IJISEM
INTERNATIONAL JOURNAL OF INNOVATIONS IN SCIENCE ENGINEERING AND MANAGEMENT

International Journal of Innovations In Science Engineering And Management

OPEN ACCESS

Volume: 5

Issue: 3

Month: August

Year: 2026

ISSN: 2583-7117

Published: 22.08.2026

Citation:

Amar Singh Verma, Neha Gupta, Akash Saxena, Amit Kumar Thakore "To Compare Various Frameworks that Integrate XAI, GANs and LLMs for Dynamic Malware Behavior Analysis" International Journal of Innovations in Science Engineering and Management, vol.5 , no.3 , 2026, pp. 426 439

DOI:

10.69968/ijisem.2026v5i3426-439

To Compare Various Frameworks that Integrate XAI, GANs and LLMs for Dynamic Malware Behavior Analysis

Amar Singh Verma[1], Neha Gupta[2], Akash Saxena[3], Amit Kumar Thakore[4]

[1]*Computer Science and Engineering Dr. K. N. Modi University India*
[2]*Computer Science and Engineering Dr. K. N. Modi University India*
[3]*Computer Science and Engineering CITM, Jaipur India*
[4]*Computer Science and Engineering Dr. K. N. Modi University* India

Abstract

Modern malware frameworks use advanced evasion techniques that bypass traditional detection methods; thus entail advanced analytical frameworks for comprehensive and robust analysis. This study provides a comparative analysis of several frameworks that utilize Explainable Artificial Intelligence (XAI), Generative Adversarial Networks (GANs), and Large Language Models (LLMs) to provide a dynamic approach to malware behaviour. The review consisted of five key metrics to assess these three frameworks: detection performance, explainability, robustness against adversarial attacks, behavioral interpretation, and automated reporting capabilities. The results indicate that Deep Learning models have attained high accuracy in detect-ing and identifying malicious code, but are not interpretable. The GAN-based frameworks are highly effective in generating adversarial samples for robustness testing. Conversely, LLM-based approaches are highly effective for generating automated forensic reports, but are not yet fully integrated into malware detection workflows. The analysis highlights a research gap pertaining to the lack of integrated frameworks for adversarial analysis, explainability and automated forensic reporting. This study proposes a unified approach of malware analysis incorpo-rating XAI, GAN and LLM to enable the development of more effective malware detection tools with more transparency, deeper analytical insight and advanced forensic decision-making.

Keywords; Malware analysis, Explainable AI, Generative Adversarial Networks, Large Language Models, Cybersecurity

INTRODUCTION

The rapid digitalization of modern societies is deeply inter-twined with interconnected computer systems and network in-frastructure. This has led to innovations and economic growth; however, it has simultaneously expanded the attack surface for cyber threats. One such example of this type of cyber threat is malware, which continues to be one of the most persistent and evolving challenges in the area of cybersecurity today. Malware refers to any type of malicious software designed to disrupt or gain unauthorized access to computer systems. In the past two decades, the progression and refinement of malware have led to the development of worms and advanced threats capable of bypassing traditional antimalware software. Modern malware uses advanced techniques such as poly-morphism, code obfuscation, encryption and sandbox evasion to conceal its malicious nature and avoid identification by traditional security measures [1].

Table 1. Evolution of Malware Threats

Time Period	Characteristics	Detection Approach
1990-2005	Basic viruses and worms	Signature-based an-tivirus
2005-2015	Trojans, botnets, spyware	Heuristic detection
2015-2020	Ransomware, poly-morphic malware	Machine learning-based detection
2020-Present	AI-enabled and eva-sive malware	Behavioral and AI-driven detection

Source: Adapted from Djenna (2023) and Harikha et al. (2024) [1]

IJISEM
INTERNATIONAL JOURNAL OF INNOVATIONS IN SCIENCE ENGINEERING AND MANAGEMENT

Antivirus solutions have relied mainly on signature-based detection to detect viruses. They use signatures to find mal-ware by matching file signatures with known patterns in a security database. Although this process continues to work well at detecting older types of viruses, it is not effective at finding new types of viruses or 'zero-day' attacks. For this reason, more and more analysts have started using Machine

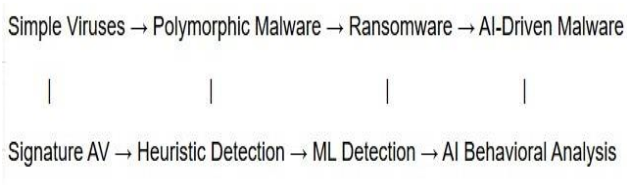

Figure 1. Malware Evolution and Detection Approaches

Learning (ML) and Artificial Intelligence (AI) to propose new malware detection approaches. Recently, Convolutional Neural Networks (CNNs) and Long Short-term Memory net-works (LSTMs) have shown strong performance in classifying malware using deep learning. The ability of these models to analyze very large amounts of behavioral and structural data allows them to detect the complex patterns that characterize malicious software [4]. However, despite their high accuracy rate, these deep learning models still operate as a black box thus making it hard for analysts to understand or interpret how they make their decisions.

To conduct malware forensics, analysts must understand the logic behind the malicious labeling of files and incidents and provide supporting evidence for their findings. To summarize, existing XAI methods enhance analysts interpretability of ML-based classification outcomes. In addition, GANs are effective tools for creating synthetic samples of malware that accurately simulate real-world attacks. Analysts can use these samples to test the effectiveness of various cyber threat detection solutions against sophisticated cyber criminals [6][7]. Recent developments have demonstrated that LLMs have several strong abilities in cybersecurity applications and are capable of analyzing complex forms of structured and textual data. Thus, they can process security logs, threat intelligence reports (TIR's) and additional documentation to generate automated forensic reports [15].

Despite these advances, modern malware analysis frame-works treat each technology as a separate entity. Deep learn-ing models prioritize detection accuracy, GAN-based models emphasize adversarial testing and LLM-based systems aim to automate analysis. However, integrating all these technologies into a single malware analysis framework remains limited; although there is a relatively small corpus of studies on integrating XAI, GAN-based adversarial analysis and LLM-based analytical functionality into a unified malware analysis framework. The objective of this study is to compare and analyze the current frameworks that combine XAI, GANs and LLMs with the strengths and limitations of each approach and highlight potential pathways for the development of a more transparent and streamlined malware analysis process. The three main contributions of this study as:

- This study compares AI-based malware detection ap-proaches across different frameworks to assess their de-tection performance, explainability, adversarial robustness and level of automation in malware detection
- Identified the research gap on basis of previous study
- This study provides directions for future work on building

AI-based integrated cybersecurity frameworks for mal-ware forensics.

BACKGROUND CONCEPTS

The study on how to analyze malware has become more sophisticated with the use of multiple artificial intelligence techniques for the purposes of detection, interpretation and automation of analytical processes. This section provides an overview of basic concepts for understanding the comparative analysis conducted in this study, including dynamic malware analysis, GANs, XAI and LLMs in the field of cybersecurity and malware forensics.

Dynamic Malware Analysis

Dynamic malware analysis involves executing a suspicious program within a sandbox environment to monitor its run-time behavior during execution. This differs from static analysis, which examines the binary representation of a program with-out running the application. In addition to examining program execution, analyst obtain behavioral artifacts (i.e., system calls, changes to the Windows registry, changes to the file system, network interaction) that occur during execution and obtain detailed insight into how the malware operates, which would otherwise remain hidden.. Sandboxing environments are also extensively used in dynamic malware analysis.

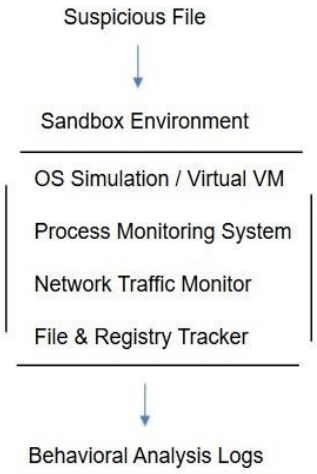

Figure 2. Basic Architecture of a Malware Sandbox Environment

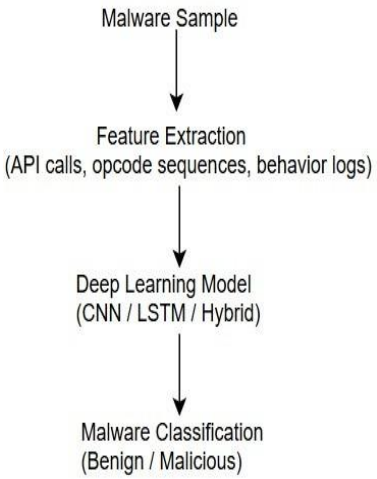

Figure 3. Dynamic Malware Analysis Workflow

The Cuckoo Sandbox, a similar platform, is used to run malicious programs in a controlled virtual environment to ob-serve how the malware interacts with system resources while executing its primary functions, such a executing malicious code.

Once this process is completed, the behavioral traces collected by the system are analyzed to determine whether the application has performed any malicious activity [3].

Many families of modern malware use encryption or pack-ing techniques, code obfuscation, to evade static detection mechanisms. Dynamic analysis has become an essential tech-nique for detecting these modern variants of malware helping us identify how advanced threats such as ransomware and file-less malware behave when executing their code on legitimate applications [1][2].

This figure illustrates the general workflow of dynamic malware analysis, in which suspicious programs are executed in a sandbox environment and behavioral features are extracted for further analysis [3][6]

Behavioral analysis

GANs consist of two deep neural networks: a Generator neural network and a Discriminator neural network. Typically, the generator generates synthetic samples that resemble real data, whereas the discriminator evaluates whether these sam-ples are real or generated. The adversarial approach enables GAN systems to understand complex data distributions and synthesize realistic samples [6]. GANs are utilized in malware forensics and cybersecurity to formulate adversarial malware variants and improve the accuracy of malware identification systems [5][7].

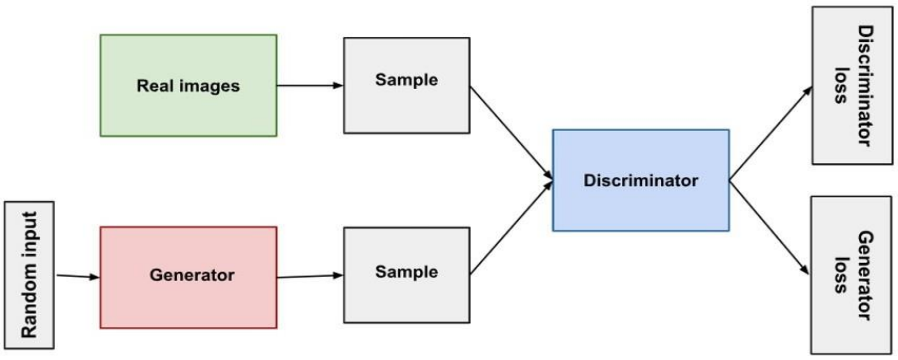

Figure 4. Generator and Discriminator

This figure depicts the GAN architecture consisting of two Deep Neural Networks: a Generator Neural Network and a Discriminator Neural Network [6][8]

Existing studies have demonstrated how GANs can be used to generate and simulate the behavior of adversarial malware and AI. One example is PlausMal-GAN, a framework that can create 'plausible' malware variants to help detect unknown threats. Adversarial training strategies can be applied to test the integrity of an ML-based approach for mitigating against adversarial attacks [7][29]. Although the benefits of using GAN-based malware analysis frameworks are extensive, they mainly focus on generating adversarial samples rather than explaining their detection decision-making processes. While GANs provide automated generation of samples and make it easier to automate testing, several challenges remain, and there is a need to integrate explainability techniques into a GAN-based analysis workflow [9].

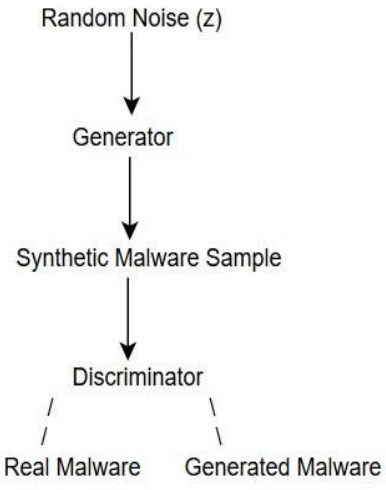

Figure 5. illustrates the adversarial learning mechanism used in GAN-based malware generation.

Explainable Artificial Intelligence in Malware Detection

Although deep learning has been effective in detecting malware, it is challenging to interpret the decision-making processes of deep learning models. Learning models often function as 'black boxes,' producing predictions without ex-plaining how conclusions are derived; this lack of transparency makes it difficult for analysts to understand why a certain pro-gram was classified as a threat. XAI mitigates this limitation by generating explanations that improve the understandability of machine learning models. Multiple approaches have been introduced to improve model transparency. Two common approaches to model interpretability are SHAP (Shapley Ad-ditive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations). These techniques provide insights into the significance of input features as a measure of their impact on the model output, and enable analysts to identify the particular behavioral indicators that contribute to malware detection decision [12][13].

Explainability methods provide a more effective approach to analyzing malicious software because they can detect behav-ioral patterns (e.g., anomalous system calls and irregular net-work traffic) through intuitive visualizations that highlight the features responsible for producing the detection results. Recent studies [10][17][18] have demonstrated that the integration of explanation techniques with malware behavioral detection systems enhances interpretability with minimal degradation in detection performance.

Large Language Models in Cybersecurity Analysis

LLMs are increasingly used to process large amounts of text and structured data. These models have been trained on extremely large datasets and support a variety of use cases such as natural language understanding, summarization and auto-mated report generation. In the field of cybersecurity, LLMs enable analysts to interpret security logs, detect threat patterns and generate structured reports that depict malware behavior. LLM-based systems can summarize highly detailed forensic artifacts and convert them into raw analysis data [14][15]. A recent study[20][21][22] identified the use of LLMs as poten-tial components of the malware analysis workflow, noting their capability to perform various tasks, including analyzing code and its workflow, vulnerability detection, reverse engineering and automated report generation.

Additionally, retrieval-augmented generation enables LLMs to leverage external threat intelligence during analysis. How-ever, more studies are required to effectively analyze large datasets and generate reports using LLM models in an auto-mated malware detection workflow [25][26].

Table 2 Comparative Capabilities of AI Technologies in Malware Analysis

Technology	Strength	Limitation
CNN/LSTM	High detection accuracy	Limited interpretability
GAN	Adversarial testing	Low explainability
XAI	Model transparency	No direct detection gain
LLM	Log analysis & reporting	Limited detection integration

RELATED WORK

The evolving complexity of modern malware has stimulated analysts to explore the use of sophisticated AI techniques for detecting malware and analyzing its behavior. Classical signature-based detection techniques are no longer effective against modern malware, which uses various techniques such as polymorphism, code obfuscation, encryption and sandbox evasion. Therefore, recent studies have focused on developing intelligent detection frameworks that can identify malicious behavior through data-driven analysis. Several AI-based mal-ware analysis approaches have been proposed, including Deep Learning Models, GANs, XAI methods and LLMs. Each of these methods provides distinct capabilities for cybersecurity analysis. This section reviews key studies in four major categories: AI-based malware detection systems, GAN-based malware analysis systems, explainable malware detection sys-tems and LLM-based cybersecurity analysis.

AI-Based Malware Detection Frameworks

Deep learning techniques are widely used in malware detection because of their ability to learn complex patterns from large datasets. The most commonly used types involve the use of Convolutional Neural Networks (CNN's), Long Short-Term Memory (LSTM) Networks and Hybrid Deep Learning frameworks. Malware detection frameworks that use CNN's typically analyze structured representations of malicious bina-ries such as opcode sequences, bytecode patterns and image-based malware signatures. Nataraj et al. demonstrated that grayscale images generated from malware binaries can be used as input to a CNN's architecture for classifying malware families with high accuracy [30]. This technique allows CNN's models to represent how code structures are visually related to each other (spatial relationships).

Conversely, LSTM networks are designed for working with sequential data and used to analyze the behavior sequences produced by running dynamic malware. Kolosnjaji et al. introduced an approach to detect malware using LSTM net-works that analyzes the sequence of system calls made by a program during execution [31]. The results demonstrated that this approach provided better detection results for new or previously unknown variants of malware, because of its effectiveness in capturing temporal relationships within the behavioral data.

Recent studies have explored hybrid frameworks using both CNN's and LSTMs to extract spatial and temporal features. Gautam et al. developed a hybrid CNN-LSTM architecture for analyzing application run-time behavior, including API call se-quences and process activities [4], resulting in higher detection accuracy than any single deep learning model but exhibiting limited interpretability because of the overall complexity of the models. Although deep learning-based detection models have demonstrated excellent performance in classification and have restricted their applicability in attack analysis and digital forensics.

Table 2 AI-Based Malware Detection Approaches [2]

Model	Key Features	Advantages	Limitations
CNN [30]	Image or opcode-based malware analysis	High pattern recognition capability	Limited temporal analysis
LSTM [31]	System call se-quence analysis	Captures tempo-ral behavior	Training complexity
CNN-LSTM Hybrid [4][34]	Combines spatial and temporal fea-tures	Improved detec-tion accuracy	Black-box inter-pretability

GAN-Based Malware Analysis Frameworks

GANs have been applied to malware analysis in several ways, including adversarial malware generation, dataset aug-mentation and robustness evaluation of detection systems. A prominent example is PlausMal-GAN, proposed by Won et al. [7], which generates samples of malware that resemble the behavioral pattern of live malware. The generated samples can be leveraged to build training datasets and enhance the detection performance of unseen threats. Adversarial attack scenarios have also been simulated using GAN-based tech-niques. Rigaki and Garcia [32] illustrated that it was possible to transform malware communication patterns with GAN-based methods and hide them without triggering detection demonstrate how GANs can be effectively used to test the robustness of malware detection systems against adversarial interference.

However, most GAN-based malware analysis systems em-phasize the performance of adversarial generation and de-tection, providing sparse information about the rationale for detection in terms of behavioral reasoning.

Explainable Malware Detection Frameworks

As deep learning models have become more developed, significant attention has been devoted to how they make internal decisions. XAI is designed to enhance the

transparency of machine learning systems by providing model predictions in an interpretable and user-centric manner. Antonio et al. [10] proposed methods to demonstrate that behavioral mal-ware detection models can assist analysts in understanding classification decisions, particularly in a cybersecurity envi-ronment, where analysts need to authenticate features during the detection process of an incident. SHAP (Shapley Ad-ditive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) are common XAI used in malware analysis [12][13]. In addition to these approaches, attention-based neural networks have also been suggested as a means to emphasize significant input features during prediction and enable analysts to visualize the behavioral features that can be used to determine malware classification results.

Table 4 GAN-Based Malware Analysis Frameworks [3]

Framework	Technique	Application	Limitation
PlausMal-GAN [7][29]	GAN-based mal-ware generation	Dataset augmen-tation	No explainability
MalGAN [5][8]	Adversarial mal-ware samples	Evasion testing	Limited behavior interpretation
GAN-based anomaly detection [9][32]	Behavioral model-ing	Network security	Computational complexity

LLM-Based Cybersecurity Analysis

Recently, LLMs have become powerful tools for analyzing unstructured data. Considering their ability to process data and understand natural language along with their competence to read structured data (i.e. system logs, LLMs are useful for applications such as threat intelligence analysis, log interpre-tation and automated documentation.

Wickramasekara et al. [15] demonstrated that LLMs can generate structured analysis reports from complex forensic artifacts to aid digital forensic analysts. Similarly, Michelet and Breitinger [16] indicated that LLMs enable analysts to generate digital forensics reports more efficiently. In this case, language models can be used to analyze execution logs and system events to generate a summary of how malware interacts with the various components of the system and how it communicates externally.

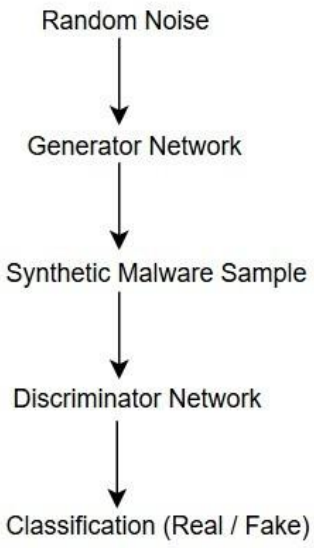

Figure 6. GAN-Based Malware Generation Framework

This figure represents a framework for GAN-based malware generation, illustrating how adversarial malware variants are created for robustness testing and detection improvement [5][7][29]

Table 5 COMMON EXPLAINABLE AI TECHNIQUES [4]

Technique	Description	Application
SHAP [12]	Uses cooperative game theory to measure the contribution of each feature to the model prediction	Feature importance analysis
LIME [13]	Creates local surrogate models that approximate complex mod-els to explain individual predic-tions	Local explanation of classification
Attention Mechanisms [10][17][19]	Highlights the most important input features within neural net-work architectures during pre-diction	Interpreting deep learning models

Source: Compiled by the authors based on existing literature

Table 6 APPLICATIONS OF LLMS IN CYBERSECURITY

Application	Description
Log interpretation	Automated analysis of system logs
Threat intelligence	Summarization of cybersecurity reports

Malware explanation	Interpretation of malware behavior
Report generation	Automated forensic documentation

COMPARATIVE ANALYSIS OF EXISTING FRAMEWORKS

As malware continues to evolve in sophistication, traditional detection mechanisms are becoming increasingly inadequate. Recent studies have emphasized the development of AI-enabled frameworks for efficient malware detection and dy-namic behavior analysis.

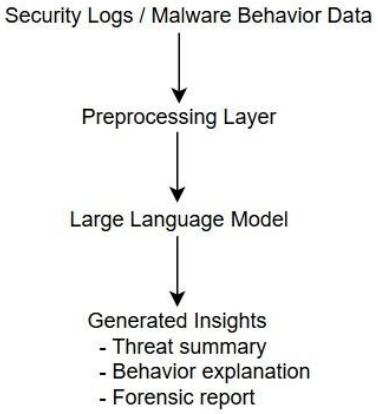

Figure 7. LLM-Based Cybersecurity Analysis Workflow

Earlier studies primarily focused on enhancing AI-based classification approaches; however, recent studies have empha-sized the integration of multiple features to achieve superior performance, including improved accuracy and efficiency, en-hanced interpretability, robustness against adversarial attacks and dynamics behavioral analysis, and automated forensic report generation. Despite these innovations, most current frameworks only assist with individual parts of the malware analysis process; that is, no framework fully supports all analytical aspects of the process. In simple terms, no single framework covers all parts of the analysis process.

The section presents a comparative evaluation of the com-mon malware analysis frameworks identified in the study to systematically highlight the key strengths and weaknesses. The evaluation encompasses various aspects such as detection capabilities, interpretability, adversarial robustness, behavioral analysis, automated reporting, dataset usage, and system lim-itations. By comparing these criteria across multiple frame-works, the evaluation reveals technology gaps that stimulate the development of combined AI-based malware analysis systems. 1) Evaluation Metrics: Analysts use the same evaluation metrics to analyze malware, so that they can comparatively evaluate different frameworks. These metrics assess the clas-sification performance and reliability of a model.

A. Accuracy This is the ratio of correctly classified samples to the total number of evaluated instances. It is defined as:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Here, TP, TN, FP and FN represent true positives, true negatives, false positives, and false negatives, respectively. Accuracy provides a general idea of how well a classification system works; however, it can be misleading when datasets are very unbalanced, which is often the case with malware detection tasks [1].

B. Precision evaluates the proportion of correctly identified malicious samples among all samples predicted as malware.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

In cybersecurity, where false alerts may affect analysis, a high precision value indicates a lower false-positive rate, which is important [2].

C. Recall Also known as the detection rate, shows how many malicious samples the model correctly detected.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

Higher recall values mean that fewer adverse samples go undetected, which makes the whole system more reliable [33].

D. F1 Score The recall and precision are combined into one metric.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

This metric is valuable, especially when evaluating models with imbalanced datasets [33].

Comparative Framework Evaluation: Selected represen-tative frameworks have recently been used to benchmark ana-lytical performance. The representational frameworks include: Deep learning models, explainable AI frameworks, Generative Adversarial Networks, Large Language Models

and Dynamic behavioral analysis frameworks. In terms of deep learning-based frameworks for malware detection, multiple frameworks have consistently demonstrated high levels of classification effectiveness. Nataraj, et al. proposed a CNN-based approach to image malware binaries as grayscale images for classifica-tion units in an automated manner [30]. Kolosanjaji et al. also discussed an LSTM-based framework that is aimed at system call sequences produced during dynamic execution [31]. It has been proposed that hybrid architectures that combine CNNs and LSTM models can capture both spatial and temporal behavioral patterns [4].

GAN-based frameworks have been developed to evaluate the adversarial robustness of malware detection mod-els. The PlausMal-GAN framework generates realistic mal-ware samples to improve zero-day malware detection performance [7] [29]. Recent studies have concentrated on adversarial malware generation methods designed to simulate evasion techniques for adversarial attacks on detection systems [32]. The inclusion of XAI into malware detection systems has significantly improved the transparency of these models. Attribution methods such as SHAP and LIME provide the ability to identify the behavioral features that played a role in the model's classification decision [10][12][13]. Feature attribution methods, like SHap and LIME, enable forensic analysts to identify behavioral characteristics that contribute to the overall classification decision. These attribution techniques are particularly useful in digital forensic investigations where analysts must be able to explain why a system has detected something. LLMs have been widely studied recently for their applications in cybersecurity analysis and forensic reporting. These systems, powered by LLMs, have exhibited impressive

TABLE 7 COMPARATIVE ANALYSIS OF MALWARE ANALYSIS FRAMEWORKS [5]

Framework	AI Model	XAI	GAN	LLM	Dynamic	Key Limitation
CNN Malware Classifier	CNN	No	No	No	Partial	Limited behavioral insight
LSTM Behavioral Model	LSTM	No	No	No	Yes	Black-box predictions
CNN–LSTM Hybrid Model	Hybrid DL	No	No	No	Yes	Lack of explainability
PlausMal-GAN	GAN	No	Yes	No	Yes	No behavioral interpretation
MalGAN	GAN	No	Yes	No	Partial	Limited robustness evaluation
XAI Malware Detection	DL + SHAP	Yes	No	No	Yes	No adversarial analysis
Explainable Behavioral Analysis	ML + XAI	Yes	No	No	Yes	Limited automation
LLM Log Analysis System	Transformer	Partial	No	Yes	No	Limited malware context
LLM Forensic Report Generator	Transformer	Partial	No	Yes	No	No dynamic analysis integration

Source: Compiled from recent AI-based malware analysis studies

proficiency in tasks such as the analysis of security logs, summarization of threat intelligence and generation of automated investigation reports [15][16][20][21]. Most existing frameworks, shown in Table VII, highlight only some of these abilities instead of combining the conceptual capability comparison into a complete system.

Framework Capability Comparison: To further illustrate the differences among malware analysis frameworks, a capability-based analysis can be done based on multiple analysis dimensions:

- Detection performance
- Explainability
- Adversarial robustness
- Behavioral analysis capability
- Automated reporting support

Traditional malware detection models have predominantly emphasized achieving high classification accuracy. Therefore, such traditional systems usually lack explanation and automa-tion capabilities. Recent study efforts have introduced the use of XAI techniques and generative models to increase the visibility and resilience of analysis procedures.

Conceptual capability comparison:

The conceptual capability comparison is that deep learn-ing frameworks can be highly efficient for detecting threats; however, they suffer from a lack of interpretability. GAN-based frameworks are effective at adversarial testing;

however, they do not provide interpretations of behavioral patterns. Deep learning frameworks are effective at detection, but lack explainability. GAN-based frameworks are effective at adver-sarial testing; however, they do not provide interpretations of behavioral patterns. XAI-based models are effective at trans-parency, but lack adversarial testing capabilities. LLM-based systems are effective at automation and report generation, but are not commonly used in the malware detection workflow.

Performance Comparison: Several studies have reported the quantitative performance of malware detection models. The key performance values reported in each of these studies are summarized in Table VIII.

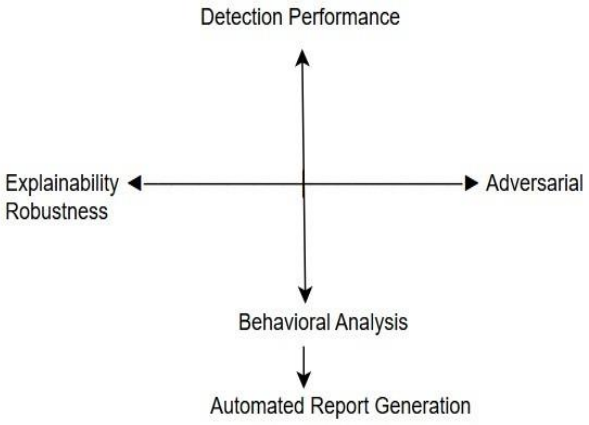

Figure 8. Conceptual Comparison of Malware Analysis Framework Capabilities [4][9][10][15][16][20][21]

TABLE 8 REPORTED PERFORMANCE METRICS OF MALWARE DETECTION FRAMEWORKS

Model	Accuracy	Precision	Recall	F1 Score
CNN-based classifier[30]	94%	0.92	0.91	0.91
LSTM behavioral model[31]	93%	0.90	0.92	0.91
CNN–LSTM hybrid model[34]	96%	0.94	0.95	0.95
GAN-assisted detection[7]	95%	0.93	0.94	0.94
XAI-enabled detection model[10]	92%	0.89	0.90	0.89

The visual representation shows that hybrid architectures (such as CNN–LSTM models) tend to outperform individual CNN or LSTM models in terms of classification accuracy. These results suggest that hybrid deep learning architectures tend to have better classification performance. However, they do not provide transparent reasoning, which limits their use in malware forensics [5].

Key Observations from the Comparative Analysis: The comparative evaluation led to the following significant obser-vations:

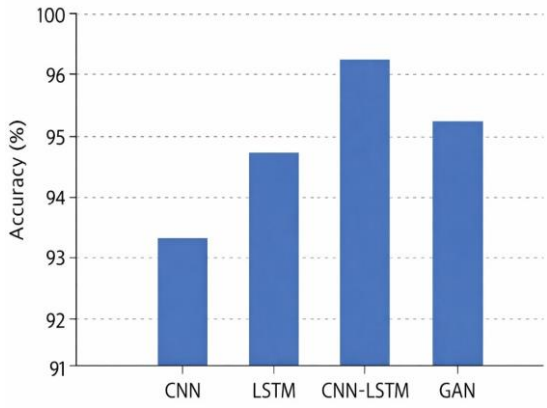

Figure 9. Comparison of Detection Accuracy Across Models

- Deep learning detection models, such as CNN and LSTM architectures, offer high classification accuracy; however, they are limited in transparency due to their black-box nature.
- GAN-based frameworks significantly enhance adversarial testing capabilities by generating synthetic malware sam-ples that mimic evasion strategies; however, they often lack interpretability.
- Explainable AI approaches improve transparency by iden-tifying the behavioral features responsible for detection decisions; although, these systems generally do not pos-sess adversarial malware generation capabilities.
- Large language model-based systems provide robust au-tomation capabilities for log analysis and forensic report generation; however, they are rarely integrated into mal-ware detection workflows.

In conclusion, these findings highlight the need for inte-grated malware analysis frameworks that combine adversarial analysis, explainable detection mechanisms and automated reporting.

V. RESEARCH GAP AND PROPOSED INTEGRATED FRAMEWORK

A. Research Gap

A comparison of the findings from previous studies on malware analysis frameworks has shown limitations regarding the effectiveness of the frameworks in real-world cyberse-curity applications. Although Al has improved the accuracy of malware detection, current systems demonstrate the use of AI to support a single analytical function rather than a comprehensive analytical workflow.

A major limitation of high-performance detection models is their lack of explainability. Deep learning architectures, such as CNN's and LSTMs, can learn complex behaviors from large amounts of data. However, it is difficult to interpret the pre-dictions made by these models. The lack of transparency leads to analysts trust in these models and limits their application in digital forensics [10][17].

Another major issue is the limited integration of adversarial analysis techniques. Recent studies have demonstrated that GANs can be utilized to generate synthetic malware variants and test detection systems against various types of evasive threats. However, many GAN-based frameworks generally aim to generate adversarial samples and expand the dataset size without incorporating interpretability tools that provide insight into how these samples affect the system's detection performance [7][29]. Additionally, the current literature lacks evidence of the comprehensive integration of AI technologies into malware analysis workflows. There are a few existing frameworks that either use deep learning to improve the accuracy of detections, provide an explainable approach, or focus solely on adversarial analysis. Generative models such as LLMs have also recently been studied for their applications in security, such as log analysis, threat intelligence summariza-tion and forensic reporting. While LLMs demonstrate strong analytical capabilities, they are rarely directly incorporated into malware detection solutions [15][20][21].

A subsequent challenge is the lack of automated documen-tation mechanisms. Malware analysis produces a large amount of historical behavioral data, including API call logs, memory artifacts and network traces. Therefore, creating investigation reports based on all this behavioral data manually is time consuming. The potential for LLM-based systems to generate structured reports from security data has recently been demon-strated; however, they are not widely included in the malware analysis workflow [16][22].

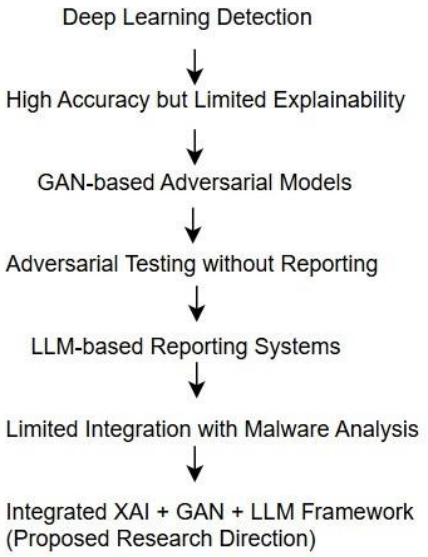

Figure 10. Research Gap in Existing Malware Analysis Frameworks

The limitations have been identified as essential factors to consider when building integrated frameworks for malware analysis that include rules for adversarial testing, explanations of detection and automated reportability as components. The integration or unification of such frameworks can greatly im-prove the transparency, effectiveness and operational efficiency of the analysis process.

Proposed Integrated Malware Analysis Framework

To address the research gaps identified in this section, we propose a conceptual framework that combines the following elements:

- GAN-based adversarial sample generation with dynamic malware analysis;
- Explainable AI and
- Large language model forensic reporting tools

in a single analytical workflow to achieve the following four objectives:

- Improve the overall detection of malware by creating new adversarial samples
- Provide users with an interpretive view of the decisions made to detect malware
- Assess malware behavior by analyzing the dynamic execution
- Automate forensic reporting, through LLMs

By linking all these functionalities, the system creates a more complete and transparent method for malware analysis.

Architecture of the Proposed Framework

The architecture of the proposed system comprised four basic layered structures, namely:

- Dynamic Malware Analysis Layer
- Adversarial Malware Generation Layer
- Explainable Detection Layer
- Automated Reporting Layer

Each of the layers supports an analytical function within the workflows.

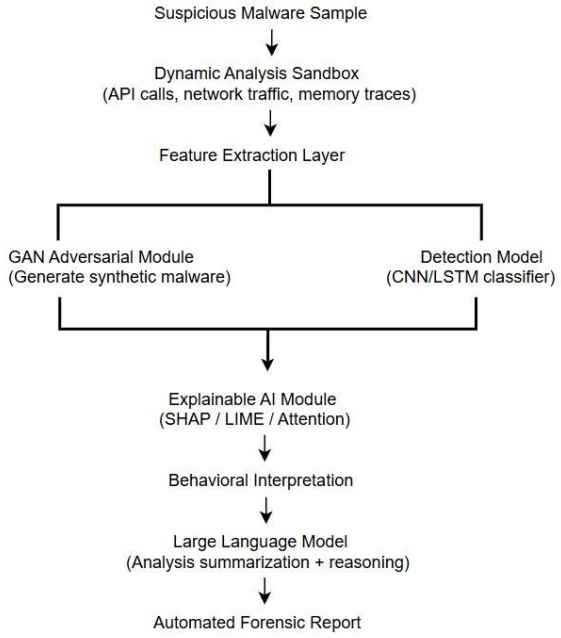

Figure 11. Integrated GAN–LLM Malware Analysis Architecture [14][23][25][26]

Framework Workflow

Step 1: Malware input & sandbox execution: Execute the suspect program within an isolated environment and record behavioral indicators, including API calls, process activities, memory usage and network communications [3].

Step 2: Behavioral feature extraction: The data collected during execution are processed to extract behavioral features that are used as feature representations for the detection models.

Step 3: Adversarial malware creation: The GAN model generates new examples of malware that exhibit evasive attack methods. This malware is used to validate the strength of detection systems and to improve the detection of new threats [5][7].

Step 4: Malware detection: The hybrid deep learning model (such as CNN-LSTM) identifies whether at sample is benign or malicious by analyzing the previously extracted be-havioral characteristics. Hybrid models at efficient in detecting both the spatial and temporal aspects of malware behavior [4].

Step 5: Detection analysis: XAI methods, such as SHAP and LIME are used to generate an interpretable rationale for a model's prediction. These techniques identify behavior attributes that contribute the most to classification outcomes and therefore help the analyst or user comprehend the rationale behind the detection decision [12][13].

Step 6: Generate an Automated Report: The LLM will generate a structured, forensic, explanatory report on malware behavior, the indicators of compromise, and the attack patterns. With automated document creation, security analysts can significantly reduce workload escalation. [15][16].

Benefits of the Proposed Framework

The proposed system would provide several benefits as an integrated framework of components, including the following:

- Resilience detection due to adversarial testing based on GAN
- Improve interpretability using XAI
- Comprehensive behavioral analysis through dynamic pro-cess monitoring
- Automated forensic reporting using an LLM-based sum-marization

The integration of all features provides a transparent and effective analysis of malware.

FUTURE RESEARCH DIRECTIONS

AI technology continues to rapidly advance and improve the capability of computer systems to detect and analyze malware. The comparative analysis presented in this study demonstrates that several significant challenges still exist in four different ar-eas: interpretability, adversarial robustness, system integration and analytical automation. Overcoming these challenges will unlock new possibilities for developing cybersecurity analysis tools by combining emerging and existing AI with advanced cybersecurity analysis frameworks.

Integration of XAI, GAN and LLM Technologies

The primary goal of this study is to construct a novel integrated framework that combines XAI, GANs and LLMs for malware analysis. Existing applications of these three technologies have been studied individually, creating multiple individual solutions that each address only a single component of the overall process of identifying and evaluating malware attacks.

An integrated system can perform various types of analytics simultaneously through a single analytic workflow. For exam-ple, GAN-based malicious AI modules can produce various types of malware that are difficult to detect using conventional detection systems; therefore, analysts can use the new forms of malware produced by GANs to test the ability of their current detection systems to identify new types of malware threats. XAI provides users with clear explanations of how different types of behavioral patterns are identified as being associated with a particular model when predicting behavior. LLMs can also take the output of an analysis and generate well-formatted intelligence reports that can be directly used for validation. Integrating these technologies will enable malware analysis systems to evolve from detection only to a more com-prehensive and transparent analytic model that assists analysts in detecting and investigating complex threat scenarios.

Explainable adversarial malware analysis

The adoption of adversarial machine learning (AML) as a tool for assessing the effectiveness of malware detection technologies in resisting attacks has significantly increased. The availability of GANs to generate synthetic malware that mimics real malware attack strategies provides an opportunity to test detection technologies against sophisticated evasion techniques. Current adversarial malware frameworks are pri-marily intended to develop adversarial samples without in-dicating the effect of these samples on detection efficiency. Therefore, future research should focus on developing new adversarial sample analysis frameworks that, in addition to producing adversarial samples, are paired with interpretable detection models for analysis. Such a detection system would provide researchers with insights into how adversarial malware behaves in ways that allow it to evade detection systems. Identifying these characteristics would help create stronger malware detection systems.

Automated forensic report generation

Malware analysis generates large quantities of data that show how a system behaved during an attack, including system logs, network communication logs and memory artifact data. The process of analyzing this behavioral data and compiling a forensic investigation report typically takes a long time and requires experienced cybersecurity professionals. The capabilities of LLMs to automate the creation of forensic documentation creates new possibilities. A study can poten-tially identify the systems that automatically create forensic reports from logs detailing how the malware acted and the findings from the analysis. The report would include details about how the malware attacked, what signs to look for if systems are compromised, and what to do to fix the problem. Automated reporting systems within digital forensics have the potential to reduce the workload of cybersecurity professionals and increase the quality of digital forensic investigations. An XAI and reporting system based on an LLM can support the production of transparent and understandable reports that align with the evidence obtained during malware analysis.

CONCLUSION

Summary of findings from framework comparison

The advancements in LLM-supporting technology provide new opportunities for automating the code-based creation of forensic documentation. Future research should explore new ways to integrate frameworks capable of converting malware behavioral logs and other analytical data into standard, struc-tured forensic reports with clear, concise descriptions of at-tack patterns, indicators of compromise and recommendations for mitigation measures. Developing reporting tools through automation will reduce the analyst workload, resulting in increased effectiveness of the digital forensic investigative process. In addition, the integration of XAI technology into the reporting mechanism utilized by LLMs will ensure that automated reporting processes align with the analytical evi-dence produced during malware analysis.

Importance of integrated Explainable AI frameworks

The findings of this study emphasize the growing impor-tance of explainability methods in modern malware analysis tools. As machine learning-based models become essential elements of cybersecurity operations,it is essential to un-derstand how machine learning-based automated detection systems operate. The use of XAI will allow the identifi-cation of behavioral indicators and feature attributions that can efficiently and accurately assist in the overall predictions produced by the model. This level of transparency is vital in high-security environments, where system users must analyze system outputs and validate detections when responding to an incident.

Motivation for developing unified malware analysis systems

From the results of the comparative analysis, a large number of existing frameworks identified in this study are aimed only at targeting certain portions of the entire malware analysis workflow. The use of explainability techniques in malware detection will improve the reliability of AI-based cybersecurity solutions. Some of the methodologies are mainly focused on the development of better classification processes, whereas others are mainly focused on the generation of hostile and their interpretation. Incorporating automation and documentation In this way, Automation and documentation have been imple-mented in LLM-based approaches until now, although rarely applied directly to malware detection methods. Distributed development limits AI-based cybersecurity systems. These limitations motivate research to develop malware analysis architectures with multiple analytical functions integrated into a single architecture.

By merging dynamic behavioral analysis and adversarial malware generation with explainable detection technology and automated reporting functions, organizations can maximize their success in simplifying the process of analyzing malware platforms. Dynamic behavioral analysis, combined with adver-sarial malware generation and automated reporting features, provides a more effective platform for users, simplifying tasks and enhancing performance. The advantages can be categorized into two key aspects,

1) The accuracy of the detection is improved, along with the provision of intelligence that can be easily inter-preted
2) Integrating forensic intelligence to assist cyber security analysts to investigate a complicated threat.

REFERENCES

[1] Djenna, "Artificial intelligence-based malware detection, analysis, and mitigation," Symmetry, vol. 15, no. 3, pp. 1–21, 2023.

[2] Faruk, M. Rahman, and A. Islam, "Malware detection and prevention using artificial intelligence techniques," in Proc. IEEE Int. Conf. Big Data, Orlando, FL, USA, 2021, pp. 4635–4642.

[3] Willems, T. Holz, and F. Freiling, "Toward Automated Dynamic Malware Analysis Using Cuckoo Sandbox," IEEE Security & Privacy, vol. 5, no. 2, pp. 32–39, 2019.

[4] Gautam, R. Kumar, and S. Gupta, "CNN–LSTM Hybrid Model for Enhanced Malware Analysis and Detection," Procedia Computer Science, vol. 233, pp. 492–503, 2024.

[5] Ghebrebrhan, H. Lee, and M. Park, "Generative adversarial networks for dynamic malware behavior: A comprehensive review, categorization, and analysis," IEEE Transactions on Artificial Intelligence, 2025.

[6] Goodfellow et al., "Generative Adversarial Networks," Advances in Neural Information Processing Systems (NeurIPS), vol. 27, pp. 2672–2680, 2014.

[7] Won, Y. Jang, and S. Lee, "PlausMal-GAN: Plausible Malware Train-ing Based on Generative Adversarial Networks for Zero-Day Malware Detection," IEEE Transactions on Emerging Topics in Computing, vol. 10, no. 3, pp. 1234–1245, 2022.

[8] P. Preeti, S. K. Sharma, and R. Singh, "Generative Adversarial Networks: Introduction, Taxonomy, Variants, and Applications," Multimedia Tools and Applications, vol. 83, pp. 10215–10248, 2024.

[9] W. Willone, M. Adams, and J. Clarke, "Future of Generative Adversarial Networks for Anomaly Detection in Network Security," Computers & Security, vol. 139, 2024.

[10] Antonio, R. Delgado, and M. Torres, "Explainability in AI-based behavioral malware detection systems," Computers & Security, vol. 141, p. 103842, 2024.

[11] Harikha, S. Reddy, and M. Patel, "Explainable Artificial Intelligence (XAI) for Malware Analysis: A Survey of Techniques, Applications, and Challenges," in Proc. Asian Conference on Innovation in Technology, 2024.

[12] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems, 2017.

[13] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016.

[14] Jelodar, M. Rahman, and T. Nguyen, "Large language models for software security: Code analysis, malware analysis, and reverse engineering," arXiv preprint arXiv:2504.07137, 2025.

[15] Wickramasekara, F. Breitinger, and M. Scanlon, "Exploring the potential of large language models for improving digital forensic inves-tigation

efficiency," Forensic Science International: Digital Investigation, vol. 52, p. 301859, 2025.

[16] G. Michelet and F. Breitinger, "ChatGPT, Llama, can you write my report? An experiment on assisted digital forensic reports written using large language models," in Proc. Digital Forensics Research Workshop (DFRWS EU), 2024.

[17] S. Gulmez, A. GorguluKakisim, and I. Sogukpinar, "XRan: Explainable deep learning-based ransomware detection using dynamic analysis," Computers & Security, vol. 139, p. 103703, 2024.

[18] Baghirov, "A comprehensive investigation into robust malware detec-tion with explainable AI," Cyber Security and Applications, vol. 3, p. 100072, 2025.

[19] Harikha, K. Ramesh, and P. Rao, "Explainable artificial intelligence for malware analysis: A survey of techniques, applications, and chal-lenges," in Proc. IEEE Asian Conf. Innovation Technology, 2024.

[20] M. Guven, "A comprehensive review of large language models in cybersecurity," International Journal of Computational and Experimental Science and Engineering, vol. 10, no. 3, pp. 507–516, 2024.

[21] W. Zhang, J. Liu, and H. Wang, "When large language models meet cybersecurity: A systematic literature review," arXiv preprint arXiv:2405.03644, 2024.

[22] Farzad, A. Khan, and M. Saleh, "Large language models in cyber-security: State-of-the-art and future research directions," arXiv preprint arXiv:2402.00891, 2024.

[23] X. Xingzhi, L. Zhang, and Y. Chen, "LAMD: Context-driven Android malware detection and classification with large language models," in Proc. IEEE Symposium on Security and Privacy Workshops, 2025.

[24] Y. Yeali, T. Kim, and H. Park, "Unleashing malware analysis and understanding with generative AI," IEEE Security & Privacy, vol. 22, no. 3, pp. 54–63, 2024.

[25] M. Mohamed, A. Al-Khalifa, and R. Hassan, "Generative AI in cy-bersecurity: A comprehensive review of LLM applications and vulner-abilities," Internet of Things and Cyber-Physical Systems, vol. 5, pp. 120–134, 2025.

[26] Al-Karaki, M. A. Khan, and M. Omar, "Exploring large language models for malware detection: Review, framework design, and counter-measure approaches," arXiv preprint arXiv:2409.07587, 2024.

[27] P. Dharani, S. Kumar, and A. Nair, "GLEAM: GAN and LLM for evasive adversarial malware," in Proc. IEEE Int. Conf. Information and Communication Technology Convergence (ICTC), 2023.

[28] R. Reza, A. Ahmed, and M. Patel, "ProveRAG: Provenance-driven vulnerability analysis with retrieval-augmented large language models," arXiv preprint arXiv:2410.17406, 2024.

[29] D.-O. Won, Y.-N. Jang, and S.-W. Lee, "PlausMal-GAN: Plausible malware training based on generative adversarial networks for analogous zero-day malware detection," IEEE Transactions on Emerging Topics in Computing, vol. 11, no. 2, pp. 1453–1465, 2023.

[30] Nataraj, S. Karthikeyan, G. Jacob, and B. Manjunath, "Malware Images: Visualization and Automatic Classification," Proc. ACM VizSec, 2011.

[31] Kolosnjaji, A. Zarras, G. Webster, and C. Eckert, "Deep Learning for Classification of Malware System Call Sequences," Australasian Joint Conference on Artificial Intelligence, 2016.

[32] Rigaki and S. Garcia, "Bringing a GAN to a Knife Fight: Adapting Malware Communication to Avoid Detection," IEEE Security & Privacy Workshops, 2018.

[33] T. Fawcett, "An Introduction to ROC Analysis," Pattern Recognition Letters, 2006.

[34] Gautam et al., "Hybrid Deep Learning for Malware Detection," Procedia Computer Science, 2024.