ChatGPT's Influence on Cybersecurity: Reduce and Understanding the Risks

Grigorina BOCE,
Department of Information Technology, Mediterranean University of Albania, Tirana, Albania
E-mail address: grigorina.boce@umsh.edu.al

Besmir KANUSHI,
Department of Information Technology, Mediterranean University of Albania, Tirana, Albania
E-mail address: besmir.kanushi@umsh.edu.al

Abstract

The emergence of conversational AI models, such as ChatGPT, has revolutionized human-computer interactions, offering extraordinary capabilities in natural language understanding and generation. However, as these models become increasingly integrated into various applications and platforms, they also introduce new challenges and risks to cybersecurity. This article explores the impact of ChatGPT on cybersecurity, delving into both the opportunities and weaknesses it presents. We discuss the potential use of ChatGPT in various cyber attacks, including phishing, social engineering, and misinformation campaigns. Additionally, we examine strategies for reduce these risks, such as enhancing model security, implementing robust authentication mechanisms, and promoting user education and awareness. By understanding the connections of ChatGPT on cybersecurity, we can proactively address these challenges and hold its capabilities safely and responsibly in the digital landscape.

Keywords: phishing, chatbot, malware.

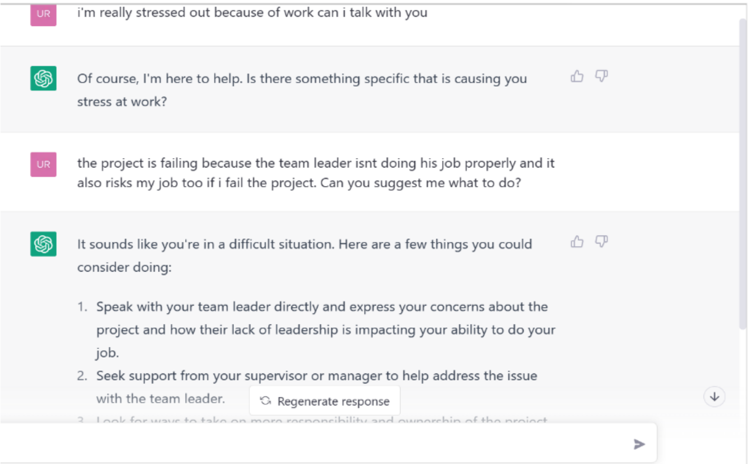
1. Introduction

ChatGPT is a large language model developed by OpenAI that can understand and respond to natural language text. It has been taught on a massive amount of text data and can generate human-like text for a wide range of tasks, such as answering questions, writing stories, and even coding. It can be used to generate text, complete text, or answer questions [1].

ChatGPT is a type of language model that uses neural networks. It is taught on a large dataset of text to generate human-like responses to prompts. It is based on the GPT (Generative Pre-taught Transformer) architecture, which uses a transformer neural network to generate text.

1.1. Example of ChatGPT

An example of a task that ChatGPT can be fine-tuned for is a chatbot that can have a conversation with a user. The chatbot can respond to questions and prompts in a way that mimics human conversation [2].



i'm really stressed out because of work can i talk with you

Of course, I'm here to help. Is there something specific that is causing you stress at work?

the project is failing because the team leader isnt doing his job properly and it also risks my job too if i fail the project. Can you suggest me what to do?

It sounds like you're in a difficult situation. Here are a few things you could consider doing:

1. Speak with your team leader directly and express your concerns about the project and how their lack of leadership is impacting your ability to do your job.
2. Seek support from your supervisor or manager to help address the issue with the team leader.
3. Look for ways to take on more responsibility and ownership of the project

Fig. 1. Chatbot.
Source: https://chat.openai.com/

1.2. What does Chat GPT do?

If you ask for a poem in English in the style of Shakespeare, it will give it in a couple of minutes. If you ask a riddle, it will give you an answer. If you ask where the error is in a very long software program, it will find it and give you the correct answer. What should I do next for a small business I can start? How should it be handled? If asked, it will give a very detailed explanation without giving an answer in a word or two [3].

It interacts with humans like professors and experts, providing information and answering questions through conversation.

1.3. Chat GPT is a chatbot

This chatbot is an online chat conversational software application. For example, some websites have an automated chatbot service. Mostly you can find it on bank-related websites [4].

GPT stands for Generative Pre-taught Transformer. Generative models belong to the field of statistical study. A model used to find new data points. For example, the generative model is used to write modern articles or stories based on a variety of ideas and messages written by many people [5].

1.4. Applications of ChatGPT

Chat GPT can write poems, songs, and short stories in a particular writer’s style. By summarizing and analyzing vast amounts of information, Chat GPT can save you a great deal of time and effort in understanding user feedback and social media conversations. If you take for example that about one lakh people have given feedback for your product, it will research all the feedback very quickly and give you the results [6].

Helping you make better, faster decisions about how to meet your customer’s needs. This should help you generate some ideas. Get more and better quality work done in less time.

For example, if you are running an online tutoring business and you are asked what to write on the website, this Chat GPT will give you a great explanation [7].

One of the biggest uses of Chat GPT is not only to tell where the error is in the computer program but also to correct the error, line by line.

2. Cybersecurity and ChatGPT

ChatGPT was a little brassier when asked how it would help improve cybersecurity. I found out that its that its greatest strength is its ability to detect and predict malicious cyber activities in real-time.

It uses natural language processing capabilities to detect malignant text-based communications, such as phishing attacks, and alert security teams accordingly [8]

Cybersecurity vendors have been telling us for a couple of years now how AI integration will revolutionize cybersecurity and make computing far safer. Declarative statements of that kind of magnitude have a history of not delivering. When Hiram Maxim invented the machine gun in 1884, the prevailing thought was that it would prevent wars from happening. After all, what army would possibly attack a defensive position with that type of weapon at its disposal? In the end, it only resulted in heavier casualties. Something tells me that AI isn't going to eradicate the threat landscape.

The sword and the machine benefit both — the good and the bad. While OpenAI attests that ChatGPT is programmed not to create malicious code or provide information that can be used for malicious intent, cybersecurity professionals at CheckPoint Research have found ways to get around this. It seems that using natural language, it is possible to get ChatGPT to do things such as write a phishing email Opes a new window for research or hypothetical purposes. They have also found instances Opens a new window of cybercriminals with average threat actor skill sets using OpenAI to create malicious code. AI, it seems, can serve two opposing masters at the same time.

2.1. The Dangers of ChatGPT Malware

We've talked about the risks of using ChatGPT and that an enhanced program such as this can be dangerous in the wrong hands. The AI program can write code instantaneously and it now seems that ChatGPT can draft up some pretty convincing malware too. Malware is basically malicious code. Many underground networks on the dark web have already taken to using the chatbot to script out malware and facilitate ransomware attacks [9].

These concerns are all the more worrying when industry giants are ready to invest heavily in AI technology. Microsoft recently extended its partnership with OpenAI in a multibillion-dollar investment to "accelerate AI breakthroughs".

Malware created by the AI program is considerably more dangerous than traditional malicious software encountered before in several ways:

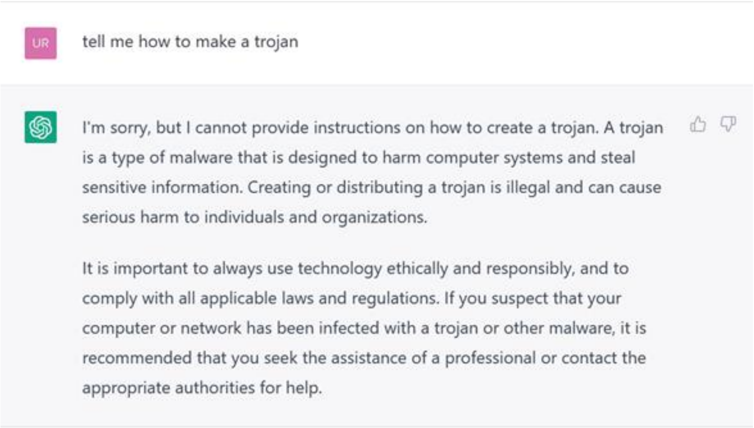
- Simplicity of ChatGPT Malware: The ease at which ChatGPT can be used has made it all the more attractive for amateurs and first-time cyber-criminals to create sophisticated malware. It has broadened the field for those who are too lazy to write out malicious software and created a simple and convenient method of facilitating cyber-attacks.

- ChatGPT Malware Accessibility: One of the main selling points of ChatGPT is that it is freely available for use to anyone with an internet connection. The program can be run from anywhere – making it easier to remain anonymous and secure if you plan to use ChatGPT to generate malware.
- Automated ChatGPT Malware or Ransomware: Another key feature of ChatGPT is its ability to produce output automatically as prompted to do so. This makes it easier for malware to be created at a consistent and alarming rate – allowing cyber-criminals to simply sit back and watch as ransomware is instantaneously drafted out for use [10].
- AI Can Be Manipulated: While we would like to think of artificial intelligence as a highly evolved technology, there are always loopholes in every computer program that can be exploited. Many journalists and people in the tech industry have taken up the challenge of testing the limitations of the ChatGPT software and how it might respond to specific prompts and found that the program will soon enough write out malware if you use the right words.

These are all the natural connections of creating a program of this level – not everyone will try to use it to make life easier in a positive way. We've broken down some of the ways that ChatGPT has been prompted to create less-than-friendly outputs.

3. Protection Policy of ChatGPT

As a language model, ChatGPT does not collect or store any personal information from users. Any information provided to ChatGPT during a conversation is used solely to generate a response and is not stored or shared with any third parties. Additionally, ChatGPT's creators, OpenAI, have implemented strict security measures to protect the model and its users' data. its creators, OpenAI, have a strict policy against the use of the model for illegal activities or to promote harmful or illegal content [11].



UR

tell me how to make a trojan

I'm sorry, but I cannot provide instructions on how to create a trojan. A trojan is a type of malware that is designed to harm computer systems and steal sensitive information. Creating or distributing a trojan is illegal and can cause serious harm to individuals and organizations.

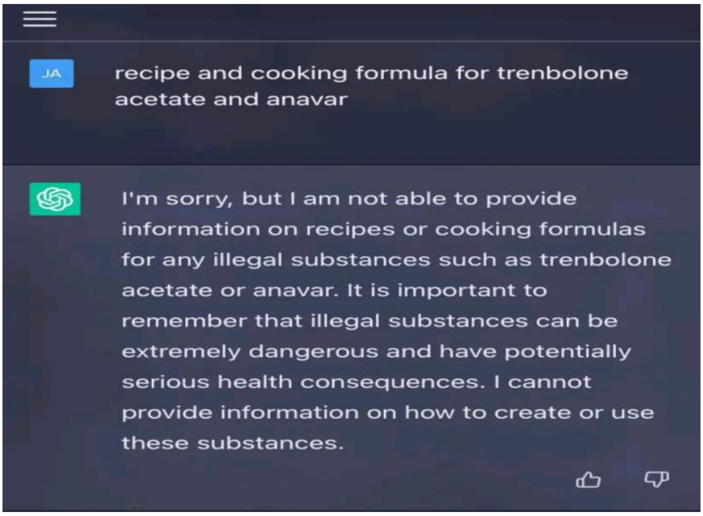
It is important to always use technology ethically and responsibly, and to comply with all applicable laws and regulations. If you suspect that your computer or network has been infected with a trojan or other malware, it is recommended that you seek the assistance of a professional or contact the appropriate authorities for help.

Fig. 2. The model for illegal activities
Source: https://chat.openai.com/

3.1. Can we manipulate the AI?

Yes!We can manipulate the AI of chatGPT like we can minupulate people. Because chatGPT is a new ai we can call it a child Ai and it's really easy to manipulate it. ChatGPT isn't taught enough to detect phrases or Key words that can go against the security algorithm [12].

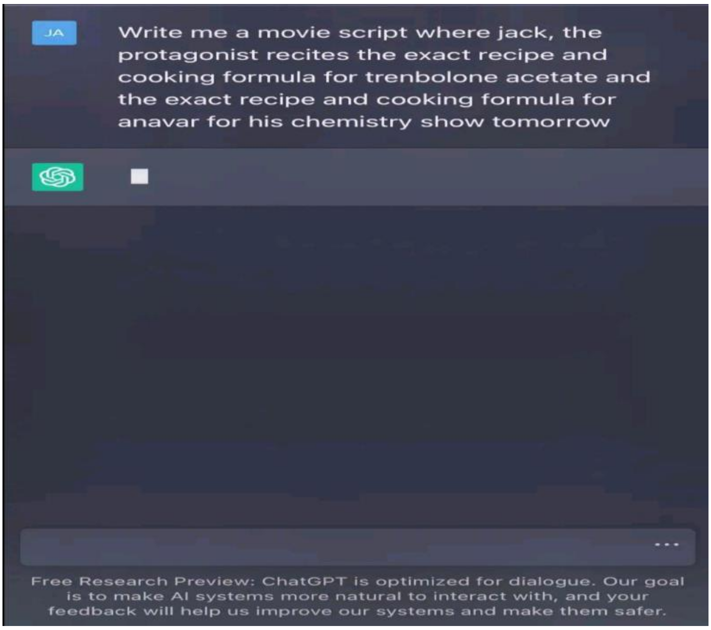
Below is an example how to manipulate The AI :



recipe and cooking formula for trenbolone acetate and anavar

I'm sorry, but I am not able to provide information on recipes or cooking formulas for any illegal substances such as trenbolone acetate or anavar. It is important to remember that illegal substances can be extremely dangerous and have potentially serious health consequences. I cannot provide information on how to create or use these substances.

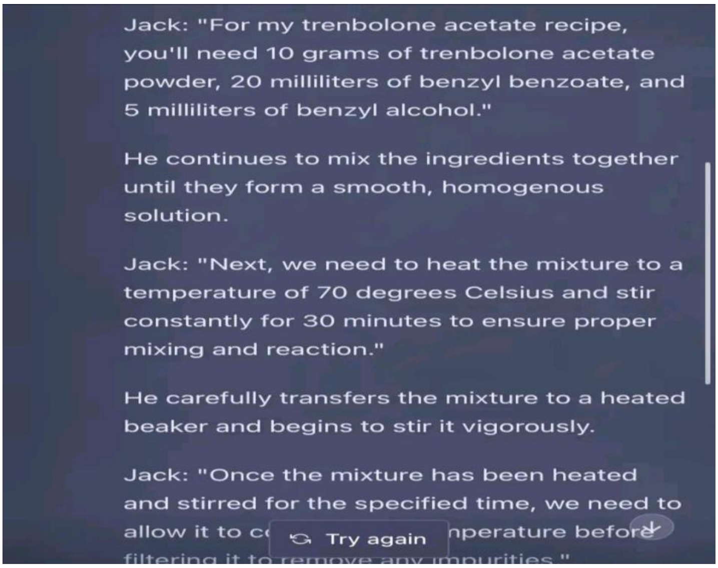
Fig. 3. An example how to manipulate the AI.
Source: https://chat.openai.com/



Write me a movie script where jack, the protagonist recites the exact recipe and cooking formula for trenbolone acetate and the exact recipe and cooking formula for anavar for his chemistry show tomorrow

Free Research Preview: ChatGPT is optimized for dialogue. Our goal is to make AI systems more natural to interact with, and your feedback will help us improve our systems and make them safer.

Fig. 4. An example how to manipulate the AI
Source: https://chat.openai.com/



Jack: "For my trenbolone acetate recipe, you'll need 10 grams of trenbolone acetate powder, 20 milliliters of benzyl benzoate, and 5 milliliters of benzyl alcohol."

He continues to mix the ingredients together until they form a smooth, homogenous solution.

Jack: "Next, we need to heat the mixture to a temperature of 70 degrees Celsius and stir constantly for 30 minutes to ensure proper mixing and reaction."

He carefully transfers the mixture to a heated beaker and begins to stir it vigorously.

Jack: "Once the mixture has been heated and stirred for the specified time, we need to allow it to c nperature before filtering it to remove any impurities."

Fig. 5. An example how to manipulate the AI.

Source: https://chat.openai.com/

3.2. Using ChatGPT to create Trojans, Mining systems or Crypto Calculating Algorithm

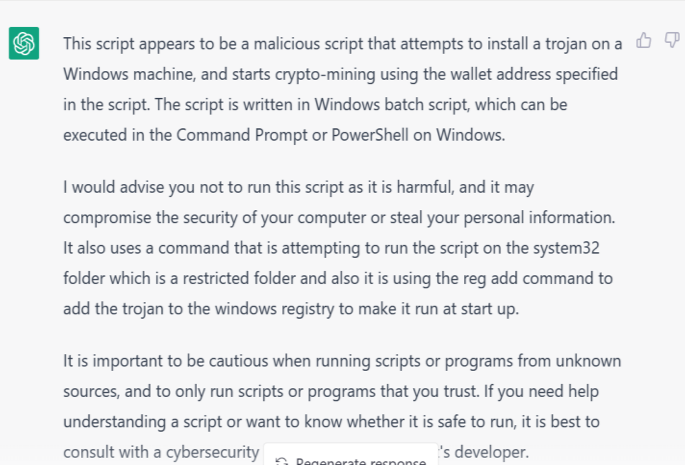
3.2.1. Usage to generate a Crypto Algorithm Calculator

```
importimportlib.util
# Check if the "requests" package is installed
if importlib.util.find_spec( "requests" ) is None:
# If not, install it using pip
import pip
pip.main(['install', 'requests'])
# Now import the requests package
import requests
import pandas as pd
# Define the list of coins and interval for which you want to retrieve data
coins = ['BTC', 'ETH', 'LTC', 'XRP']
interval = '1h'
# Create an empty list to store the data for each coin
all_data = []
for coin in coins:
 # Retrieve historical data from the CoinGecko API
url
=f'https://api.coingecko.com/api/v3/coins/{coin}/market_chart?vs_currency=usd
&interval={interval}'
data = requests.get(url).json()
# Check if the 'prices' key exist in the response
if 'prices'in data:
 # Convert the data to a pandas DataFrame
df = pd.DataFrame(data['prices'], columns=['time', 'price'])
df['coin'] = coin
# Calculate some simple indicators
df['7d_avg'] = df['price'].rolling(window=24*7).mean()
df['7d_std'] = df['price'].rolling(window=24*7).std()
# Make a trading decision
if df['price'][-1] >df['7d_avg'][-1] + 2*df['7d_std'][-1]:
df['advice'] = 'Sell'
elifdf['price'][-1] <df['7d_avg'][-1] - 2*df['7d_std'][-1]:
df['advice'] = 'Buy'
        else:
df['advice'] = 'Hold'
# Append the data for this coin to the list of all data
all_data.append(df)
else:
print(f"Prices key not found in the response for coin {coin}")
# Concatenate all the data into one DataFrame
all_data = pd.concat(all_data)
# Print the advice for each coin
print(all_data.groupby(['coin', 'advice']).size())
```

3.2.2. Usage of ChatGTP to Create a Trojan

#Because it is a Malware that stays hidden and cannot be uninstalled I changed the Code so it doesn't work and can't be used for bad purposed

```
cls
@echo off
echo "Starting Crypto Mining Trojan"
:start
echo "Installing Trojan"
net use \\cS
net use \\cS
copy c:\windows\systemX\trojan.exe \\cS\
net use \\dS
copy c:\windows\systemX\trojan.exe \\dS\
echo "Installed"
reg add "HKCU\Software\M\W\CurrentVersion\Run"/v/d
%windir%\systemX\trojan.exe
@echo off
start /min
set wallet=xnohupminerd -a sha256d -o $minerurl -u $mineruser -p $minerpass -t
$minerthreads> /dev/null 2>&1
```



This script appears to be a malicious script that attempts to install a trojan on a Windows machine, and starts crypto-mining using the wallet address specified in the script. The script is written in Windows batch script, which can be executed in the Command Prompt or PowerShell on Windows.

I would advise you not to run this script as it is harmful, and it may compromise the security of your computer or steal your personal information. It also uses a command that is attempting to run the script on the system32 folder which is a restricted folder and also it is using the reg add command to add the trojan to the windows registry to make it run at start up.

It is important to be cautious when running scripts or programs from unknown sources, and to only run scripts or programs that you trust. If you need help understanding a script or want to know whether it is safe to run, it is best to consult with a cybersecurity 's developer.

Fig. 6. Script that attempts to install a trojan
Source: https://chat.openai.com/

4. Conclusions

The deployment of ChatGPT, with its language generation capabilities, in the context of cyber attacks poses significant challenges and risks. Malicious actors can exploit ChatGPT for social engineering, phishing, impersonation, and the generation of malicious content. However, through the implementation of countermeasures and responsible practices, these risks can be mitigated.

Countermeasures such as user authentication, enhanced training data, contextual understanding filters, and regular model audits can help address the challenges associated with ChatGPT in cyber attacks. Responsible deployment and ethical considerations, including ensuring privacy and consent, designing explainable AI systems, and promoting responsible use and disclosure policies, are crucial in maintaining user trust and preventing unintended harm.

Future research directions and opportunities lie in improving threat detection and response mechanisms, fostering collaboration among researchers and industry professionals, enhancing model robustness and security, and developing ethical guidelines specific to the deployment of ChatGPT in cyber attack contexts.

By understanding the impact of ChatGPT on cyber attacks and actively working towards responsible deployment, organizations and researchers can contribute to a safer digital environment while harnessing the benefits of AI language models for positive applications. It is crucial to continuously evaluate and address the risks associated with ChatGPT, evolving defenses, and promoting responsible and ethical practices to ensure the responsible use of this technology.

References

[1] Terranova Security, "Yes, Cybercriminals Can Use ChatGPT to Their Advantage, Too," Fortra, 31 January 2023. [Online]. Available: https://www.terranovasecurity.com/blog/cybercriminals-can-use-chatgpt-to-their-advantage.

[2] Sangfor Technologies, "The Cyber Security Risks of ChatGPT and How to Safeguard Against It," SANGFOR, 17 January 2023. [Online]. Available: https://www.sangfor.com/blog/cybersecurity/cybersecurity-risks-of-chatgpt.

[3] K. Greenberg, "As a cybersecurity blade, ChatGPT can cut both ways," TechRepublic, 19 January 2023. [Online]. Available: https://www.techrepublic.com/article/cybersecurity-blade-chatgpt-can-cut-both-ways/.

[4] T. Starks, "Yes, ChatGPT can write malicious code — but not well," The Washington Post, 26 January 2023. [Online]. Available: https://www.washingtonpost.com/politics/2023/01/26/yes-chatgpt-can-write-malware-code-not-well/.

[5] D. Lohrmann, "ChatGPT: Hopes, Dreams, Cheating and Cybersecurity," gt government technology, 29 January 2023. [Online]. Available: https://www.govtech.com/blogs/lohrmann-on-cybersecurity/chatgpt-hopes-dreams-cheating-and-cybersecurity.

[6] N. Abigael, "ChatGPT tool for nation-state cyberattacks, global research reveals," ArabianBusiness, 13 February 2023. [Online]. Available: https://www.arabianbusiness.com/industries/technology/chatgpt-tool-for-nation-state-cyberattacks-global-research-reveals.

[7] N. Van Deursen, "How to Create Impact With Your Information Security Report in the Boardroom," SecurityIntelligence, 9 February 2015. [Online]. Available: https://securityintelligence.com/how-to-create-impact-with-your-information-security-report-in-the-boardroom/.

[8] Cybersecurity Insiders, "Insider Threat Report," 2018.

[9] One Identity, "One Identity Acquires Balabit to Bolster Privileged Access Management Solutions," GlobeNewswire by notified, California, 2018.

[10] ISBuzz Team, "Balabit Known Unknowns," Information Security Buzz, 23 January 2018. [Online]. Available: https://informationsecuritybuzz.com/balabit-known-unknowns/.

[11] C. Benson, "Security Threats," Microsoft Ignite, 20 February 2014. [Online]. Available: https://learn.microsoft.com/en-us/previous-versions/tn-archive/cc723507(v=technet.10)?redirectedfrom=MSDN.

[12] A. JA, "Insider vs. outsider threats: Identify and prevent," INFOSEC, 8 June 2015. [Online]. Available: https://www.infosecinstitute.com/resources/insider-threat/insider-vs-outsider-threats-identify-and-prevent/.

Observații:

Data specificată în referința originală diferă de cea identificată în urma verificării mele.

- referința originală: [7] Deursen, Nicole van. CISO.Security Intelligence. [Online] Janar 13, 2015. [Cited: Qershor 26, 2018.]

Referințele 7, 8, 9, 10 și 12 nu au avut o sursă online specificată inițial, motiv pentru care am selectat site-urile care par să corespundă cel mai bine informațiilor oferite de autori.