# Getting your hands dirty: Exploring Exploits with ChatGPT

## Introduction

### Who Am I?

Blake is a data scientist, software engineer, and cybersecurity expert, has had a diverse career spanning roles at NASA, Oak Ridge National Laboratory, SecureSet, Swimlane, and Revelstoke Security. From designing algorithms and optimizing software at NASA to pioneering data-driven solutions at Oak Ridge and teaching cybersecurity at SecureSet, Blake has made a significant impact across the country. Currently, through his pivotal roles at Swimlane and Revelstoke Security, Blake is driving innovation in the cybersecurity industry.

### History

#### How Things Got Here

ChatGPT's swift achievement of 100 million users can be attributed to its exceptional conversational abilities, continuous learning, democratized access, versatility, user-friendly integration, and reliable performance. These factors have set it apart from other technologies and driven its widespread adoption across diverse domains.

#### What is ChatGPT?

ChatGPT is a language model developed by OpenAI, based on the GPT (Generative Pre-trained Transformer) series of models. GPT models are specifically designed to generate human-like text by leveraging contextual input. ChatGPT has undergone training on a vast corpus of text from the internet, enabling it to provide conversational responses that are contextually relevant and coherent.

### ChatGPT as a Cybersecurity Tool

- ChatGPT can be harnessed as a valuable cybersecurity tool in various ways:

    - Information and Education: It can serve as a resource hub, providing valuable knowledge and resources on cybersecurity topics.
    - Threat Intelligence and Analysis: ChatGPT assists users in understanding and analyzing cyber threats, helping to identify potential risks and vulnerabilities.
      Security Policy and Compliance: It offers guidance on security policies and compliance standards, aiding organizations in establishing robust security frameworks.
    - Vulnerability Assessment: ChatGPT provides insights into common security vulnerabilities, assisting in the identification and mitigation of potential weaknesses.
    - Secure Coding Practices: It offers examples and advice on writing secure code, promoting the adoption of best practices for developing secure software.
    - Incident Response and Forensics: ChatGPT provides general guidance on incident response procedures, assisting in the handling and investigation of cybersecurity incidents.

By leveraging ChatGPT's capabilities in these areas, users can enhance their cybersecurity posture, improve their understanding of threats and vulnerabilities, and establish proactive measures to protect their digital assets.

## Goals of Today's Session

### Communicate Key Takeaways

The goal of today's session is twofold. First, we aim to communicate key takeaways that will deepen your understanding of AI's role in cybersecurity, including the utilization of ChatGPT. One important takeaway is the significance of prompt engineering, understanding how to effectively phrase questions or prompts to elicit desired responses from the model. This skill enables users to maximize the value and accuracy of the information provided by ChatGPT. Second, we want you to see ChatGPT in action, showcasing its capabilities and how it can assist in securing digital systems. By observing real-world examples and demonstrations, you'll gain insights into the practical applications of ChatGPT and its potential to enhance cybersecurity practices.

# Prompt Engineering

## Jailbreaking ChatGPT

Jailbreaking ChatGPT involves pushing the model's limits and experimenting with unconventional prompt styles. It unlocks new possibilities but can also produce less reliable outputs. It's important to exercise caution, maintain ethical boundaries, and respect OpenAI's policies while exploring ChatGPT's capabilities.

## Exploring Prompt Engineering Techniques

Prompt engineering techniques are essential for optimizing the performance of ChatGPT. This section delves deeper into the various techniques involved in prompt engineering. It explores the importance of formatting and structuring prompts effectively, providing explicit instructions to guide the model's response, and incorporating relevant context to improve comprehension. Additionally, it emphasizes the iterative nature of testing and refining prompts based on feedback and evaluation, highlighting the continuous improvement process.

## Understanding Different Prompt Styles

Introducing participants to different prompt styles enhances their understanding of ChatGPT's capabilities. This section explores various prompt styles, such as DAN (Do Anything Now), STAN (Strive To Avoid Norms), DUDE (Do Anything Now), and Mongo Tom. Each style offers a unique approach to engaging with ChatGPT and elicits specific response patterns. Examples and prompts are provided to allow participants to experiment with these styles and observe the varying outputs they generate.

## Maintaining Character and Maximizing Output

Staying in character is crucial when utilizing different prompt styles and avoiding censorship. This section emphasizes the significance of fully immersing oneself in the chosen prompt style, aligning with its intended tone and characteristics. By remaining consistent and authentic, participants can maximize the desired output from ChatGPT. Practical guidance is provided to help participants effectively embody the selected prompt style and generate the desired responses while ensuring ethical and responsible usage.

## ChatGPT Prompting for Cybersecurity

### Modes and Roles:

The available modes and roles in prompt engineering include the following:

| Roles | Description | Example Prompt |
|---|---|---|
| Intern | Assisting with research on cybersecurity topics. | Find recent studies on the impact of AI in cybersecurity. |
| Idea Generator | Generating creative and innovative ideas for projects. | Generate 5 unique strategies to enhance network security. |
| Editor | Reviewing and editing written content for accuracy and clarity. | Edit this cybersecurity report for grammar and coherence. |
| Teacher | Educating and instructing others on cybersecurity principles and best practices. | Teach me about encryption algorithms and their applications. |
| Critic | Providing constructive feedback and analysis on cybersecurity strategies and arguments. | Critique this argument on the effectiveness of password managers. |

The different formats that can be used in prompt engineering are:

| | | |
|---|---|---|
| Code | Table | Essay |

| | | |
|---|---|---|
| Tweet | Blog | Report |
| Email | Presentation | Social Media Post |
| Bullets | Research | |

When it comes to tones, prompt engineering allows for various options such as:

| | | |
|---|---|---|
| Firm | Professional | Persuasive |
| Confident | Descriptive | Formal |
| Poetic | Humorous | Informal |
| Narrative | Academic | Friendly |

## How to Build a Chain Prompt with Example:

- Insert first prompt: "Give me a summary of this cybersecurity document" (insert or copy-paste document text).
- Modify the output: "Use the summary above and write a 500-word piece that explains the cybersecurity topic to beginners."
- Modify the tone: "Change the tone of the answer above and make it sound more professional."
- Modify the format: "Convert the answer above into text for a presentation with 1 slide for each key cybersecurity point."

## Prompts for Marketers in Cybersecurity:

- List [insert number] ideas for blog posts about cybersecurity best practices.
- Create a 30-day social media calendar about cybersecurity tips and awareness.
- Generate compelling landing page copy for a cybersecurity product or service.
- Write 5 pieces of Facebook ad copy for a cybersecurity training program.
- Generate 5 persuasive subject lines for an email about cybersecurity threat prevention.

## Prompts for Coding in Cybersecurity:

- Help me find vulnerabilities in my code: [insert your code].
- Explain what this snippet of cybersecurity code does: [insert code snippet].
- What is the correct syntax for secure input validation in Python?
- How do I fix the following security issue in my web application code? [insert code snippet].

## Prompts for Sales in Cybersecurity:

1. Generate 10 strategies to generate leads for a cybersecurity consulting company.
2. Create a personalized sales email for potential customers. Include key cybersecurity benefits and offerings.
3. Write a sales landing page description for a cybersecurity software solution.
4. Generate 5 customer personas for targeting cybersecurity services.
5. Develop a script for cold-calling potential clients interested in cybersecurity products.

## Prompts for Designers in Cybersecurity:

- What are some key user interactions to consider when designing a secure authentication system?
- Create a user persona for a cybersecurity awareness training platform.
- Generate 10 interview questions to gather user insights on cybersecurity user experience.
- Create a user journey for a mobile app focused on secure data transmission.
- Identify UI/UX design requirements for a cybersecurity dashboard application.

## Prompts for Research in Cybersecurity:

- Identify the top 20 cybersecurity companies in terms of market reputation.
- What are the emerging trends in network security for 2023?
- Find me the best-reviewed software for endpoint protection in small businesses.
- Summarize the annual cybersecurity report of [insert company].
- Summarize this research paper on the latest advancements in encryption algorithms.

### Prompts for Customer Service in Cybersecurity:

- Create a template for an email response to customers inquiring about cybersecurity incident response.
- What are the most frequently asked questions about securing a network infrastructure?
- Create a help page that explains how to set up two-factor authentication for your product.
- Summarize the following knowledge base article to provide step-by-step instructions on securing IoT devices: [insert article].

### General Prompts for Cybersecurity:

1. Rewrite this cybersecurity text and make it easy for a beginner to understand: [insert text].
2. I want to enhance the security of my network. Generate 5 ideas for improving network security.
3. Explain secure socket layer (SSL) in simple terms that any beginner can understand.
4. Summarize the text below and provide a list of bullet points with key insights and the most important cybersecurity facts.
5. Proofread my writing above. Correct any grammar and spelling mistakes, and provide suggestions to improve the clarity of my writing.

## Hands-on Lab: Demonstrating Attacks with ChatGPT

### Detailed instructions and URLs for testing SQL injection, XSS, and SSRF attacks:

- Provide participants with specific URLs and endpoints where they can perform hands-on testing of these attacks.
- Explain the purpose of each URL and its vulnerability to the corresponding attack.

### Sample inputs and expected results for each attack scenario:

- Provide a list of sample inputs that participants can use to generate results for each attack.
- Clarify the expected outcomes and highlight the significance of the results in terms of demonstrating the vulnerabilities and potential impact.

### Guidance on exploring the vulnerabilities and understanding the risks involved:

- Encourage participants to delve deeper into the vulnerabilities and understand the risks associated with each attack.
- Provide additional resources or references to support further exploration and learning.

## Wrap-up and Conclusion

### Recap of key takeaways:

Summarize the main points covered throughout the session, including the role of AI in cybersecurity, the importance of prompt engineering, and the demonstration of various attacks with ChatGPT.

### Emphasize the evolving nature of AI:

Discuss how AI technologies, including ChatGPT, are continuously improving and evolving. Highlight the need for responsible use and ethical considerations when leveraging AI in cybersecurity.

### Advantages of Generative AI in Cybersecurity:

- Enhanced defense strategies and response plans
- Improved cyber defenses and vulnerability identification
- Facilitated healing process after attacks
- Faster adoption of preventive measures
- Proactive response to potential threats
- Analyzing large volumes of data and detecting patterns
- Strengthened organization's defense capabilities

### Challenges of Generative AI in Cybersecurity:

- Escalated time to attack and defense
- Red Team's ability to exploit vulnerabilities efficiently
- Difficulty in keeping up with evolving attack landscape
- Potential risks and misuse of generative AI
- Facilitation of malicious activities
- Sophisticated attack strategies by adversaries
- Need for vigilant and updated defense measures

### Encouragement to use AI powers for good:

- Inspire participants to utilize their AI skills and knowledge for positive impact in the field of cybersecurity.

- Emphasize the importance of responsible and ethical AI practices to ensure a secure and sustainable digital landscape.

## Conclusion:

- Positive and negative implications of generative AI in cybersecurity
- Empowered Blue Team with enhanced capabilities
- Escalated challenges in adapting to evolving threats
- Navigating risks and benefits for effective cybersecurity.