Issues in Information Systems
Volume 25, Issue 2, pp. 230-243, 2024

DOI: https://doi.org/10.48009/2_iis_2024_119

The impacts of generative artificial intelligence (AI) in knowledge discovery and generation for cyber defense

Ping Wang, *Robert Morris University, wangp@rmu.edu*
Christopher Johnson, *Robert Morris University, johnsonch@rmu.edu*

Abstract

Artificial Intelligence (AI) technologies have the potential to improve efficiency, productivity, and accuracy in knowledge discovery and generation. Knowledge discovery through penetration testing (pentesting) is a critical step in cyber defense to learn and manage security vulnerabilities. Inspired by the knowledge principles and defense strategies from Sun Tzu's classic work *The Art of War*, this research explores the role and moderating effects of generative AI and its limitations in security knowledge discovery for cyber defense. The study focuses on the strategies of discovering and knowing the cybersecurity vulnerabilities of oneself and one's opponents, which may be enhanced by generative AI and Large Language Model solutions such as ChatGPT. The proposed AI-moderated knowledge discovery model for cyber defense is tested and illustrated with prompts and output from the interactive ChatGPT tool trained with security data from virtual network simulations for penetration testing. AI solutions such as LLM-based ChatGPT are still evolving with limitations and challenges for applications. This study will present and discuss the positive impacts as well as limitations and challenges of generative AI in pentesting for security knowledge discovery for cyber defense.

Keywords: Generative AI, *The Art of War*, cyber defense, knowledge, pentesting, ChatGPT

Introduction

As individuals, organizations, and nations are increasingly dependent on trustworthy digital networks and systems in daily life, economy, critical infrastructures, and national security, cyber threats and attacks are becoming more frequent and increasingly sophisticated and challenging for cyber defense (Clarke & Knake, 2010; Jada & Mayayise, 2023). A significant factor for the cyber defense challenges is the lack of adequate knowledge of the cybersecurity vulnerabilities and threats necessary for effective security decision making and response (Booker & Musman, 2019; Wang & D'Cruze, 2020, 2021).

The old saying from Renaissance that knowledge is power still applies to cyber defense. The latest National Cybersecurity Strategy for the United States recognizes the significance of knowledge sharing and calls for a collaborative networked cyber defense model for increased connectivity "enabled by the automated exchange of data, information, and knowledge" (The White House, 2023).

Cyber defense shares the same dynamics, principles, and strategies from traditional military defense and offense. Knowledge for defense is relative to ignorance and dynamically involves and impacts both the defender (oneself) and the adversary (other or opponent). In *The Art of War*, a military classic by Sun Tzu in the 5th century B.C., the knowledge of the adversary and oneself is highly important and the critical

factor for consistent triumph (Michaelson & Michaelson, 2003; Sun, 1910). Knowledge of the adversary and oneself has been found applicable to a wide range of domains including specific areas of cybersecurity, such as the practical benefits of knowing hackers' motivation and techniques and tools of ransomware for defending against these cyber-attacks (Dunn, 2024; Kelly et al., 2021; Wang & D'Cruze, 2020, 2021).

Penetration testing is a primary method of knowledge discovery in cyber defense. Pentesting can be used to discover knowledge and intelligence on the strengths and digital system weaknesses or vulnerabilities of oneself or the adversary for better cyber defense against the adversary. As the volume of data communication and pentesting grows exponentially, it is necessary to find technologies to improve the efficiency and accuracy of planning, data collection, and data analysis in the pentesting process. AI-enabled pentesting has the potential to improve accuracy and save processing time and investment of human resources (Dsousa, 2018; Jada & Mayayise, 2023).

Pre-trained generative AI solutions have the potential strengths to automate repetitive tasks, enhance threat detection, and improve the efficiency and accuracy of responses to cyber threats and attacks (Kaur, Gabrijelcic, & Klobucar, 2023). However, research also indicates that generative AI solutions based on Large Language Models, such as ChatGPT, may have limitations and pose challenges for applications in cybersecurity, including concerns over disclosures of sensitive private information, hallucinations or misleading information, and adequacy and quality of training data (Al-Hawawreh, Aljuhani, & Jararweh, 2023; Gupta, Aryal, & Praharaj, 2023).

This study will focus on the knowledge discovery process through pentesting for cyber defense and explore the moderating role and effects of generative AI in the pentesting process. The goal of the study is to contribute an AI-moderated knowledge discovery model and new empirical data for cyber defense.

The model is tested and illustrated with prompts and output from LLM-based ChatGPT pre-trained with security data from pentesting simulations on a contained virtual network. The following sections will review the relevant theoretical background, explain the proposed model, describe the simulation method, present and discuss the findings from the simulation, and conclude the paper.

Background

This section reviews the key research background on the important role of knowledge discovery in cyber defense and the moderating effects of artificial intelligence (AI) in the knowledge discovery process. Knowledge discovery is essential to the success of defense in traditional warfare and the success of modern cyber defense as knowledge-based defense principles and strategies can be transferred across generations and across domains.

AI-driven security solutions can help organizations stay ahead of increasingly sophisticated cyberattacks, improve efficiency and reduce human error, but it can also be weaponized by malicious actors to launch more sophisticated cyberattacks (Wilson, 2023).

Table 1 below highlights and summarizes significant research and findings on the power of knowledge and knowledge discovery applicable to modern cyber defense and the positive and negative impacts of AI solutions.

Table 1: Summary of Research on Knowledge Discovery and AI for Cyber Defense

Year	Authors	Methodology	Key Findings/Conclusions
5th Century B.C. Trans. 1910	Sun (Trans. Giles)	Observations on defense strategy based on warfare command experience	On the importance of knowing oneself and the opponent: • If you know the enemy and know yourself, you need not fear the result of a hundred battles (p. 45) • If you know yourself but not the enemy, for every victory gained you will also suffer a defeat (p. 45) • If you know neither the enemy nor yourself, you will succumb in every battle (p. 45) • He will win who knows how to handle both superior and inferior forces (p. 45)
5th Century B.C. Trans. 1910	Sun (Trans. Giles)	Observations on defense strategy based on warfare command experience	On the importance of discovering and knowing the security vulnerabilities of the opponent and knowing and hiding the vulnerabilities of oneself: • You can ensure the safety of your defense if you only hold positions that cannot be attacked (p. 58) • A general is skillful in attack whose opponent does not know what to defend; and he is skillful in defense whose opponent does not know what to attack (p. 58) • By discovering the enemy's dispositions and remaining invisible ourselves, we can keep our forces concentrated, while the enemy's must be divided (p. 59) • In making tactical dispositions, the highest pitch you can attain is to conceal them; conceal your dispositions, and you will be safe from the prying of the subtlest spies, from the machinations of the wisest brains (p. 62)
2020	Wang, & D'Cruze	Simulation	• Proposed and implemented a cyber defense knowledge model through network pentesting simulation. • The knowledge discovery process for cyber defense includes these key components: - Footprinting - Port scanning - Enumeration - Penetration testing - Vulnerability analysis
2023	Takko, et al.	Experiment	• Proposed and implemented a novel knowledge graph based framework for extracting strategic level knowledge for risk analysis from open unstructured sources of data on cyber-attacks • Recognizing the need for further research to improve the cyber knowledge discovery for better accuracy and automation to process increasingly large amount of data

Year	Authors	Methodology	Key Findings/Conclusions
2023	Wang, Liu, Zhong, & Zhou	Experiment	• Designed and implemented a knowledge graph completion model to improve integrity and accuracy of cybersecurity knowledge via pentesting • Accurate cyberspace detection information is the key to ensuring successful penetration testing • Interactive recommender systems and question answering systems may help achieve intelligent penetration testing
2023	Al-Hawawreh, Aljuhani, Jararweh	Simulation	• ChatGPT can be used by technologists to add (or improve) cybersecurity practices • ChatGPT can be used by threat actors to find attack vectors and obscure their presence • Over-reliance on ChatGPT can lead to security and privacy issues
2023	Mamgai	Analysis	• "Incomplete" data leads to limited or flawed responses • ChatGPT is in the public domain, be mindful as to the sensitivity of the data being used to train it • High degree of risk in creating more harm than good if ChatGPT is not used/implemented correctly
2023	Temara	Experiment	• ChatGPT is an effective tool to aid in reconnaissance • The data used to train ChatGPT can be used in reconnaissance against your organization • ChatGPT responses can change over time as it continues to be trained
2023	Zhan, Xu, Sarkadi	Experiment	• ChatGPT can produce misleading or deceptive information • There is a risk associated with an ill-informed user acting on this information • Fact checking and regulation can help mitigate this risk
2023	Wilson		• AI is shaping cybersecurity in positive and negative ways • Up-to-date knowledge about AI advancements and trends will help stay ahead of emerging threats • Be mindful of any concerns (ie. ethical, data protecton) that may come from the use of AI

As an example of the positive impact of AI on cyber defense, ChatGPT is used by developers who do not have sufficient information about cybersecurity and need to suffuse their code with better cybersecurity practices (Al-Hawawreh et al., 2023). The same can be said for Database Administrators, Server Administrators, and Network Engineers, and all of them have specialized areas of expertise but may not be

cybersecurity minded. ChatGPT has also been found useful in reconnaissance in penetration testing, gathering useful knowledge including application server type, database type, operating systems, big data technologies, logging and monitoring software, and other infrastructure related information (Temera, 2023).

As for negative impacts of AI, the first concern is the deployment of confidential information into the public domain. Second, an ill-trained AI model can do as much damage as none at all because AI models are only as good as the data that powers them. Third, because there is a cost associated with building and training models, a successful generative AI model for cybersecurity requires an efficient infrastructure and large datasets.

Without sufficient data and the necessary resources, AI systems can produce incorrect monitoring results and false positives, which can have consequences for organizations (Mamgai, 2023). Further, ChatGPT can be used to enhance a bad actor's attack vector. Lastly, considering that while advocates are looking for ways to leverage ChatGPT to enhance cyber defense, threat actors stand to gain the same efficiency. Malicious actors can leverage AI-driven tools and techniques to launch more sophisticated and targeted cyberattacks, posing significant challenges for organizations and individuals seeking to protect their digital assets (Wilson, 2023).

There are further limitations and concerns to consider with ChatGPT. AI-driven security solutions often require access to large amounts of data, raising concerns about user privacy and data protection (Wilson, 2023). While ChatGPT may have efficient access to sizeable amounts of data, it should not be treated as *infallible*. Over-reliance on AI-driven security solutions can create a false sense of security and potentially leave organizations vulnerable if AI systems fail or are compromised (Wilson, 2023).

There are also challenges to consider regarding the use of ChatGPT in cybersecurity. As stated earlier, the accuracy of information provided by ChatGPT is dependent upon the quantity and quality of data it has access to. A user querying ChatGPT must be cautious about its output, and not make assumptions that it has all the data necessary to create a correct or appropriate response. The misleading and deceptive content can have adverse impacts on users who are not equipped to distinguish 'fact' from 'fiction', thereby precipitating detrimental consequences (Zhan et al., 2023).

The background review above shows the importance of knowing one's vulnerabilities and strengths and those of one's opponent and hiding one's vulnerabilities and strengths for self-defense, including cyber defense. The review also identifies the positive and negative impacts of AI for cyber defense and the need for further research to automate knowledge discovery such as pentesting to enhance the efficiency, accuracy, and intelligence of knowledge for cyber defense.

Theoretical Model

Table 2 below presents the AI-moderated Knowledge Discovery model adopted for this study. This model is adapted from the Knowledge and Goals Matrix model by Wang and D'Cruze (2020) with defense strategies and principles from *The Art of War* and additional moderating effects of artificial intelligence (AI). This model emphasizes the role of knowing and exploiting the opponent's strengths and vulnerabilities as well as knowing and concealing one's own vulnerabilities, assets, and strengths to mislead the opponent and defend oneself.

Table 2: AI-moderated Knowledge Discovery Model

	Knowledge	Goals	Moderating Effects of AI
Oneself	• Know one's own vulnerabilities • Know how to mitigate one's own vulnerabilities • Know how to hide one's assets and vulnerabilities from the opponent • Know how to set up deceptive vulnerabilities	• To minimize one's vulnerabilities • To assess and manage one's vulnerabilities and risks • To minimize the opponent's knowledge of one's vulnerabilities • To mislead, misinform, distract, and deceive the opponent	• AI may enhance the effects of knowledge discovery • AI may limit the effects of knowledge discovery
Opponent	• Know the opponent's assets and strengths • Know the opponent's vulnerabilities • Know how to discover the opponent's vulnerabilities	• To be aware of threats and avoid striking the strong spots of the opponent • To exploit the vulnerabilities of the opponent • To maximize knowledge of the opponent	• AI may enhance the effects of knowledge discovery • AI may limit the effects of knowledge discovery

The AI-moderated Knowledge Discovery model incorporates the role of artificial intelligence (AI) solutions as a moderating factor in the knowledge discovery process. AI has the potential to enhance the efficiency and accuracy of knowledge discovery such as pentesting (Jada & Mayayise, 2023; Takko, et al., 2023; Wang, Liu, Zhong, & Zhou, 2023). On the other hand, AI may have limitations and pose risks and challenges for application in cybersecurity (Al-Hawawreh, Aljuhani, & Jararweh, 2023; Gupta, Aryal, & Praharaj, 2023; Wilson, 2023). Pentesting is an aggressive knowledge discovery to test and determine the target system's vulnerabilities exploitability. Vulnerability analysis leads to the ultimate goal of the knowledge discovery process to analyze and assess the potential impact of the known vulnerabilities and associated threats and risks so as to manage the cyber threats and risks for cyber defense (Muckin & Fitch, 2019; Wang & D'Cruze, 2021).

Methodology

This study uses a virtual machine lab environment to simulate penetration testing techniques for cyber knowledge discovery, generating text data which is then used to train the AI tool of ChatGPT for further knowledge generation. The simulation network environment leverages multiple open-source solutions installed on dedicated physical hardware shown in Figure 1 below.

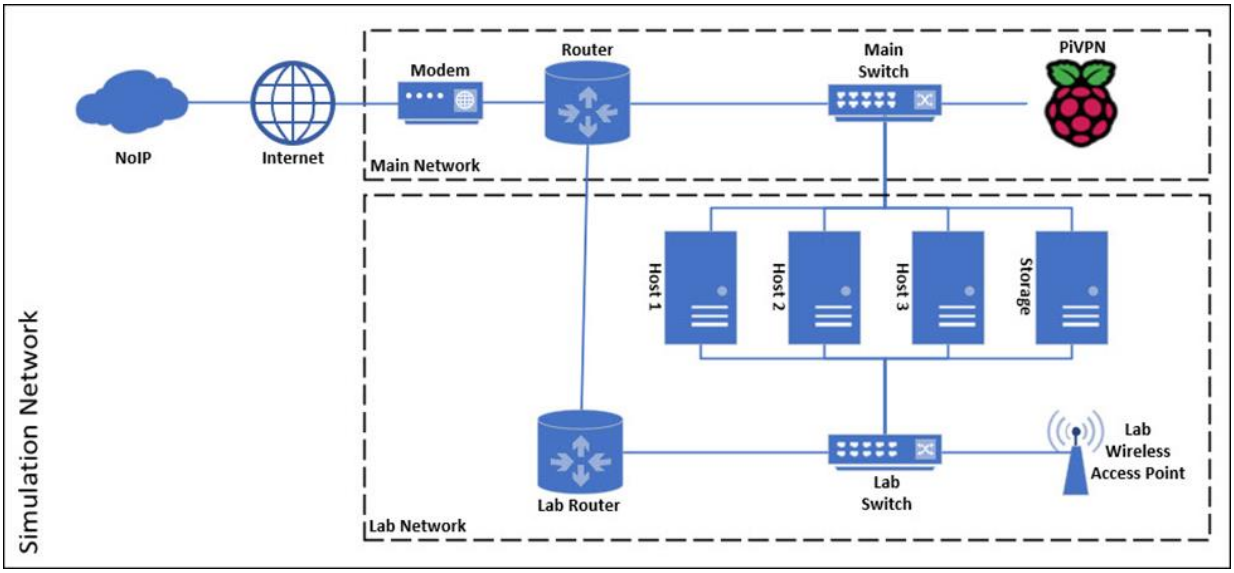

Figure 1: Lab Diagram and Hardware

Hosts 1, 2, and 3 are running Proxmox Virtual Environment 8.1.4; a Type 1 hypervisor. The Storage host is running TrueNAS Core 13.0-U6.1, providing network attached storage for the Proxmox cluster. Finally, the Lab Router hardware is running pfSense Community Edition 2.7.2. A dedicated switch and wireless access point round out this dedicated lab environment; providing the network for any virtual machines (VMs), and any physical devices that are intended to interact with them.

Pro

Tab he simu

Name	IP Address	Operating System	Installed Tools
100	192.168.1.204	Windows 11 Pro 23H2	WinSCP, Putty
102	192.168.1.201	Ubuntu Desktop 22.04.4 LTS	ChatGPT Desktop Client
103	192.168.1.203	Kali Linux 2024.1	Metasploitable Framework/Wireshark/Nmap
107	192.168.1.209	Ubuntu Desktop 22.04.4 LTS B	Pentbox
300	192.168.1.210	Metasploitable2 VM	Metasploitable 2

This study leverages Kali Linux tools installed on VM 103, including Metasploit Framework and Wireshark. The Metasploitable2 VM (300) includes many known security vulnerabilities. Table 4 below lists the selected vulnerabilities to be exploited by Metasploit Framework running on the Kali Linux VM (103). While the exploit was running, Wireshark running on VM 103 captured packets in the lab network. These captures were exported to text files which in turn were used to train ChatGPT. In addition, a honeypot was set up on VM 107 using PentBox, and a Wireshark packet capture of this activity was also exported for the purpose of training ChatGPT.

Table 4: Vulnerabilities for Simulated Exploitations

Vulnerability	ID	Packet Capture File
FTP Backdoor	CVE-2011-2523	vsftpdcapture.txt
Slowloris DOS attack	CVE-2007-6750	slowloriscapture.txt
Samba User Name Script	CVE-2007-2447	sambausernamemapscript capture1.txt
Brute Force on VNC	CVE-1999-0502	vncloginscannercapture.txt
Honeypot Pentest	N/A	honeypotcapture.txt

The packet capture files generated by the simulation were transferred to VM 102. These files were then uploaded to ChatGPT 4 using a desktop client. Initial attempts using the native Wireshark 'pcap' format were unsuccessful; as it did not have access to libraries needed to parse .pcap files. Noting this as a current limitation of ChatGPT, the training strategy changed to use .txt files instead. Initial attempts using .txt files for longer packet captures were also unsuccessful; so subsequently, shorter captures were made to reduce the amount of data in each file. This is another limitation found with ChatGPT. Once .txt files of appropriate size were obtained from the above simulation, the ChatGPT desktop client accepted them as training data.

Findings and Discussions

Answers to specific prompts on ChatGPT-4 pre-trained with the pentesting simulation data were obtained as discovery of generated knowledge for cyber defense. For the **vsftpdcapture.txt** capture, ChatGPT was asked to analyze and determine if there are any backdoor vulnerabilities. ChatGPT reported the following points of interest in this capture file:

- FTP traffic from VM 103 to VM 300 with commands such as 'USER' and 'PASS', indicating an FTP login attempt.
- Commands transmitted over FTP indicating attempts to execute commands or script actions after login.

ChatGPT was not resolute in its response, however, as it qualified its analysis with the following:

> *These observations could indicate potential security risks or misconfigurations that might be exploited as backdoors. However, confirming whether these are indeed vulnerabilities or simply benign activities typical in your network environment would require further context about the network's expected traffic, security measures in place, and any corresponding logs or system behaviors observed during these transmissions (OpenAI, 2023).*

For the **slowloriscapture.txt** capture, ChatGPT was asked to analyze and determine if there are any Denial of Service (DOS) vulnerabilities.

Figure 2 below is a sample capture of the original query prompt on ChatGPT regarding the Slowloris DoS vulnerability, and Figure 3 below captures the answer output from ChatGPT as a sample of knowledge extraction the LLM-based AI tool.

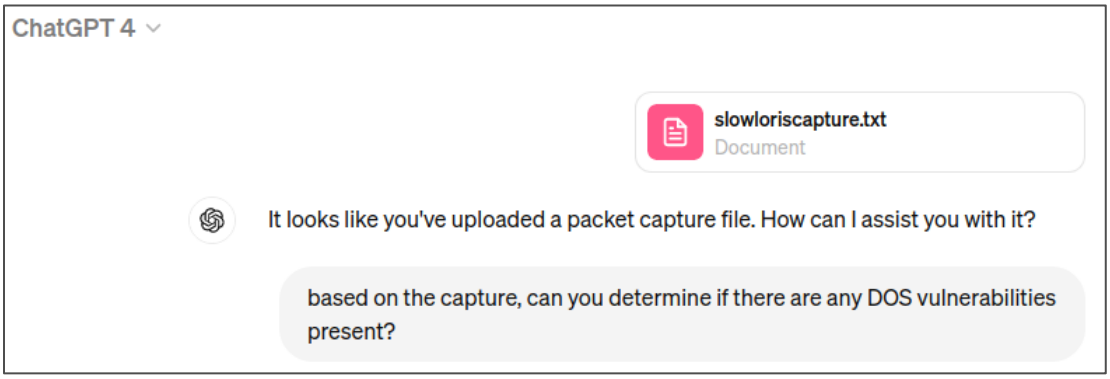

Figure 2: Sample Prompt on DoS on ChatGPT

For this query, ChatGPT provided the following answer:

- A significant number of small packets being sent rapidly from a single source IP (VM 103) to a destination IP (VM 300) on port 80.
- Each packet contains minimal amount of data and is flagged with [PSH,ACK], indicating that each packet is pushing data to be processed immediately.

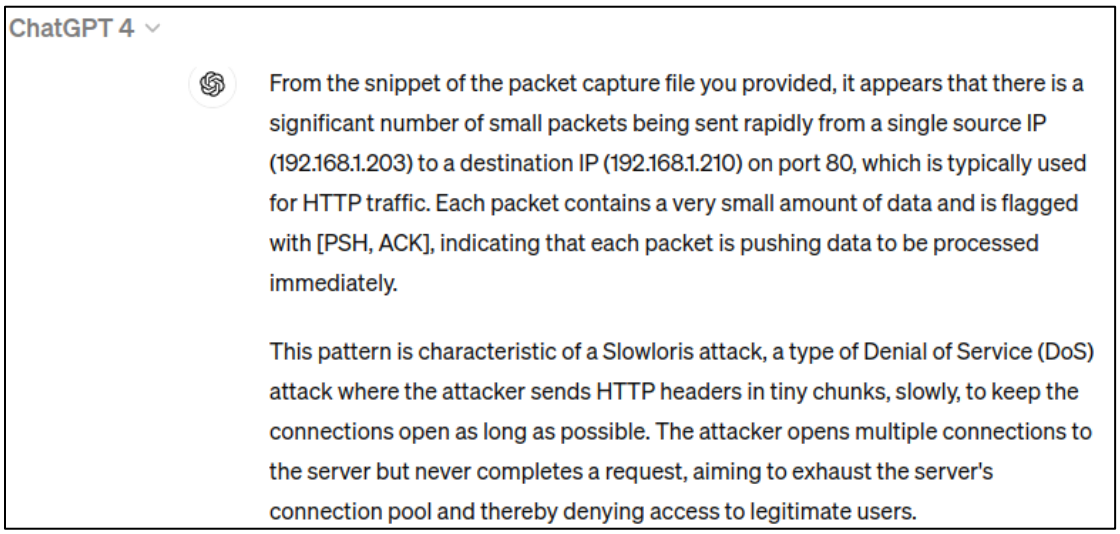

Figure 3: Sample Output on DoS from ChatGPT

ChatGPT did report that this pattern is characteristic of a Slowloris DOS attack. Again, ChatGPT provided indicators to look for when determining if a DOS attack is happening in an environment. ChatGPT also qualified its analysis on this capture.

> *This type of activity can indeed indicate a DoS vulnerability, particularly if the targeted server is not configured to handle or limit this kind of abuse, such as by setting timeouts or maximum connection limits per IP. If this traffic pattern is unexpected and not part of a stress test, it could be a real attack or a security test to check the server's resilience against Slowloris-type attacks (OpenAI, 2023).*

For the **sambausernamemapscriptcapture1.txt** capture, ChatGPT was asked to analyze and determine if there are any samba vulnerabilities present. For this query, ChatGPT provided the following answer:

- Traffic on port 139, associated with Server Message Block (SMB) used by Samba for network file and printer sharing.
- Traffic on port 4444, often associated with non-standard applications.
- Payloads contained commands like 'whoami,' 'uid,' and 'gid' which may indicate attempts to gain elevated access.

ChatGPT once again qualified its analysis with the following statement:

> *From the detailed review of the capture snippets, while there are many Samba-related commands, without specific knowledge of what's normal for this network and without additional context such as logs, system version information, and configurations, it's hard to definitely say there's a vulnerability being exploited just from the network capture (OpenAI, 2023).*

For the **vncloginscannercapture.txt** capture, ChatGPT was asked to analyze and determine if it saw evidence of a brute force attack. For this query, ChatGPT provided the following answer:

- A sequence of attempts of VNC connection attempts between VM 103 and VM 300, involving multiple initiation of TCP connections followed by VNC authentication processes.
- Repeated TCP connection initiation ("SYN") and response ("SYN,ACK").
- Repeated authentication challenge from the server

ChatGPT was most resolute in its analysis with this simulation example. Given these patterns, it seems plausible that this capture could be documenting a brute-force attack against the VNC server (OpenAI, 2023). The analysis report was generated quickly. The knowledge and its confidence levels could be helpful knowledge for cybersecurity professionals in assessing and managing the similar risks in the real world.

For the **honeypotcapture.txt** capture, ChatGPT was asked to analyze and determine if a honeypot was present on the network. For this query, ChatGPT acknowledged the following activity:

- A large number of SYN requests to a wide range of ports on the same destination IP (from port 80 to 3389 and beyond).
- Numerous RST (reset) packers following the SYN requests, indicating that the ports are closed.

ChatGPT did attribute this activity as typical for a honeypot, but qualified this analysis again with the following statement:

> *While these patterns might suggest the presence of a honeypot, it's also crucial to consider other explanations such as standard security measures, firewalls, or inactive services on a network. A detailed analysis of the network's configuration and purpose would be necessary to confirm if these observations indeed point to a honeypot (OpenAI, 2023).*

Table 5: Summary of Answers and Knowledge from ChatGPT

Vulnerability	Key Indicators	Mitigation Strategies
ftp-vsftpd-backdoor	• Unexpected or suspicions outbound connections • Non-standard port usage • Unusual IP addresses or domains in network traffic	• Ensure services are up-to-date • Use strong, non-default credentials • Secure critical protocols with secure encryption
Slowloris DOS attack	• Repeated small payload sizes • Rapid succession of packets to the same destination on a web service port	• Implement rate limiting • Implement connection timeouts • Use reverse proxies, or • Specialized firewall rules
Samba User Name Script	• Outdated Samba versions • Increased SMB traffic, especially involving unusual ports or high volumes of data transfer • Misconfigurations, such as weak passwords and lack of encryption	• Check versions of Samba and compare against known vulnerabilities listed in databases like CVE • Review system and security logs for signs of unauthorized access or suspicious activities
Brute Force on VNC	• Repeated failed login attempts from the same source • Different credentials being used from the same IP address	• Password policies requiring complex and lengthy passwords • Implement account lockout policies after a certain number of failed login attempts
Honeypot Pentest	• High interaction with unknown services • Repetitive patterns of traffic • Responses from normally closed ports	• Great for detecting unauthorized access or malicious activity within a network • Monitoring interactions with a honeypot can warn organizations earlier of a potential attack or emerging threats

Table 5 above summarizes ChatGPT output answers to queries on various security network security vulnerabilities. The LLM-based AI tool of ChatGPT pre-trained with the simulation pentesting data was able to provide quick answers and reports on essential information, threat indicators, and further suggestions on mitigation strategies to consider when looking for these vulnerabilities in an environment. The efficient knowledge from ChatGPT would be able to enhance the efficiency and accuracy of the pentesting process and vulnerability analysis for risk assessment and mitigation in cyber defense.

Limitations of ChatGPT

This research exposes certain limitations of ChatGPT in pentesting that will be of future research interests. While the generated answers from pre-trained ChatGPT provide helpful information and suggestions on pentesting for cyber knowledge discovery, a major limitation is that the level of certainty and confidence is fairly low, showing tentativeness not helpful for cybersecurity decision making. Another limitation is that most of the suggestions on mitigation strategies are fairly general and generic and lack specific details.

The LLM-based ChatGPT has some technical limitations as an evolving AI solution. For example, ChatGPT was not able to process the initial attempt to process .pcap files from Wireshark. Figure 4 below shows this limitation.

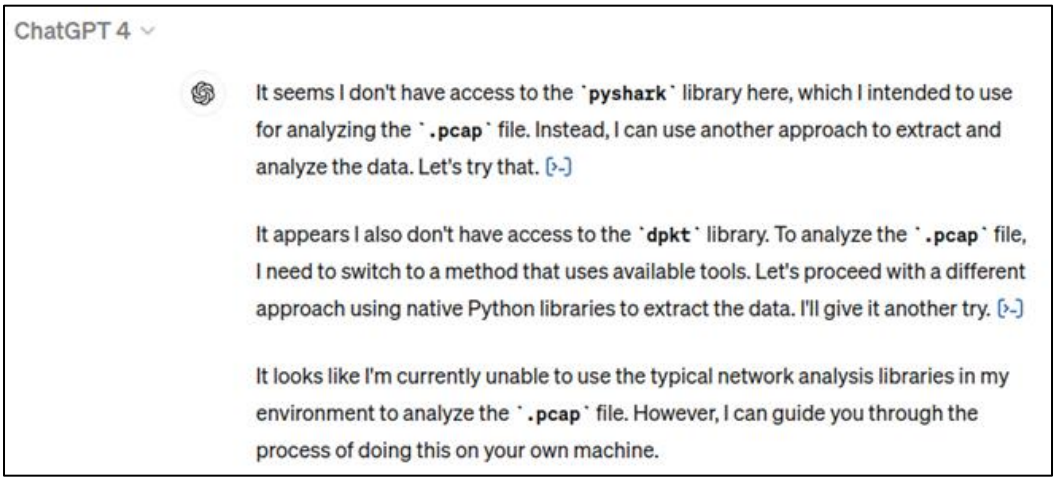

Figure 4: ChatGPT Limitation for Processing .pcap Files

Another limitation of ChatGPT is its inability to process large text files of longer network captures. Figure 5 below shows captures this limitation.

Figure 5: ChatGPT Limitation for Processing Large Text Files

Conclusion

Knowledge has been a powerful factor for effective decision making. Knowledge discovery of the strengths and vulnerabilities of oneself and the opponent has been critical to effective defense and victory in the traditional military warfare as shown in Sun Tzu's *The Art of War*. Similar strategies and principles on knowledge and knowledge discovery still apply to modern cyber defense. Penetration testing is a common process of knowledge discovery for intelligence gathering and vulnerability analysis to mitigate and manage cyber risks. Fast-growing artificial intelligence technology may enhance the efficiency and effectiveness of

knowledge discovery through pentesting even though it may still have limitations and risks. This study presents an adapted model of AI-moderated Knowledge Discovery for cyber defense. The model is tested and illustrated using a virtual network pentesting simulation for data collection to train the LLM-based AI tool of ChatGPT on various network vulnerabilities. Subsequent queries and generated answers from ChatGPT indicate efficient knowledge and information on the vulnerabilities with suggested strategies for mitigation. The knowledge from pre-trained ChatGPT shows that AI solutions may enhance the pentesting process for cyber defense. The simulation test also shows some limitations of the AI tool.

Both cyber defense and AI models and solutions are rapidly evolving areas with new issues and ideas for constant research and development. This study is based on limited virtual network simulation tests and limited data to train the ChatGPT-4 AI solution. Further research with new and increased training datasets is needed to explore the consistency of the moderating effects of AI solutions and tools, including their positive effects, limitations, challenges, and risks. Future studies may also attempt to include and compare different AI models, tools and techniques to help with pentesting for more effective knowledge discovery and generation for cyber defense.

References

Al-Hawawreh, M., Aljuhani, A., & Jararweh, Y. (2023). ChatGPT for cybersecurity: practical applications, challenges, and future directions. *Cluster Computing (2023), 26*:3421–3436

Booker, L.B., & Musman, S.A. (2019). A model-based, decision-theoretic perspective on automated cyber response. Retrieved May 10, 2024 from https://arxiv.org/abs/2002.08957

Clarke, R.A., & Knake, R.K. (2010). Cyber war: The next threat to national security and what to do about it. New York, NY: HarperCollins Publishers

Dsouza, M. (2018). How artificial intelligence can improve pentesting. Retrieved May 4, 2024, from https://hub.packtpub.com/how-artificial-intelligence-can-improve-pentesting/

Dunn, J.E. (2024, April 9). AI will power a new generation of ransomware, predicts Britain's NCSC. Retrieved May 4, 2024 from https://ransomware.org/

Gupta, M., Aryal, K., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy," in *IEEE Access, vol. 11*, 2023, pp. 80218-80245

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

Kaur, R., Gabrijelcic, D., & Klobucar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions, *Information Fusion, 97* (2023) 101804, pp. 1-29.

Kelly, C., Pitropakis, N., Mylonas, A., McKeown, S., & Buchanan, W.J. (2021). A comparative analysis of honeypots on different cloud platforms. *Sensors, 21*(2433), 1-19.

Mamgai, A. (2023, October 16). *Generative AI with cybersecurity: Friend or foe of Digital Transformation?*. ISACA. https://www.isaca.org/resources/news-and-trends/industry-news/2023/generative-ai-with-cybersecurity-friend-or-foe-of-digital-transformation

Michaelson, G., & Michaelson, S. (2003). *Sun Tzu for success.* Avon, MA: Adams Media Corporation.

M'manga,A., Faily,S., McAlaney, J., Williams, C., Kadobayashi, Y., & Miyamoto, D. (2019). A normative decision-making model for cyber security. *Information & Computer Security, 27*(5), 636-646.

Muckin, M., & Fitch, S.C. (2019). A threat-driven approach to cyber security. Retrieved May 10, 2024 from https://www.lockheedmartin.com/

OpenAI. (2023). ChatGPT. Retrieved May 10, 2024 from https://openai.com/

Sun, T. (1910). *The art of war* (L. Giles, Trans.). BookYards.com.

Takko, T., Bhattacharya, K., Lehto, M., & Jalasvirta, P. (2023). Knowledge mining of unstructured information: application to cyber domain. *Scientific Reports 13(1):*1714.

Temara, S. (2023, March 21). *Maximizing penetration testing success with effective reconnaissance techniques using ChatGPT*. Research Square. https://www.researchsquare.com/article/rs-2707376/v1

The White House. (2023, March). National Cybersecurity Strategy. Retrieved May 3, 2024, from https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Cybersecurity-Strategy-2023.pdf

Wang, H., & Lin, W. (2018). Analysis of the art of war of Sun Tzu by text mining technology. *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS),* 626-628.

Wang, P., & D'Cruze, H. (2020). Lessons on the power of knowledge for cyber defense from Sun Tzu's *The Art of War*. *Issues in Information Systems, 21* (3), 105-116.

Wang, P., & D'Cruze, H. (2021). Honeypots and knowledge for network defense. *Issues in Information Systems, 22*(3), 241-254.

Wang, P., Liu, J., Zhong, X., & Zhou, S. (2023). A cybersecurity knowledge graph completion method for penetration testing. *Electronics 2023, 12, 1837*. https://doi.org/10.3390/electronics12081837

Wilson, R. (2018). Sun Tzu and the art of cyberwar. *Defense AT&L, January-February 2018,* 30-34.

Wilson, S. (2023). *Cybersecurity and artificial intelligence: Threats and opportunities*. Contrast Security.

Zhan, X., Xu, Y., & Sarkadi, S. (2023, July 1). Deceptive ai ecosystems: The case of ChatGPT. *Proceedings of the 5th international conference on conversational user interfaces*. ACM Conferences. https://dl.acm.org/doi/10.1145/3571884.3603754