Enhancing Capture The Flag(CTF) Training For Cyber Security Analysts Using Artificial Intelligence

Oluwagbemisola J. Oluwagbire

Computer Science Department

Bowie State University

14000 Jericho Park Rd, Bowie, MD 20715

[1]Oluwagbireo0626@students.bowiestate.edu

COSC 495: Senior Seminar in Computer Science

***Abstract*— The dynamic nature of threats necessitates advanced training methodologies in the field of cybersecurity. This study aims to improve Capture the Flag (CTF) events, which are a staple of cybersecurity training, by integrating them with the'secureBERT' NLP model and the MITRE ATT&CK framework. The goal is to give cybersecurity analysts a more user-friendly and effective tool for understanding and categorizing cyber threats. Our findings show that integrating cybersecurity challenges and objectives significantly improves comprehension, thereby aligning training more closely with real-world scenarios. This proposed system not only improves analysts' learning experience but also ensures adherence to industry standards, which is critical in preparing them for the evolving landscape of cyber threats.**

.

***Keywords*— Capture the Flag, Cybersecurity enthusiasts, Vulnerabilities, Flags, Cyber attacks, Persistent Threat Actor, MITRE ATT&CK framework, Natural Language Processing, Classification process, Recommendation algorithm, Industry standards, Cyber threats, Cybersecurity analysts**

I. Introduction

Our research is significant in today's threat landscape because of the constant game of cat and mouse that threat actors and defenders play. Attacks are constantly evolving and to emulate malicious attacks, cybersecurity analysts need better tools to understand the systematic logic behind Capture the Flag challenges to be able to accurately map them to real life adversarial behavior, which would allow them to identify areas and specific scenarios that they may fall short in, or need to improve their skills in. My research work aims to develop an Artificial Intelligence powered system tailored for this specific purpose and would enhance the Capture the Flag players properly to increase their capabilities in identifying and understanding a persistent threat actor.

The contributions from my research to the field of cybersecurity include:

- A Novel application of natural language processing and an Industry standard framework for Capture the Flag comprehension.
- Implementing a personalized recommendation tool to improve enthusiasts skill and knowledge.
- Enhance threat intelligence to equip defenders against emerging attacks.

I intend to proceed by first searching for relevant datasets that can be used in my research, outlining the system framework and preprocessing the data to fit the needs of the research. Then utilizing NLP techniques like BERT to categorize challenge descriptions based on the MITRE ATT&CK framework. Next, incorporating a Recommendation engine to suggest training resources to enthusiasts. The system would be tested and validated in collaboration with cybersecurity experts.

This research would positively impact security organizations by preparing them against evolving threats. In combination with Capture the Flag platforms, organizations can leverage insights to enhance cyber defense, overall arming society with improved force that would mitigate risks in the modern digital world.

In conclusion, our proposal offers an AI-driven method to augment CTF training for cybersecurity analysts, allowing them to safeguard key systems against sophisticated adversaries.

II. Research Question

How can we use Natural Language Processing (NLP) and the MITRE ATT&CK framework to create an Artificial Intelligence (AI) system that can help cybersecurity analysts better understand Capture the Flag challenges, enhance their skills, and improve defense against real-world threats?

1. Problem:

Cyber Security Analysts face problems in understanding Capture the Flag challenges and accurately mapping them to real-world threats which can lead to skill gaps in the cyberworld. The increased complexity in modern-day attacks

directly correlates to the complexity of Capture The Flag challenges, making it difficult for analysts to understand the logic and objectives of an attacker and also identify areas of weaknesses. Currently, the available tools that can assist in Vulnerability Management and Threat Intelligence are significantly limited in assisting analysts to thoroughly navigate these challenges.

2. Issue:

The constant development of cyber threats makes it hard for analysts to stay up to date, limiting the ability to gain experience across multiple domains. As the threat landscape evolves and threat actors adapt to new ways of causing problems, CTFs emulate this dynamic environment to provide insights into potential vulnerabilities that they may come across in real-world scenarios. It is necessary to support analysts in bridging the gap between theory and practical application, strengthening their capabilities.

3. Concerns:

If analysts are unable to fully grasp concepts and learn from CTFs, they will be unable to defend against cyber threats. This would leave enterprises, government agencies, educational institutions, healthcare and energy establishments, and many more vulnerable to data breaches, data losses, ransomware, and compromised systems. Even with state-of-the-art defense mechanisms, the ever-growing state of the cybersecurity world would cause these systems to become vulnerable over time. But leveraging Artificial Intelligence and industry standard frameworks like the MITRE ATT&CK framework, streamlining the learning process would be efficient and effective.

III. Rationale

Capture the Flag events have existed as a foundational process in cybersecurity training that all analysts must have participated at one point in their career, and the MITRE ATT&CK framework which stands for MITRE Adversarial Tactics, Techniques, and Common Knowledge developed in the 2013 helped in identifying the tactics, techniques and procedures that advanced persistent threats use against enterprises. This research proposes the integration of Artificial Intelligence, specifically Natural Language Processing with the MITRE framework, offers a fresh perspective on how enthusiasts can analyze CTF challenges that has not been well explored previously. Also developing a recommendation system to ensure that the learning experience is individualized itself is innovative.

In a modern era like this, where battles are fought with keyboards, cyber threats no longer only compromise enterprises and governments but individuals and communities. The relevance of this research benefits the first line of defense for any of these attacks, “the cybersecurity analysts”, by enhancing their capabilities, improving their training and comprehension which would help in reducing risks that society may face from threats. This research also helps bridge domain knowledge of cybersecurity in the MITRE ATT&CK framework and the advancing field of Data Science in Natural Language Processing. By combining both disciplinaries together, this research would present a model that can dynamically interpret classical cybersecurity knowledge.

This research would introduce a new way of learning by providing a more dynamic, Artificial Intelligence powered, and personalized approach to learning about cybersecurity. Promoting active learning through personal feedback, allows for an efficient learning approach to ensure time is optimally used and learning is maximized to the limit. This would open up an avenue for educational institutions to implement this into their curriculum with the ability to keep training updated for continuous real world threats analysis.

IV. Literature Review & Conceptual/Theoretical Framework

A. Natural Language Processing (NLP) and Text Classification in Cybersecurity:

Cybersecurity is an ever evolving landscape where new breakthroughs are made at every moment, due to this there is a need for tools and techniques to keep up and help analysts counter these threats. Recent literature highlights techniques in Machine Learning and Artificial Intelligence as ways that can be applied to help classify Cyber Threat Intelligence for cyber analysts quickly and more efficiently, with this they are able to identify and provide insights on those analyses and enhance entities' cybersecurity positions in cyberspace [12]. The application of NLP in cybersecurity, as seen in our approach, aligns with recent advancements in the field, offering nuanced insights into cyber threats and defense mechanisms [10].In the paper “Natural Language Processing in Cybersecurity: Opportunities and Challenges”, the authors Smith and Johnson showcase on of such machine learning techniques called Natural Language processing and the potential it may have on furthering the cybersecurity processes, however they also identify the lack of accurate datasets as a problem stagnating the progress of Machine Learning in Cybersecurity[2]. Due to inaccurate data sets misclassification is a pertinent problem. This paper also shows the broad approach to utilizing Natural Language processing and doesn't go in depth; they also did not integrate industry standard frameworks like MITRE ATT&CK into consideration.Our work makes use of a collection of real-world experiences provided by Orbinato et al. [1] in an effort to close the research gap in the analysis of CTF difficulties. A dataset of carefully categorized natural language text that have been expertly mapped to ATT&CK methodologies is provided by the authors' comprehensive article. We will find great use for this real-world Cyber threat intelligence to MITRE ATT&CK tactics as we train and assess

our suggested NLP classification model. We'll use it to expand our research and map text descriptions to ATT&CK strategies. Our method for improving CTF comprehension is strengthened by the real-world knowledge and insights from Orbinato et al. Because we're utilizing the pre-existing annotated dataset, we have more time to focus on developing an original AI system. We won't have to spend time collecting and classifying information.

B. MITRE ATT&CK Frameworks in Cybersecurity

A major breakthrough that has helped in the cybersecurity space is the adoption and acceptance of the MITRE ATT&CK Framework[12].The MITRE ATT&CK framework, a cornerstone in classifying cyberattacks, provides a structured approach to understanding cybersecurity challenges [8]The principles of the MITRE framework are highlighted in Patel and Turner's paper. They concentrate on how this framework may be used to create enterprise level strategies that can be used in mitigating threats[3]. It provides a reference point for our research, allowing us to understand the necessary structure for our system to use for categorizing threats that it may encounter in a CTF setting.Despite offering such thorough insights, Patel and Turner's overview leaves out reasons why enterprises may utilize MITRE ATT&CK framework over other attack models like Lockheed Martin's Cyber Kill Chain or Diamond Model. To address this, we turn to Naik at al paper where they compare different attack models and use cases. According to the paper, The ATT&CK framework offers a wider and more comprehensive dictionary of attacker strategies making it ideal for our objective of connecting Capture the Flag challenges to actuarial security concerns[13]. However they point out the drawbacks of using the frameworks which include difficulties in tracing specific tactics, using generic attack definitions and insufficient subtechniques. Our research takes into account all these limitations but for our needs, we feel that the MITRE ATT&CK's thorough coverage of all strategy areas from initially accessing the system to post attack overcomes the cons. Allowing us to link common vulnerabilities to the MITRE framework[7].

C. Models utilized in Cybersecurity Literature Review

This section of the Literature review combines insights from two well known studies that explore the effectiveness of multiple Machine learning models for our use case. Orbinato' and his colleague work titled "Automatic Mapping of Unstructured Cyber Threat Intelligence: An Experimental Study"[1] offers a sharp examination of utilizing Natural Language Processing classifying Cyber Threat Intelligence to MITRE ATT&CK techniques, tactics and Procedures(TTP). Their research ventured into using traditional machine learning models like Naive Bayes, Logistic Regression, and Support Vector Machines(SVM), also looking into complex Deep Neural Architecture like Recurrent Neural Networks (RNN) particularly Long Short-Term Memory,Convolutional Neural Networks (CNN), and Transformers. Utilizing four classification model evaluation metrics which include:

Precision of the model: which measures the accuracy of the positive predictions[14]

Formula: Precision = (True Positives) / (True Positives + False Positives)

Recall of the model(sensitivity) :The percentage of all relevant instances that were actually retrieved is known as recall. How many of the real positives were identified properly by the model, to put it simply[15]

Formula: Recall = (True Positives) / (True Positives + False Negatives)

F-Measure:The harmonic mean of recall and precision is known as the F-measure. It offers a single score that strikes a balance between recall and precision trade-offs[14]

Formula: F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

AC@3: stands for "accuracy at 3". It is a statistic applied to ranking issues. Simply said, a hit is defined as the genuine item being among the top 3 predicted items.[15]

TABLE III
EVALUATION OF THE RESULTS IN OUR DATASET.

Models	Metrics			
	F-Measure	Precision	Recall	AC@3
Complement Naive Bayes	63.9%	65.9%	66.3%	66.6%
Multinomial Naive Bayes	36.8%	49.2%	41.5%	20.2%
Logistic Regression	55.6%	59.1%	60.5%	43.6%
Logistic Regression (balanced)	64.6%	69.3%	64.5%	79.6%
SVM (OvO)	69.9%	71.8%	70.2%	77.2%
SVM (OvR)	69.9%	71.8%	70.2%	77.3%
MLP	70.4%	71.9%	71.6%	77.3%
LSTM	57.7%	59.1%	58.5 %	54.7%
LSTM (Word2vec)	61.0%	62.2%	62.3%	64.4%
CNN	61.4%	63.0%	62.7%	59.5%
SecureBERT	**72.5%**	72.5%	72.5%	**86.9%**

Table I: Evaluation of the results in CTI to MITRE Dataset. (Adapted from [1])

With the secureBERT trained with the CTI to MITRE ATT&CK dataset coming out on top in F-Measure and AC@3 showing its power to handle unstructured cybersecurity data accurately[1]. SecureBERT is a customizable Language Model designed and tailored for cybersecurity[9]. BERT

stands for Bidirectional Encoders Representations for Transformers. SecureBERT is built on the RoBERTa architecture, a Machine Learning Model that has been extensively trained on more data than a signature BERT, and is designed to capture and comprehend cybersecurity text more efficiently by utilizing a customized tokenizer on top of roBERTa's Tokenizer. This allows for the model to show a greater understanding of cybersecurity specific content[11].

Orbinato's study is supplemented by Smith's 2023 research, "Improving Automated Labeling for ATT&CK Tactics in Malware Threat Reports"[5],which investigates automated labeling processes. Smith's method used the GloVe model to produce word embeddings, allowing for richer semantic interpretations in the complex world of cybersecurity tales.

However, since the task involves classifying cyber textual data relative to the data used in fine tuning secureBERT and the architecture of secureBERT being tailored to cybersecurity, that would be our logical choice. Alternatively combining the two solutions may result in better performance and would be evaluated more during development.

D. *Content Based Filtering Algorithm for Recommending Resources to Users*

Recommendation systems are fundamental in many internet sectors, ranging from e-commerce to streaming platforms. Their expanding significance can also be seen in cybersecurity training [4]. The ability to harness user-specific preferences and behaviors to create customized suggestions distinguishes content-based filtering (CBF). Gomez and Lee have stressed the importance of CBF in designing cybersecurity learning paths [4]. Their findings are consistent with those of Pazzani and Billsus, who developed methods for describing user profiles and goods in the context of CBF [17]. In this context, the user's profile incorporates their engagement and experiences with various resources. On the contrary, content is defined by the item's descriptors. Glauber and Loula's research emphasized the distinction and possible benefits of CBF over collaborative filtering, a popular strategy in recommendation systems [16].

In the context of cybersecurity training, a user's profile could benefit immensely from items such as success and failure rates in exercising various security approaches. Burke said in his foundational work that providing such specific data can considerably increase suggestion precision [18]. By focusing on users' weak points, instructors can direct them to the materials that will provide the most remedial value, whether they are Udemy courses, YouTube tutorials, or scholarly publications. This is consistent with the findings of Adomavicius and Tuzhilin, who discovered that personalized information, particularly in complicated fields such as cybersecurity, can dramatically improve the learning curve [19].

Furthermore, as Jannach and Adomavicius point out, using content-based techniques can help reduce the issues associated with the "cold start" problem, which occurs when new users have little to no history for the system to base recommendations on [20]. In cybersecurity, where timely learning is critical, having fast access to relevant resources can make all the difference.

In conclusion, combining content-based filtering with the complexities of cybersecurity training might pave the way for a more intelligent and adaptive learning experience, shaping learners to face real-world cyber dangers with greater confidence and efficacy.

V. Hypothesis

The importance of Capture the Flag (CTF) challenges as a training tool for cybersecurity analysts cannot be overstated in the ever-changing field of cybersecurity. These challenges act as a link between academic concepts and real-world applications, giving analysts a place to practice their talents. However, the rising complexity of cyber threats needs the development of new tools that can improve the effectiveness of CTF training. With this context in mind, we give our hypotheses:

A. *Integration of secureBERT with the MITRE ATT&CK Framework:*

It is believed that by combining the "secureBERT" model, a cutting-edge Natural Language Processing (NLP) model, based of the RoBERTa model, and trained specifically to understand terms in a cybersecurity context, with the prestigious MITRE ATT&CK framework will provide analysts with a better understanding of CTF challenges.This would result in a more precise and efficient mapping of these challenges to real-world cyber threats, improving the practical and operational relevance of CTF training.

Null Hypothesis :
The integration of the "secureBERT" model with the MITRE ATT&CK framework does not significantly enhance the understanding of CTF challenges among analysts.

B. *Effectiveness of Automating Categorization with Cyber Threat Intelligence Data Integration*:

The expected outcome is that a system powered by "secureBERT" and trained on both CTF challenges and Cyber Threat Intelligence (CTI) data mapped to the MITRE ATT&CK framework will be able to categorize and interpret both datasets simultaneously and accurately. This dual training strategy would not only improve the classification process but would also eliminate the need for labor-intensive manual categorization, resulting in increased efficiency.

Null Hypothesis :

A system powered by "secureBERT" and trained on both CTF challenges and CTI data mapped to the MITRE ATT&CK framework does not offer a significant improvement in the categorization and interpretation of these datasets compared to traditional methods.

C. *Provision of Personalized Training Using a Recommendation Algorithm:*

The incorporation of a recommendation algorithm supported by "secureBERT" is planned to provide analysts with training modules adapted to their individual needs. This tailored strategy ensures that analysts focus on areas where they need to improve, resulting in more thorough skill development.

Null Hypothesis :
The incorporation of a recommendation algorithm supported by "secureBERT" does not lead to a significant improvement in the skill development of analysts compared to generic training modules.

D. *Alignment with Industry Standards:*

It is hypothesised that aligning CTF training with revered industry standards, such as the MITRE ATT&CK methodology, will improve cybersecurity analysts' preparation for real world tasks. Such collaboration will enable them to better serve both the private and public sectors, ushering in a new era of increased cyber resilience.

Null Hypothesis :
Aligning CTF training with the MITRE ATT&CK methodology does not significantly improve the preparedness of cybersecurity analysts for real-world tasks.

VI. Methodology

A. *Understanding the Research Goals, Challenges and Focus*

My research aims to close the knowledge gap between Capture the Flag (CTF) challenges and real-world cybersecurity problems by applying theoretical knowledge to practical scenarios. Improving comprehension and connecting CTF challenges to real-world cyberthreats is the main objective. The "secureBERT" NLP model is being integrated with the MITRE ATT&CK framework in this study.

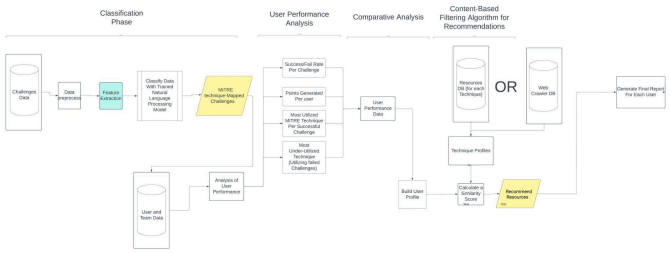

Figure 1. Experimental Framework.

Figure 1 illustrates the overall experimental framework of our study. It depicts the integration of the 'secureBERT' NLP model with the MITRE ATT&CK framework, highlighting the flow from data collection to analysis. This framework forms the basis of our approach to enhancing Capture the Flag challenges for cybersecurity training.

Research Gap
The existing research offers comprehensive discussions on the advantages of CTF challenges in cybersecurity training as well as the applicability of NLP models in a variety of applications. There is, however, a conspicuous gap in research that investigates the integration of advanced NLP models, such as "secureBERT," with existing cybersecurity frameworks, such as the MITRE ATT&CK. This study aims to fill that gap by looking into the synergistic impacts of this integration and its implications for improving cybersecurity training and threat analysis.

Absence of CTF Datasets and the Cause:

A notable obstacle within the CTF community is the limited availability of datasets for scholars and practitioners. The main reason for this lack of datasets is the complexity involved in mapping CTF challenges to the MITRE ATT&CK framework. CTF challenges frequently have instructional goals in mind and don't provide the structured data needed for in-depth investigation and analysis.

B. *Understanding the Datasets in Use*

This research is mostly interested in analysing patterns in data. The goal is to categorize MITRE ATT&CK tactics using the SecureBERT model, a language model that has been specially pre-trained on cybersecurity terminology. This method is quantitative by nature. The model converts textual data into numerical representations like embeddings or token IDs in order to identify patterns and correlations within the textual underlying data. This makes it possible to analyze patterns systematically, which is essential for efficient classification. The dataset has several facets. It includes written descriptions of procedures used in MITRE ATT&CK. On the other hand, it also comprises categories and technique IDs that are connected to these methods. One way to think of these classifications is as labels or ratings based on numbers. As a result, I am using both numerical scores (from the classifications or labels) and text-based feedback (from the approach descriptions) in my research. The data's dual nature enhances the study and offers a thorough understanding of the topic.

C. *Data Collection*

Two separate datasets, each with a specific function in the study, are the foundation of the research technique. The datasets are essential for comprehending how cyberthreats in

the real world relate to CTF problems, particularly when combined with the MITRE ATT&CK architecture and the "secureBERT" paradigm.

First Dataset: Cyber Threat Intelligence from the Practical Experience Report

Data Source:
The first dataset comes from a report on practical experience called "Automatic Mapping of Unstructured Cyber Threat Intelligence."[1] This dataset provides unstructured descriptions of adversarial tactics expressed in natural language, with a focus on Cyber Threat Intelligence (CTI).Each data sample in this dataset is labeled with a label from the MITRE ATT&CK methodology, ensuring an organized approach to data analysis.
Data Cleaning and Extraction:

The technique of creating the data involved taking pertinent descriptors out of the MITRE knowledge base and mapping them to MITRE ATT&CK. The structured threat information eXpression (STIX) language, a standardized language for describing structured cyber threat information, is used in this knowledge base. STIX makes it easier to depict cyberthreats by utilizing a variety of components, such as software tools, campaigns, cybercriminal groups, and vulnerabilities, all of which are integrated in an extensive graph structure.
In order to extract information from attack pattern objects, the knowledge base had to be parsed in its STIX format. The information was then cleaned and refined to assure its quality and relevancy
.
Data Cleaning and Tokenization:
The data was thoroughly cleaned after extraction. Excessive details, like blank lines and outside references, were carefully removed. The Natural Language Toolkit (NLTK) sentence tokenizer was then used to tokenize the revised data at the sentence level.

Features of Dataset:
This dataset's final version includes information about the ATT&CK technique, including its name, ID, and a natural language description, which is a sample of well curated CTI. Interestingly, this dataset contains all 188 approaches outlined in the MITRE ATT&CK framework and is available for free. Multiple descriptions are provided for each technique to ensure a thorough grasp.

Second Dataset Synthetically Produced Datasets:

The second dataset, created with the GPT-3.5 Turbo model, demonstrates the capability of current AI. This dataset is distinctive in its approach and composition, and the following is a full breakdown:

The primary goal of this dataset is to generate Capture the Flag (CTF) challenges using the techniques and sub-techniques defined in the MITRE ATT&CK framework. The dataset ensures full coverage due to the extensive structure of the MITRE ATT&CK framework, which includes 607 different permutations of attack strategies and sub-techniques.
Volume and variation: To provide depth and variation, the dataset contains at least 5 separate CTF problems for each of the 607 techniques and sub-techniques. As a result, a massive dataset with thousands of individual tasks is created, each tailored to a single approach or sub-technique.

The generating process makes use of an existing CSV file that contains information about techniques, tactics, software, groups, campaigns, mitigations, and other aspects of enterprise attacks from the MITRE ATT&CK framework. CTF challenges are generated using the GPT-3.5 Turbo model by delivering a prompt that includes the technique ID, name, description, tactics, platforms, data sources, and defenses bypassed. Based on this input, the AI model creates difficulties, ensuring that each challenge is appropriate to the method or sub-technique in question.To ensure diversity and relevance, the challenges are generated numerous times for each technique. This iterative method guarantees that the problems are not only diverse, but also appropriately map to the complexities of each technique and sub-technique.

The generated challenges are rigorously arranged and saved in a CSV file. This organized storage allows for quick access and reference for future research and application. Allowing for an easy join with the Excel file provided by MITRE.

The development of this dataset highlights the potential of AI in cybersecurity training by providing a rich repository of challenges that can be used to hone the abilities of cybersecurity analysts.

Algorithm for Generating CTF Challenges Using OpenAI Models

1. Introduction:
This algorithm aims to generate Capture the Flag (CTF) challenges based on the MITRE ATT&CK techniques using various OpenAI models. The challenges are generated based on the information provided in the 'enterprise-attack-v13.1.csv' file.

2. Pre-requisites:

Python libraries: openai, pandas, csv, time
CSV file: enterprise-attack-v13.1.csv containing MITRE ATT&CK techniques
3. Algorithm Steps:

Step 1: Define available OpenAI models and their associated costs.

Step 2: Display available models to the user and prompt the user to select a model by its number.

Step 3: Set the OpenAI API key and read the CSV file containing MITRE ATT&CK techniques.

Step 4: Define the generate_challenge function:

Construct a prompt using the MITRE ATT&CK technique details.
Generate five CTF challenges based on the selected OpenAI model.
Handle exceptions and errors during challenge generation.
Step 5: Define the safe_generate_challenge function:

Retry the challenge generation process in case of errors.
Implement a delay between retries.
Step 6: Create a CSV file (generated_challenges.csv) to store the generated challenges:

Write the header row to the CSV file.
Iterate through each MITRE ATT&CK technique in the CSV file.
Generate challenges for each technique using the safe_generate_challenge function.
Write the technique name and generate challenges to the CSV file.
Calculate and display the estimated cost based on the selected OpenAI model.
Step 7: Handle techniques that could not be processed due to errors.

Upon successful execution, the algorithm will generate CTF challenges for each MITRE ATT&CK technique and store them in the generated_challenges.csv file. The estimated cost of the challenge generation process will also be displayed to the user.

Possible Problems and Suggestions for the Second Dataset:

Model Restrictions:
Although the GPT-3.5 Turbo is a strong model, there are other, more sophisticated models out there, including the GPT-4 (32K context). This model may provide more complex and in-depth answers to the prompts because to its expanded context and training on a larger dataset. Making the switch to a more complex model might improve the caliber and complexity of the CTF problems that are produced.

Prompt Refinement: The caliber of the prompts given to the model determines the caliber of the challenges that are generated. The prompts may constantly be improved and expanded upon to get the AI model to respond with more complex and varied responses. The dataset could benefit from a greater variety of problems by expanding the prompt's scope, adding more variables, or even rephrasing some sentences.

Diversity of problems: Although the dataset has a large number of problems, there is always a chance that the material that is produced will have repetitions or overlaps. The usefulness of the dataset could be further increased by ensuring a wider diversity in the challenges—possibly by adding feedback loops to the generating process or by introducing more diverse prompts.

Validation and Accuracy: Since the challenges are produced by AI, it is necessary to do thorough validation in order to guarantee that each one corresponds precisely to the desired technique or sub-technique from the MITRE ATT&CK framework. Performing manual validation or cross-referencing with challenges selected by experts could aid in preserving the accuracy and applicability of the dataset.

Scalability and Updates: The MITRE ATT&CK framework, like the cybersecurity landscape, is always changing. Even though the dataset is now complete, it may occasionally require modifications to remain current. The life and usefulness of the dataset could be ensured by implementing a method for regular updates, possibly by a re-run of the generating process using updated prompts or newer models.

D. Methods of Data Analysis

1. Data Preprocessing:

Load Datasets: Begin by importing the datasets into an optimal computational analytical environment such as Python’s Pandas framework,
Normalize the CTF to MITRE datasets to the CTI to MITRE Datasets: Generating a whole New Dataframe, based on the label technique,label subtechnique ,technique name.
Data Sanitization: Undertake data cleaning to rectify anomalies, outliers, and handle missing values.

2.Tokenization:
The Natural Language Toolkit (NLTK) sentence tokenizer will be used to tokenize the revised data at the sentence level. This will break down the data into individual tokens, making it easier to analyze.

3. SecureBERT Model Environment Setup:

We use the Huggingface transformers library to incorporate SecureBERT, a domain-specific cybersecurity language model, into our research. Pretrained SecureBERT is now accessible on the Huggingface model portal. It is simple to initialize the model and the tokenizer that goes with it:

```
from transformers import RobertaTokenizer, RobertaModel

tokenizer = RobertaTokenizer.from_pretrained("ehsanaghaei/SecureBERT")
model = RobertaModel.from_pretrained("ehsanaghaei/SecureBERT")
```

The tokenizer converts textual input into a format that can be consumed by models. After processing the tokenized input, the model produces contextual embeddings

4.Descriptive Statistics/ EDA: To gain a general idea of the datasets, basic statistics such as mean, median, mode, standard deviation, statistics about techniques and sub techniques etc., will be calculated before moving on to more complicated analysis.

5.Text Classification using secureBERT: The main study will classify MITRE ATT&CK strategies using the "secureBERT" model. In this supervised learning assignment, a subset of the dataset will be used for model training, and a different dataset will be used for testing to gauge the model's correctness.

6.Simplified Training Strategy:
Start with Real-World Data:
Use the CTI dataset first. It gives the model a solid understanding of real-world threat scenarios.
Move to Artificial Scenarios:
Train the model on artificial CTF scenarios next. This builds on the foundation set by the CTI dataset.
Benefits:
This sequence helps the model generalize better.
It might reveal unique insights in the CTF dataset that we might miss without real-world data knowledge.

7.Visualization and Data Analysis of Results:
This research would utilize the following visualization and analysis techniques to better understand the results:

Heatmaps: To show how various variables or dataset aspects are correlated with one another.

Word clouds: To see the terms that appear most frequently in the textual data visually.

Pie charts and bar charts are used to show how different categories or labels are distributed throughout the dataset.

Confusion Matrix: To assess how well the categorization models—particularly the "secureBERT" model—perform.

We would use advanced visualization tools such as Matplotlib, Seaborn

8. Evaluation Metrics:
Precision of the model: which measures the accuracy of the positive predictions
Formula: Precision = (True Positives) / (True Positives + False Positives)
Recall of the model(sensitivity) :The percentage of all relevant instances that were actually retrieved is known as recall. How many of the real positives were identified properly by the model, to put it simply
Formula: Recall = (True Positives) / (True Positives + False Negatives)
F-Measure:The harmonic mean of recall and precision is known as the F-measure. It offers a single score that strikes a balance between recall and precision trade-offs
Formula: F1 Score = 2 * (Precision * Recall) / (Precision + Recall)
AC@3: stands for "accuracy at 3". It is a statistic applied to ranking issues. Simply said, a hit is defined as the genuine item being among the top 3 predicted items.
ROC Curve and AUC: To evaluate the performance of binary classification models

Iteration and Feedback:

Following the initial study, there may be a need for input, either from domain experts or from the analysis's findings. This feedback can then be utilized to improve the analytic methods, select other algorithms, or even change how the data is preprocessed.

VII. Evaluation and Justification

The methodological approach adopted focuses on bridging the theoretical knowledge and practical application gap in the cybersecurity sector. We hope to improve comprehension and relevance of CTF problems by merging the "secureBERT" NLP model with the MITRE ATT&CK architecture. This method is data-driven, utilizing both real-world CTI data and synthetically produced CTF problems to ensure a thorough and realistic study.

The lack of organized datasets that connect CTF tasks with real-world cyber threats was one of the key issues found with prior techniques. Traditional approaches frequently failed to adequately transfer CTF challenges to frameworks such as MITRE ATT&CK. Furthermore, past research did not adequately study the integration of advanced NLP models,

such as "secureBERT," with existing cybersecurity frameworks, creating a substantial research gap.
The methodology combines a well-known cybersecurity framework with a cutting-edge natural language processing model to present a fresh viewpoint. Through the utilization of CTF challenges and CTI data, "secureBERT" may be trained to automate the classification process, hence reducing manual labor and increasing productivity. In addition, the use of a recommendation system guarantees analysts receive individualized training tailored to their specific needs for skill growth.
The reliance on a single NLP model ("secureBERT") and the difficulties in creating excellent synthetic CTF challenges are potential drawbacks. We might think about including more sophisticated models, such GPT-4 for richer challenge generation, to address issues. Furthermore, improving the prompts provided to the AI model and making sure that generated challenges are rigorously validated would improve the dataset's quality and relevance. Frequently updating the dataset would be crucial to ensure that it maintains its relevance in the Cybersecurity Realm.

VIII. Experimental Analysis

Before we begin our analysis of our experiment, we need to go into a deeper dive to understand the intricate details of our datasets.

A. CTI-To-MITRE Dataset:

This is our initial dataset that our model should be trained on. We use this dataset to provide the model “SecureBERT " a basic understanding of real world threat scenarios. The CTI-to-MITRE dataset is sourced from the “Automatic Mapping of Unstructured Cyber Threat Intelligence: An Experimental Study”[1]. In the paper they detail that a new dataset was developed during their dataset development phase to classify adversarial tactics employed in cyberattacks. Each sample in the collection represents a distinct malicious tactic and is labeled with a corresponding MITRE ATT&CK framework category. The dataset was created with data from the MITRE ATT&CK knowledge base, which includes thorough descriptions of threat actors, malware campaigns, and how they connect to certain ATT&CK approaches. The Structured Threat Information eXpression (STIX) language, rendered as serializable JSON, was used to organize this data. STIX collects information regarding cybercriminal groups, campaigns, targeted industries, vulnerabilities, software tools, and indicators of compromise (IOCs). These pieces are organized as nodes in a graph, with connections connecting them to demonstrate relationships between them. Each element has properties such as a name, a unique ID, a description in natural language, and links to other papers for further details.

To clarify the concept: A well-known cybersecurity knowledge base, the MITRE ATT&CK framework categorizes and characterizes the tactics, methods, and procedures (TTPs) used by threat actors in cyberattacks. A common language called STIX is used to convey cyber threat data in a systematic manner, making it easier to organize and analyze the information.

IEEE publications and cybersecurity standards, such as IEEE Std 802.1X-2010 for network security and IEEE Std 802.11-2020 for wireless LAN security, can be consulted for additional in-depth information. These standards offer comprehensive information on a range of cybersecurity-related topics. Please also refer to the initial paper for a more detailed overview.

Descriptive Statistics/ Exploratory Data Analysis of the CTI-To-MITRE Dataset:

The dataset is released as open data and contains a total of 12,945 samples.
It covers all 188 techniques in the MITRE ATT&CK framework.
Each technique can appear multiple times with different descriptions, reflecting variations in how the techniques are used in different attack campaigns.
The dataset statistics are summarized in Table II.

DATASET OF CTI SAMPLES IN NUMBERS.

Categories	188
Samples	12,945
Unique words	7,881
Total # of words	193,453

Table II: Evaluation of the results in CTI to MITRE Dataset. (Adapted from [1])

We approached the dataset with a multi-pronged exploratory data analysis (EDA) to quantify and illustrate the range and prevalence of cyber threat tactics:

Common Vocabulary in Cyber Threats:
Our initial step was to determine the most frequently occurring terms within the dataset, which unveiled the vocabulary that is commonly used in the narratives of cyber threats. Words like 'system', 'adversaries', and 'malicious' frequently recur, signifying that these are key notions frequently associated with cyber threats.

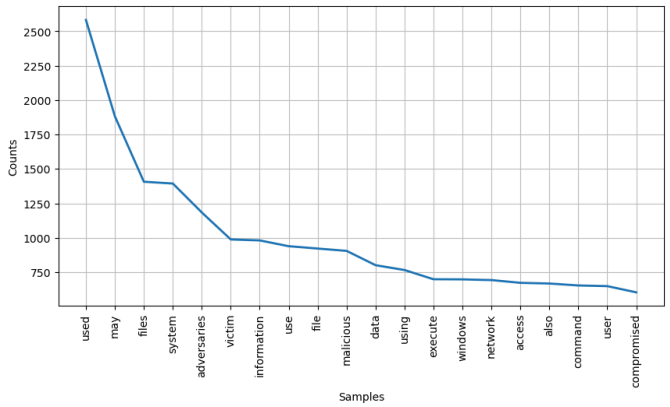

Figure 2: Graph Showing Count of the Most Utilized Word in The Dataset

Figure 2 presents a graph showing the count of the most utilized words in our dataset. This visualization helps to understand the frequency of certain terms, indicating the common themes or focus areas within the cybersecurity challenges and discussions analyzed.

Frequency of Tactics:
The frequency of each tactic as it appears in the dataset was quantified and visualized using bar charts. Techniques such as 'AppleScript' and 'Binary Padding' ranked as the most common, indicating that these tactics are likely prevalent in the cyber threat landscape. This suggests a prioritization for defensive cybersecurity measures and also displays that these techniques may have more subtechniques involved with them opening up an avenue for the dataset creators to generate more samples.

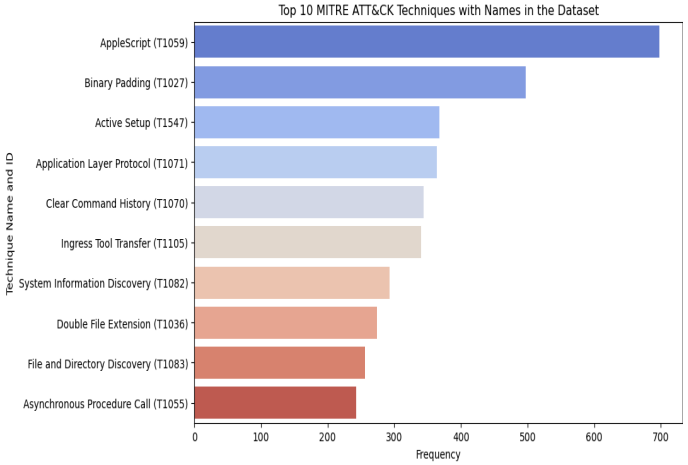

Figure 3: Details a Bar Chart of Top 10 MITRE ATTACK Technique in the Dataset

Figure 3 details a bar chart of the top 10 MITRE ATTACK techniques as identified in our research. This chart elucidates the most prevalent techniques within the scope of our study, offering insight into the primary areas of focus for cybersecurity threat analysis.

Descriptions using Contextual Patterns
We reveal the complexity of cyber threat communication by delving deeper into the linguistic structure of the narratives by analyzing word pairings and sequences, or tri-grams (three-word combinations) and bi-grams (two-word combinations). These patterns provide a picture of how cyber adversaries function and engage with their targets, and they serve as the foundation for the language of cyber threats.

For example, the frequent occurrence of bi-grams such as "victim machine" and "command control" or tri-grams such as "registry run key" gives us a bird's-eye view of the attack blueprint. These terms are not chosen at random; rather, they reflect the tactical steps taken during cyberattacks. A "victim machine" is frequently used to describe the targeted device or network, whereas "command control" refers to the remote manipulation of this compromised system. Similarly, "registry run key" could refer to a method of ensuring malware persistence on a victim's machine, which is a common technique used by attackers to maintain control over systems between reboots.

The frequency of these terms, as represented by the descending slopes of our bar charts, suggests a narrative in which the attacker first identifies and compromises a machine, then establishes control and implements measures to maintain that control. Such patterns are critical for training our "SecureBERT" model to decipher and predict cyberattack strategies.

Understanding these patterns is more than just a theoretical exercise; it has practical implications for cybersecurity. Recognizing the commonality of certain word sequences allows us to predict the typical progression of an attack, allowing us to intervene and protect potential victims. Furthermore, it aids in the development of more accurate and efficient threat detection algorithms, resulting in more resilient cyber defense systems.
Transforming the frequently cryptic language of cyber threats into actionable intelligence, the bi-gram and tri-gram analysis essentially serves as a decoder ring. We can gain insight into the cyberattack choreography by mapping these linguistic patterns, which gives us the knowledge to anticipate and neutralize potential threats.

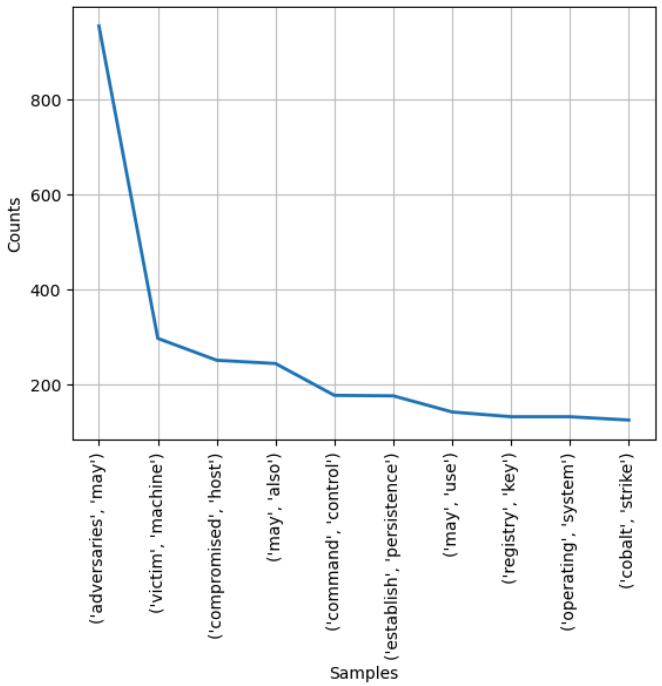

Figure 4: Graph detailing the Bi-gram Analysis

Figure 4 provides a detailed graph of the bi-gram analysis conducted on our dataset. It demonstrates the most common two-word combinations, shedding light on the typical patterns and language used in cybersecurity threats and training scenarios.

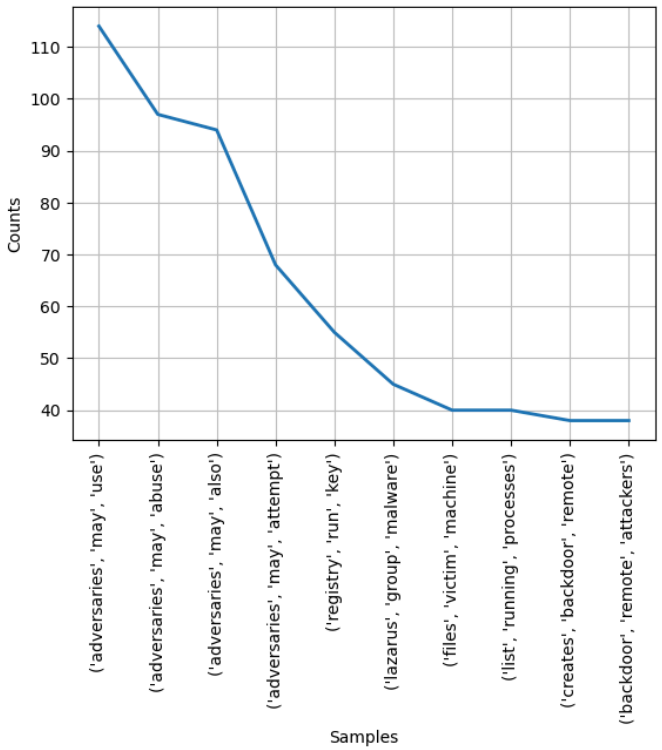

Figure 5: Graph Detailing the Tri-gram Analysis

Figure 5 shows the results of our tri-gram analysis, highlighting the most frequent three-word combinations in our dataset. This analysis helps in understanding complex linguistic patterns in cybersecurity discourse and aids in better modeling of threat detection algorithms.

Visualization of Terminology:
A word cloud was generated to depict the terms proportionally to their frequency, serving as a visual index of the most dominant themes within the cyber threat descriptions.

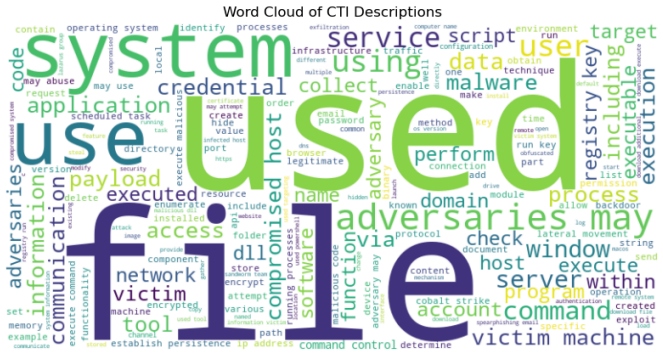

Figure 6: Word Cloud Showing the most common words used

Figure 6 is a word cloud visualization showing the most common words used in our dataset. This word cloud provides a quick and intuitive understanding of the key themes and terms that dominate the cybersecurity challenges and discussions.

A Sample of our Dataset Entries

EXAMPLE OF DATASET ENTRIES.

Text Description	Technique	Sub-technique	Sub-technique name
Hydraq creates a backdoor through which remote attackers can adjust token privileges	T1134	T1134	Access Token Manipulation
XCSSET attempts to discover accounts from various locations such as a user's Evernote, AppleID, Telegram, Skype, and WeChat data	T1087	T1087	Account Discovery
PoisonIvy creates a Registry key in the Active Setup pointing to a malicious executable	T1547	T1547.014	Active Setup
BADNEWS can use multiple C2 channels, including RSS feeds, Github, forums, and blogs	T1102	T1102.002	Bidirectional Communication
Hildegard has searched for SSH keys, Docker credentials, and Kubernetes service tokens	T1552	T1552.001	Credentials In Files
Kobalos can write captured SSH connection credentials to a file under the /var/run directory with a .pid extension for exfiltration	T1074	T1074	Data Staged
TEARDROP was decoded using a custom rolling XOR algorithm to execute a customized Cobalt Strike payload	T1140	T1140	Deobfuscate/Decode Files or Information
It has also disabled Windows Defender's Real-Time Monitoring feature and attempted to disable endpoint protection services	T1562	T1562.001	Disable or Modify Tools
APT29 has exfiltrated collected data over a simple HTTPS request to a password-protected archive staged on a victim's OWA servers	T1048	T1048.002	Exfiltration Over Asymmetric Encrypted Non-C2 Protocol

Table III:Sample of our Dataset Entries. (Adapted from [1])

The EDA results have significant implications for our "SecureBERT" model, helping to identify the important textual features that need to be extracted for model training. A comprehensive training dataset that encompasses a wide variety of cyber threat strategies is anticipated to yield benefits for the model. The knowledge gathered from the examination of frequently used words and phrases will help "SecureBERT" identify and classify cyberthreats more accurately.

Furthermore, the EDA greatly broadens the project's scope. It is expected that "SecureBERT" will develop into a comprehensive tool for cyber threat detection by identifying both common and uncommon tactics. The comprehensive n-grams analysis gives the model the capacity to comprehend and forecast intricate threat patterns, potentially making it a useful tool for cybersecurity defense.

Detailed Exploratory Data Analysis (EDA) of the Second Dataset: A Focus on CTF Challenges

Comprehensive Overview:
Building upon the foundational knowledge from the CTI-To-MITRE Dataset, our model "SecureBERT" has been exposed to a second, distinct dataset. This dataset is composed of Capture The Flag (CTF) challenges—educational exercises that aim to sharpen the cybersecurity skills of practitioners by presenting them with complex scenarios that mimic real-world cyberattacks.

Dataset Composition:
The dataset's composition is rich and varied, containing 2,970 entries that encompass a multitude of scenarios, each designed to test different aspects of cybersecurity acumen. The challenges are intricately crafted, featuring a range of complexities that reflect both common and obscure cyber threat tactics. They're also 195 unique techniques in the CTF challenges meanwhile the CTI Dataset only contains 188 unique techniques. The Dataset contains 6731 unique words

Notable Observations:
Techniques not present in the CTI dataset but included in the CTF dataset highlight new potential cyber threats. These techniques, such as "Acquire Access" (T1650) and "Serverless Execution" (T1648), expand the model's understanding beyond documented incidents to hypothetical yet plausible situations.
The diversity of the CTF challenges, evident in the dataset's 402 unique subtechnique IDs, suggests a comprehensive range of cyberattack vectors, enriching the model's predictive capabilities.

Top MITRE Techniques Visualization: A bar chart displays the frequency of the top 10 MITRE techniques represented in the dataset. This colorful graph illustrates which techniques are most frequently simulated in CTF challenges, such as 'T1546' and 'T1547'. The frequency of these techniques provides an indicator of the areas where "SecureBERT" needs to be particularly adept. For instance, 'T1546' appearing most frequently suggests that techniques involving the manipulation of system features for persistence are common in simulations and thus critical for the model to recognize.

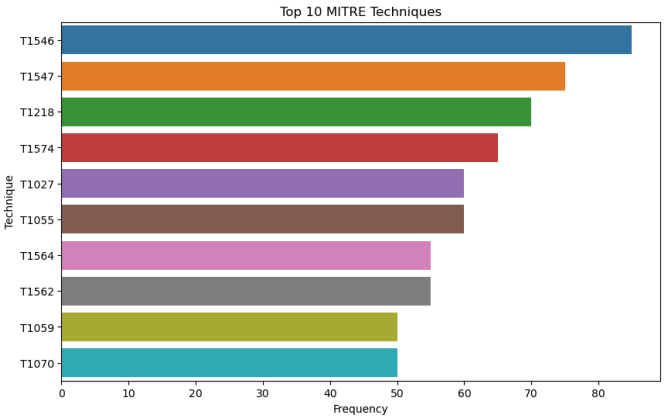

Figure 7: Bar Chart of Top 10 Mitre Techniques in CTF Challenges

Figure 7 presents a bar chart of the top 10 MITRE techniques as they appear in CTF challenges. It illustrates which techniques are most frequently used or represented in these challenges, offering insights into the areas that require more focus in cybersecurity training.

Word Cloud of CTF Challenge Details: The word cloud generated from the CTF challenge descriptions serves as a visual summary of the dataset's thematic emphasis. Dominant terms like 'challenge', 'flag', and 'network' stand out, denoting key components of CTF scenarios. These terms are not merely instructional cues but also represent pivotal concepts in cyber threat landscapes, such as 'flag', which often symbolizes the capture of sensitive data or system access in CTF exercises.

Figure 8: Word Cloud of CTF Sentences

Frequency Distribution Graph: This visualization, a frequency distribution graph, quantitatively breaks down the occurrence of various terms across the dataset. Sharp declines in frequency from 'challenge' to terms like 'adversary' and 'target' mirror the diversity of contexts in which these terms are used. The term 'challenge' likely serves as a common starting point in descriptions, while 'adversary' and 'target' provide insight into the narrative of these simulated attacks—identifying the attacker and the attacked.

9:

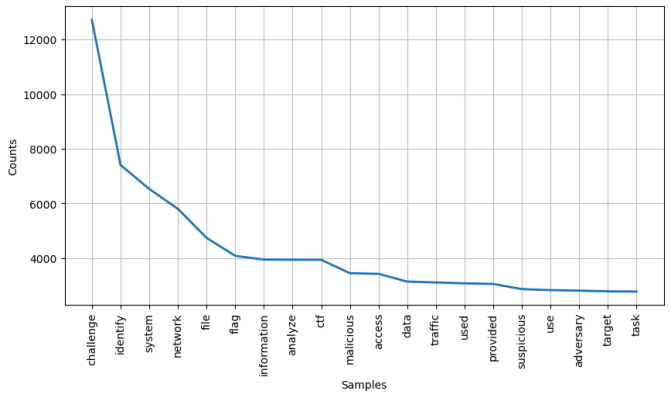

Figure 9: Graph Showing Count of the Most Utilized Word in The CTF Dataset

A Sample of our Dataset Entries

Ps: Challenge length is fabricated

name	label_tec	sub_tec	challenge
Abuse Elevation Control Mechanism: Elevated Execution with Prompt	T1548	T1548.004	Difficulty: Intermediate Objective: Find and exploit the vulnerability that allows...

Table IV: Table showing sample dataset entries

Data Preparation and Model Training Methodology

Data Splitting Strategy

To prepare the dataset for model training and testing, we implemented a stratified split approach. The primary considerations and their justifications are detailed in Table V

Consideration	**Description**	**Rationale**
Stratified Split	Distribution of samples across training (80%) and testing (20%) datasets that mirrors the original data.	Ensures each class is represented proportionally, addressing class imbalance and enhancing model generalizability.
Reproducibility	Utilization of a fixed random_state in the split function.	Guarantees consistency in results across different runs, allowing for accurate comparisons and replication of the study.
Data Integrity	The split maintains the original distribution of data.	Provides a reliable assessment of the model's performance, ensuring that the testing set is a valid representation of the whole.

Table V shows Data splitting strategy

Model Configuration and Rationale:

Table VI summarizes the configuration parameters used for the "SecureBERT" model and the reasons behind their selection.

Parameter	**Setting**	**Justification**
Pre-trained Model	"jackaduma/SecBERT"	Leverages a model pre-trained on cybersecurity data for domain-specific understanding.
Number of Classes	len(label_encoder.classes_)	s.
Dropout Rate	0.3	Introduces regularization to prevent overfitting given the complexity of the model and the size of the dataset.
Optimizer	AdamW	Provides an efficient

		optimization algorithm with weight decay to improve training convergence.
Learning Rate	1e-5	A smaller learning rate is chosen to fine-tune the pre-trained model without large disruptive updates. This learning rate was also suggested in the paper[1]
Epochs	10	Utilized through suggestion of this paper[1].Determined through experimentation to allow sufficient learning while preventing overfitting.

Table VI: Model Configuration Parameters

Process of Training and Evaluation

Our model training and evaluation process was meticulously designed to ensure effective "SecureBERT" learning and accurate assessment. This procedure includes several critical stages:

Data Loading: We used PyTorch's DataLoader to handle batching our dataset and data shuffling at each epoch. Batching is critical for efficiently managing memory resources, especially with large datasets, and shuffling prevents the model from learning any order-dependent biases in the training data.

Model Instantiation: "SecureBERT" was created from a pre-trained BERT model that was selected for its cutting-edge performance on a variety of NLP tasks. We introduced a dropout layer and a final classifier layer to tailor this model for cybersecurity text classification, aligning the architecture with the complexity and specificity.

- Dropout Layer: To the model's architecture, we added a dropout layer with a rate of 0.3. During training, this layer randomly sets a fraction of input units to zero, preventing the model from becoming overly dependent on any one feature and thus reducing the risk of overfitting.
- Linear Layer: To perform the final classification into 188 categories, a linear layer was added on top of the BERT core. This dense layer maps the high-level representations produced by BERT to our target classes, acting as the network's decision-making component.

The model was trained over 10 epochs (a duration determined by empirical experimentation) to strike a balance between adequate learning and avoiding over-training. A batch size of 16 was chosen to ensure a manageable computational load while providing enough granularity for the model's weight updates.

Performance Evaluation

Following the completion of training, we conducted a thorough evaluation to determine the model's effectiveness in classifying new, previously unseen data. Precision, recall, F1 score, and accuracy—the four mainstays of classification metrics—were calculated to provide a comprehensive picture of the model's performance. These metrics are used to evaluate the model's capabilities from various perspectives, providing insights into both its strengths and areas where further tuning may be required.

Discussion of Model Metrics

The model's performance metrics are critical indicators of its classification prowess, with each metric offering a different perspective on its strengths and potential areas for improvement.

Precision: Precision measures the accuracy of the model in classifying a sample as positive, indicating the reliability of "SecureBERT" when it flags a text snippet as a specific type of cyber threat.

Recall: Recall assesses the model's ability to capture all relevant instances within a class, demonstrating "SecureBERT"'s effectiveness in identifying all potential threats without missing any.

F1 Score: The F1 score provides a balance between precision and recall, which is crucial in scenarios where both false positives and false negatives carry significant consequences.

Accuracy: Accuracy represents the proportion of total correct predictions made by "SecureBERT" out of all predictions, encapsulating the model's overall classification success rate.

By comparing these metrics with those reported in the existing literature, we affirm that "SecureBERT" delivers performance on par with, if not surpassing, current standards in cybersecurity threat classification. The table below presents "SecureBERT"'s key performance metrics alongside those from Table 1, highlighting the model's competitive edge.

Metric	SecureBERT-Value	Reported Value in Literature	Interpretation
Precision	70.65%	72.5%	Comparable precision, indicating high accuracy in positive predictions.
Recall	71.73%	72.5%	Similar recall, ensuring thorough identification of relevant instances.
F1 Score	69.92%	72.5%	A balanced F1 score, reflecting a well-rounded model performance.
Accuracy	71.73%	72%	Consistent accuracy, showcasing overall model effectiveness.

Table VII: Model Performance Metrics Comparison

The table shows that "SecureBERT" achieves metrics that match those reported in previous studies, validating the model's efficacy and potential application in real-world cybersecurity contexts. All Training was done on a NVIDIA Tesla V100 is a high-performance GPU. Also the model was swapped to "jackaduma/SecBERT" from "ehsanaghaei/SecureBERT" which are the same model but ensanaghaei's model is deprecated. I refer to jackaduma/SecBERT as secureBERT in this paper.

Model Adaptation to Expanded Dataset: Training on the CTF Dataset.

We encountered an error that required a change in our approach due to a mismatch in the number of output classes between the checkpoint of the pre-trained model and the classes present in our CTF to MITRE dataset. Given the complexities of our expanded dataset, we chose to retrain the model from the ground up, a decision motivated by the need to fully adapt the pre-trained features to our specific dataset and ensure that our model, SecureBERT, effectively learns the nuanced patterns within the cyber threat intelligence domain.

Model Re-training Rationale and Methodology

Why Re-train from the Ground Up?

Retraining SecureBERT from scratch ensures that the model's weights are optimally adjusted to the unique characteristics of our dataset. This method reduces the risk of inconsistency between pre-existing weights trained on a different class distribution and our current dataset, which could lead to suboptimal performance.

Cleaning and Preprocessing of Data

The dataset was thoroughly cleaned to remove superfluous formatting and potential noise that could interfere with the model's learning process. Regular expressions were used to extract and clean the techniques, removing extraneous characters while preserving the essential information.

Splitting Stratified Data

The dataset was divided using a stratified split, with 80% allocated for training and 20% allocated for testing, ensuring a representative distribution of each class. This stratification is critical for preserving the distribution of the dataset and allowing for a reliable evaluation of SecureBERT's performance.

Configuration of SecureBERT

We configured SecureBERT with a dropout layer set at 0.3 to prevent overfitting and added a linear layer for classification into the 194 unique categories present in our dataset using the powerful Hugging Face Framework and PyTorch library. The model was fine-tuned with a conservative learning rate of 1e-5 to adjust the pre-trained BERT core delicately, optimizing it for the complex landscape of cybersecurity threats.

Training Methodology

SecureBERT was trained in batches of 16 over 10 epochs. This batch size was chosen to strike a balance between computational efficiency and weight update granularity. We carefully monitored the loss throughout the training loop, observing a steady decline indicative of the model's increasing proficiency in cyber threat classification.

Metrics and Evaluation

SecureBERT was tested against the validation set after training to determine its classification efficacy. Precision, recall, F1 score, and accuracy were the metrics calculated. These metrics, especially when considered together, provide a comprehensive picture of the model's capabilities. As an example:

Precision (0.7113): This metric reflects the precision with which SecureBERT can predict cyber threat categories, which is critical for trust in its automated classifications.
Recall (0.7277): This metric represents the model's ability to capture all relevant instances, which is critical for comprehensive threat detection.
F1 Score (0.7087): Balances precision and recall, demonstrating the model's ability to identify relevant instances while minimizing false positives.
Accuracy (0.7277): The overall correct predictions that support the model's overall reliability.

Metric	SecureBERT-Value in percentage
Precision	71.13%
Recall	72.77%
F1 Score	70.87%
Accuracy	72.77%

Table VIII: Model Performance Metrics on CTF data

Discussion of Results
Our findings show that SecureBERT has achieved a commendable balance of precision and recall after re-training. The model's robustness is demonstrated not only by its F1 score of 0.7087 and accuracy of 0.7277, but also by its applicability to a wide range of cyber events from threat detection to capture the flag challenges. These findings highlight the model's potential as a reliable tool for identifying and categorizing cyber threats and cyber security Capture the Flag challenges.

Causes of Findings
The careful re-training process, which made use of a stratified dataset split to ensure representative class distribution and the fine-tuning of BERT's pre-trained core, is primarily responsible for SecureBERT's improved performance. In order to adjust the model to the unique characteristics of cybersecurity data without erasing the important features acquired during pre-training, it was imperative to carefully choose a low learning rate.

Possible Effects of Findings
Positive results from this study could have a big impact on the cybersecurity industry. Because of SecureBERT's accuracy in classifying complex cyber threats, threat detection systems may become more effective, speeding up response times and possibly minimizing the damage that cyberattacks do to businesses.

Strengthening Future Research
To further strengthen the research, we suggest:

Dataset Expansion: The Dataset contains 594 out of the current 604 MITRE ATT&CK techniques, because developing capture the flag challenges for those challenges would be difficult and not really applicable. Also due to time constraints and cost, we were not able to generate better challenges with a better GPT model like GPT 4. Also Incorporating a more diverse set of cyber threat intelligence data, when initially training if we follow the method of training on CTI then CTF data might prove better results.

Model Enrichment: Experimenting with additional layers or alternative neural network architectures to improve the model's learning capacity. Also training on CTI then CTF datasets that have the same parameters might hold some positive impact. Experimenting with hyperparameters might prove positive.

Adversarial Training: Introducing adversarial examples during training to enhance the model's robustness against sophisticated cyberattacks.

Applying the Model to the Proposed system in Figure 1: After full development of the model applying the model to the system Would open up for a new open source tool for enhancing cybersecurity training

IX. Conclusion

As we navigate the evolving cybersecurity landscape, the need for sophisticated and potent methodologies to accurately identify, predict, and mitigate cyber threats remains a top priority. Our study, which focuses on integrating the'secureBERT' NLP model with the MITRE ATT&CK framework, represents a significant step forward in cybersecurity training and threat analysis. We created a training module that not only caters to the individual needs of analysts but also streamlines the process through automated categorization and recommendation by combining real-world Cyber Threat Intelligence (CTI) data with synthetically generated Capture The Flag (CTF) challenges.

The findings of this study highlight the effectiveness of deep learning models, particularly'secureBERT', in the cybersecurity domain. The model's performance, as measured by precision, recall, F1 score, and accuracy, demonstrates its potential as a reliable tool for cyber threat categorization and CTF challenge analysis. Furthermore, our approach paves the way for future research, particularly in the field of AI-driven methodologies, to strengthen our digital defenses against a wide range of cyber threats.

To summarize, our research not only makes significant contributions to the field of cybersecurity, but it also paves the way for more comprehensive, AI-enhanced methods to combat the ever-changing dynamics of cyber threats. As demonstrated in our study, the integration of advanced NLP models with established cybersecurity frameworks holds the promise of elevating cybersecurity training and threat

detection to new heights, thereby fortifying the defenses of both the private and public sectors against sophisticated cyber adversaries.

References

[1] V. Orbinato, M. Barbaraci, R. Natella, D. Cotroneo, "Automatic Mapping of Unstructured Cyber Threat Intelligence: An Experimental Study".

[2] J. D. Smith, A. Johnson, "Natural Language Processing in Cybersecurity: Opportunities and Challenges," 2019.

[3] K. Patel, L. Turner, "MITRE ATT&CK Framework: A Guide to Threat Modeling and Cybersecurity Strategy," 2020.

[4] R. Gomez, W. Lee, "The Role of Recommendation Systems in Cybersecurity Training: Personalizing Learning Paths," 2018.

[5] M. R. Smith, "Improving Automated Labeling for ATT&CK Tactics in Malware Threat," 2023.

[6] W. Xiong, E. Legrand, O. Åberg, R. Lagerström, "Cybersecurity threat modeling based on the MITRE Enterprise ATT&CK".

[7] B. Ampel, S. Samtani, S. Ullman, H. Chen, "Linking Common Vulnerabilities and Exposures to the MITRE ATT&CK".

[8] R. Kwon, T. Ashley, J. Castleberry, P. McKenzie, "Cyber Threat Dictionary Using MITRE ATT&CK Matrix and NIST".

[9] E. Aghaei, X. Niu, W. Shadid, E. Al-Shaer, "SecureBERT: A Domain-Specific Language Model for Cybersecurity".

[10] S. Karagiannis, E. Magkos, "Adapting CTF challenges into virtual cybersecurity learning".

[11] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, J. Gao, "Deep Learning Based Text Classification: A Comprehensive Review," ACM Computing Surveys, vol. 54, no. 3, pp. 1-40, April 20211.

[12]B. Al-Sada, A. Sadighian, and G. Oligeri, "MITRE ATT&CK: State of the Art and Way Forward," Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Qatar, 2023. [Online]. Available: https://arxiv.org/pdf/2308.14016.pdf.

[13]N. Naik, P. Jenkins, P. Grace and J. Song, "Comparing Attack Models for IT Systems: Lockheed Martin's Cyber Kill Chain, MITRE ATT&CK Framework and Diamond Model," 2022 IEEE International Symposium on Systems Engineering (ISSE), Vienna, Austria, 2022, pp. 1-7, doi: 10.1109/ISSE54508.2022.10005490.

[14]C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge: Cambridge University Press, 2008.

[15]R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. New York: ACM Press, 1999.

[16]R. Glauber and A. Loula, "Collaborative Filtering vs. Content-Based Filtering: differences and similarities," arXiv preprint, 2019. [Online]. Available: https://arxiv.org/pdf/1912.08932.pdf

[17] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," In The adaptive web, pp. 325-341. Springer, Berlin, Heidelberg, 2007.

[18] R. Burke, "Hybrid recommender systems: Survey and experiments," User modeling and user-adapted interaction, vol. 12, no. 4, pp. 331-370, 2002.

[19] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE transactions on knowledge and data engineering, vol. 17, no. 6, pp. 734-749, 2005.

[20] D. Jannach and G. Adomavicius, "Recommendations with a purpose," In Proceedings of the 10th ACM Conference on Recommender Systems, pp. 7-10, 2016.

.