Cybersecurity 2030: The Synergy between Machine Learning and Generative AI

Mário Marques da Silva [1,2,3 [0000−0002−5498−3220],] Fernando Correia[1,2,3 [0000-0002-9530-1986]], Héctor Dave Orrillo Ascama [1,2,3[0000−0003−1555−6821],] Laercio Cruvinel Júnior[1,2,3[0000−0002−7672−3243]], 3243], Guilherme Lopes Fernandes [1[0009 − 0001 − 4317 − 4751]], Tomás Viana [1[0009 − 0005 − 6198 − 9195]] and Afonso Frasquilho[1[??????????????]]

[1] Department of Engineering and Computer Sciences, Universidade Autónoma de Lisboa, 1169-023 Lisboa, Portugal
[2] Autónoma TechLab, 1169-023 Lisboa, Portugal
[3] Ci2—Centro de Investigação em Cidades Inteligentes, 2300-313 Tomar, Portugal
mmsilva@autonoma.pt, fjcorreia@autonoma.pt, horrillo@autonoma.pt, lcruvinel@autonoma.pt, 30010398@students.ual.pt, 30010623@students.ual.pt, 30010929@students.ual.pt

Abstract. The rapid evolution of technology, characterized by the 4th Industrial Revolution, has reshaped the cybersecurity landscape. This paper explores the implementation of Generative AI in the context of cybersecurity, highlighting their applications in data analysis procedures and automatic control systems. By integrating machine learning (ML) for real-time detection and generative AI for simulating advanced attack scenarios, we can detect cyberattacks at an early stage and minimize the impact on the systems functioning. We conclude by discussing the implications for the future of cybersecurity and the anticipated dominance of AI-driven solutions by 2030.

Keywords: Cybersecurity and Supervised Learning · Generative AI · Machine Learning · Hybrid Defense Strategy · Anomaly Detection · Synthetic Data.

1 Introduction

The proliferation of digital systems in the 4th Industrial Revolution has introduced unprecedented cybersecurity challenges. This paper explores the implementation of Machine Learning (ML) and Generative Artificial Intelligence (GEN AI) in the context of cybersecurity, highlighting their applications in data analysis procedures and automatic control systems. By integrating ML for real-time detection and GEN AI for simulating advanced attack scenarios, we can detect cyberattacks at an early stage and minimize the impact on the systems functioning. We conclude by discussing the implications for the future of cybersecurity and the anticipated dominance of AI-driven

solutions by 2030.With the advent of interconnected devices and the Internet of Things (IoT), cyber-attacks have become more sophisticated, necessitating advanced defence mechanisms. Artificial Intelligence (AI) has emerged as a cornerstone in this domain, offering tools for automation, real-time threat detection, and strategic simulation (Samuel, 1959). This paper examines the use ML methodologies,

which rely on labelled datasets (Hinton et al., 2006), and GEN AI, which leverages unstructured data to simulate and preempt cyber threats (OpenAI, 2024). We conclude by discussing the implications for the future of cybersecurity and the anticipated dominance of AI-driven solutions by 2030.

This paper is organized as follow: section 2 relies on the background and literature review; section 3 presents the methodology used in this paper; section 4 shows the simulation results and discussion, while section 5 concludes this paper.

2 Background and Literature Review

According to Capodieci et al. (Capodieci et al., 2024), there are different types of AI, such as Machine Learning (ML), and Deep Learning (DL), and GenAI. While ML depends on the human factor to classify and contextualize data, DL is a system with greater autonomy in processing unstructured data. Also, Large Language Models (LLM) represent another type of AI technology, based on GenAI, capable of understanding and generating written language (Lee, 2023).

In cybersecurity, given the volume of data that needs to be analysed to identify possible threats, the human factor can be a limiting factor. However, ML has been a fundamental technique in cybersecurity, excelling in detecting known attack patterns and automating routine tasks such as log analysis (Abadi et al., 2016), and to do that, human interaction is required.

DL, an advanced subset of ML, enhances these capabilities through multilayer neural networks (Goodfellow et al., 2016). DL is part of unsupervised learning, which is effective in anomaly detection and clustering, identifying previously unseen threats without the need for labeled data.

GEN AI, powered by large language and multimedia models, extends these capabilities by creating synthetic data, simulating attack scenarios, and even mimicking social engineering tactics (OpenAI, 2024).

However, its potential misuse, such as generating deepfakes or creating malware variants, highlights the dual-edged nature of this technology (Brynjolfsson, 2014). According to the 2024 report by Capgemini (Capgemini, 2024), in the universe of 1000 organizations analysed, around 45% have encountered deepfake attacks in the last 2 years, and in 43% there have been financial losses due to the exploitation of deepfake technology.

GEN AI operates along two main vectors: one focused on attacks and the other on testing and strengthening defences. In the attack vector, as mentioned earlier, GEN AI is used to create new forms of attacks on systems by exploiting unknown vulnerabilities. On the other hand, the same vector can be proactively applied to generate simulated attack scenarios, which enables the training and continuous improvement of detection

systems. This duality not only expands the understanding of emerging threats but also strengthens cyber defences by anticipating and mitigating potential future attacks. In this way, it is possible to leverage the capabilities of ML in combination with GEN AI. While ML can learn to detect known attacks, GEN AI uses this acquired knowledge to infer patterns and generate simulated attack scenarios. This combined approach contributes to minimizing the impact of AI-driven cyberattacks (Siam et al., 2025).

3 Methodology

To explore the integration of ML and GEN AI in cybersecurity, we analyzed:

- Applications of ML, including classification, regression, and anomaly detection. While Supervised Learning (SL) clearly identifies an attack type, Unsupervised Learning (UL) can be used to detect an event that is outside of the normal activity.
- GEN AI can be applied in cybersecurity to simulate phishing attacks, malware behaviors, and potential intrusion scenarios, enhancing the ability to predict and mitigate cyber threats.
- A hybrid defense strategy combining the strengths of both paradigms.

SL can be used as a classier, where labels are assigned to data to identify the categories to which that data belongs. For example, the classification of animal images by identifying which animal each image belongs to. It can also be used in regression when it involves the continuous prediction of numerical values based on the input data. Finally, object detection which focuses on identifying and locating multiple objects in images or video. Examples of SL can be found at car park license plate identification processes for automatic payment systems, weather forecasting, credit card fraud detection, among others. On the other hand, UL can be employed to identify anomalous patterns in real-time events. This normally requires the analysis of an operator, to identify or not as a threat.

ML makes it possible to collect data and produce data resulting from previous experiences, improving information quality. In other words, with increased knowledge, performance criteria can be optimized. ML makes it easier to solve various computational problems involving classification and regression procedures. Generically, ML allows for better control over data classes in AI training.

Currently, TBytes of data are produced, so classifying them using human supervision is challenging. Training an ML model requires a lot of computing time and supervision time. ML cannot perform all the complex tasks in the field of Machine Learning, which can be a disadvantage in the case of applying this model. The employment of ML and GEN AI must be weighed up and applied according to the volume of data to be processed and the results to be obtained.

Thus, a hybrid security system can be established, integrating not only classification, regression, and anomaly detection processes but also proactive mechanisms. These mechanisms enable the simulation of attack scenarios generated by GEN AI, which can

be used to retrain and enhance the ML detection system, improving its ability to identify and respond to emerging threats.

Our study involved simulated environments using real-world datasets and generative models trained to mimic cyberattack scenarios. Metrics for evaluation included detection accuracy, response time, and the ability to pre-emptively counter novel threats.

4 Simulation studies

This section aims to demonstrate that supervised models like Random Forest are effective in classifying multiple attack classes. Unsupervised models like Autoencoder are effective in detecting anomalous patterns, complementing the supervised models. The proposed hybrid approach improves detection accuracy and reduces response times, enabling organizations to mitigate threats before they escalate.

The use of ML for intrusion detection (Intrusion Detection Systems [IDS]) has become prominent in the field of cybersecurity, enabling the analysis of large volumes of network traffic and the identification of anomalous or malicious behavior. This approach facilitates early attack detection and real-time automated monitoring.

In this context, it is pertinent to consider how IDS can benefit from the integration of multiple ML approaches. The use of a hybrid system, combining supervised and unsupervised models, emerges as an effective response to the growing sophistication of cyberattacks. Supervised models excel in accurately detecting known patterns, thanks to training with previously labeled data. However, they reveal limitations when faced with novel attacks or variants of existing threats. It is in such situations that unsupervised models play a crucial role, allowing the identification of anomalous patterns in real time, even without prior knowledge, acting as additional sensors for unexpected behaviors within the network.

4.1 Characterization of the simulation scenario

A complementary hybrid approach is illustrated in Fig. 1, combining the accuracy of the supervised model with the adaptive capabilities of the unsupervised model. Initially, network traffic is analyzed by a supervised model trained to recognize patterns associated with previously known attacks, immediately generating an alert whenever a threat is identified. If the supervised model does not detect any threat, the traffic is then analyzed by an unsupervised model, which clusters network behavior into similar patterns. Significant deviations from the established patterns may indicate anomalous behaviors, enabling the detection of unknown or newly emerged attacks. When an anomaly is detected, the probability of the traffic being malicious is estimated, and only the cases that exceed a defined threshold are forwarded for manual verification, thereby reducing the occurrence of false positives and minimizing the manual analysis workload.

Having introduced the concept of a hybrid approach to intrusion detection, it becomes necessary to select and prepare appropriate data to support its development and future validation. In this context, the present study is based on the “CSE-CIC-IDS2018” dataset, developed by the Canadian Institute for Cybersecurity (CIC) in collaboration

with the Communications Security Establishment (CSE) of Canada for Cybersecurity (2018). The dataset was generated from real network traffic captured in a controlled environment, encompassing both benign communications and a variety of attack types (brute-force, DoS, DDoS, botnets, SQL injections), distributed across several CSV files, each corresponding to a day of simulation.

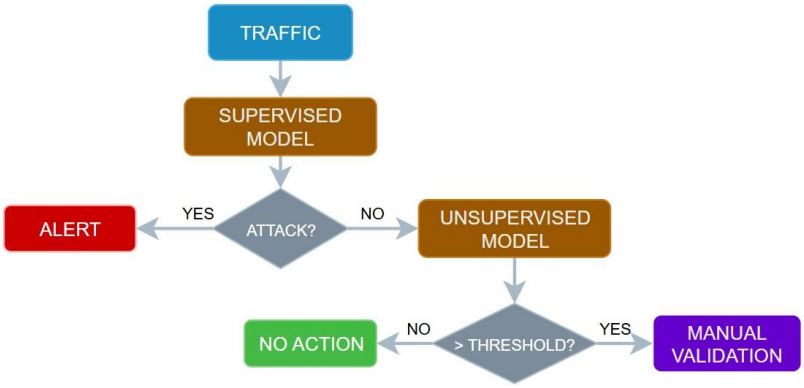

Figure 1 - Hybrid IDS operation fl ow

4.2 Results

A consolidated version was constructed using data from six specific days (February 14, 15, 21, 22, 23, and March 2, 2018), totaling 162,658 records and 78 numerical features. The integration and preprocessing steps were conducted using Python scripts, leveraging libraries such as “pandas” for data loading, transformation, and unification.

Among the available features, several stand out for their relevance in characterizing network traffic:

- Flow Duration: total flow duration in microseconds;
- Tot Fwd Pkts / Tot Bwd Pkts: total number of packets sent in the forward and backward directions;
- TotLen Fwd Pkts / TotLen Bwd Pkts: total packet length (in bytes) in both directions;
- Fwd Pkt Len Mean / Bwd Pkt Len Mean: average packet length, essential for understanding client-server communication patterns;
- Flow Byts/s / Flow Pkts/s: average transmission rate in bytes and packets per second;
- Flow IAT Mean / Std / Max / Min: inter-arrival times between successive packets within a flow, useful for identifying irregular temporal patterns;
- Pkt Len Min / Max / Mean / Std / Var: descriptive statistics of packet length, useful for detecting uncommon or suspicious flows;

- SYN Flag Cnt / ACK Flag Cnt / FIN Flag Cnt: TCP flag counters, important for detecting typical attack behaviors such as SYN Floods;
- Idle Mean / Max / Min: idle time between communications, which may indicate evasion attempts.

All features were standardized into numerical formats (int64 and float64), ensuring consistency across files.

The statistical analysis revealed several noteworthy patterns. Flow Duration shows a mean of 105,457 µs and a median of 37,634 µs, with strong skewness (-236.88) and high kurtosis (59,458.29), indicating the presence of outliers. Tot Fwd Pkts displays negative minimum values and a skewness of -2.05. Meanwhile, Tot Bwd Pkts shows a skewness of 82.61, suggesting bursts of intensive traffic. These calculations were obtained using the "describe()" function and additional statistical tools from the "NumPy" library. These characteristics emphasize the need for normalization and rigorous data validation.

The class distribution is illustrated in Fig. 2. The "Benign" class includes 100,000 records (61.48%). Classes 2 to 7, with 10,000 records each (6.15%), represent the primary attack types. Classes 8 to 11 are less represented: 1,730 (1.06%), 611 (0.38%), 230 (0.14%), and 87 (0.05%) records, respectively.

The "Label" variable identifies each type of traffic:

1. benign;
2. FTP-BruteForce;
3. SSH-Bruteforce;
4. DDOS attack-HOIC;
5. Bot;
6. DoS attacks-GoldenEye;
7. DoS attacks-Slowloris;
8. DDOS attack-LOIC-UDP;
9. Brute Force-Web;
10. Brute Force-XSS;
11. SQL Injection.

This multiclass structure enables the training of models capable of distinguishing between normal traffic and multiple attack patterns. The intentional imbalance reflects real-world network conditions, where benign traffic predominates. This favours model training by reinforcing the recognition of normal patterns. The distribution was visualized using "matplotlib.pyplot.bar" on the values of the "Label" variable.

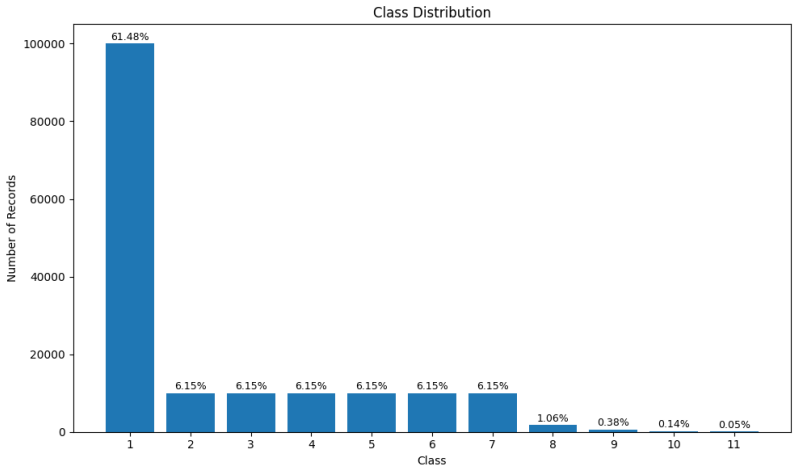

Figure 2 - Class distribution

Correlation analysis was conducted using the 15 features with the highest variance (Fig. 3), focusing on the most informative ones. Strong correlations were observed among time-based metrics related to inactivity, such as Idle Max, Idle Mean, and Idle Std (all near 1.00), as well as among inter-packet timing features like Fwd IAT Std with Flow IAT Std (0.94) and Fwd IAT Mean (0.94). The correlation matrix was generated with "seaborn.heatmap" using the Pearson method on numerical features. These redundancies suggest overlapping information, and their removal or consolidation during feature selection is recommended to optimize model performance and reduce complexity.

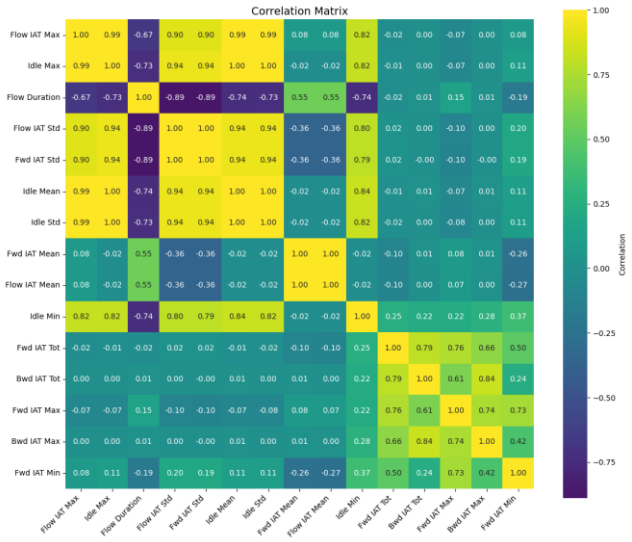

Figure 3 - Correlation matrix

4.3 Results Analysis

After defining the hybrid architecture and preparing the data, ML models were selected to ensure robustness and adaptability in intrusion detection. For the supervised component, three models were tested: Multilayer Perceptron (MLP), Random Forest, and Support Vector Machine (SVM). In the unsupervised domain, Isolation Forest, K-Means, and Autoencoder were analyzed.

The SVM model was included due to its effectiveness in separating classes in high-dimensional spaces. After normalization with StandardScaler and splitting the data into 70% for training and 30% for testing, it was trained using the SVC classifier. The evaluation on the test set revealed a weighted F1-score of 0.99, along with equally high values for precision and recall (both 0.99). The confusion matrix (Fig. 4) shows an almost perfect performance across the first seven main classes. However, the minority classes (9, 10, and 11) exhibited lower recall values: 68%, 45%, and 29%, respectively. Ten-fold cross-validation confirmed the model's consistent performance, with an average of 98.50%.

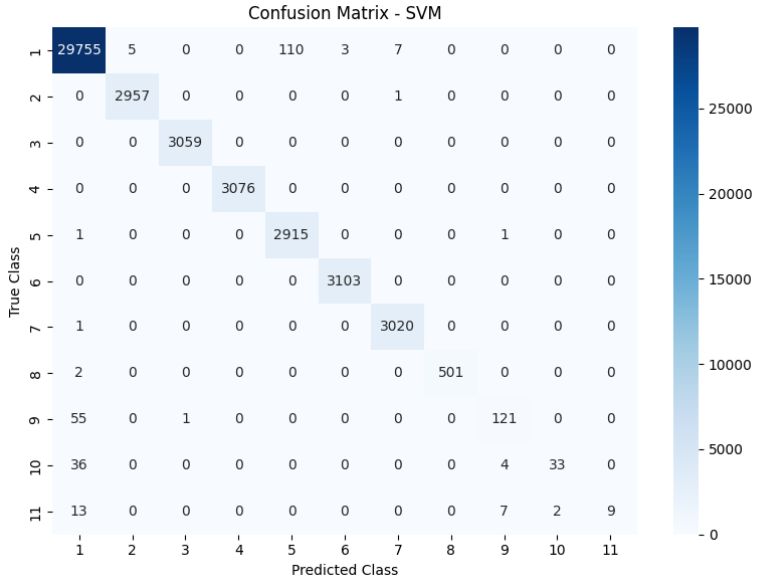

Figure 4 - SVM Confusion Matrix

Despite the model's strong performance, its limitation in detecting rare attacks led to the exclusion of SVM as the final model. The Multilayer Perceptron (MLP) was selected for its ability to capture nonlinear patterns. It was trained with a single hidden layer of 100 neurons and 500 iterations, aiming to balance performance and computational cost. The evaluation showed weighted precision, recall, and F1-score of 1.00. The confusion matrix (Fig. 5) demonstrates strong results for the main classes and improvements over SVM in classes 9, 10, and 11, with recall values of 88%, 71%, and 29%, respectively.

Cross-validation confirmed the model's stability (99.12% ± 0.48%). However, increasing the number of neurons or layers could potentially improve the results, but at the cost of longer training times and higher computational resource consumption, which would not be justifiable compared to the next model.

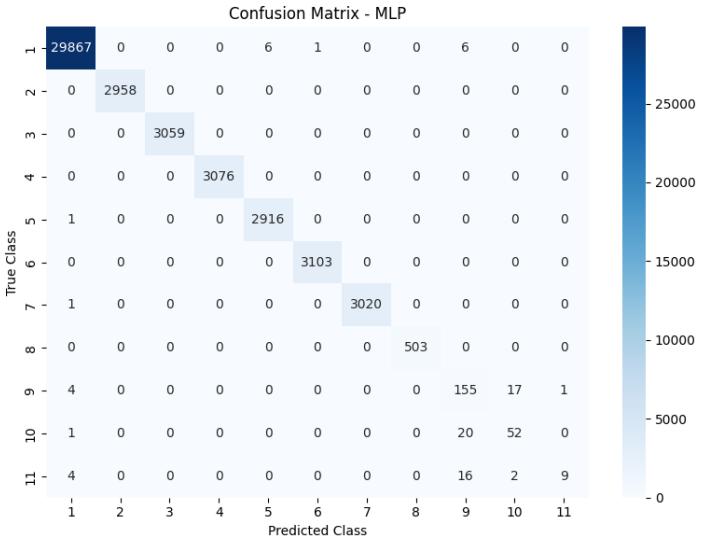

Figure 5 - MLP Confusion Matrix

The Random Forest model was included due to its effectiveness in multiclass classification and its generalization ability resulting from the combination of multiple decision trees. Using default parameters, it achieved a weighted F1-score of 1.00. The confusion matrix (Fig. 6) shows excellent performance in the less represented classes, with recall values of 92%, 79%, and 58% for classes 9, 10, and 11, respectively - outperforming both the MLP and SVM models.

Ten-fold cross-validation confirmed this robustness, with an average F1-score of 99.11% ± 0.57%. Considering the results obtained, Random Forest proved to be the most suitable choice for the supervised component, offering the best balance between accuracy, generalization capacity, and performance in detecting underrepresented classes, without compromising computational requirements.

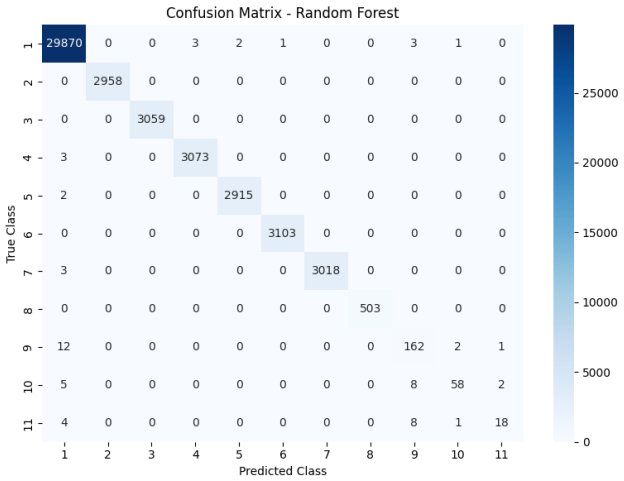

Figure 6 - Random Forest Confusion Matrix

The Isolation Forest model was used as a classical unsupervised approach for anomaly detection, leveraging its ability to isolate outliers through randomly constructed binary trees. The data were converted into a binary classification (0-benign traffic, 1-attack), and the model was trained using a contamination rate matching the proportion of attacks in the training set. Despite being computationally efficient, its performance was unsatisfactory due to its inability to capture the specific characteristics of the traffic in the dataset. The model assumed an anomaly distribution that did not reflect the real complexity of the data, flagging unusual but legitimate behaviours as attacks while ignoring subtle malicious patterns. On the test set, it achieved an F1-score of only 0.32, with equally low precision and recall, failing to reliably distinguish between normal and anomalous traffic. The confusion matrix (Fig. 7) shows a high number of false positives and false negatives, seriously limiting the model's applicability in real-world scenarios.

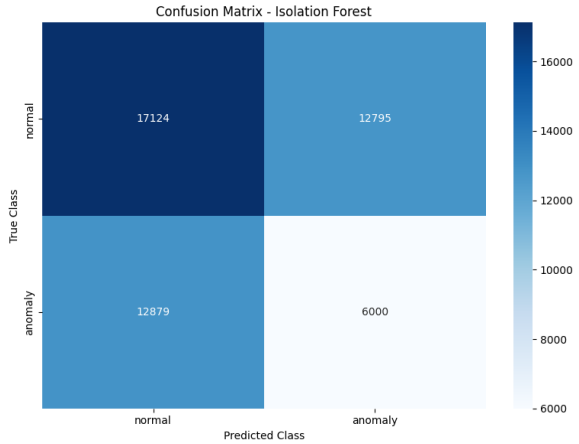

Figure 7 - Isolation Forest Confusion Matrix

Given these results, and despite the advantages of Isolation Forest in terms of simplicity and speed, its poor performance led to the rejection of this model.

The K-Means model was tested as an unsupervised approach for intrusion detection, based on the assumption that normal and malicious traffic would naturally form two distinct clusters. After converting the classes to binary (0- benign, 1 _ attack), the model was trained with two centroids. To align the output with the actual classification, it was necessary to invert the predictions so that the clusters correctly matched the "normal" and "anomaly" labels. The results were unsatisfactory. The model achieved an F1-score of only 0.43, with highly unbalanced performance across classes: the _normal_ class had a recall of just 26%, while the _anomaly_ class reached 60%. The confusion matrix (Fig. 8) highlights a significant number of false positives and false negatives, with over 22,000 benign samples incorrectly classified as anomalies and more than 7,500 attacks going undetected.

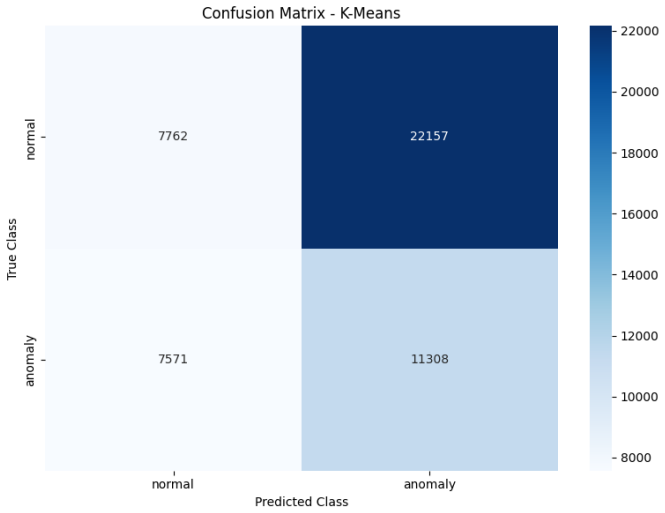

Figure 8 - K-means Confusion Matrix

These values demonstrate K-Means' inability to capture the complexity and overlap of patterns present in network traffic. The algorithm's linear nature - assuming spherical clusters with similar variance - proved incompatible with the actual structure of the data. This limitation justifies the rejection of K-Means as a model for the unsupervised component of the hybrid system.

The Autoencoder model was implemented with the goal of learning to reconstruct normal traffic patterns, if anomalies would yield significantly higher reconstruction errors. It was trained exclusively on benign data, using a symmetric neural network architecture with L1/L2 regularization and optimization via EarlyStopping. Evaluation was conducted on data containing both benign and malicious traffic, converted into a binary classification (0 – benign, 1 – attack).

The distinction was based on the mean squared reconstruction error (MSE), and the optimal threshold was determined using the precision-recall curve, selecting the point

that maximized the F1-score. The resulting value was 0.000003, which may seem extremely low at first glance. However, this is justified using MinMaxScaler, which normalizes the data to the [0, 1] range, naturally reducing the scale of reconstruction errors. The results showed excellent anomaly detection capability: the model achieved an F1-score of 0.64, with a precision of 0.48 and recall of 1.00. The confusion matrix (**Erro! A origem da referência não foi encontrada.**) highlights 18,788 attacks correctly detected and only 91 false negatives, indicating high sensitivity to anomalous patterns. In contrast, 20,650 false positives were recorded, representing a considerable rate of incorrect alerts. Despite the high number of false alarms, the model's sensitivity makes it particularly suitable as a complementary layer in a hybrid system, where it can act as an anomaly sensor for traffic that deviates from known patterns.

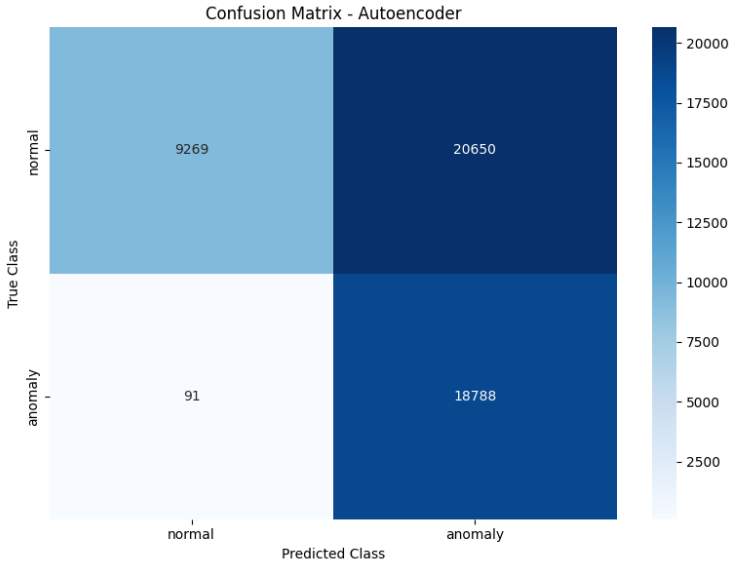

Figure 9 - Autoencoder Confusion Matrix

Although it does not reach the precision levels of the supervised models, the Autoencoder proved to be the most effective approach within the unsupervised domain. Its ability to adapt to new patterns and detect previously unknown traffic makes it an ideal complement to the supervised model, strengthening the hybrid system in scenarios involving uncatalogued attacks. Moreover, since it is trained exclusively on benign traffic, it can be updated periodically or even continuously with new unlabeled network data, allowing it to evolve in real time. It is also important to emphasize that all anomalies detected by this model are subject to manual validation, ensuring that potential false positives have no direct impact and preserving the overall reliability of the system.

5 Conclusions

The integration of ML and GEN AI presents a transformative approach to cybersecurity, offering a robust defense strategy against increasingly sophisticated threats. ML excels in real-time anomaly detection and classification of known attack patterns, while GEN AI enhances these capabilities by simulating novel attack scenarios and generating synthetic data for proactive defense training. This hybrid model not only improves detection accuracy but also reduces response times, enabling organizations to mitigate threats before they escalate. As cyberattacks grow in complexity, the synergy between these technologies will be critical in maintaining resilient and adaptive security systems.

However, the dual-edged nature of GEN AI cannot be overlooked. While it strengthens defenses, its potential misuse - such as creating deepfakes or advanced malware - poses significant risks. The findings of this study underscore the importance of ethical oversight and continuous human supervision to ensure these technologies are deployed responsibly. Organizations must balance innovation with caution, implementing safeguards to prevent malicious exploitation while leveraging GEN AI's potential to anticipate and neutralize emerging threats.

Looking ahead, the dominance of AI-driven solutions in cybersecurity is inevitable. By 2030, AI is expected to resolve up to 95% of cybersecurity incidents, revolutionizing how threats are identified and countered. To stay ahead in this arms race, institutions must invest in both ML and GEN AI, fostering collaboration between human expertise and automated systems. The future of cybersecurity lies in adaptive, intelligent frameworks that combine the precision of supervised learning with the creativity of generative models, ensuring a safer digital landscape for all.

References

M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore,D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng. Tensor_ow: A system for large-scale machine learning, 2016. URL https://arxiv.org/abs/1605.08695.

E. Brynjolfsson. The second machine age : work, progress, and prosperity in a time of brilliant technologies. W.W. Norton & Company, 2014.

Capgemini. New defenses, new threats: What ai and gen ai bring to cybersecurity. Technical report, Capgemini Research Institute, 2024.

N. Capodieci, C. Sanchez-Adames, J. Harris, and U. Tatar. The impact of generative ai and llms on the cybersecurity profession. In 2024 Systems and Information Engineering Design Symposium (SIEDS), pages 448_453, 2024. https://doi.org/10.1109/SIEDS61124.2024.10534674.

C. I. for Cybersecurity. A realistic cyber defense dataset (cse-cic-ids2018) was accessed on apr2025 from https://registry.opendata.aws/cse-cic-ids2018., 2018.

I. Goodfellow, Y. Bengio, and A. Courville. Deep Learning. The MIT Press, 2016. ISBN 9780262035613.

G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. Neural Computation, 18(7):1527 1554, 2006. https://doi.org/10.1162/neco.2006.18.7.1527.

A. Lee. What are large language models and why are they important? [online] Available: https://blogs.nvidia.com/blog/2023/01/26/whatare-large-language-models-used-for/, Jan. 023. URL https://blogs.nvidia.com/blog/2023/01/26/what-are-large-language-modelsused-for/.

OpenAI. Gpt-4 technical report, 2024. URL https://arxiv.org/abs/2303.08774.

A. L. Samuel. Some studies in machine learning using the game of checkers. IBM Journal of Research and Development, 3(3):210-229, 1959. https://doi.org/10.1147/rd.33.0210.

A. A. Siam, M. M. Hassan, and T. Bhuiyan. Artificial intelligence for cybersecurity: A state of the art. In 2025 IEEE 4th International Conference on AI in Cybersecurity (ICAIC), pages 1-7, 2025. https://doi.org/10.1109/ICAIC63015.2025.10848980.