (2025, March 27). *AI-driven threat detection: Revolutionizing cyber defense.* Zscaler. https://www.zscaler.com/blogs/product-insights/ai-driven-threat-detection-revolutionizing-cyber-defense

Microsoft (2024, May 21). *What is Microsoft Sentinel?* Microsoft Learn. https://learn.microsoft.com/en-us/azure/sentinel/overview?tabs=defender-portal

OWASP (n.d.). *OWASP machine learning security top ten*. OWASP. https://owasp.org/www-project-machine-learning-security-top-10/

Paloalto (n.d.). *What Is Adversarial AI in Machine Learning?* Paloalto. https://www.paloaltonetworks.com/cyberpedia/what-are-adversarial-attacks-on-AI-Machine-Learning#:~:text=Versatility%20and%20Risks&text=Defending%20against%20Adversarial%20AI%20requires,considers%20vulnerabilities%20across%20different%20systems.

Ross, C. F. (2025, May 2022). AI in cybersecurity: How AI is impacting the fight against cybercrime. Akamai. https://www.akamai.com/blog/security/ai-cybersecurity-how-impacting-fight-against-cybercrime#:~:text=AI%20enables%20cybercriminals%20and%20hackers,sophisticated%20attacks%2C%20and%20scale%20their

Stanham, L. (2024, November 2). Adversarial AI & adversarial Machine Learning. CrowdStrike. https://www.crowdstrike.com/en-us/cybersecurity-101/artificial-intelligence/adversarial-ai-and-machine-learning/#:~:text=By%20teaching%20an%20ML%20model,defend%20against%20attacks%20such%20as