ChatGPT for Cybersecurity: Automate, Defend & Secure with AI

www.labcyber.com

In today's rapidly evolving cybersecurity landscape, professionals must analyze threats, respond to incidents, and automate security tasks faster than ever. ChatGPT is changing the game, offering powerful AI-driven assistance to streamline security workflows, enhance threat intelligence, and improve response strategies.

This guide book will show you how to leverage ChatGPT for real-world cyber security applications—from threat intelligence analysis and vulnerability management to incident response automation and phishing simulations. You'll learn how to use AI to extract Indicators of Compromise (IoCs), summarize security reports, generate risk assessments, and even assist in writing Python scripts for cybersecurity automation

Table of Contents

How CHATGPT Works

- ChatGPT is an AI-powered chatbot developed by OpenAI, designed to process natural language and generate human-like responses.

- It is based on a large language model (LLM) that has been trained on diverse text data to assist users with various tasks, including research, content creation, and cybersecurity applications.

- ChatGPT is trained using a mix of supervised learning and reinforcement learning from human feedback (RLHF).

- It learns language patterns by analyzing massive datasets from books, articles, and other internet sources.

- ChatGPT predicts the most likely next words based on context.

- It does not "think" or "understand" but generates responses based on patterns.

Training Process

- Phase 1 - ChatGPT is first trained on a massive dataset containing books, articles, websites, scientific papers and code repositories.

- During pre training, the model is not explicitly programmed with rules. Instead, it learns by predicting the next word in a sentence. This is known as the self-supervised learning approach

- Phase 2 - ChatGPT learns to break words into a sequence of tokens allowing it to analyze meaning based on context.

- “Cybersecurity is important !” will be broken into ["Cyber", "security", " is", " important", "!"]

- Phase 3 - Reinforcement Learning from Human Feedback (RLHF)
    - Human reviewers provide example conversations and rank different responses based on quality.
    - A reward model is trained to rank responses.
    - The AI is further trained using reinforcement learning to generate better responses over time.

- Phase 4 - Continuous Learning (Fine-Tuning)
    - OpenAI may periodically fine-tune the model using user interactions.

    - However, ChatGPT does not learn from individual user sessions—it does not retain memory of past chats unless specifically fine-tuned by OpenAI.

Prompt Engineering for CHATGPT

ChatGPT's effectiveness depends on how well you phrase your queries. Proper prompt engineering ensures clear, accurate, and useful responses.

- Be Specific and Detailed
    - Instead of: "Tell me about malware."
    - Use: "Explain different types of malware and their attack vectors in a corporate network."

- Define the Format
    - Instead of: "List cybersecurity threats."
    - Use: "Provide a table comparing the three most common cyber attacks, including attack methods and mitigation strategies."

- Ask Step by Step
    - Instead of: "Explain incident response."
    - Use: "Break down the five stages of the incident response lifecycle with real-world examples."

- Refine Responses with Follow-ups
    - If ChatGPT gives a vague or generic response, refine it:

    - "Can you provide more details on how SIEM tools integrate with incident response?

Ensuring Accuracy & Reliability

- Always verify ChatGPT's responses using:
    - Official cybersecurity resources (e.g., NIST, MITRE, OWASP, CIS).
    - Security vendor reports (e.g., Cisco, Palo Alto, CrowdStrike).
    - Threat intelligence platforms (e.g., VirusTotal, Shodan, Recorded Future).

- ChatGPT should assist in drafting reports, policies, or scripts, but a human should review and approve them.

- *Example: If ChatGPT generates a Python script for log analysis, test it in a sandbox environment before deploying*.

- ChatGPT sometimes "hallucinates" (creates fake information).

- Red flags:
    - Incorrect cybersecurity terms.
    - Fabricated vulnerability identifiers (e.g., fake CVE numbers).
    - Outdated or misleading attack mitigation steps.

- Best Practice: Ask ChatGPT to provide sources or compare its response with verified industry documentation.

- Pair ChatGPT with Threat Intelligence Platforms
    - Use Shodan, VirusTotal, or AlienVault OTX for real-time threat intelligence.
    - Ask ChatGPT to explain results, but always verify through external sources.

- Use ChatGPT to Enhance Security Automation
    - AI can assist in writing scripts for SIEM automation, log analysis, or firewall configurations.
    - However, all AI-generated code must be reviewed and tested before deployment.

- Avoid Sharing
    - Personally Identifiable Information (PII) (employee/customer names, phone numbers, passwords)
    - Company security policies or classified reports
    - Real-world network configurations or incident logs

- Replace real IP addresses, usernames, and file paths with placeholders.

- Remove timestamps if they indicate internal event timing.

- Use dummy data that mimics real attack patterns but does not expose actual network details.

Using CHAGPT for Threat Intelligence

Summarizing Threat Reports

- *Summarize the following threat intelligence report into 5 key bullet points, focusing on the target, attack method, malware behavior, identified IoCs, and mitigation recommendations*

- *From the same report, Extract all IoCs (IP addresses, domains, file hashes) and display them in a structured format*

- *Based on the same threat intelligence report, suggest a step-by-step mitigation plan to address the identified threats*

- *Summarize the report in no more than 300 words in plain language that will be easy to understand by a non-technical executive.*

Threat Actor Profiling

- *Analyze the behavior and known tactics of a ransomware threat actor named ‘BlackFog’. List their common tools, attack patterns, and motivations*

- *Compile all the information about 'Conti' into a clear and structured threat actor profile suitable for a cybersecurity report*

- *Based on Conti's techniques and tools, suggest practical countermeasures for an organization to defend against their attacks.*

- *Using the MITRE ATT&CK framework, map Conti's known attack techniques*

Identifying Patterns and Trends in reports

- *From the following report, Identify trends in attack frequency and types from the report. Summarize the findings and suggest actions to mitigate future attacks*

- *Analyze the following two security incident reports. Identify:*

    - *Recurring IP addresses and the attack types they are associated with.*

- *Commonly targeted servers across both reports.*
- *Trends in attack frequency for each attack type over the week.*
- *Any unique or distinct findings in each report.*

Incident Response with CHATGPT

Using ChatGPT for Preparation

- *Create an Incident Response Plan outline for a medium-sized organization, focusing on handling ransomware attacks.*

- *Create a comprehensive incident response plan for a financial institution to handle data breaches. The plan should include objectives, roles and responsibilities, communication protocols, incident detection, containment, eradication, recovery and post-incident analysis. Ensure the plan follows best cybersecurity practices and regulatory compliance guidelines.*

- *Write a playbook for responding to a phishing attack that includes steps for containment, employee*

communication, and recovery.

- *Create a training scenario where employees must respond to a data breach caused by unauthorized access.*

- *Create a tabletop exercise scenario for a data breach caused by a phishing attack. Include questions for participants to discuss during each phase of incident response.*

- *Provide a checklist to prepare an organization for potential malware attacks, including preventive measures.*

- *Write an email template to notify employees about a phishing email targeting the company. Include steps they should take.*

- *Draft a press release informing customers about a data breach and the steps the company is taking to resolve it.*

Identification with ChatGPT

- *Analyze this server log and identify any unusual patterns.*

- *Cross-check these IP addresses with known malicious IPs:*

203.0.113.15, 198.51.100.23, 192.168.1.10

- *Analyze this email header and determine if it's a phishing email.*

- *Analyze these security alerts and identify any potential threats.*

Containment with ChatGPT

- *A suspected malware infection is spreading in our network. What immediate containment steps should we take?*

- *What are the best containment strategies for stopping a ransomware attack before it encrypts more files?*

- *We detected unauthorized access to a privileged account. What are the containment steps to prevent further damage?*

- *An employee's credentials were exposed in a data breach. How can we contain the risk before damage occurs?*

- *Help me draft containment actions for an ongoing DDoS attack on our company's website.*

- *Generate a Python script to automatically block malicious IPs on our firewall*
- *Write a PowerShell script to disable compromised user accounts in Active Directory*
- *Generate a containment playbook for a phishing attack targeting employees*
- *Create a decision tree for containment actions based on different attack types (e.g., malware, SQL injection, DDoS)*
- *What immediate actions should be taken if an attacker is exploiting a zero-day vulnerability in our web application*
- *Generate an internal security alert email instructing employees on containment measures for a phishing attack*
- *Draft a management briefing explaining the containment actions taken for a malware outbreak*

Eradication and Recovery with ChatGPT

- *Provide a step-by-step guide to remove malware from a Windows system.*

- *Write instructions for patching a vulnerability in Apache servers.*

- *Create a checklist for restoring a system from a backup after a ransomware attack.*

- *Write an email to notify employees that systems are restored and operational after a malware attack.*

- *Create a plan to verify that systems are secure and operational after recovering from a data breach.*

Post-Incident Analysis with ChatGPT

- *Write a post-incident report for a ransomware attack, including a summary, timeline, root cause, and recommendations.*

- *Analyze this Incident Response timeline and suggest improvements to reduce response time.*

- *Provide recommendations to improve email security and employee training to prevent future malware attacks.*

- *Summarize the lessons learned from this brute-force attack and suggest preventive measures.*

- *Create a comprehensive Incident Response Plan for a financial institution to handle data breaches. The plan should include objectives, roles and responsibilities, communication protocols, incident detection, containment, eradication, recovery, and post-incident analysis. Ensure the plan follows best cybersecurity practices and regulatory compliance guidelines.*

Vulnerability Management with ChatGPT

- *Explain the following CVE description in simple terms: 'Buffer overflow in XYZ component allows remote attackers to execute arbitrary code*

- *Summarize this CVE in plain language CVE-2025-23780*

- *Explain CVE-2025-23780 and its impact on a WordPress website using WooCommerce*

- *From the following Nessus report summarize the top three critical vulnerabilities and their potential impacts*

- *Explain the most critical vulnerability in simple terms for a non-technical manager*

- *Generate an executive summary for a COO explaining these three vulnerabilities that require immediate action. Use non-technical language, focusing on business impact and risk mitigation. Summarize in under 250 words*

    - *1. CVE-2023-54321 (Critical) - Remote code execution affecting payment systems.*
    - *2. CVE-2022-98765 (High) - Unauthorized access flaw in HR systems.*
    - *3. CVE-2021-45678 (Medium) - Outdated SSL/TLS affecting secure login.*

- *Given a CVSS score of 8.5, explain the severity and suggest prioritization steps for a database server vulnerability.*

- *A high-risk vulnerability exists on a rarely used admin portal, while a medium-risk vulnerability is found on a customer-facing website. Which should be addressed first?*

- *Prioritize the following vulnerabilities considering that the database server holds customer financial data, the web server is public-facing, and the Windows server is used internally*

    - *1. CVE-2022-87654 (CVSS 8.5) - Remote code execution on an internal Windows server.*

    - *2. CVE-2021-65432 (CVSS 6.0) - Outdated Apache version on a public-facing website.*
    - *3. CVE-2023-98765 (CVSS 7.8) - SQL Injection vulnerability on a customer database.*

- *For a healthcare organization dealing with sensitive patient data, prioritize these vulnerabilities based on HIPAA compliance risk:*

    - *1. CVE-2023-54321 (CVSS 7.8) - Unencrypted patient data transmission.*
    - *2. CVE-2022-98765 (CVSS 8.3) - Unauthorized access vulnerability in patient portal.*
    - *3. CVE-2021-45678 (CVSS 6.7) - Outdated SSL/TLS certificates used for login authentication.*

- *From the following Nessus report, for each vulnerability, provide a step-by-step remediation plan*

- *Provide a step-by-step remediation plan for CVE-2023-12345*

- *Generate a compliance report for CVE-2023-65432, including risk assessment, mitigation measures, and patching details*

- *Compare the remediation timelines for vulnerabilities in Microsoft Exchange vs. OpenSSL*

Automation & Scripting with ChatGPT

- *Generate a Python script that reads system logs and detects multiple failed login attempts from the same IP address.*

- *Generate a Python script that checks if specific services (SSH, FTP, HTTP) are running on a remote server and logs the results.*

- *Create a Python script that monitors file changes in a specific directory and logs any modifications or new files*

- *Create a script that integrates with VirusTotal API to scan a list of file hashes and return their threat status*

- *Write a Python script that automatically generates a daily cybersecurity tip and sends it to employees via email*

- *Modify this script to include better error handling and logging for security purposes*

- *Optimize this security script to run faster and use fewer system resources*

- *Rewrite this automation script to integrate with Slack and send real-time security alerts*

Security Awareness and Training Content with ChatGPT

- *Generate a phishing email from fake IT support asking employees to reset their passwords urgently*

- *Generate a conversation script where an attacker impersonates an IT technician and tries to convince an employee to reveal their login credentials.*

- *Write a scenario where an employee encounters a stranger trying to enter a restricted office area without a badge*

- *Generate a ransomware message demanding Bitcoin payment in exchange for decrypting company files*

- *Write a 2-minute explainer video script on how employees can protect their accounts from phishing attacks*

- *Write a step-by-step tutorial script for a video explaining how to enable 2FA on Microsoft 365*

- *Create a cybersecurity challenge where employees must analyze an email and determine if it is a phishing attempt.*

- *Create badge descriptions for a security awareness program where employees earn rewards for completing different levels of training*

- *Generate 10 cybersecurity trivia questions for a workplace competition*

- *Create a phishing simulation email to test employees' ability to recognize fraudulent messages*

- *Analyze employee security training results where 30% of employees fell for a phishing test. Provide recommendations for improving awareness*

- *Write a 10-question anonymous survey to assess employee perceptions of cybersecurity training.*
- *Generate a daily cybersecurity tip for employees to receive via email or Slack.*

- *Generate an automated warning message for employees who receive a phishing email*

- *Generate a friendly reminder for employees to update their passwords every 90 days*

- *Write an automated message reminding employees to enable 2FA*

- *Create a chatbot response for an employee who asks how to report a security incident*