**Topic -** **Fake Cyber Threat Intelligence and Chat GPT- 4**

**Research how to fine-tune GPT model. Train the GPT model using collected FAKE CTI (cyber threat intelligence).**

**Stage 1**

**Introduction:**

The integration of ChatGPT into cybersecurity has been the focus of numerous discussions because of the likely applications and challenges it poses. Late studies (Al-Hawawreh et al., 2023; Chowdhury et al., 2023) have disclosed the capability of ChatGPT in cyber-attacks pointing at its expected influence on the CIA Set of three of Cyber Security (Chowdhury et al., 2023). While the use of ChatGPT is growing, it becomes more critical to perceive its clouded side, threats, and implications for cybersecurity experts and organizations. The research objective is to uncover the secret threats posed by ChatGPT and to give the necessary information to users to actually battle emerging cyber threats.

**Motivation**

The objective behind this study is to uncover potential digital threats to ChatGPT and to raise client mindfulness significantly further. If we have any desire to construct viable countermeasures, we want to take a gander at the cloudy side of ChatGPT, say Alawida et al. (2024). To stay in front of various new online protection risks, it is significant to find out about the ongoing purposes, troubles, and likely future improvements of ChatGPT (Al-Hawawreh et al., 2023) in the field. Since digital dangers and weaknesses are continuously developing, it is important to reconsider the present status of network safety considering the inescapable use of models in view of AI like ChatGPT. The motivation behind this examination is to assist with overcoming any barrier among hypothesis and practice by exploring the crossing point of ChatGPT and online protection. The expectation is that this will prompt better safeguard cautious strategies.

What the Task Involves

An intensive assessment of the expected purposes, troubles, and risks of involving ChatGPT for network safety is the objective of this undertaking. Potential considerations incorporate strategies to increment client mindfulness and security, cyberattacks coordinated by ChatGPT and their effect on the CIA triplet, and comparative drives. To give a total image of what man-made intelligence based models mean for network safety methods, specialists are concentrating on true occasions and utilizing ChatGPT-created material to recreate digital risks.

Information Gathering

To direct an intensive examination, a few Open Source Knowledge, or operating system sources have given text based material relating to network protection. Dataset components remember subtleties for genuine digital risks, assault vectors, and security issues. The unique tweaking for the ChatGPT system for applications in network protection depends on this assortment of different data sources. To comprehend the intricacy and assortment of digital risks, the examination utilizes checked data. Along these lines, we can look at ChatGPT's job in moderating online protection worries in more profundity.

Model Adjusting

The information gathered is utilized to adjust the model of the GPT-2 transformer. Crothers et al. (2023) found that network protection faces exceptional issues from machine-made messages. Digital Danger Knowledge (CTI) text that parodies network protection specialists and frameworks might be produced by calibrating the model utilizing the online protection corpus. This study expects to work on the model's capacity to produce logically applicable and convincing CTI by iterative boundary tuning and the consolidation of area explicit data. The objective is to empower a more precise assessment of the model's impact on network protection rehearses.

Evaluation

Language precision, rationale, and the capacity to emulate genuine CTI are a portion of the actions used to assess the presentation of the refined ChatGPT model. This study will further develop network protection safeguard frameworks by giving trial results on the qualities and shortcomings of ChatGPT in making fake CTI, as well as by efficiently contrasting the model's exhibition with pre-set benchmarks and well-qualified assessment.

Incorporation

The essential goal is to show this present reality impacts of ChatGPT-made content on security frameworks all through the reconciliation stage. As indicated by the review, fake CTI might think twice about wellbeing of CKG and obstruct independent direction.

**Documentation and Communication**

A critical component of this venture is to record each step of the process, from information assortment and model adjustment to assessment and integration. The documentation, based on the standards of Gupta et al. (2023), guarantees transparency and reproducibility. In addition, correspondence of the findings is urgent for sharing information, creating awareness and bringing together the security local area in the battle against the growing security threats. Research outcomes will be disseminated through scholarly publications, presentations, and workshops fully intent on contributing to a cooperative understanding of ChatGPT's impact on cybersecurity and helping stakeholders make very much informed decisions.

**Background**

Foundation research covers comprehensive survey including threat models and location methods connected with machine created text as referenced by Crothers et al. (2023). The investigation of the ongoing landscape will give you significant information about the expected risks and vulnerabilities the ChatGPT can bring to cybersecurity. Synthesizing the existing writing and identifying the missing links in the ongoing understanding builds a strong base for the study of the impact of ChatGPT on the cybersecurity practices.

**Tasks for Next Meeting**

Our next meeting will be equipped towards improving the model fine-tuning process based on the initial results that we have obtained. Addressing the challenges and utilizing input from cybersecurity specialists during the assessment phase will be critical. Moreover, the practical demonstration of faking CTI using ChatGPT4, as indicated by the tutor, will be anticipated during the meeting. Through the continuous improvement of the research philosophy and the arrangement of research goals with the expectations of the stakeholders, this research seeks to give the practical solutions and useful insights in request to handle the emerging cybersecurity threats.

**Problems Faced**

During the work on the undertaking, it was sometimes challenging to make the synthesized fake CTI correspond to genuine cybersecurity scenarios. Toeing the fine line among authenticity and likely misinformation, as discussed by Falade (2023), has remained an enduring trouble. Based on the input obtained during the initial meetings, the task is being immediate toward the tutor's expectations, which are lined up with the created work. Proactive way to deal with manage challenges and by collecting criticism from project stakeholders, this research would defeat challenges and convey excellent outcomes that would work with the headway of cybersecurity information and practices.

**Stage 2**

**Project Problem**

The project aims at dealing with the rising threat from the growing use of ChatGPT in cybersecurity systems, a peculiarity that has gained a ton of consideration as of late because of its capability to change the basic principles of cybersecurity, specifically the CIA Group of three - Secrecy, Integrity, and Availability. Ongoing progress in regular language processing has prompted the making of perplexing language models such as ChatGPT, which have changed the scope of cybersecurity. While the widespread use of PC text created information, especially in Cyber Threat Intelligence (CTI), results in a ton of complexities and risks.

Network safety scientists, for example, Alawida et al. (2024) and Chowdhury et al. (2023) have found a weakness in ChatGPT that could think twice about frameworks. The venture's all-encompassing objective is to take apart this knotty issue by researching the association among engineered CTI with the heartiness of network safety limits. By diving into the intricacies of this point, the undertaking tries to give essential information that might help with the advancement of impervious cautious components, actually fortifying network safety frameworks despite steadily evolving dangers.

Point

A broad examination of ChatGPT's possible impacts on network safety is the essential target of this exploration. All the more particularly, the's undertaking will likely explain how digital danger data that utilizes machine-composed text could imperil the wellbeing of these frameworks. The objective of this examination is to point out the field of network protection and have a beneficial outcome by exploring the shortcomings of ChatGPT-created material. Along these lines, preplanned protective strategies to decrease the perils of CTI brought about by machines might be created.

Objectives

The objective of this examination is to incorporate an exhaustive investigation of the digital danger knowledge climate in accordance with ChatGPT. This requires figuring out how the model could impact network safety experts' independent direction and surveying how well it repeats genuine CTI.

The subsequent stage of the exploration is to build a Network protection the information chart and reproduce an information harming attack utilizing the refined ChatGPT model. To show the genuine impacts of these attacks and the hardships they cause to network protection frameworks, the exploration will bring misleading data into portrayal of information frameworks.

The task's general objective is to foster safeguard procedures that increment client information on the network safety risks presented by ChatGPT. To assist with bracing the network protection biological system, the venture is likewise intending to give powerful guarded components and correspondence strategies to battle machine-created text attacks.

**Equipment and Tools**

Every essential apparatus and gear should be close by for the task to be finished to the end. The preparation of the ChatGPT model and the administration of enormous datasets both need admittance to HPC assets. The model will be assessed utilizing specific techniques that action verbal rightness, soundness, and the ability to copy genuine CTI. Moreover, creating apparatuses to reproduce information harming assaults and survey their effect on network protection foundation expects admittance to pertinent data and network safety information charts. We will follow best practices and base our hardware and apparatus choice on the proposals of momentum research (Greshake et al., 2023; Gupta et al., 2023).















**Explanation**

Utilizing a dataset of Digital Danger Insight (CTI), this Python code makes it simpler to calibrate a GPT-2 model and produce counterfeit CTI reports. The principal part deals with the imports that are required, including tqdm for estimating preparing progress, the Transformers library for GPT-2 models, and PyTorch for profound learning.

To preprocess and tokenize CTI information, a particular PyTorch dataset class called CTIDataset was created. The code characterizes this class. To work on the information texts' consistence with the GPT-2 model, this class ensures that they are suitably tokenized utilizing a predefined tokenizer.

The primary part of the code is the fine\_tune\_model capability, which is accountable for the genuine adjusting process. Significant activities including choosing a gadget (computer processor or GPU), placing the model in preparing mode, fabricating a DataLoader for compelling information stacking, beginning an AdamW enhancer, and running the preparation circle for the foreordained number of ages are completely remembered for this capability. Inside every age, the technique comprises of repeating over the dataset in clusters, figuring the misfortune, doing backpropagation, and altering the model boundaries as the need should arise.

The code test additionally tells the best way to utilize genuine CTI information utilizing the fine\_tune\_model capability and the CTIDataset class. Initialised with the tokenizer, greatest succession length, and accumulated CTI information is the CTIDataset class case. The pre-prepared GPT-2 model, dataset, tokenizer, and hyperparameters, (for example, cluster size and number of ages) are utilized when the fine\_tune\_model capability is run.

The upgraded GPT-2 model and tokenizer are saved in the suitable organizers toward the finish of the code. However long the model has been prepared on the specific elements of the provided CTI information, this ensures that it could be saved and utilized again to deliver counterfeit CTI reports from here on out. In light of everything, the code offers a deliberate and careful method for changing a GPT-2 model to deliver counterfeit CTI reports utilizing the information that has been assembled. To redo the framework for specific CTI creating assignments, you might change the model design and hyperparameters.

**Literature Review**

The writing on the integration of ChatGPT into cyber practices is enormous and diverse, reflecting the increasing nervousness among the cybersecurity local area about the possible dangers and consequences of machine-produced text. Crothers et al. (2023) dig into threat models and identification methods relating to machine-created text, which enlightens the various approaches used to distinguish and forestall risks associated with chatGPT-produced text.

Al-Hawawreh et al. (2023) investigate the practical applications, challenges, and future directions of ChatGPT in cybersecurity. This emphasizes the prerequisite for strong defense mechanisms to counter the unique threat landscape. The study emphasizes the necessity of proactive measures to secure cybersecurity infrastructure against the spread of machine-produced counter-threat intelligence (CTI).

In the same vein, Falade (2023) scrutinizes on the threat landscape posed by ChatGPT, FraudGPT, and WormGPT in social engineering attacks thus urging cybersecurity professionals to be more cautious and mindful. The study demonstrates the capacity of text created by machines to deceive and control the users, stressing the basic of devising successful countermeasures.

Moreover, Scanlon et al. (2023) investigate the use of ChatGPT in computerized forensic investigation and offer some insights on its strengths and weaknesses in this field. This study provides insights to the cyber security local area by examining the impact of machine-produced text on advanced forensic practices which will inform the improvement of investigative techniques that will be robust.

The writing audit emphasizes the intricate interconnection among ChatGPT and cybersecurity which points to the requirement for imprecise precautions in request to kill the risks associated with machine-created text. Through the synergy of various studies, the writing survey provides an overall outline of the flow situation of research on the topic, setting the stage for the observational studies of this venture.

**Problem Faced**

Many difficulties and obstacles have been experienced during the undertaking, featuring the multifaceted design of the review interaction. One of the greatest impediments is assembling and tidying up every one of the fundamental informational collections to prepare and test the ChatGPT model. Simply by doing thorough arrangement and approval might the information's quality and honesty at any point be guaranteed, which are the essential worries. In conclusion, doing online protection research utilizing machine-produced text is full of moral problems that requirement for smart conversation and severe regard to laid out standards and norms. The venture is focused on proceeding with its work and making progress in network safety in spite of these obstructions.

**References**

Alawida, M., Abu Shawar, B., Abiodun, O.I., Mehmood, A., Omolara, A.E. and Al Hwaitat, A.K., 2024. Unveiling the dark side of chatgpt: Exploring cyberattacks and enhancing user awareness. *Information*, *15*(1), p.27.

Al-Hawawreh, M., Aljuhani, A. and Jararweh, Y., 2023. Chatgpt for cybersecurity: practical applications, challenges, and future directions. *Cluster Computing*, *26*(6), pp.3421-3436.

Chowdhury, M.M., Rifat, N., Ahsan, M., Latif, S., Gomes, R. and Rahman, M.S., 2023, May. ChatGPT: A Threat Against the CIA Triad of Cyber Security. In *2023 IEEE International Conference on Electro Information Technology (eIT)* (pp. 1-6). IEEE.

Crothers, E., Japkowicz, N. and Viktor, H.L., 2023. Machine-generated text: A comprehensive survey of threat models and detection methods. *IEEE Access*.

Falade, P.V., 2023. Decoding the threat landscape: Chatgpt, fraudgpt, and wormgpt in social engineering attacks. *arXiv preprint arXiv:2310.05595*.

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T. and Fritz, M., 2023, November. Not what you've signed up for: Compromising real-world llm-integrated applications with indirect prompt injection. In *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security* (pp. 79-90).

Gupta, M., Akiri, C., Aryal, K., Parker, E. and Praharaj, L., 2023. From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access*.

Kalla, D. and Kuraku, S., 2023. Advantages, Disadvantages and Risks associated with ChatGPT and AI on Cybersecurity. *Journal of Emerging Technologies and Innovative Research*, *10*(10).

Prasad, S.G., Sharmila, V.C. and Badrinarayanan, M.K., 2023, May. Role of Artificial Intelligence based Chat Generative Pre-trained Transformer (ChatGPT) in Cyber Security. In *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 107-114). IEEE.

Scanlon, M., Breitinger, F., Hargreaves, C., Hilgert, J.N. and Sheppard, J., 2023. ChatGPT for digital forensic investigation: The good, the bad, and the unknown. *Forensic Science International: Digital Investigation*, *46*, p.301609.

**Java code:**

import java.util.ArrayList;

import java.util.List;

public class FakeCTIGenerator {

 // Method to generate fake CTI reports

 public static List<String> generateFakeCTI(int numReports) {

 List<String> fakeCTIReports = new ArrayList<>();

 // Logic to generate fake CTI reports using ChatGPT-4

 // This part would involve fine-tuning the model on a cybersecurity corpus

 // and then generating text based on that fine-tuned model

 // Sample logic (replace with actual code using OpenAI API):

 for (int i = 0; i < numReports; i++) {

 String fakeReport = generateFakeReportFromGPT4();

 fakeCTIReports.add(fakeReport);

 }

 return fakeCTIReports;

 }

 // Method to generate fake report using ChatGPT-4 (replace with actual implementation)

 private static String generateFakeReportFromGPT4() {

 // Placeholder method to generate fake report using GPT-4

 // This would involve interacting with the OpenAI API

 // and utilizing the fine-tuned model for generating text

 // Sample implementation:

 String fakeReport = "This is a fake CTI report generated by ChatGPT-4.";

 return fakeReport;

 }

 // Main method for demonstration

 public static void main(String[] args) {

 // Generating 5 fake CTI reports for demonstration

 List<String> fakeReports = generateFakeCTI(5);

 // Displaying the generated fake CTI reports

 for (int i = 0; i < fakeReports.size(); i++) {

 System.out.println("Fake CTI Report " + (i + 1) + ": " + fakeReports.get(i));

 }

 }

}

output

java -cp/tmp/RtPYZh7C1S FakeCTIGgenerator

Fake CTI Report 1: This is a fake CTI report generated by ChatGPT-4.

Fake CTI Report 2: This is a fake CTI report created by ChatGPT-4.

Fake CTI Report 3: This is a fake CTI report created by ChatGPT-4.

Fake CTI Report 4: This is a fake CTI report created by ChatGPT-4.

Fake CTI Report 5: This is a fake CTI report generated by ChatGPT-4.

**Explanation:**

**Code:**

**Import statement:**

Import the File class from the java.io package.

**Fine Tune GPTM model class:**

This class contains functionality related to fine-tuning the GPT model using CTI data collected.

**Fine Tune GPTM model method:**

* This method takes a File object representing the collected CTI data as input.
* A host provides the actual implementation of fine-tuning logic, which typically involves using libraries or frameworks for machine learning and natural language processing.

**The method includes sample logical steps:**

* Load a pre-built GPT model (eg ChatGPT-4).
* Prepare CTI data for fine-tuning.
* Adjust hyper parameters for fine-tuning (eg, training speed, batch size).
* Refine the model using CTI data.
* Save the refined model for future use.

**Basic method:**

* This method serves as the entry point into the program and is used for demonstration purposes.
* Create a file object representing the CTI data (replace with the actual file path).
* It calls the fine Tune GPT Model method to start the fine-tuning process.

**Output**

* The output consists of several lines, each representing a false CTI (Cyber ​​​​Threat Intelligence) report generated by a Java application.
* Each line begins with the label "Akeasama CTI Report" and the number that indicates the index of our report.
* The index starts at 1 and increments from 1 for the next report.
* After the index, there is a colon (:) with the contents of the dummy CTI report.
* The content of each fake CTI report is the same, in this case it is a host created by the display target code.
* Fake CTI Report Content: "This is a fake CTI report generated by ChatGPT-4." This is the host text generated by the generateFakeReportFromGPT4 method.
* Each output line represents a fake CTI report, and in the "main" method there are a total of five reports (5 reports in this case) as specified in the parameters given to the "generateFakeCTI" method.