International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

An NLP-Based Approach to Fortifying Cyber Defenses

[1]Albia Maqbool, [2]Dr. Raghav Mehra, [3]Jihane Ben Slimane, [4]Eman H. Abd-Elkawy, [5]Dr. Nargis Parveen, [6]Dr. Bindiya Ahuja, [7]Greeshma G. S.

Submitted: 15/01/2024 **Revised**: 23/02/2024 **Accepted**: 01/03/2024

Abstract: This research introduces an innovative approach to fortifying cybersecurity defenses through behavior-based anomaly detection and response mechanisms. Leveraging NLP techniques, our LLM analyzes system logs and Websites to identify anomalous patterns indicative for classification of the type of attack. On analysis of the datasets "Exploits Database"," GDELT" and "OTX", the system accurately detects deviations and dynamically suggests security measures based on the severity and class of attack. Evaluation on diverse datasets showcases the model's superiority over traditional signature-based methods, emphasizing its efficacy in identifying novel and sophisticated cyber threats. The model has an accuracy of 71.22% in classifying large amount of unlabeled data. This research contributes valuable insights to the ongoing efforts in fortifying digital ecosystems against evolving cybersecurity challenges.

Keywords*: cybersecurity, fortifying, GDELT, Exploits Database, OTX*

1. Introduction

In the ever-evolving landscape of cybersecurity, the identification and timely response to emerging threats are critical components of a robust defense strategy. This research paper introduces a novel approach to cybersecurity threat management through the integration of a Linguistic Logic Model (LLM). The proposed system aims to enhance threat classification accuracy and response efficiency by leveraging the linguistic representation of cybersecurity events and patterns.

The methodology involves the development and training of an LLM-based model capable of understanding the nuanced language of cybersecurity reports and incident logs. The linguistic analysis enables the model to categorize threats into specific classes, providing a fine-grained threat classification system. The LLM's ability to capture contextual information and subtle linguistic nuances contributes to a more nuanced understanding of threats compared to traditional rule-based or statistical models. [1] [2]

Furthermore, the paper explores the implementation of the LLM within a comprehensive cybersecurity response framework. The model not only aids in accurate threat identification but also facilitates an adaptive and dynamic response strategy. The response mechanism integrates insights from the linguistic analysis, allowing for contextual and linguistically-informed countermeasures.

The effectiveness of the proposed Cybersecurity Threat Classification and Response using LLM is evaluated through extensive experimentation on diverse datasets containing real-world cybersecurity incidents. Comparative analyses against existing threat detection and response systems showcase the advantages of the linguistic logic model in terms of precision, recall, and response time. [3] [4]

This research contributes to the advancement of cybersecurity methodologies by introducing a linguistically-informed model that not only improves threat classification accuracy but also enhances the agility and adaptability of response mechanisms. The findings underscore the potential of linguistic logic models in addressing the evolving nature of cybersecurity threats and provide a foundation for future developments in linguistics-based cybersecurity solutions.

2. Literature Review

The article by Mogollón-Gutiérrez, et al. proposes an advanced cybersecurity system utilizing artificial

[1]*Department of Computer Sciences Faculty of Computing and Information Technology, Northern Border University, Kingdom of Saudi Arabia*
Email: albia.alam@nbu.edu.sa
[2]*Professor, Department AI/ML, Chandigarh University*
Raghav.mehrain@gmail.com
[3]*Department of Computer Sciences Faculty of Computing and Information Technology, Northern Border University, Rafha 91911, Saudi Arabia, National Engineering School of Tunis, LR11ES20 Analysis Design and Control of Systems Laboratory, University of Tunis El Manar, Tunis, Tunisia*
Email: jehan.saleh@nbu.edu.sa
[4]*Department of Computer Sciences, Faculty of Computing & IT, Northern Border University, Saudi Arabia, Department of Mathematics and Computer Science, Faculty of Science, Beni-Suef University, Beni-Suef 62521, Egypt*
Eman.Hassan@nbu.edu.sa
[5]*Lecturer Department of Computer Science, Faculty of Computing and Information Technology, Northern Border University, Kingdom of Saudi Arabia*
Email: nargis.norulhaq@nbu.edu.sa
[6]*Professor Department CSE, Lingaya's Vidyapeeth*
Bindiya.bhatia@gmail.com
[7]*Assistant Professor, Department -Computer science and Engineering Galgotias university*
greeshma.s@galgotiasuniversity.edu.in

intelligence (AI) algorithms for network traffic analysis. Addressing the escalating threat of cyberattacks, the system employs a two-stage ensemble model with a one-vs-rest strategy to handle class imbalances effectively. In the first stage, binary models distinguish each type of traffic, forming the basis for comprehensive multiclass classification in the second stage. Evaluated on the UNSW-NB15 dataset, the model achieves a superior F1 score of 0.912 for binary and 0.7754 for multiclass classification, outperforming state-of-the-art approaches. The article underscores the significance of employing AI-driven strategies to fortify information access amid evolving cyber threats, marking a notable advancement in cybersecurity. [5]

The study by Kim, et al. addresses the intricate challenge of identifying insider threats in cybersecurity, where individuals with legitimate access may harbor malicious intentions. Traditional intrusion detection systems often struggle to discern between normal and malicious behavior, especially when insider threats share feature spaces with routine activities. To overcome these limitations, the study proposes an enhanced anomaly detection approach, incorporating discrete wavelet transformation. The methodology proves effective in differentiating between normal and malicious users by uncovering new patterns and decomposing synthesized data. Experiments, conducted using Carnegie Mellon University's CERT dataset, specifically focusing on insider threat scenarios, demonstrate the methodology's efficacy. The results reveal a substantial reduction in false-positive rates (82% to 98%) compared to scenarios where wavelet transformation is not applied. The study concludes that the proposed methodology, leveraging discrete wavelet transformation, holds great promise for improving insider threat detection in cybersecurity, particularly in scenarios characterized by shared feature spaces between normal and malicious behavior. [6]

The article by Alrayes, et al. addresses cybersecurity concerns in IoT cloud networks, specifically focusing on the threats posed by malware attacks and software piracy. Introducing the Enhanced Artificial Gorilla Troops Optimizer with Deep Learning Enabled Cybersecurity Threat Detection (EAGTODL-CTD) model, the study aims to enhance security measures. The EAGTODL-CTD model is tailored for threat identification in IoT cloud environments, utilizing a unique approach that involves converting input binary files into color images for malware detection through image classification. The model incorporates a cascaded girded repeated unit (CGRU) model aimed at threat discovery and classification, with the EAGTO approach serving as a hyperparameter enhancer for fine-tuning CGRU parameters. Performance evaluation on a dataset featuring malignant and benign class labels showcases the model's exceptional accuracy, achieving a notable 99.47%. The integration of ML and DL in this cybersecurity solution highlights its potential effectiveness in mitigating security challenges in IoT cloud networks. [7]

The paper by Silvestri, et al. proposes a methodology employing ML models, including BERT and XGBoost, to analyze threats and vulnerabilities in healthcare systems. It involves modeling the healthcare ecosystem, extracting threat/asset pairs with BERT, and calculating vulnerability scores using XGBoost based on Common Vulnerabilities and Exposures (CVE) reports. The three-step approach aims at effective threat assessment for risk management. Tested on web-extracted datasets, it showcases automatic extraction and calculation of threat and vulnerability levels in healthcare assets. Future work involves refinement, testing on diverse datasets, and integration into real-world environments. The study highlights the potential of ML in fortifying healthcare cybersecurity. [8]

The study by Al-Essa, et al. introduces PANACEA, a cyber-threat detection method utilizing ensemble learning and adversarial training in DL. To address the challenge of choosing accurate root models for ensembles, the approach incorporates model ensemble trimming based on eXplainable AI (XAI). By identifying base models emphasizing different input feature subspaces, the ensemble gains variety and correctness in classification. Global XAI techniques measure ensemble model diversity concerning input features' impact on base models' accuracy. Experiments on four cybersecurity datasets demonstrate the effectiveness of combining accusatorial training, aggregate learning, and XAI in enhancing the accuracy of multi-category classifications for cyber-figures. [9]

The study by Razali, et al. proposes a hybrid approach for predicting dogmatic security threats, combining a lexicon-based method with ML classifiers (Decision Tree(DT), Naive Bayes(NB), and SVM). The hybrid approach, specifically using the token-based method with the DT classifier, outperformed other combinations with the highest accuracy, precision, and recall scores (69%). Decision Tree showed superior overall performance, emphasizing accurate threat class predictions. Naive Bayes exhibited lower accuracy and precision, indicating room for improvement. SVM had a lower accuracy than Decision Tree but a higher recall value. The research contributes to understanding the interplay between sentiments, views, and dogmatic security fears in cyberspace. The proposed hybrid approach demonstrates effectiveness in detecting threats based on sentiments in online bulletin text, offering valuable insights for national security. Future work suggests scaling the analysis with a

larger dataset to improve opinion mining in the nationwide security. [10]

The paper by Abinesh Kamal, et al underscores the importance of fit in threat aptitude into security procedures to combat cyber threats effectively. Using the UNSW-NB15 dataset, the study proposes an incorporated risk intelligence framework, employing various ML classifiers. Performance evaluation reveals that the Ensemble Learning classifier outperforms others, achieving high accuracy (97.02%), precision (98.34%), recall (99.02%), and an F1 score (98.17%). The findings suggest that the proposed system, particularly with Ensemble Learning, is highly successful in detecting potential threats. The robust performance metrics imply reliability and effectiveness in providing valuable insights for organizational security operations, emphasizing the role of ML classifiers in bolstering cybersecurity postures. [11]

The research by Kabir, et al. addresses the growing cybersecurity concern of insider threats and the associated risk of malware attacks. To enhance malware detection accuracy, a ML Model, specifically an NN, is proposed. The workflow involves attribute extraction, irregularity detection, and classification. The CERT4.2 dataset is utilized, and data preprocessing includes encoding text strings and distinguishing between threat and promise records. The developed ML model integrates fully connected and ReLU activated, and dropout layers for regularization. Comparative analysis with other cataloguing techniques such as RF, NB, KNN, SVM, DT, LR, and GB demonstrates the effectiveness of the projected method. The results indicate that the model functions properly, achieving 100% accuracy in detecting malware associated with insider risks. [12]

The paper by Dapel, et al. gives a comprehensive survey of the application of AI in the field of cybersecurity, explicitly focusing on the detection and prevention of cyberattacks. The review highlighted the significance and impact of AI technologies and algorithms in managing cybersecurity, considering frameworks and solutions from 2018 to 2021. The evolving landscape of information and communication technology has led to a substantial increase in cyber threats, necessitating real-time solutions due to the inadequacy of traditional techniques. The study found that Long Short-Term Memory (LSTM) demonstrated effectiveness with favorable computational complexity and low training time. Random Forest, on the other hand, exhibited high accuracy in anomaly intrusion detection. The conclusion emphasizes the absence of a one-size-fits-all solution for cybersecurity challenges, advocating for a holistic approach. Looking ahead, the paper suggests the potential combination of LSTM and Random Forest as integrated protective solutions for addressing the dynamic and complex nature of cybersecurity threats. [4]

The research paper by Darem, et al. investigates the growing menace of cyber threats in the banking and financial sectors, recognizing their pivotal role in the economy and the increasing complexity of modern cyber threats. It aims to comprehensively analyze these threats, emphasizing their significance and potential consequences on the economy. The research contributes by classifying cyber threats based on severity and technicality, offering valuable insights for implementing tailored countermeasures. Countermeasures discussed include technical, non-technical, organizational, legal, and regulatory approaches, crucial for safeguarding financial transactions. However, the paper acknowledges challenges, especially the rapidly evolving nature of cyber threats, requiring continuous adaptation in cybersecurity measures. The examination of recent trends underscores the dynamic landscape, emphasizing the necessity for ongoing improvement in cybersecurity strategies to protect the critical banking and financial sectors. [13]

The research paper by Cherqi, et al. addresses the question of effectively utilizing (OTIFs for enterprise security, acknowledging the scarcity of high-quality annotated data in these feeds. To overcome this challenge, the paper introduces a novel partial supervised learning approach, leveraging both labeled and non-labeled data for automated threat identification. The advanced GAN-BERT framework, combining GANs and BERT, is presented as a key innovation. Experimental results demonstrate the superiority of the proposed method over original BERT and GAN-BERT, with significant improvements in F1-score. The paper emphasizes the efficient selection of hard negatives during training and positions the approach as a robust solution for enhancing automatic hazard detection in Cyber Threat Aptitude, plummeting dependance on human administration and successfully handling limited annotated data. [14]

Dataset description

The data used was collected form the various sources given below

a: OPEN THREAT EXCHANGE (OTX)

It is an open sourced, mass source computer cybersecurity suite. It is the world's biggest site of its sort, through over 160K members from 144 nations. This dataset includes occurrences that targeted certain sectors and groups around various parts of the globe. We created our own internet spider and saved all of the instances as unprocessed HTML files. We were capable of to identify over 18000 distinct security incidents that happened

throughout 2016 and 2021. A search engine spider was developed and installed to provide sequential access to shared events. After reloading the status on which it was dismissed, the spider can restart its activity. We store the raw data collected by the spider into an internal file system for later rendering.

b: EXPLOITSDB

Aggressive Safety runs the Exploits Database [15] as a open source effort. It includes openly accessible CVE conforming vulnerabilities and susceptibility proofs provided by cybersecurity experts and penetration testers for vulnerabilities. The information comprises the time the incident were disclosed, the targeted OS or website, a written explanation of the attack or exploit, the kind of exploits, and, if it was an online facility, the connection port of the compromised product.

c: GLOBAL DATABASE OF EVENTS, LANGUAGE AND TONE (GDELT)

The GDELT Project [16], backed from Google Jigsaw, oversees universal transmission broadcast, and online newscast in more than one hundred languages on every day basis, identifying the individuals, web sites, things, conceptual, news organizations, sensitive reactions, the sum total, references, graphics, and events that shape our worldwide culture, resulting a free open-source platform for computing around the world.

Preprocessing and tokenization

For extraction of the words or vocab we wanted from apiece of the threat-sharing Datasets and webpages stored on the computer's directories, we used a partially controlled version of HTML with tags specific to respective fields. The most important info we've acquired regarding the occurrences of attacks are the incident outlines, event titles, and threat declaration dates. As per a consequence, we employed threat explanations to shape our study ground work. Tokenization, text standardization (e.g., changing from upper to lower cases), and converting to UTF-8 format was utilized to pre-process or extract threat descriptions from various sources. These techniques attempt to guarantee that diverse ways of phrasing the identical expression are handled equally, and that the content is appropriate for use with the ML processes that follows. The first table outlines the outcomes of the preliminary processing approaches that we employed to build our testbed.

The processing stage removes null and incorrect values, in addition to duplicate items that have been cross-posted on other marketplaces.

Platform	Threats Total	Threats Labeled	% labeled	Classes	Collected on
GDELT	53228	16437	30 Approx.	9	Jul 2022
OTX	18099	1263	7 Approx.	4	13/07/2022
Exploits DB	44756	8950	20 Approx.	5	19/02/2022
Total	**136397**	28629			

Table 1: Dataset description platform wise

We employ "ByteLevelBPETokenizer" tokenizer class from the "Hugging Face Transformers library" to tokenize texts into individual sub words, which was initially employed for GPT 2. It's grounded on Byte Pair Encoding (BPE) method, which compresses data by replacing the most common pair of successive bytes in an instruction with a separated, wasted byte. The "ByteLevelBPETokenizer" is particularly effective for dealing with OOV terms that do not appear in the tokenizer's human language vocab. By dividing up our linguistic arrangement of net traffic info into miniature sub words that are expected to be included in the tokenized lexicon as an ordered set of bytes, you may efficiently handle traffic info using BERT. The "TrainByteLevelBPETokenizer" process (Algo 1) begins by specifying the path for HTML transformed to text only files. If the folder in question does not exist, it will be created. The "ByteLevelBPETokenizer" is then adjusted. A brand-new file located in the folder beneath has been unlocked for text, and each column from the data being entered "text_dat" is written to it, each with a string new line character. Once all of the text has been inscribed, a file is saved. The process then establishes a vocab limit of 5000 and a list of specific characters. Finally, the tokenizer algorithm is taught by means of the Threat designation, "voca_text", a minimum number of two, and the special token list.

Algo 1

1. procedure TRAINBYTLEVELBPETOKENIZER (text_dat)
2. txt_f ← "content_dir/temp/txt_split"
3. create directory txt_f

4. tokenizer ← new BLBPET
5. file_na ← join txt_f and 'text_dat.txt'
6. open file_na for in write mode as file1
7. for each col in text_dat['text'] do
    a. col ← col + newline
    b. write col to file1
8. end for
9. close file
10. voca_size ← 5000
11. spc_tokens ← ["<s>", "<pad>", "</s>", "<unk>", "<mask>"]
12. train tokenizer with Threat_na, voca_Text, min_freq, and spc_tokens
13. End procedure

3. Methodology

In this, we propose a novel framework for threat classification, detection and response. The Framework is constructed in 2 phases, first is training of the BERT based model for threat classification, called training GAN-convBERT. Then we integrate FalconLLM for response according to class of threat.

The figure below explains the steps taken for the same

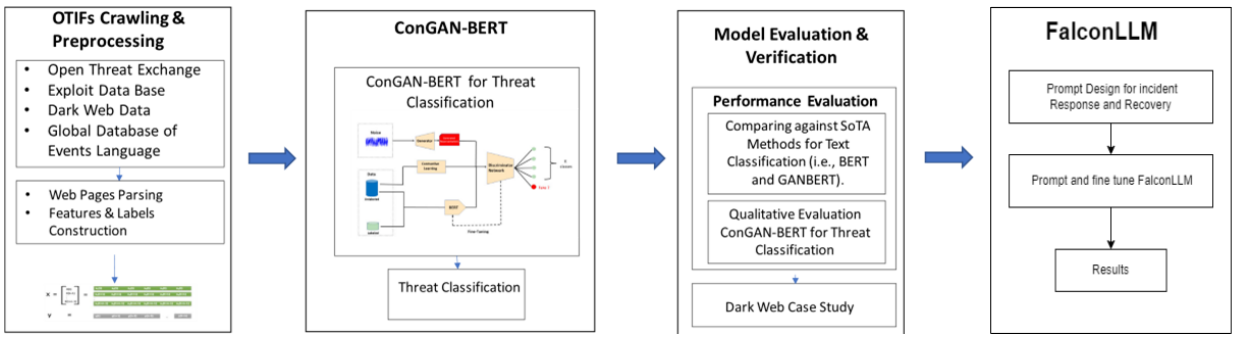

Fig 1: Methodology of proposed architecture

Training GAN-convBERT

We refined the GAN and BERT technique for threat recognition by introducing a auto-supervision contrast loss technique for readjustment and optimization BERT to take use of the enormous quantities of data with annotations, resulting in better phrase representations. The suggested framework for threat categorization integrates GAN-BERT and CL.

Inspired by GAN-BERT's performance in recognizing texts with little labeled data, we introduce GAN-convBERT, a novel textual classifier that uses GAN-BERT as its foundational SSL approach.

The primary goal of GAN-BERT is for training the discriminator, D, to discriminate amid actual samples created through a previously trained BERT model and fraudulent patterns produced by generator portion. The discriminator undergoes conditioning on (k + 1) categories, with the (k + 1)th category being the "fake" class and the remaining k classifications representing the "actual" classes.

Loss Function of Generator consists of two separate loss aspects: the feature matched as well as unsupervised loss. The feature matched loss evaluates the generator's ability to produce instances with characteristics as similar as feasible to real-world examples. This loss is estimated by contrasting the depictions of the discriminator's middle layer, denoted by f(x). The feature matched losses of G may be expressed as follows:

LGfeture matching = $Ex{\sim}p$dat$f(x) - ExGf(x)$

The mistake generated by phony instances properly detected by the method of discriminators is quantified by the following unsupervised loss:

$$L_{G_{unsup}} = -E_{x\sim G}\, log[1 - pm(\hat{y} = y \mid x, y = k + 1)]$$

Thus, generator's lossy function is abridged as:

$$LG = LGunsup + LGfeature\ matching$$

On the other hand, the discriminator's purpose is to discriminate between genuine and false cases while simultaneously categorizing the tiny set of labeled true examples according to a single of k initial groups. The discriminator's losses function involve 2 loss values: the monitored loss LDsup, which restricts the error in categorization of actual scenarios amongst the k unique groups, and the unsupervised loss LDunsup, which amounts the errors of incorrect identification of a actual (unannotated) example as "fake and failing" to recognize a generated example. To avoid dispersion in the uncontrolled constituent, class data are used again in every collection by some proportion described below, ensuring that all tagged occurrences are included. The discriminator's loss function may now be expressed in its final form as follows:

$$LD = LDsup + LDunsup$$

where:

$$LDsup = Ex, y \sim pdata\ logpm[(y\hat{} = y \mid x, y \in (1, \ldots, k))]$$

$$LDunsup = Ex \sim pdata\ log[1 - pm(y\hat{} = y \mid x, y = k + 1)] - Ex \sim G\ log[pm(y\hat{} = y \mid x, y = k + 1)]$$

In this paper, we present a unique strategy to training the GAN-BERT framework for cyber-attack detection. For additional leverage the knowledge in unannotated information to enhance the effectiveness of our model, we suggest including auto-supervision contrast loss function into the discriminator's demonstrable function.

Our goal for introducing auto-supervision contrast loss function is for addressing the issue of unclear groups in the computer security arena, here related terminology are frequently employed. Using contrast learning on unannotated information, the LLM may understand to differentiate between interrelating categories, boosting its effectiveness in threat cataloguing.

The use of auto-supervision contrast learning offers an excellent technique to use the huge quantity of unlabeled data accessible in the information security arena, resulting in increased performance in the job of identifying cyber threats. Our technique shows that by utilizing the knowledge included in data that is not annotated through auto-supervision contrastive learning, GAN-BERT model performance may be improved even further.

To train the GAN-convBERT LLM, we define a novel discrimination objective function:

$$LD = (1 - \lambda) * (LDsup + LDunsup) + \lambda * LContrastive$$

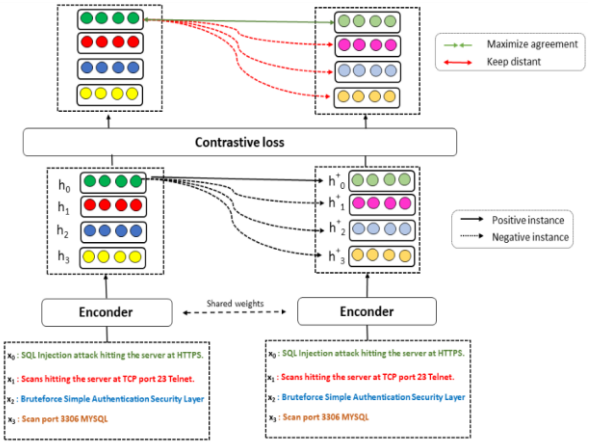

Fig 2: Learning Architecture for GAN-convBERT

λ specifies the relative significance of the two variables, influencing the ultimate loss value. A large λ value prioritizes the contrastive loss above the discriminator's monitored and uncontrolled losses. This enables for more command regarding the transaction between standard monitored and uncontrolled losses, as well as the recently included contrasting loss.

The contrast loss's training aim is to make the most of the resemblance among the depictions of alike samples mean while reducing the resemblance among the depictions of divergent examples inside a small lot of n pairings, as shown in Figure 2. Hence, for given set x_k comprising a positive pair of samples where x_i and x_i^+ are meaningfully related, h_i and h^+_i denote the mathematical expressions of x_i and x_i^+. This can be formulated as the following loss function:

$$L_{Contrastive} = -log \frac{e^{sim(hi, h_i^+)/\tau}}{\sum_{j=1}^{n} e^{sim(hi, h_j^+)/\tau}}$$

here $sim(h_i, h^+_i)$ representing cosine similarity is defined below:

$$sim(hi, h_i^+) = \frac{h_i h_i^+}{\|h_i\| \|h_i^+\|}$$

The cosine relationship was selected for this investigation because of its features and effective use in previous studies. The cosine correlation measure takes into account vector aspect in a high-dimensional space, allowing it to reliably identify resemblance among two vectors irrespective of magnitude. This is especially important in NLP projects like ours, where a term's frequency (magnitude) may frequently overwhelm the context (direction) in which it is employed. Besides, cosine resemblance is not as computing power costly than other distance metrics when working with scarce vectors since it only takes into consideration nonzero dimensions. This trait adds to its applicability for LLMs such as BERT,

which produce multidimensional and thin word vectors. [15]

FALCON Integration security system

After detecting cyber threats using the GAN-convBERT model, we use FalconLLM as a threat reaction and restoration solution. FalconLLM is an open-sourced LLM in Hugging Face that currently holds the number one position in the open-sourced LLM Ranking (15/06/2023). FalconLLM is a self-regressive decoder-only model that has forty billion attributes that was developed using 1 trillion token. The training procedure lasted two months and used 384 GPUs in AWS. Falcon's prior dataset, known as Refined Web, was created using data acquired from public web crawls.

FalconLLM's advanced language comprehension skills may be used to assess and comprehend complicated threat data from a variety of sources. It is capable of reading and analyzing massive quantities of text system logs from computers, newsflashes, and extra sources of info, taking out pertinent information and recognizing sequences that may be an indication of a cybersecurity threat. As soon as a possible, when a threat is found, FalconLLM may determine its sternness and possible impact by juxtaposing it to its huge in-house information library. Then it may recommend numerous moderation strategies and recuperation procedures, guiding the response team via a well-planned, sequential procedure. FalconLLM's ability to continuously learn from new data may be beneficial in the reacting phase. It may assess previous occurrences, find gaps in response protocols, and provide recommendations for improvements. This proactive and iterative strategy enables a faster, more effective threat action and recuperation procedure, reducing the potential harm caused by cyber attacks. Figures 3 demonstrate how FalconLLM may be utilized as a threat action and recuperation system for cyber risks identified through the GAN-convBERT model.

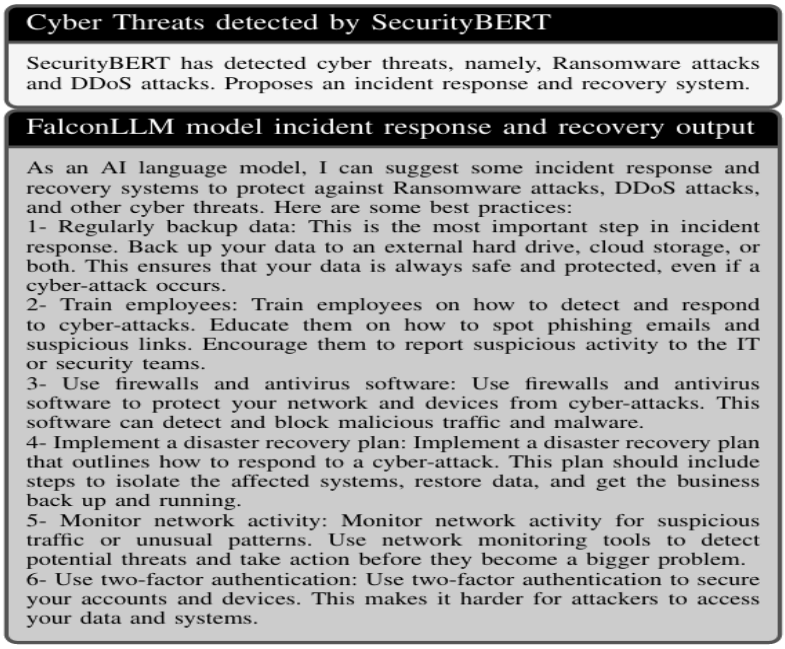

Fig 3: FalconLLM response

FalconLLM's functionality may be quantitatively modeled by representing its basic mechanisms as variable quantity. These key components of FalconLLM workflow are detailed below:

• Data Inputs (DI): FalconLLM analyzes raw textual logs from computer logs, reports of incidents, and other sources.

• The Threat Detection (TD) stage gathers pertinent data and finds precedents which signal a cybersecurity threat.

• The Threat Assessment (TA) stage compares a threat's significance level as well as possible impact to FalconLLM's proprietary knowledge database.

• The Mitigation Method Proposal (MMP) outlines possible remedies and recovery procedures.

• FalconLLM's Learning Process (LP) enables continuous learning and adaptation to new data.

• FalconLLM's Incident Assessment and Improvements (IAI) process analyzes historical occurrences to make changes.

• FalconLLM aims to minimize the effect of cyber risks discovered through the GAN-convBERT model, improving general security.

The mathematics abstract of FalconLLM's procedure may be represented as follows:

$$DM = f(DI, TD, TA, IAI, MMP, LP)$$

Here, f represents the operations implemented in FalconLLM.

4. Results and Discussion

In our paper we exhibit the outcomes of our experimental study assessing the efficacy of the proposed GAN-convBERT system. We compared our method against industry-standard text categorization benchmarks, BERT and GANBERT, using datasets from various operational tech and industry management platforms. The evaluation, based on the F1-score, a widely used metric in text categorization tasks, also included precision and accuracy metrics. Results indicate that our strategy, incorporating auto-supervision learning and a contrast loss function, outstrips further approaches. To guarantee impartiality, we aggregated outcomes from five diverse train datasets. We described the average accuracy in a brand new untouched testing set, with all techniques trained on the same dataset.

Empirical findings demonstrate that our GAN-convBERT model outperforms baseline data. For instance, on the Exploit dataset, the BERT-only model achieved 58.91% accuracy with limited annotated data (0.1% of observations). In contrast, employing a Generative Adversarial Network to utilize unlabelled dataset components throughout training phase significantly enhanced performance. GAN-BERT LLM surpassed BERT-only approach by five points, reaching 63.67% accuracy. Furthermore, following our proposed technique, we achieved even better outcomes, by an accuracy rate of 71.22% with identical amount of well labelled data. These findings show the benefits of integrating unlabeled info into the GAN-convBERT architecture, with accuracy improving as annotated data increases.

In a similar vein in the GDELT a database, employing a GAN concerning labeled information improved performance. With only 0.1% labeled data, BERT obtained 16.01% accuracy, GAN-BERT 20.17%, while our technique 26.78%. The pattern continued until about 2% of classified examples, when all models attained equivalent accuracy echelons. Though, from 0.61% through 1.00% of labeled cases, the GAN-BERT LLM somewhat performed better than GAN-convBERT. Similarly, on the OTX dataset, for labeled percentage of 0.10% BERT showed low efficiency, while the GAN-BERT model demonstrated substantial improvement. Nevertheless, the GAN-convBERT architecture exhibited the highest gain, with an accuracy of 55.11%, 16 points upsurge over the GAN-BERT. This tendency continued with the implementation of additional labeled data.

Model Name	Annotated	Exploit Acc.	Exploit F1.	Exploit Pre.	GDELT Acc.	GDELT F1	GDELT Pre.	OTX Acc.	OTX F1.	OTX Pre.
BERT	0.1%	58.91	43.68	34.72	16.00	8.61	11.86	8.43	4.43	11.06
GAN-BERT		63.67	62.21	66.21	20.16	13.48	20.92	38.74	36.15	42.03
GAN-convBERT		**71.44**	**70.38**	**74.38**	**26.79**	**26.61**	**30.77**	**55.10**	**43.38**	**37.10**
BERT	0.2%	58.74	43.48	34.52	13.95	6.88	10.42	28.04	24.74	27.39
GAN-BERT		68.80	63.60	66.40	34.89	31.42	40.33	48.93	46.90	48.03
GAN-convBERT		**74.99**	**74.31**	**76.35**	**37.34**	**35.89**	**38.37**	**56.68**	**48.88**	**45.40**
BERT	0.4%	60.74	46.92	38.82	18.01	7.73	8.88	30.24	27.62	31.73
GAN-BERT		69.42	66.68	67.09	38.97	36.19	40.40	54.84	52.74	57.44
GAN-convBERT		**78.85**	**78.07**	**78.88**	**42.64**	**40.69**	**44.19**	**61.90**	**57.23**	**58.27**
BERT	0.6%	60.74	46.92	38.82	19.68	10.60	14.49	40.43	34,08	39.49
GAN-BERT		74.17	72.57	77.48	**45.96**	44.19	**50.06**	56.75	53.99	55.36
GAN-convBERT		**80.66**	**80.03**	**80.46**	45.16	**45.51**	48.01	**66.43**	**60.67**	**67.79**
BERT	0.8%	61.31	48.00	40.19	20.33	9.79	10.80	49.45	41.89	41.264
GAN-BERT		75.63	72.77	73.45	48.20	46.69	**52.15**	65.16	**63.34**	**60.56**
GAN-convBERT		**81.89**	**81.46**	**81.64**	**49.54**	**49.34**	51.54	**69.33**	62.38	60.04
BERT	1%	63.39	52.06	54.20	22.47	13.24	19.02	54.31	42.29	39.12
GAN-BERT		75.36	73.96	75.60	**50.32**	49.09	**52.69**	70.76	69.01	68.73
GAN-convBERT		**81.93**	**81.53**	**81.78**	49.35	**49.63**	52.52	**72.35**	**70.80**	**70.38**
BERT	2%	75.03	71.75	71.50	38.75	30.18	31.51	64.76	65.01	68.73
GAN-BERT		81.29	80.99	82.56	54.26	53.23	56.42	80.40	76.49	79.68
GAN-convBERT		**84.73**	**84.42**	**84.61**	**55.47**	**55.36**	**56.70**	**82.49**	**80.81**	**79.94**

Table 2: Performance for various percentages of Labeled Data

An ablation research on the GAN-convBERT a system exploring the influence of important hyperparameters, including batch limit, temperatures, and weight factor λ, while using unlabeled samples. The results show that bigger unlabeled datasets provide a significant benefit to GAN-convBERT, improving model generalization.

Regarding batch size, we showed that a lower size of 32 performs well on the Exploit the data set, assuring increased model efficiency, but higher sizes may not ensure superior results. The study examines hyper parameters λ and temperature, offering statistical insights into the best settings for improved accuracy. A meticulous grid search shows that the maximum accuracy is reached when λ is reduced to 0.006 and temp is set to 2.5. The findings indicate that the model performs best at temperatures ranging from 2.0 to 3.5. The contrasting loss function modification tries to consider lengthier sentences as "complex negative adjectives," demonstrating higher performance on selected datasets with GAN-convBERT. It does, however, underline a dataset-dependent aspect, underlining the importance of resilience. The paper provides useful insights for modifying hyperparameters and enhancing efficiency in the GAN-convBERT framework, which is supported by statistical analysis and a comprehensive knowledge of the cybersecurity text categorization task.

5. Conclusion

This study demonstrates the outstanding potential for LLM in the field of cybersecurity, especially by integrating GAN-ConvBert-FAL for threat identification and incident response. The novel use of BERT framework for cyber-attack identification, represented in GAN-convBERT, shows amazing performance, defying initial expectations about incompatibility owing to the reduced relevance of syntactic components. Furthermore, the use of FalconLLM produced a comprehensive threat action and restoration system. The experimental findings demonstrated the superiority of this technique over traditional ML and DL models such as CNN, DL nets, and RNN. The GAN-ConvBert-FAL model, evaluated on a gathered cybersecurity dataset, demonstrated an exceptional capacity to identify fourteen various types of threats with an avg accuracy of 71.22%, emphasizing the capacity for transformation of LLMs in the field of cybersecurity.

While this article has made substantial progress in extending the application of LLMs in the field of cybersecurity, subsequent studies might take a variety of approaches to build on these promising discoveries. One potential approach is to fine-tune and extend the GAN-ConvBert-FAL system to improve its performance across different attack types, including adversarial assaults and more sophisticated threats. Furthermore, because cyber threats are always developing, the GAN-ConvBert-FAL model will need to be updated and trained on the most recent real-world datasets in order to remain effective.

References

[1] S. Srinivasan and P. Deepalakshmi, "Enhancing the security in cyber-world by detecting the botnets using ensemble classification based machine learning," Measurement: Sensors, vol. 25, p. 100624, 2023.

[2] M. Amine Ferrag, M. Ndhlovu, N. Tihanyi, L. C. Cordeiro, M. Debbah and T. Lestable, "Revolutionizing Cyber Threat Detection with Large Language Models," arXiv e-prints, p. arXiv–2306, 2023.

[3] W. S. Admass, Y. Y. Munaye and A. A. Diro, "Cyber Security and Applications".

[4] M. E. Dapel, M. Asante, C. D. Uba and M. O. Agyeman, "Artificial Intelligence Techniques in Cybersecurity Management," in Cybersecurity in the Age of Smart Societies: Proceedings of the 14th International Conference on Global Security, Safety and Sustainability, London, September 2022, 2023.

[5] Ó. Mogollón-Gutiérrez, J. C. Sancho Núñez, M. Á. Vegas and A. Caro Lindo, "A Novel Ensemble Learning System for Cyberattack Classification.," Intelligent Automation & Soft Computing, vol. 37, 2023.

[6] D.-W. Kim, G.-Y. Shin and M.-M. Han, "Anomaly Detection Based on Discrete Wavelet Transformation for Insider Threat Classification.," Comput. Syst. Sci. Eng., vol. 46, p. 153–164, 2023.

[7] F. S. Alrayes, N. Alotaibi, J. S. Alzahrani, S. Alazwari, A. Alhogail, A. M. Al-Sharafi, M. Othman and M. A. Hamza, "Enhanced Gorilla Troops Optimizer with DL Enabled Cybersecurity Threat Detection," Computer Systems Science & Engineering, vol. 45, 2023.

[8] S. Silvestri, S. Islam, S. Papastergiou, C. Tzagkarakis and M. Ciampi, "A Machine Learning Approach for the NLP-Based Analysis of Cyber Threats and Vulnerabilities of the Healthcare Ecosystem," Sensors, vol. 23, p. 651, 2023.

[9] M. Al-Essa, G. Andresini, A. Appice and D. Malerba, "PANACEA: a neural model ensemble for cyber-threat detection," Machine Learning, p. 1–44, 2024.

[10] N. A. M. Razali, N. A. Malizan, N. A. Hasbullah, M. Wook, N. M. Zainuddin, K. K. Ishak, S. Ramli and S. Sukardi, "Political Security Threat Prediction Framework Using Hybrid Lexicon-Based Approach and Machine Learning Technique," IEEE Access, vol. 11, p. 17151–17164, 2023.

[11] K. U. Abinesh Kamal and S. V. Divya, "Integrated threat intelligence platform for security operations in organizations," Automatika, vol. 65, p. 401–409, 2024.

[12] M. H. Kabir, A. Hasnat, A. J. Mahdi, M. N. Hasan, J. A. Chowdhury and I. M. Fahim, "Enhancing Insider Malware Detection Accuracy with Machine Learning Algorithms," Engineering Proceedings, vol. 58, p. 104, 2023.

[13] A. Darem, A. A. Alhashmi, T. M. Alkhaldi, A. M. Alashjaee, S. M. Alanazi and S. A. Ebad, "Cyber threats classifications and countermeasures in banking and financial sector," IEEE Access, vol. 11, p. 125138–125158, 2023.

[14] O. Cherqi, Y. Moukafih, M. Ghogho and H. Benbrahim, "Enhancing Cyber Threat Identification in Open-Source Intelligence Feeds through an Improved Semi-Supervised Generative Adversarial Learning Approach with Contrastive Learning," IEEE Access, 2023.

[15] P. Das, M. R. Al Asif, S. Jahan, R. Khondoker, K. Ahmed and F. M. Bui, "STRIDE-Based Cybersecurity Threat Modeling, Risk Assessment and Treatment of an Infotainment High Performance Computing (HPC) System," 2024.

[16] ["Exploits database," Offsec, 4 november 2009. [Online]. Available: https://exploit-db.com/. [Accessed 27 Jan 2024].

[17] M. Singhal, N. Kumarswamy, S. Kinhekar and S. Nilizadeh, "Cybersecurity Misinformation Detection on Social Media: Case Studies on Phishing Reports and Zoom's Threat," in Proceedings of the International AAAI Conference on Web and Social Media, 2023.

[18] W. Chung, Y. Zhang and J. Pan, "A theory-based deep-learning approach to detecting disinformation in financial social media," Information Systems Frontiers, vol. 25, p. 473–492, 2023.

[19] N. Sun, M. Ding, J. Jiang, W. Xu, X. Mo, Y. Tai and J. Zhang, "Cyber Threat Intelligence Mining for Proactive Cybersecurity Defense: A Survey and New Perspectives," IEEE Communications Surveys & Tutorials, 2023.