AI IN ETHICAL HACKING: DETECTING AND EXPLOITING VULNERABILITIES

AI impacts cybersecurity throughout its life cycle improving automation, threat intelligence and cyber defense. AI brings challenges like adversarial attacks and the need for high-quality data, leading to its inefficiency. There is a positive influence of AI on cybersecurity, enhancing its effectiveness and resilience (Jada and Mayayise, 2023).

RESEARCH QUESTIONS

Q1: What AI models are effective in automating vulnerability discover during penetration tests and how do they compare to conventional methods?

Q2: How will ethical hackers ensure that AI systems are used for penetration testing and don't inadvertently introduce new vulnerabilities and exploits?

Why does this matter?

Q1: This compares and assesses how effective AI is compared to current industry standards (LinkedIn, 2024).

Q2: This addresses AI misuse risks and the potential of new exploits and vulnerabilities (Department for Science, Innovation & Technology, 2024).

METHODOLOGY

Research Approach: Papers used will be from existing literature, case studies, trusted articles or government whitepapers.

Literature Review: Analysis of AI models and tools used in pen-testing such as Machine Learning, Deep learning, and Generative Adversarial Networks (GANs) (Subramanian and M. Chinnadurai, 2024).

Simulation: Tests of AI tools vs conventional in a virtual machine

Ethical Analysis: Finding frameworks that mitigate the risk of AI misuse

KEY FINDINGS

Answers to Q1: AI models such as Random Forests and Deep Neural Networks detect vulnerabilities faster with less false positives than conventional methods.

“Generating a GAN-based synthetic attack dataset, by training a GAN model on real and fake attack samples. This synthetic data can help to train web application layer defensive devices, such Web Application Firewalls (WAFs), against sophisticated attacks, like APT” (Chowdhary, Jha and Zhao, 2023)

Answers to Q2: Validation and verification of AI models before they get released, ensure ethical and legal standards are adhered to, and implementing human oversight of AI to protect against reputational damage. (Fernández Peñalver, 2024)

IMPORTANCE OF FINDINGS

Why it matters: Research findings can help enhance the effectiveness of pen-testing.

Showcases current and future risks of AI misuse and the need for proactive mitigation.

Broader Impact: “By leveraging AI responsibly, ethical hackers can significantly improve an organization's security posture while addressing the complex ethical landscape of modern cybersecurity” (Pradeep Sambamurthy, 2024).

CONCLUSION

“Incorporating AI solutions into organisational cybersecurity has a predominantly beneficial effect. Essentially, AI usage offers an effective, advanced, and heightened level of cyber protection.” (Jada and Mayayise, 2023)

AI models such as GANs and DNNs improve speed and accuracy of vulnerability detection incomparison to conventional methods. Combining human input and expertise with AI tools will show greater results than just AI by itself (Lim et al., 2024).

AI systems introduce new risks and ethical issues that are extensive.

AI is a very useful tool and has many use cases in ethical hacking. However, needs to follow rules and regulations to prevent misuse and maintain trust (Díaz-rodríguez Et Al., 2023).

Q & A

FIRE AWAY

REFERENCES

Chowdhary, A., Jha, K. and Zhao, M. (2023). Generative Adversarial Network (GAN)-Based Autonomous Penetration Testing for Web Applications. *Sensors*, [online] 23(18), p.8014. doi:https://doi.org/10.3390/s23188014.

Department for Science, Innovation & Technology (2024). *Cyber security risks to artificial intelligence*. [online] GOV.UK. Available at: https://www.gov.uk/government/publications/research-on-the-cyber-security-of-ai/cyber-security-risks-to-artificial-intelligence [Accessed 16 Jan. 2025].

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E. and Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, [online] 99(101896), p.101896. Available at: https://www.sciencedirect.com/science/article/pii/S1566253523002129 [Accessed 17 Jan. 2025].

Fernández Peñalver, M. (2024). *Keeping AI in Check: The Critical Role of Human Agency and Oversight*. [online] www.nemko.com. Available at: https://www.nemko.com/blog/keeping-ai-in-check-the-critical-role-of-human-agency-and-oversight [Accessed 17 Jan. 2025].

Jada, I. and Mayayise, T.O. (2023). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management*, [online] 8(2), pp.100063–100063. doi:https://doi.org/10.1016/j.dim.2023.100063.

Lim, W., Yong, K.S.C., Lau, B.T. and Tan, C.C.L. (2024). Future of generative adversarial networks (GAN) for anomaly detection in network security: A review. *Computers & Security*, [online] 139(2), p.103733. doi:https://doi.org/10.1016/j.cose.2024.103733.

LinkedIn (2024). *Here's how you can measure your performance as an AI professional against industry standards.* [online] Linkedin.com. Available at: https://www.linkedin.com/advice/3/heres-how-you-can-measure-your-performance-fqgte [Accessed 16 Jan. 2025].

Pradeep Sambamurthy (2024). *The Integration of Artificial Intelligence in Ethical Hacking: Revolutionizing Cybersecurity Predictive Analytics* **********. [online] ResearchGate. Available at: https://www.researchgate.net/publication/383398507_The_Integration_of_Artificial_Intelligence_in_Ethical_Hacking_Revolutionizing_Cybersecurity_Predictive_Analytics [Accessed 16 Jan. 2025].

Subramanian, G. and M. Chinnadurai (2024). Hybrid quantum enhanced federated learning for cyber attack detection. *Scientific Reports*, [online] 14(1). doi:https://doi.org/10.1038/s41598-024-83682-z.