Open Access ISSN: 3007-1437 (Online), ISSN: 3007-1429 (Print)

IRAQI Academic Scientific Journals

Alkadhim Journal for Computer Science (KJCS)

Journal Homepage: https://alkadhum-col.edu.iq/JKCEAS

KJCS ALKADHIM JOURNAL FOR COMPUTER SCIENCE

Cyber-Defense Powered by Generative AI: A Comprehensive State of the Art Review

[1]Ahmed Ali Alsamman*, [2]Najla Badie Al Dabagh

[1] College of Computer Science and Mathematics, University of Mosul, Mosul, 41002 – IRAQ

[2] College of Computer Science and Mathematics, University of Mosul, Mosul, 41002 – IRAQ

Article information

Article history:
Received: May, 11, 2026
Accepted: April, 28, 2026
Available online: June, 25, 2026

Keywords:
Cybersecurity,
Cyber defense,
Generative AI

Corresponding Author:
Ahmed Ali Alsamman
ahmedalialsamman@uomosul.edu.iq

DOI:
https://doi.org/10.61710/9jq4y549

Abstract

Cyberattacks are becoming more advanced, challenging the traditional defenses based on signature-based and rule-based approaches. The current literature on generative artificial intelligence (GenAI) in the context of cybersecurity is disjointed based on model family or area of application, and it does not combine technical performance indicators with practical implementation limitations, making a comprehensive picture impossible. To fill this gap, the study conducts a comprehensive review of peer-reviewed articles and conference papers (2021-2026) indexed in Scopus and ScienceDirect, and Springer databases, involving the use of GenAI as a defensive cybersecurity tool. These papers are divided into (I) GenAI model family: generative adversarial networks (GAN), large language models (LLM), variational autoencoders (VAE), diffusion models, and hybrid GenAI; (II) application domain: intrusion detection, malware detection, anomaly detection, threat intelligence, privacy-preservation, vulnerability detection, and phishing and spamming detection; and (III) defense strategy: reactive, proactive, and adversarial. GenAI typically increases the accuracy of detection and data efficiency and provides active defense. Nevertheless, the practical validation is usually limited to offline tests that apply imprecise metrics. The paper provides the performance–efficacy trade-off model, which relates technical standards and realistic implementation limitations. It also identifies a research roadmap in the future focused on creating autonomous, privacy-protective, and trustful GenAI-powered cyber defenses and suggests a living review platform to keep track of advances in the fast-changing area.

1. Introduction

The cybersecurity environment has experienced unprecedented changes, where more advanced techniques are used by attackers through the aid of artificial intelligence. This dynamic threat environment is an obstacle to the effectiveness of traditional rule-based and signature-based security mechanisms which creates a need for

adaptive defenses. Models used to produce synthetic data examples, which replicate the statistical distributions of their training sets, are also widely called generative artificial intelligence (GenAI). Generative models, conversely, learn complex patterns/structures [1], [2]. Discriminative models are concerned with classification activities. This capability is beneficial for cybersecurity, enabling the development of normal behavior models, generate synthetic attack vectors to test, and develop data representations with privacy [3], [4].

The use of GenAI represents a paradigm shift in the process of analyzing and deterring cybersecurity threats, providing new methods to simulate attacks, anomaly detectors, and creating adaptive responses that advance along with new threats [5], [6]. This shifts the field from reactive defense measures toward dynamic security frameworks. Generative models excel at representing complex data distributions, making them useful for modeling both normal system behavior and complex attack vectors [3], [4]. By synthesizing diverse attack environments and training on heterogeneous datasets, these models improve threat detection and build resilience against novel threats [5], [6].

GenAI is applied across diverse cybersecurity subfields. Generative models used in intrusion detection systems (IDS) allow the detection of unusual network patterns which are not conforming to current baselines [4], [5]. Generative techniques can be used in the vulnerability detection process to generate real patterns of the code, which helps determine possible security vulnerabilities [6]. These models significantly preserve privacy in that they produce artificial sets of data that are statistically acceptable and protect sensitive data [7].

GenAI technologies applied to malware analysis generate controlled variations of malicious software, which can be introduced to a rigorous test and extensive analysis [1], [8]. Code analysis, threat intelligence, and policy development using GenAI have once again transformed the forward-moving power of planning by interpreting various code structures, finding hidden susceptibilities, and generating the actionable results in an approachable way [6], [9]. Similarly, GenAI models have significant potential to produce various high-quality synthetic cybersecurity data [1], [5]. This transition is necessary because cyber threats are constantly evolving because of the risk posed by AI-centric attacks, advanced persistent threats (APTs), and a growing attack surface. The conventional methods of mitigation cannot keep up with the pace, sophistication, and variety of contemporary threat landscapes [2], [10]. These limitations hinder the interception of zero-day exploits, APTs, and polymorphic malware that have specifically been crafted to bypass the use of static defenses [7], [8], [11].

While there have been prior studies focusing on particular aspects of AI in cyber security, a comprehensive review of GenAI's complete impact has yet to be conducted. This paper seeks to fill this gap by unifying current studies and offering insights into future research direction.

1.2 Contributions

The main contributions of this research paper include the following aspects that focus on methodological advancements, general analytical synthesis of the research, and practical advice to the cybersecurity community:

- It provides a multidimensional, unified classification of GenAI applications in the field of cybersecurity, which is defined in terms of the type of model, field of application, and strategy of protection, thus enabling practical deployment and theoretical analysis.
- It conducts an extensive literature review of peer-reviewed papers (2021-2026), compares generative models, such as generative adversarial networks (GAN), variational autoencoders (VAE), diffusion models, and large language models (LLM), and points to their respective strengths, weaknesses, implementation issues, and effectiveness in the real world in relation to various defensive areas.
- It provides qualitative summary comparison tables and graphs that combine research developments, model potential, and unresolved issues, as a useful resource to researchers, analysts, and policy-makers.
- It develops a new performance–efficacy trade-off that could be used to turn the model selection and bridge the gap between technical benchmarks and operational cyber defense requirements.
- It also presents a strategic roadmap for future research based on autonomous cyber defense, privacy-preserving federated learning, ethical AI, and explainable, trustworthy, and adaptable AI security solutions that can be applicable in the real world.
- It suggests a unique live review structure that helps dynamically improve the results of the review, guaranteeing its relevance in the dynamic conditions of constantly changing trends in the generation AI technology and a state of threat.

1.3 Scope and Methodology

The semantic search involved using ScienceDirect, Springer, and Scopus to find the literature regarding GenAI-based cyber-security. The search query was the following: ("Generative AI" OR GAN OR VAE OR Diffusion OR LLM OR GPT) AND (Cybersecurity OR Threat OR Malicious OR Intrusion OR Malware OR Vulnerability OR Phishing OR Spam).
In order to ensure the relevance and quality of the studies, we applied various filters: Field (Computer Science), Document Type (Article or Conference Paper), Language (English), and Publication Year (2021-2026). The duplicates were removed, and the rest of the records were filtered.
Title screening removed articles that were not on GenAI-driven cyber-security, had no full-text access, or were on offensive instead of defensive approaches. The inclusion criteria in abstract screening were the necessity to research the use of generative AI techniques in defensive cybersecurity and provide in-depth empirical data. Papers with insufficient methodological coverage or those that used non-cyber-defense methods are waived from the comprehensive review of the text. We selected 99 studies for final analysis.

1.4 An Overview

This section provides a foundational overview of the primary GenAI model architectures discussed in this review. The core mechanics and cybersecurity applications of GAN, VAE, LLM, and diffusion models are delineated to establish the necessary technical context for the subsequent analysis.

GenAI Models

GenAI models are diverse machine learning frameworks designed to produce new data samples that resemble real-world data without replicating it exactly. These models play a crucial role in cybersecurity by enabling the creation of synthetic datasets for testing, enhancing threat detection through anomaly identification, and supporting secure data sharing. Their flexibility across modalities—text, images, and network traffic—makes them indispensable tools for simulating and anticipating evolving cyber threats [12], [13].

GAN

GAN use two competing neural networks—a generator that creates data and a discriminator that distinguishes real from fake samples. Through this adversarial process, GAN learn to produce highly realistic outputs [14]. They have seen wide use in image generation and data augmentation, and in cybersecurity, they help simulate attack traffic, design adversarial examples, and support intrusion detection system testing [13].

VAE

VAE combine standard autoencoders with probabilistic latent variables, allowing them to reconstruct inputs and to generate novel data by sampling from their learned distributions [15]. Their probabilistic design makes them particularly suitable for anomaly detection, where unusual or malicious activity can be revealed by poor reconstruction quality compared to known normal patterns [16].

LLM

LLM are transformer-based generative systems trained on large text datasets, enabling them to produce coherent and contextually relevant text outputs [17], [18]. In cybersecurity, they increasingly support automated vulnerability detection, the drafting of security reports and policies, and conversational tools for security analysts [19]. While powerful, their reliability and potential misuse (e.g., generating phishing content) remain concerns.

Diffusion Models

Diffusion models generate data by reversing a gradual "noising" process, where information is first destroyed and then reconstructed step by step with learned denoising functions [20]. Unlike GAN, they are stable to train and now achieve state-of-the-art results in high-quality synthesis [21]. In cybersecurity, their ability to produce realistic synthetic traffic and privacy-preserving data makes them promising for intrusion detection, adversarial defense, and controlled malware simulation [22].

2. Related Review Studies

This synthesis structures 14 reviews (2024-2026) into three main analytical categories, each having subcategories, to clarify key concepts, determine trends, and reveal gaps in the review landscape itself.

2.1 Domain-Focused Reviews

These reviews look at the implementation of GenAI in certain operational settings or technical sub-areas of cybersecurity.

- **Intrusion & Network Security**: [23], [24] concentrate on the application of GAN and VAE to improve detection through the creation of synthetic traffic and detection of malicious behavior. Nevertheless, real-time deployment and new transformer-based models are usually overlooked in these reviews.
- **IoT & Embedded Security:** [4], [25] discuss GAN in tasks related to the IoT, including lightweight IDS and device authentication. The major weakness is that the analysis of computational overhead in resource-constrained settings is not done.
- **Autonomous Systems & Cyber-Physical Security:** [26] examines the use of GAN and transformers to improve the security of drones and self-driving vehicles by augmenting the data but does not provide much information on scalability in practice.
- **Digital Forensics & Incident Response:** [27] discusses the use of LLM to automate the process of forensic analysis and evidence reconstruction in the cloud but does not adequately cover the legal admissibility of AI-generated evidence.

The reviews analyze the architecture of the familiar GAN and VAE models to address domain-specific and data-oriented security challenges. They possess a stable and highly detailed technical concentration on algorithmic applications that are applicable in tasks like anomaly synthesis. Nevertheless, these analyses do not fit the operational needs because they are still in the silos of analysis and do not give much consideration to the problem of real-world integration and scalability. Consequently, the primary conclusion is that such studies are authoritative technical manuals in certain fields of issues. One of the most important gaps is the lack of coverage of new architectural paradigms and cross-domain application directions.

2.2 Model & Capability-Centric Surveys

Such surveys consider GenAI in a transversal way, separating analysis by model family or general functional capability.

- **LLM Evaluations:** [28] provides specific benchmarks, comparing 42 LLM in tasks like malware detection. [29] discusses ChatGPT and DALL-E about password security and threat intelligence. The two studies point at the danger of obsolescence.
- **Dual-Use & Offensive/Defensive Taxonomies:** [30] classifies the dual purpose of GenAI and provides a hierarchical taxonomy of 154 studies of models that involve GAN, GPT, and RL. Autonomous security improvements are analyzed in [7]. A recent study provides a taxonomy-based analysis of the dual-use of GenAI in the vulnerability assessment and risk management [31]. It provides a framework of categorizing LLM, GAN and reinforcement learning (RL) approaches in the cybersecurity lifecycle on a structured and model-focused basis. The taxonomy is quite comprehensive, but the synthesis remains largely descriptive, and it does not offer a critical, comparative analysis of the relative effectiveness, limitations, and implementation tradeoffs of the surveyed methodologies. The main weakness of these works is that they are theoretically oriented and there is little discussion on the actual implementation.
- **Empirical Tool-Based Assessments:** [32] is the only one to measure the usefulness of ChatGPT-3.5 in various steps of penetration testing and provides practical application information, but the results are limited due to the single model.

This category evaluates GenAI in a model-focused perspective, measuring specific families and defining their dual-use capabilities in the sphere of cybersecurity. These surveys always provide organized, advanced descriptions of the field. They are conflicted by the tension between broad theoretical taxonomies and narrow empirical assessments. They contributed in providing a crucial, macro-level description of dual-use potential; however, there is still a substantial gap in the form of a substantive gap because they do not fully cover the generative model continuum and are not dynamic but rather analytic.

2.3 Holistic & Governance-Focused Surveys

These studies are more inclined toward macro-level structures, ethics, and strategic execution rather than technical specifications.

- **Systematic Lifecycle Frameworks**: [9] presents a five-stage lifecycle of responsible AI deployment informed by a PRISMA-based analysis, but this is mostly theoretical and has not been empirically validated.
- **Strategic & Implications Overview**: [33] explores the transformative potential of GenAI and related ethical issues but does not concentrate on the particular implementation issues**.**

Taken together, these contributions shift the focus away from technical mechanisms to macro-level governance, ethical, and strategic frameworks that are necessary in responsible GenAI adoption. They are consistent in raising the discussion to high non-technical levels, like ethics and compliance, but also point out a conflict between high-level, model-agnostic principles and the need to provide actionable, model-specific advice. Their fundamental value is to establish principles upon which integration is to be based; however, there is a big gap between these principles and specific governance protocols and legal standards that are architecture specific.

Collectively, these review studies chart the potential of GenAI and demonstrate a fragmented and incomplete landscape. A consistent trend is the overwhelming presence of analyses about GAN, VAE, and LLM, which leaves a significant blind spot regarding diffusion models and other emerging architectures. The main contradiction is between highly specialized application studies of existing models and more general, but still incomplete, model surveys. There are still critical synthetic gaps, such as the lack of coherent frameworks between models, domains, and defense strategies; the lack of emphasis on trade-offs between computational resources; a theoretical focus that is loosely connected with operational realities; and an overall tendency toward descriptive landscapes that offer weak comparative evaluations of the dependence of data sets and model cross-applicability.

3. Classification Framework and Taxonomy

This paper identifies shortcomings in previous reviews whose main peculiarity was predominantly divided subdomain scopes and taxonomies described. In turn, it proposes a unified, multi-dimensional classification. As depicted in Figure (1), a three-core axis framework is simultaneously involved in this framework:

- **GenAI Family**: The types of models described include the GAN, VAE, diffusion models, LLM, and hybrid GenAI.
- **Applications Domain**: Intrusion detection, malware analysis, privacy preservation, vulnerability detection, and phishing, and spamming detection.
- **Defense Strategy**: Covers reactive, proactive, and adversarial resilience.

This taxonomy provides a great support to decision-making by providing explicitly attuned capabilities of models to domain-specific operational demands and hypotheses to defense paradigms, providing enhanced comparability, and directing future investigations into scalable and credible systems.

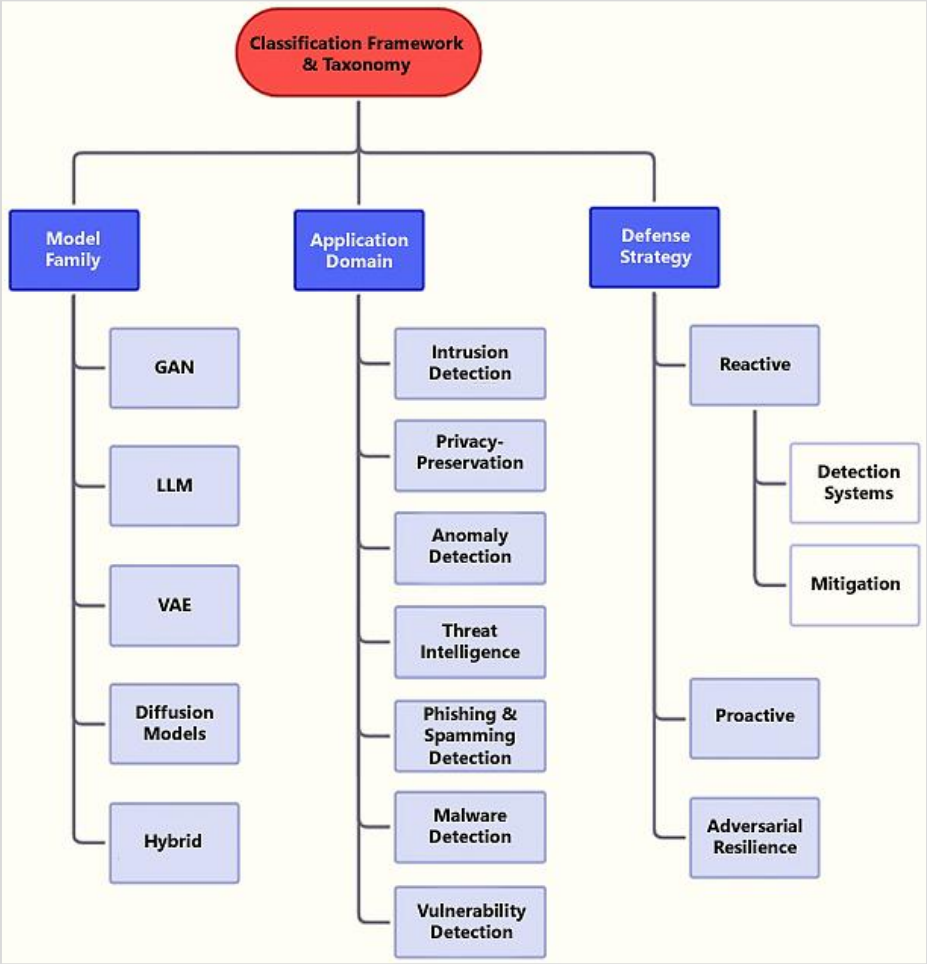

Figure (1): Multi-dimensional classification framework

4. Literature Review

The body of research on GenAI for cyber-security has expanded substantially in recent years, driven by advancements in generative architectures and the evolving landscape of cyber threats [1], [2]. GenAI models—GAN, VAE, diffusion models, and LLM—enable the generation and synthesis of new data instances mimicking complex real-world cyber behaviors [7], [8].

4.1 Thematic Trends

Most of the available literature focuses on the following domains:

Intrusion Detection

Recent GenAI breakthroughs have transformed IDS research, as they have provided models capable of generating synthetic network data, improving the detection quality, and allowing resilience to new cyber threats.

- **GAN-Based Frameworks**: They have proven to be very effective in solving the problem of data imbalance and underrepresentation of attacks in intrusion detection systems. They create high-fidelity synthetic examples, which increase the diversity of datasets and significantly enhance key detection metrics. Experimental comparisons of ACGAN, WGAN, and CTGAN models indicate that the models achieve precision, recall, and F1-scores of up to 98.0-99.0% [34], [35], [36]. Hybrid approaches like AE-WGAN and WGAN-AE also increase recall and training effectiveness of IoT-based detection, despite the recognized computational cost and stability issues [37], [38]. Semi-supervised systems such as CL-GAN, GPIDs, and GUIDE have demonstrated the ability to identify zero-day and contextual attacks in IoT, vehicular, and UAV systems with acceptable latency, but only specific types of attacks, such as DoS and

botnets, can be detected [39] [40], [41]. Privacy-sensitive methods like FEDGAN-IDS can also be used to perform federated learning at high accuracy but at a substantial computational cost [42]. Recent advances such as federated quantum GAN (QGAN) demonstrate encouraging faster convergence, but they can become worse with noise [43]. These developments are said to achieve high detection rates with latencies of sub-milliseconds in edge and IoT applications [44]. Exceptional performance metrics are consistently reported in these model families, but a major conflict remains between these high-performance claims and the consistent reports of significant constraints, such as computational overhead, training instability, and inference latency [35], [37], [38]. A recent study [45] proposed a hybrid architecture of deception, GAN-AII Pot, a honeypot based on a BERT model that uses a GAN generator to generate intelligent interactive decoys to IoT devices. The system has a high F1-score of 99.4% but is very costly in terms of computation and has a lower capability of preventing denial-of-service (DoS) attacks. The main methodological innovations to improve GAN in IDS have been architectural hybridization and domain adaptation. In spite of these developments, there are still significant open gaps, such as computational cost, mode collapse, the requirement to keep up with new attacks and DoS attacks, and a reliance on offline validation that invalidates claims of real-time robustness.

- **LLM-Based Frameworks**: Are being developed as anomaly detectors through the synthesis of complex attack examples of minority classes. The authors in [46] show that GPT-based models, namely GReaT and Real Tab Former, can produce tabular data to detect intrusion in the industrial internet and that their contextual-learning model is more robust to classifiers than the conventional approaches, including SMOTE. This method, as always, increases the performance of classifiers and improves the macro F1-score from 81.0% to more than 90.0%. The recent works [47], [48] have shown signs of a twofold progress of partnering and powerful AI-based cybersecurity. New federated learning architecture that integrates LLM with Federated Retrieval Augmented Generation (RAG) to detect intrusions has been suggested [47]. This framework improves the F1-score of the base model from 86.3 % to 95.5% by augmenting alerts with privacy-sensitive threat intelligence that is shared. Another study examines the dual-model systems, including ChatGPT integration with traditional machine learning methods [48] . The results show that adversarial and prompt-based attacks can decrease F1 -scores by up to 50% but defensive strategies such as adversarial training and prompts to enhance robustness limit the drop to about 5. In a related field, an LLM-APTDS framework proposed [49], where fine-tuned LLM are used to identify Advanced Persistent Threats (APTs); the system achieves an F1-score of up to 99.2% and produces explainable reports which are mapped to the MITRE ATT&CK framework. However, the method is hindered by the high computational cost and overdependence on the underlying data. Taken together, these studies highlight the need to support the secure cooperation between architectures and proactively protect AI security systems against the new threats. LLM have proven to be of significant use in the cybersecurity domain, especially in synthetic data generation and supporting federated intrusion detection systems. Such applications have led to a significant improvement in performance as shown by F1-scores improvement. However, there are still a number of critical constraints. First, these methods have been mostly validated in single-data, single-epoch experiments. Second, the issue of fidelity of the synthesized data is not resolved. Three, there is a gap between the semantic benefits of the LLM-generated data, on the one hand, and the methodological limitations that, on the other hand, limit their maximum effectiveness.
- **Hybrid GenAI Frameworks**: A hybrid structure addresses class imbalance when detecting intrusions through a dual-stage structure. It generates synthetic attack traffic with the help of a hybrid VAE-WGAN and uses a fused Stacked LSTM and Multi-Scale CNN to obtain temporal and spatial features and classify them. When tested on both NSL-KDD and AWID, the model attains between 83.5% and 98.9% accuracy and 83.7% and 98.9% F1-score which proves that generative augmentation is effective in enhancing performance. Some of the major ones are label-guided data synthesis and spatiotemporal feature fusion. Still, there are limitations: poor generalization with limited or noisy data, large computational costs, no ability to run in real time or function as a lightweight solution--a gap between laboratory performance and practical deployment remains [50].

Out of all these studies, it is clear that there is one common thread: although an extremely high detection accuracy (above 98.0% is typically asserted), very few verify their models to be used in edge device deployment and measure the inference time.

Privacy Preservation

Generative models support diverse privacy-preserving applications since these models allow the transfer and learning of information without sensitive data being lost. These models reduce the aforementioned risks through the application of privacy-preserving, anonymization, and federated adaptation methods and maintain downstream task utility.

- **GAN-Based Frameworks:** Differently private GAN Architectures Differentiated private GAN, including RDP-CGAN and Wasserstein GAN, have shown superiority in processing algorithmic datasets in health and IoT applications, and have formed considerable privacy assurances. Context GAN and Fed-GAN versions of federated variants can be used to generate medical and financial data in a secure way and at a disseminated scale, which allows utility to be preserved in a heterogeneous population of clients [51], [52], [53], [54]. Consistently, these models provide tunable privacy with high accuracy (e.g., 98.6%). However, a conflict exists between achieving strong privacy and maintaining model stability and utility. The primary innovation is integrating differential privacy (DP) and federated learning into GAN frameworks. Key open gaps include the privacy-utility trade-off, discriminator instability, high complexity, and limited domain applicability.
- **VAE-Based Frameworks:** De-Identification digital forensics VAE and analogous methods have been found to anonymize sensitive data by disentangling the presented data to generate privacy-compliant clinical images and artificial network traffic that can be used with intrusion detection algorithms without storing personal information. Such strategies highlight the possibility of representation disentanglement and synthetic generation of data as an enhancement and explanation of privacy and interpretability [55], [56]. Consistently, they enable feature disentanglement for privacy. A conflict arises between generating private data and preserving the fidelity and utility of the original data, especially for complex modalities. The contribution is using conditional and disentangled VAE. Open gaps include low output realism, recall loss in synthetic traffic, and a general lack of formal differential privacy guarantees.
- **Diffusion-Based Systems:** They have shown the capability to protect privacy by using anonymization and safe data synthesis. Anonymous diffusion affects use systems based on a blockchain design to secure client inputs and model parameters and provide powerful tailoring against re-identification and inference of attacks, [57]. Consistently, they offer strong anonymity with high data fidelity (SSIM ≈1). A conflict is the significant system overhead (4-19%) introduced by privacy mechanisms like blockchain. The contribution involves integrating diffusion models with secure, decentralized frameworks. Major open gaps include reliance on specific infrastructure (e.g., blockchain), the "honest-server" assumption, and validation primarily on synthetic tasks.
- **LLM-Based Frameworks:** LLM like Security BERT adopt dynamic encoding models to handle sensitive IoT information on explicit security policies, trade-offs between privacy, and other performance indicators like classification accuracy. In addition, GPT and Llama are used to verify privacy policies, which results in stronger compliance and interpretability under regulated settings [58], [59]. Consistently, LLM achieve high performance (F1 up to 98.0%) in privacy-aware tasks. A conflict exists between their high capability and practical constraints like API costs, low determinism, and overfitting. The contribution is applying LLM to policy analysis and privacy-aware encoding. Open gaps include limited adversarial robustness, evaluation on single datasets, and high computational costs.
- **Hybrid GenAI Frameworks:** VAE-WGAN Classification Hybrid architectures are another subcategorization within such research, applying additional privacy mechanism, such as privatization and feature preservation modules, to maintain anonymity in the medical case-based system and retain the relevant features [60]. Consistently, hybrids aim to balance multiple objectives like privacy and utility (Acc. > 90.0%). The conflict is the added complexity from combining architectures. The core contribution is the fusion of models like VAE and WGAN. Open gaps involve challenges in scaling to multimodal data and ensuring the fidelity of privatized features.

Threat Intelligence

GenAI adds to the threat intelligence paradigm, through which realistic attack situations can be generated in a systematic way, and modeling of adaptive adversary behaviors can be achieved, to prompt proactive and resilient defense capabilities that can adapt to the changing cybersecurity threats.

- GAN-Based Frameworks: GAN are able to generate realistic attack data, which can be used to support automated threat intelligence and detecting accuracy [61]. Other frameworks, including EAC-GAN and GAN+XGBoost, have reached high precision (96.8%) and almost perfect accuracy (99.2%) [62], [63], which has consistently enhanced data-driven monitoring. There is a gap between simulated performance and real-time or enterprise validation; a solution has been proposed to use GAN along with interpretable machine learning (XGBoost, SHAP). The gaps that are left are that the models have limited generalizability and interpretability.
- VAE-Based Frameworks: VAE-based models, specifically VAE-CNN models and AdMVAE hybrids, can be useful in malware classification and adversarial defense with high recall and robustness [64]. Some adversarial defense models like ND-VAE use noise filtering to protect against adversarial samples, achieving accuracies of up to 95.0% [65] and being highly effective at extracting features to detect anomalies. Nonetheless, strong performance comes at the cost of poor inference time and scalability problems. One refinement is hierarchical, noise-filtering, adversarial VAE architectures, but interpretability, scalability, and inference latency issues persist.
- Diffusion Models-Based Frameworks: Lightweight diffusion models such as LW-Diff generate malicious traffic on edge devices with a 92.3% accuracy, sacrificing little performance [66]. These models can always provide quality synthesis in low-resource environments, but there is a trade-off between the quality of generation and lightness of computation. The key contribution is designing DDPM-based models for edge computing, though lacking points such as the limited scope of attacks and no real-time synthesis validation.
- LLM-Based Frameworks: LLM can be used to simulate attacks and identify anomalies automatically; retrieval-augmented and dual-LLM systems can be used to improve proactive cloud security and scenario realism [67], [68], [69], providing context-dependent adaptability. Their great computation complexity and sensitivity to tuning are the sources of conflict; the innovation is using retrieval-augmentation and multi-agent models. The open gaps entail scalability, data bias, and the necessity of federated learning strategies [67], [69],.
- **Hybrid GenAI Frameworks**: Multi-layered defenses based on GAN and VAE provide highly accurate (97.0%) and low false-positive defenses against polymorphic attacks in the cloud [8], [70], and provide scalable, adaptive learning. The high level of computational intensity needed to perform such integration is a conflict. These studies are a blend of architectural generative elements, yet there are still gaps in relation to major computational requirements and the problem of misuse of synthetic data, which is ethically questionable.

Anomaly Detection

Generative models offer a scalable baseline for anomaly detection in complex cyber-physical systems through the generation of balanced training and learning robust latent representations of normative behavior.

- GAN-Based Frameworks: GAN can increase the performance of detection in network security, finance, and the Internet of Things by balancing the distribution of classes. R-GAN, T-GAN, and PCA-GAN architectures have almost perfect accuracy and F1-scores (up to 99.5%) of fraud and intrusion detection [71], [72], [73]. They keep on enhancing the performance of classifiers through adaptive synthesis. However, there is still a conflict between achieving high metrics and suppressing natural instability (mode collapse) and computational limits. One potential solution is to combine GAN with transformers, PCA, or automated machine-learning pipelines to enhance stability and efficiency. The current gaps include lower recall when training on small datasets, instability in training, and the black box nature of the models.
- LLM-Based Frameworks: Context-aware, real-time anomaly detection in cloud systems can be done with LLM like GPT and Llama with high accuracy (e.g., 96.3%) [74]. They always provide scalable detection based on contextual reasoning. Nevertheless, there is a challenge of high computational cost and variable false-positive rates. One of the possible solutions uses retrieval-augmented generation to enable explanations and spatial analysis. Existing issues that are open are variability in the false-positive rates and the need to constantly update the models.
- Hybrid GenAI Frameworks: Models like VADAD incorporated VAE and Diffusion model for better reconstruction fidelity and stability in noisy situations, and have high ROC-AUC scores (e.g., 92.1%) [75].

They always offer powerful data synthesis. The first disadvantage is that advanced augmentation is quite expensive in terms of computation. Another approach that can be considered is to enrich data in the latent space by integrating diffusion models with VAE and Synthetic Minority Over-sampling Technique (SMOTE). Open gaps include the preponderance of tabular data, high computational costs, and the need for more extensive validation.

4.2 Malware Detection

The malware detection domain is a very sensitive area where the GenAI is used to complement traditional defense tools by solving the lack of data and one-day variants, and protecting the classifiers by generating synthetic data and implementing adversarial systems.

- **GAN-Based Frameworks:** GAN, specifically the DCGAN and MCOGAN models, are used to convert malware binaries to color images and to create synthetic samples, which aid in optimizing the classifier. The current development of DCGAN algorithms has made zero-shot learning possible in previously unknown malware families with classification rates reaching up to 98.9% [76]. Similarly, an optimizer based on GAN can be used to retrain classifiers adversarial, and thus increase the accuracy of convolutional neural networks by 87.5 to 96.0% [77]. In both methods, synthetic data generation is successful in reducing the imbalance in classes. However, there is still a major trade-off between achieving high accuracy and incurring a high cost of computation. The key weaknesses of these methods are that they require a static analysis, intensive computation, and the lack of dynamic behavioral properties.
- **VAE-Based Frameworks**: Conditional VAE (CVAE) are synthetic malware image generators that help to boost detection accuracy (up to 99.0%) and macro-F1 scores (91.0%) on Android and multi-family settings [78], [79]. Latent distributions of minority classes are continuously modeled in the models. Nevertheless, their computational complexity and insufficient augmentation to negligible malware families is a conflict. The solution is based on conditional generation and hybrid VAE-GAN designs. The unresolved issues include high computational expense, training difficulty, and limited sample fidelity.
- **Hybrid GenAI Frameworks**: Hybrid models like GLEAM combine Copula GAN with GPT-Neo to generate a variety of evasive malware samples that reduce the true-positive of black-box classifiers [80]. These methods are consistent in increasing adversarial sample diversity and evasion robustness. However, they are susceptible to model collapse and are limited to fixed feature analysis. Its fundamental mitigation approach is to combine statistical generative models with language-model-based contextual synthesis. Unresolved questions include the lack of dynamic integration of features, the instability of GAN, and the fact that they require a large amount of fine-tuning.

Although the application of generative models in the overcoming of class imbalance in malware detection works well, studies are still highly reliant on the use of static analysis of binary images or opcode sequences, without considering the dynamic analysis of malware behavior in the execution environment.

4.3 Vulnerability Detection

The development of LLM has a huge positive impact on vulnerability detection by improving semantic understanding, explain-ability, multi-vulnerability identification, representation learning, and hybrid generative-discriminative pipelines. A study [81] is the first to propose a BERT-based vulnerability detector that uses explain-ability interfaces such as SHAP and LIME with a high accuracy rate of 91.9% and high F1-scores on DiverseVul, overcoming existing transparency risks. A different work presented Smart Guard [82], a retrieval-augmented GPT-3.5. The Turbo chain-of-thought model identifies vulnerabilities in smart contracts with a recall of 95.1%, which is better than the conventional static analysis systems. GRACE is a model proposed [83] that combines GPT-4 with graph-structured contexts and semantic retrieval and increases the F1-score by 28.7% in detecting software vulnerabilities, but the model is limited by the risk of data leakage. Another study compared various fine-tuned and base LLM [84] and found that CodeLlama-7b has the best F1-score, but cross-dataset generalization is still a problem. A suggested hybrid pipeline consists of GPT-3.5. Turbo generation and fine-tuned BERT-CNN classifiers provide automated and accurate CVSS scoring with almost perfect accuracy (up to 99.0%) on key metrics [85]. The study also investigates embeddings based on Llama, Qwen, and other models to cluster semantically CVEs and finds about 50% accuracy in kNN accuracy on CWE classification in complex multi-class problems. GPT3.5-Turbo [86] is used to detect vulnerabilities in

continuous mode and generate record-fix, with a satisfactory accuracy of 77.0%, prioritizing the true positive results in OWASP benchmarks.
Regularly, the models using the LLM prove to be better in semantic analysis and more accurate than the conventional ones. There is a primary conflict between their sophisticated features and realistic limitations, including narrow context windows, high API prices, and a possibility of data leakage. The current trend is the creation of hybrid architectures combining LLM with retrieval-augmented generation, graph-based code analysis, or discriminative classifiers. Critical open challenges remain, such as overfitting, sensitivity to prompt and template design, and lack of generalization to different datasets, as well as intrinsic to particular programming languages.

The LLMS are very efficient with respect to vulnerability detection, but because of their dependency on API calls, they become practically challenging due to the cost implicated, privacy concerns, and reproducibility. The lack of generalization across datasets with the LLMS, also means that there is an overfitting issue.

4.4 Phishing and Spamming Detection

Phishing and spamming detection practice is becoming increasingly based on GenAI to enhance detection and address the issue of class imbalance. These models increase the pool of available data, which increases the representativeness of training sets. Researchers aim to mitigate the problem of obsolescence in traditional supervised learning pipelines by adding synthetic artifacts, namely textual content, uniform resource locators (URLs), and related metadata.

- **GAN-Based Frameworks**: GAN like Leak GAN and DCGAN create synthetic phishing emails with high F1-scores (up to 99.8%) despite input modality limitations [87], [88]. Models such as CTGAN and CGAN generate high-quality data, which allows an accuracy rate of more than 98.0% [89], [90]. They are, again, effective in enhancing training information. However, there is a tradeoff between high performance and high computational cost or scalability constraints. Main point is the integration of GAN models and BERT classifiers, positive-unlabeled (PU) learning, and explainable AI (XAI). The open gaps include binary label constraints, metadata lack, and significant computational cost [87], [88], [90].

- **VAE-Based Frameworks**: VAE, which combine deep and convolutional neural networks, learn latent representations to detect phishing URLs with high accuracy (up to 97.9%) and low false-positive rates and can be used in real-time [91], [92]. Regularly, they offer effective and scalable feature extraction. There is a trade-off between real-time performance and explain-ability. The contribution entails hybrid incorporation of the VAE with the supervised classifiers. The only missing features are inference latency (~1.9 s), privacy issues, and model explain-ability.
- **LLM-Based Frameworks**: (LLM like GPT 3.5/4 and hybrid meta-models (PhishEmailLLM) can be used to perform dynamic and high-precision detection (up to 99.0%) [93], [94]. Real-time spam filters are enhanced by integrations with ChatGPT [95]. Regularly, such methods take advantage of sophisticated semantic knowledge. These works involve dynamic model choice, prompt engineering, and hybrid architectures. There is a conflict of dependency on API calls, token limits, and prompt design sensitivity. Privacy risks related to calls that are dependent on the internet, the use of non-fine-tuned models, and operational restrictions (such as token limits) are considered open gaps.

There is an interesting contradiction in this area: API calls via the internet that are required by LLM-based detectors are highly accurate (up to 99.0%), but also present privacy threats. This balance between data confidentiality and detection performance needs to be taken into account when implementing the trade-off in the enterprise.

4.5 Critical Gaps and Controversies in Literature

Although the reviewed studies yield valuable insights, various issues are still not properly addressed and represent major limitations to progress:
IDS: Frameworks are associated with computational resources, volatile GAN dynamics, and retraining costs when implemented in the enterprise [37], [43].

- Privacy Preservation: Trade-offs hinder the progress of the area between data utility and privacy, as well as the issue of computational complexity and heterogeneity of federated systems [53], [54], [57].
- Cyber Threat Detection: Hybrid combinations have better explain-ability but are hampered by low enterprise validation and huge amounts of resources required [62], [63].
- Anomaly detection: These detectors though they greater fidelity, face instability in model convergence and biases in datasets and cannot operate well when faced with real-time scenarios [71], [96].
- Malware Detection: Generative models increase data diversity but have problems with the aspects of dataset bias, synthesis realism, and dynamically analyzing data [76], [77].
- Vulnerability assessment: VLLM-based scanners promote better reasoning performance, but there are still limited context windows and high computing costs [81], [82].
- Phishing and spamming detection: Systems promise improvement in phishing reduction but are limited to partial metadata, the issue of fidelity, and resistance to evasion strategies [87], [94].

4.6 Defensive Application Domains

GenAI is changing the cybersecurity field by having a significant impact in four key areas of defense:

- **Intrusion Detection:** Synthetic AI methods help detect intrusions through IDS when based on a variety of GenAI models, such as GAN, VAE, diffusion models, or LLM-enhanced frameworks [39], [41], [43], [44]. The technologies are efficient in reducing the imbalance of data and detecting advanced zero-day attacks and ensure high and steady quality even in the limited conditions provided by IoT and edge devices, by privacy-safe strategies like federated and quantum GAN models.
- **Malware Analysis**: Malware analysis combines GAN, VAE, diffusion, and Hybrid GenAI Frameworks to facilitate the creation of variants of malware approaches that evade training, trains opcode datasets, and dramatically enhances the classification of novel types of malwares with high detection schemes [76], [78], [80].
- **Privacy-Preservation**: Privacy-preserving synthetic data generation uses sophisticated constructions that are based on differential and federated GAN, VAE, and diffusion models [52], [57]. These systems ensure high levels of anonymization and confidentiality of sensitive data in the area of healthcare and IoT, while they manage to preserve functionality for downstream tasks successfully.
- **Vulnerability Assessment**: Automated vulnerability detection and patch generation is based on the idea of adding semantic reasoning, high recall, and improved explanation ability to the existing multi-vulnerability set discovery and rating through LLM and by hybrid generative-discriminative pipelines [82], [97].

4.7 Different GenAI Models and Capabilities

Key Distinctions:

- **GAN:** GenAI-based architectures generate high-fidelity and diverse synthetic data, which significantly increase the intrusion detection system (IDS) and cyber threat detection accuracy, which are commonly over 98%. Nevertheless, these architectures are still computationally expensive and prone to training instability [87], [90].
- **VAE**: Is effective in detecting anomalies as well as anonymous data in privacy-preserving and URL-dependent phishing detection. However, their explain-ability and applicability in the real-time detection setting have their limitations [60], [64].
- **Diffusion Models**: Diffusion-based systems offer stable data generation and are applicable on edge devices in the context of IDS. However, the cost of these models, characterized by significant computational requirements, restricts their real-time usage in resource-constrained systems [66], [98].
- **LLM**: They can provide more context-based and semantic insights into intrusion and phishing detection, and the accuracy is very high in dynamic settings. Their interpretation and large resource needs remain as a big challenge [68].

4.8 Summary Tables

The section consists of seven detailed overview tables from Table (1) to Table (8), reflecting the major results of a comprehensive overview of GenAI applications in different cybersecurity domains. The key points captured in the tables include the types of models used, data used to evaluate the models, evaluation metrics, contributions

of the models, constraints of the system, and the research gaps. The structure of this information contributes to the fact that the tables are a clear and brief comparative reference that helps to understand the contemporary potential, obstacles, and research directions in GenAI-powered cybersecurity.

Table (1): Summary of GenAI-Based Intrusion Detection Studies

Ref	Year	GenAI Model	Datasets	Key Metric	Contributions	Limitations/ Gaps
[35] [36]	2024 2025	WGAN, Vanilla GAN, CTGAN	CIC-IDS2017	Prec./Rec. up to 100.0%/82.0%; Prec./F1 perfect on synth	Improved synthetic data fidelity; Augmentation efficacy (49×)	Computation overhead; Mode collapse; Needs updates for evolving attacks
[37] [38] [99]	2025 2025 2024	AE-WGAN; WGAN-AE; WGAN-DL-IDS	NSL-KDD, CICIDS2017, CICIDS2017, 5G-NIDD, IDSIoT2024	F1 up to 99.6%; PR-AUC 99.87%; Acc. 98.0%, F1 92.2%	Denoising + imbalance reduction; Hybrid for resource-constrained devices; Oversampling + feature reduction;	Computational cost/overhead; Latency increase; Stability issues;
[100] [101]	2023 2025	BEGAN; BEGAN + BiLSTM + XAI	NSL-KDD, UNSW-NB15, IoT-23; NSL-KDD, UNSW-NB15	Acc. up to 93.2%; Acc. 93.8%, F1 83.0%	Autoencoder discriminator; Combined GAN balancing with explainable AI	Synthetic data quantity dependent; Long training time
[102]	2025	SA-WGAN	Mississippi State pipeline	Acc. 98.5%	Self-attention for feature emphasis	Single dataset; Offline batch
[39]	2024	CL-GAN (Semi-supervised)	NSL-KDD, CICIDS2018, Bot-IoT	Acc./F1 +~5.0%	Semi-supervised GAN;	Focused on DoS/botnet
[34] [42]	2021 2022	ACGAN+CNN; FEDGAN-IDS (ACGAN-like)	KDDCUP99, UNSW-NB15, CICIDS17, AAGM17; NSL-KDD, KDD-CUP99, UNSW-NB15	F1 up to 98.7%; Acc. >99%, F1 99.0%	Balanced training with GAN-generated attacks; Federated training with minority augmentation	Multi-class limitation; Computational burden; Instability
[103] [104]	2025 2023	5DGWO-GAN + CNN-AE; TDCGAN	UGR'16; NSL-KDD, UNSW-NB15, IoT-23	Acc. 99.2%, RMSE 0.09; Acc. 95.0%, F1 94.0%	Gray wolf optimizer for tuning; Triple-discriminator	High training complexity; High computation/scalability
[105]	2025	GAN-LSTM	SWaT, WADI	Acc./F1 up to 91.0%	Temporal attack sequences	Latency, Mode collapse
[40] [106]	2024 2024	Semi-sup. GAN; Dual-discriminator GAN	Real vehicle CAN; Honda, KIA vehicles	Acc. >93.0%, F1 >93.0%; Detection rate 99.9%	Contextual pattern-aware for CAN; High acc., low latency on vehicular nets	Per-vehicle models; Fixed thresholds; Predefined attacks
[41] [107]	2024 2024	GUIDE (text-GAN); 1D convolutional GAN	RT-IoT2022 UAV; Custom SCADA	F1 up to 91.0%; Acc. 99.1%, F1 99.4%	Text-GAN for sequence augmentation; Realistic flow generation in industrial IDS	Diversity optimization needed; Dataset scarcity
[108]	2025	SYN-GAN	UNSW-NB15, NSL-KDD, Bot-IoT	Acc.>90.0%	Synthetic-only training feasibility	Needs retraining for new attacks
[109] [110]	2023 2023	GAN-IF; SGAN-IDS	Simulated data; CICIDS2017, NSL-KDD	Acc. 98.8%; Detection rate -15.90%	Robust vs. zero-days; Adversarial example generation for IDS eval	Simulated data; Black-box; Single attack types
[44] [111]	2025 2023	CGAN + VGG16; GAN adversarial training	RT-IoT2022; Edge-IIoTset	Acc. 97.6%; Acc. ~95.0%, Rec. 96.0%	Augmentation + transfer learning; Robust feature training	Dataset diversity; Invalid samples; Limited testing
[112]	2025	GAN + XGBoost	NSL-KDD, UNSW-NB15, CICIDS2017	Accuracy: 99.9%	GAN-XGBoost IDS to solve class imbalance	High computational cost; parameter sensitivity
[113]	2025	GAN-based	NSL-KDD, CICIDS2017	FID: 8.2	Validated GAN for synthetic data to mitigate scarcity & improve IDS.	Potential mode collapse; hyperparameter sensitivity.
[50]	2024	VAE-WGAN	NSL-KDD, AWID	Acc: 83.45-98.97%; F1: 83.69-98.90%	Label-guided augmentation; LSTM-MSCNN for spatiotemporal features	Poor gen. on scarce/noisy data; high compute; not real-time/lightweight
[45]	2026	GAN + BERT (GAN-AIIPot)	IoT HTTP, NSL-KDD, ToN_IoT	F1: 99.4%	GAN-based intelligent honeypot for dynamic deception against ML attackers	High computations; false positives; less effective vs. DoS
[43]	2025	Federated Quantum GAN (QGAN)	NSL-KDD	Accuracy: 91.3%, F1-score: 90.3%	Hybrid quantum-classical GAN, fed. training for scalability and robustness, noise simulation analysis	Decline in performance under noise, complexity of training, need for noise mitigation and further tuning
[98]	2025	NT-DDPM;	CTU, USTC-TFC, ISAC, UJS-IDS2024	Acc: 99.9%; F1: 99.4%	Multi-layer encoding, superior stability	High inference latency; No drift adaptation
[46]	2024	SMOTE, SMOTE-NC, GReaT (GPT), REalTabFormer	Edge-IIoTset	Macro F1: 81.0% → >90.0% (RF)	Systematic comp. of trad. vs GPT-based tabular aug. for IIoT	Single-epoch GPT tuning; single dataset; GPT data quality issues
[47]	2026	LLM (Gemma, Zysec, Llama, Ministral)	CIC-IDS-2017	F1: 95.50%	Privacy-preserving FL+LLM+RAG for alert enrichment	Risk of poisoning; computational bottlenecks; uses older dataset
[48]	2026	ChatGPT-based	CRÈME, CICIDS2017, CelebA	F1 drop: ~5.0% (with defense)	Analysis of adversarial threats/defenses for LLM-based dual-model systems	ChatGPT constraints; focused on specific attacks
[49]	2026	Fine-tuned LLM (Qwen, Deepseek)	DARPA TC-E3	F1: up to 99.2%	LLM for precise, explainable APT reports via MITRE ATT&CK mapping	Computational overhead; single dataset reliance.
[114]	2026	MiniLM + GPT-4	UNSW-NB15 (82k flows)	F1: 88.9%	Edge-cloud LLM for IoE: explainable analysis & automated rules	No adversarial testing; dataset not IoE-specific

Table (2): Summary of GenAI-Based Privacy-Preservation Studies

Ref	Year	GenAI Model	Datasets	Key Metric	Contributions	Limitations/ Gaps
[54] [51] [52] [115] [116]	2021 2021 2024 2022 2021	RDP-CGAN; WGAN-DP; ContextGAN; DP-CTGAN/PATE; BGAN + DP	Medical; MNIST; Healthcare/Security/Financ e; Fitbit; Smart-health	F1:.44.0%; Acc: 98.6%; EMD: 0.12; TU: >80.0%; KS: p≤0.98	Tight privacy guarantees; DP optimization; Domain-constraint enforcement; Use-case privacy metrics; Tunable privacy	Privacy-utility trade-off; Discriminator instability; High complexity; Dataset-specific; Small/noisy data
[53] [117]	2025 2022	Fed-GAN; AMT-GAN	MNIST, Fashion, CelebA; CelebA-HQ, LADN	Acc: +22.4%; ASR: up to 89.6%, FID: 34.44	Server-only gen.; pHash-KT transfer; High adversarial transferability	Image-only domain; Model inversion; Gender/attribute issues
[118] [119]	2021 2025	DP-GAN; CryptoGAN + Paillier	Yelp, trajectories; ISIC, Brain MRI	RMSE: 0.07-0.14; Acc: 78.6-89.7%	Pan-privacy in streams; Homomorphic encryption vs leakage	Slow convergence; High compute; Accuracy reduction
[60] [55]	2021 2024	CVAE+WGAN-GP; Disentanglement AE + VAE	Iris v2.1, CheXpert, CelebAMask-HQ	Acc: >90.0%; SSIM: 0.59-0.65	Privacy-preserving explanations; Feature disentangling & counterfactuals	Low image realism; Multimodal scaling
[56]	2024	Conditional VAE	IoT-23, CSE-CIC-IDS2018, Kitsune	F1 loss: 0.5%-12.9%	Scalable traffic generator; Publicly released traces	Poor benign traffic rep.; Recall loss; No formal DP
[57] [120]	2025 2025	Stable Diffusion+Blockchain, Split UNet	Stable Diffusion v1.4 generated Images	Overhead: 4-19%; SSIM: ≈1	Neural split; Blockchain anonymity; No encryption loss	Relies on blockchain; Synthetic tasks only
[58]	2024	SecurityBERT (BERT)	Edge-IIoTset	Acc: 98.2%; F1: 98.0%; Inf. time: <0.15s	PPFLE privacy encoding; lightweight for IoT	Misclassifies ransomware; Single dataset; Adversarial robustness
[59]	2024	ChatGPT, Llama 2-70B	MAPP, OPP-115, APP-350	F1: up to 93.5%	LLM outperform symbolic/statistical in policy analysis	Overfitting; API/cost constraints; Low determinism
[121]	2025	GPT-2, GPT-3, GPT-4	500 code samples (Python/Solidity)	BLEU: 0.81; Throughput: ↑69-87%	Automated smart contract, optimization for blockchain	Simulated env.; High compute cost; Data variability

Table (3): Summary of GenAI-Based Threat Intelligence Studies

Ref	Year	GenAI Model	Datasets	Key Metric	Contributions	Limitations/ Gaps
[61] [63]	2024 2025	GAN; GAN+XGBoost	Multi-source sim.; Kaggle attacks	Acc: 92.3%; Acc: 99.2%	Auto threat intel; GAN aug. + explainable ML	Simulated results; No enterprise validation
[62]	2025	EAC-GAN	Worldline-ULB fraud	Prec: 96.8%, AUC: 96.4%	Novel EAC-GAN for minority aug.	Little generalizability; No RT testing; Lacks interpretability.
[8]	2024	GAN + VAE	Malware, Phishing datasets	Acc. 93.0-95.0%	Realistic threat gen., reduces blind spots,	High compute demand; Ethical dual-use concerns

					improves detection time	
[64] [122]	2024 2022	VAE-CNN; AdMVAE (VAE-GAN)	Real-time traffic; Fashion/CIFAR/ImageNet	Recall (high); Acc (up to 90.0%)	Real-time anomaly detection; Robustness vs attacks	Interpretability; High inference time
[65]	2023	ND-VAE	MNIST/Fashion	Acc (up to 95.0%)	Noise filtering defense	Scalability issues
[66]	2024	LW-Diff (DDPM-based)	USTC-TFC2016	Acc. 92.3% (vs. Base-Diff 94.0%)	Lightweight model for high-quality synthetic data on edge devices	Slight perf. trade-off, limited attack scope, No RT synthesis
[67]	2025	GPT-3, Llama	CIC-IDS, ISOT-CID	Adaptive resilience	Proactive cloud sec. framework; dynamic IDS	High compute; Large input capacity; Scalability
[68]	2024	Llama-2	N/A	Human & LLM eval.	Dual-LLM system for realism & creativity	Limited expert validation; Data biases; RT integration
[69]	2025	GPT-4 variants	Synthetic charging data	Acc. 77.0%, F1 70.0%	Context-aware anomaly detection; reduces FPs	Sensitive to tuning; Needs multi-agent & FL
[70]	2025	GAN-VAE Integrated	CICIDS2017, Cloud logs	Acc. 97.0%, FPR<5.0%, TTD<15ms	Cloud-native integration; scalable adaptive learning	Computationally intense; Ethical synthetic data issues

Table (4): Summary of GenAI-Based Anomaly Detection Studies

Ref	Year	GenAI Model	Datasets	Key Metric	Contributions	Limitations/ Gaps
[71]	2025	R-GAN	Credit Card Fraud, Ecoli3, Yeast	Acc: 99.5%, F1: 99.5%	Hybrid arch. & AutoML	Recall drops on small data
[72]	2025	LSTM/1D-CNN/ TCN-GAN	Orion, CIC-DDoS2019, CIC-IDS2017	F1: up to 98.6%	Fast DDoS/port scan mitigation	Training instability; Mode collapse
[73]	2024	PCA-GAN	Credit Card Fraud, Ecoli3, Yeast; Kaggle net. logs	Acc: 99.1%, F1: 97.4%	PCA for dim. reduction & low latency	Black-box nature; Compute constraints
[96] [123]	2024 2025	Cond. WGAN-GP; Cloud-GAN	EM signals; Arrhythmia, Cardiotocography	AUC: 99.6%; F1: up 13.0% (e.g., 77.3%)	Synthetic signal gen.; Adaptive loss & cloud-model latent	Single device; Lower gains on low-dim/small data
[124]	2025	IGAN (Inverse GAN)	Tabular benchmarks, billing data	AUROC: 97.7%, AUPRC: 94.2%	Efficient scoring bypassing inversion; Fast inference	Unexplored adversarial robustness; High-dim redundancy
[125]	2025	GAN (TS-GAN)	Real-world (40k records)	Anomaly detection: 25.0% (97.0% high-severity)	GAN for synthetic data + real-time analysis with diff. privacy (ε=1.0).	Occasional misclassification; GAN training overhead.
[126]	2025	WGAN-GP	IES Sim (IEEE 118-bus)	F1: 95.5%	GenAI for RT anomaly detection in cyber-physical energy grids.	Relies on simulation data; may not generalize to unseen attacks.
[75] [127]	2025 2023	VADAD (VAE-SMOTE+Diffusion); Quantum VAE (LWE+RL+QSVD)	ADBench tabular; Quantum-Anomaly net. traffic	AUC-ROC: 92.1%; AUC-ROC: 97.6%	Advanced aug. in latent space; Quantum-resistant secure pipeline	Tabular-only focus; High computations; Needs optimization/validation
[74]	2025	GPT, Llama	Multi-cloud datasets	Acc: 96.3% Latency.: 1.3s/1.6s	Scalable LLM-based detection for multi-cloud	High compute; FP rate variability; needs continuous updates
[128]	2025	Llama 3 Korean Blossom 8B	South Korea National Air Quality Monitoring Network	Acc: 93.0%; Recall 84.0%	Retrieval-augmented anomaly detection, spatial DTW, LLM explanations	Limited to univariate time series; sensitive to initial segmentation; computational overhead

Table (5): Summary of GenAI-Based Phishing And Spamming Studies

Ref	Year	GenAI Model	Datasets	Key Metric	Contributions	Limitations/ Gaps
[87] [88]	2022 2024	LeakGAN+PU+BERT; DCGAN	CODASPY, PhishBench; Kaggle Phishing	F1: 99.6-99.8%; F1: 83.2-83.7%	Adaptive phishing content gen.; Validates synthetic data utility	No email metadata, BERT limits; Feature vector capped
[89]	2025	CTGAN	NSL-KDD	Acc: 98.0-98.3%; KS: 0.92	GAN/XAI framework; High fidelity synthetic data	Binary labels only; Scalability limits
[90]	2024	CGAN	HSPAM, SMS Collection	Acc: 98.2%, F1: 93.0%	GAN oversampling + hybrid classifier	High computational cost
[91] [92]	2023 2025	VAE+DNN; VAE+CNN	ISCX-URL-2016, Kaggle; Labeled/Unlabeled	Acc: 97.5-98.0%; F1: 95.9-97.5%	Automatic latent features; Scalable hybrid framework; Real-time	Latency (~1.9s); Privacy concerns; Lack of explainability
[93] [94]	2024 2025	GPT-3.5/4; PhishEmailLLM (Llama/Gemma/Qwen/Mistral)	Enron, Synthetic; Enron, SpamAssassin, Phishing Pot	Acc: 95.0-96.0%; Prec: up to 99.0%	Dynamic model selection; Hybrid meta-model arch. reducing hallucination	Privacy from internet calls; Reliance on offline, non-fine-tuned models
[95]	2025	ChatGPT 3.5/4.0	Real email	Acc improvement	Real-time spam filtering	Token limits; Prompt design

Table (6): Summary of GenAI-Based Malware Detection Studies

Ref	Year	GenAI Model	Datasets	Key Metric	Contributions	Limitations/ Gaps
[76]	2025	Progressive DCGAN-ZSL	dumpware10, Malimg	Acc: 96.2-98.9%	Image norm.; color mapping; zero-shot for unseen malware	Static images only; High compute; No dynamic analysis
[77]	2024	MCOGAN	Figshare Malware	Acc: 96.0%, 8% improvement	High-fidelity opcode embeddings	High computations, Ignores registers
[78] [79]	2024 2023	CVAE; CVAE+CNN	Malimg; Drebin, AMD	Acc: 98.0%; Macro-F1: 91.0%, Acc: 99.0%	Image synth. for minority balancing; Strong gains in multiclass	Training complexity; Poor aug. for tiny classes; High computational cost
[80]	2023	CopulaGAN + GPT-neo-1.3B	PE Malware Dataset	TPR reduction (GAN: 57.0-83.0%; LLM: 59-6%)	Hybrid static feature integration for diverse adversarial samples	Static analysis only; GAN collapse risk; No dynamic features; Needs fine-tuning
[129] [130]	2024 2021	CVAE, Dense VAE, WGAN-GP; Conv. VAE	Malicia, VirusShare; Microsoft Malware	Min. samples for 95.0% acc; Acc NB: 89.7%	Sys. VAE vs GAN comp.; Latent features for NB/SVM	Limited sample fidelity; Not benchmarked vs non-VAE
[131]	2025	Diffusion	Malicia, VirusShare	F1: 96.0%	AMS arch. with YARA for variants	PE image focus

Table (7): Summary of GenAI-Based Vulnerability Detection Studies

Ref	Year	GenAI Model	Datasets	Key Metric	Contributions	Limitations/ Gaps
[81] [84]	2024 2024	BERT; RoBERTa/ CodeBERT	DiverseVul; Devign/PrimeVul	Acc: 91.9%; F1: up to 97.0%	Holistic transparency; Systematic LLM comp	Context window; Token sensitivity; Overfitting
[82] [83]	2025 2024	GPT-3.5/ChatGLM/ Qwen; GPT-4 + Graphs	SolidiFI; Devign/Reveal/Big-Vul	Rec: 95.1%; F1: 53.8%	Autonomous CoT; Multi-vul detection; Graph code integration	Context window; API costs; Data leakage; C/C++ only
[88]	2024	GPT-3.5Turbo + FNT-BERT-CNN	NVD + Synthetic (6,370 descriptions)	Acc: up to 99.0% (AC, UI, S parameters)	Template-driven generative-discriminative pipeline; Synthetic dataset methodology; Near-perfect CVSS metric prediction	Template dependence; Computational demands; No non-textual features; Needs continuous updates
[97] [86]	2025 2023	Llama/Qwen/ Mistral Ens; GPT-3.5	NVD CVEs; OWASP Benchmark	kNN Acc: ~50.0%; Acc: 77.0%	LLM embedding for semantic grouping; Continuous monitoring	Incomplete classification; Bias; Prompt engineering

Table (8): SWOT Analysis of GenAI Models for Cybersecurity Applications

Model	Strengths	Weaknesses	Opportunities	Threats
GAN	High-fidelity synthetic data generation; Improves detection accuracy beyond 98.0%; Effective for imbalanced datasets; Enhances intrusion detection and phishing detection	Mode collapse; Training instability; High computational cost and overhead; Less effective against DoS attacks; Black-box nature	Federated GAN for privacy-preserving IDS; Quantum GAN for faster convergence; Integration with XGBoost for explainability; Self-attention mechanisms for feature emphasis	Adversarial attacks targeting GAN-based systems; Dual-use for generating malicious content; Evasion techniques exploiting mode collapse
VAE	Stable anomaly detection; Privacy-preserving feature extraction via disentanglement; Effective for phishing URL detection; Good for malware	Lower reconstruction fidelity compared to GAN; High computational expense; Poor augmentation for tiny/negligible malware families; Recall loss in synthetic traffic	Conditional VAE for malware family classification; Hybrid VAE-GAN architectures; VAE-SMOTE combined with diffusion models; Integration with CNN for real-time detection	Privacy leakage from latent vectors; Reconstruction-based attacks; Lack of formal differential privacy guarantees

	classification; Scalable traffic generation			
LLM	Superior semantic understanding for vulnerability detection; Explainable reports mapped to MITRE ATT&CK; High F1-scores (up to 99.2%); Automated patch generation; Contextual threat intelligence	High computational cost; Hallucination risks; API cost dependency; Prompt sensitivity; Limited context windows; Data leakage risk; Overfitting to single datasets	Federated LLM with RAG for intrusion detection; Retrieval-augmented generation for threat intelligence; Autonomous security agents; Multi-agent systems for proactive defense	Adversarial and prompt-based attacks reducing F1 by up to 50.0%; Privacy risks from internet-dependent API calls; Data leakage through model inversion; Dependency on non-fine-tuned models
Diffusion Models	State-of-the-art synthetic data quality; Stable training; No mode collapse; High accuracy (up to 99.9% F1); Lightweight variants available for edge devices	Very high inference latency; No drift adaptation; Poor for real-time deployment on resource-constrained devices; High computational requirements for training and inference	Lightweight diffusion models for edge devices; Privacy-preserving anonymous diffusion with blockchain; Adversarial defense; Privacy-preserving synthetic data generation	Real-time deployment infeasibility on constrained IoT/edge devices; High inference latency preventing time-critical

Table 8 summarizes the key trade-offs of each of the GenAI categories. While GAN and Diffusion Models exhibit high data quality, they come with certain constraints on stability/latency. On the other hand, VAE and LLMs showcase excellence in stability/semantics, but they fall short on efficiency/fidelity. It is this very point that dictates the efficacy-performance trade-off model suggested by this paper.

5. Meta-Analysis, Results, and Discussion

The literature reviewed collectively places more emphasis on accuracy and F1-score as the main metrics, which can disadvantageously report computational cost, inference latency, or even real-time feasibility. This bias leads to a positive publication bias favoring high-performing, but computationally expensive models, which can give a false indication of deployment readiness to practitioners.

5.1 Yearly Publication Trends

Publication trends show an initial exploratory phase from 2021, focused on foundational models and proof-of-concept applications. This expanded around 2023 into diverse applications in detection, privacy, and anomaly challenges. Recent years to 2025 reflect maturing research emphasizing practical deployments, hybrid models, and addressing scalability and ethics. The trend in Figure (2) follows a canonical maturation path, including an exploratory phase (2021-2022), a fast growth phase (2023), and a consolidation phase where sophisticated AI-based cybersecurity solutions are going to be created (2024-2025). Finally, and given that we are at early 2026, related research output is very low. This trend reflects the field's growing academic interest and maturity.

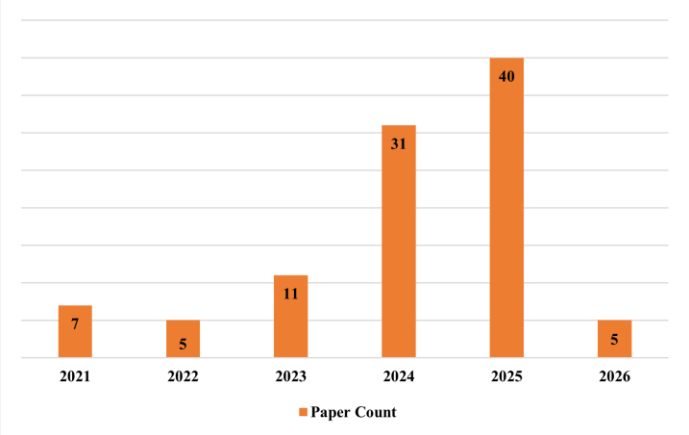

Figure (2): Annual Publication Trends

5.2 Research Trends and Model Evolution

The distribution across cybersecurity application domains in Figure 3 shows that intrusion detection is the most active research area, while emerging applications of vulnerability detection show increasing interest. Figure (3) visualize the distribution of GenAI-powered cyber defense domains with total works appear next to each legend.

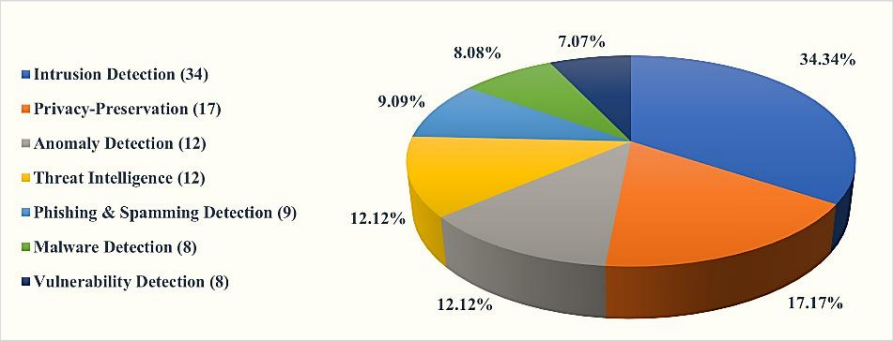

Figure (3): Distribution of GenAI-Powered Cyber Defense Domains

GAN dominate cybersecurity applications as their adversarial training aligns naturally with threat modeling and detection. LLM are growing rapidly from 2023 onward, powered by transformer advances and large datasets, achieving success in code analysis and automated threat intelligence. VAE significantly contributes to anomaly and malware detection through probabilistic modeling. Diffusion models are becoming a promising method for synthetic data generation, which offers greater diversity, privacy, and strength, especially in resource-limited settings. Such a landscape confirms the key position of GAN and increases the role of big LLM and diffusion models. The GenAI model distribution was presented in Figure (4) and shows the number of works per type of the models in the legend.

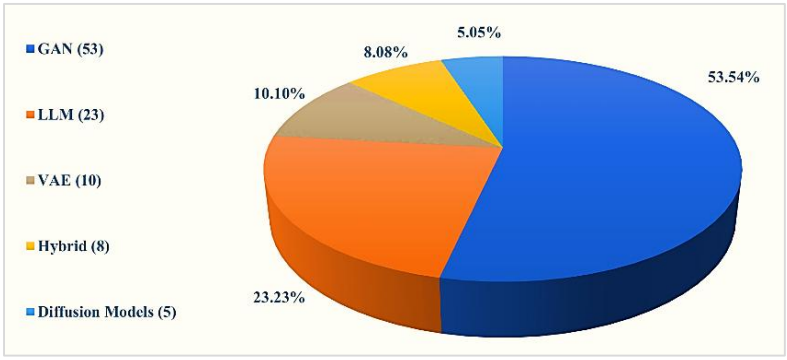

Figure (4): Distribution of GenAI Models

5.3 Cybersecurity Issues and Defense Mechanisms

GenAI models are exposed to a wide range of cyber threats, including polymorphic malware, zero -day, Advanced Persistent Threat (APT), and privacy issues. The principal applications of their particular aptitudes consist of the following:

- GAN help in reducing the problem of data shortage by adversarial synthesis.
- VAE uses normalcy learning to find anomalies.
- Diffusion models can be used to create realistic cyber traffic that can support privacy-sensitive uses.
- LLM aid improve automated vulnerability detection and security policy formulation.

5.4 Deployment Challenges and Emerging Opportunities

Nevertheless, despite the significant developments, serious issues exist regarding the stable training of models, the practicability of real-time implementation, scalability, and uniformity in benchmarking. New hybrid architectures and dynamic training strategies seem to be the key to reducing these barriers.

5.5 Performance and Efficiency Analysis

Table (9) shows a comparative analysis of the GenAI models in terms of several key performance indicators, including accuracy in detection, latency, complexity of training, and consumption of computing resources. GAN have higher detection accuracy, but high computational cost and sensitivity to training dynamics prevent their use. VAE can provide fairly efficient and consistent inference at the cost of significant sacrifices in

reconstruction fidelity. Diffusion models and LLM are often notably more profitable and resilient overall; however, both require significant investments and resources, thus limiting their functionality in real-time working settings.

Table (9): Comparative GenAI Model Performance Benchmarking

GenAI Model	Primary Strength	Primary Limitation	Best-Suited Application
GAN	Extremely real in the artificial data (images, traffic).	Mode collapse (instability), and high computational overhead.	Data augmentation and adversarial attack simulation.
VAE	Constant training of the stable conditions and the effective baseline learning of latent representation.	Lower-fidelity and blurrier outputs compared to GAN.	Anomaly detection amd feature extraction applications
LLM	Higher semantic interpretation of a programming code as well as natural language.	Increased demand of resources with hallucination.	Vulnerabilities identification, reports generation, and analysis of phishing cases.
Diffusion Models	State-of-the-art in terms of sample quality and diversity.	The training and inference processes are significantly heavy in terms of the computational load involved..	Synthetic data generation of high fidelity for privacy-preservation and cyber threat detection.

Table (10) also clarifies the concept of computational efficiency and operating trade-offs, with the emphasis on the necessary balance between more accurate detection performance on the one hand and realistic deployment constraints on the other hand.

Table (10): Computational Efficiency of GenAI Models

GenAI Model	Training Cost	Inference Cost	Example Efficiency Gain
GAN	High	Medium-High	~25% faster training (AE-WGAN)
VAE	Medium-High	Medium	Limited gains
LLM	Very High	High	<0.15s inference (SecurityBERT)
Diffusion Models	Very High	Very High	80% MAC reduction (LW-Diff)
Hybrid	High	High	Varies widely

5.6 Identification of Cross-Cutting Trends and Gaps

The emergence of GenAI creates systemic changes in the sphere of cybersecurity and introduces new defensive strategies, which is accompanied by an explanation of persistent gaps. The discussion that follows outlines significant emerging patterns and existing issues that form the basis of modern research directions.

Pervasive Trends

The Data Imbalance Solution: In all subfields of cybersecurity, there has been a consistent motif of the application of GenAI models, especially GAN and VAE, as a primary solution to the classical issue of class imbalance.

- The Data Imbalance Solution: In all sub fields of cybersecurity, there has been a consistent motif of the application of GAI models, especially GANs and VAEs, as a primary solution to the classical issue of class imbalance.
- The Empowerment of Hybrid Constructions: A growing direction is associated with the notion of hybrid architectures (e.g., GAN-VAE) designed and shaped to combine the capabilities of the heterogeneous models in a bid to deal with their distinct weaknesses.

Critical Research Gaps

Dataset Obsolescence: Most modern studies still use older standards, like NSL-KDD and KDDCUP99, thus strongly limiting the applicability of the results to the real-world context.

- No time like the Real Time Deployment Chasm: Many papers do not talk about the computational latency part but show high accuracies but fail to illustrate the capability of actual real-time performance in resource-constrained hardware environments, such as on edge devices.
- Provided there is no set of standard measures and evaluation criteria, it is difficult to compare the results of studies directly, which discourages the use of standard evaluation measures on the community level.

5.7 The Trade-Off between Efficacy and Performance

The major meta-level lesson that the synthesis comes out with is the presence of a universal performance–efficacy trade-off. There is a strong inverse relationship between the attained accuracy or fidelity of a mode and the practicality of its operation:

- High-efficacy, low-performance, like high-performance models, these achieve state-of-the-art performance, such as diffusion models or LLM, and are, as a result, very expensive in terms of computational intensity; therefore, they cannot be deployed in real-time.
- Lightweight VAE or distilled LLM are built to be efficient and can thus be expected to yield a clearly noticeable but acceptable drop in performance in exchange for faster inference time at a reduced set of resources.

This trade-off is the key strategic dilemma discovered in the existing literature and used to make a wise choice of a suitable generative model applicable to a given cybersecurity operational situation.

5.8 Outlining Reactive and Proactive GenAI Use in Cybersecurity

Although a significant portion of the current studies in the field of GenAI as applied to cybersecurity are dedicated to enhancing reactive capabilities, including better intrusion detection, anomaly detection, and malware classification, an important and more recent trend is the shift to proactive defense strategies. Reactive applications largely respond to a threat after it has been launched to detect or prevent it with the help of AI-enhanced anomaly recognition and synthetic data augmentation methods.

Active GenAI applications, in comparison, aim to forecast, imagine, and avoid cyber threats before they occur in the field of operation. These include:

- Automated vulnerability detection and patch generation, in which generative models of the type of LLM and hybrid pipelines are used to predict and generate fixes to software vulnerabilities.
- Generated adversarial attack scenario with GAN and diffusion models to capture realistic threat behaviors and continuously adapt the defense mechanisms.
- Self-learning and self-defending systems that enable self-defense systems to learn continuously and respond in a fashion that does not result in human intervention but rather increases the resilience of the system.
- Privacy-conservative collaborative intelligence sharing systems that can be used to share intelligence among various parties to make joint predictions and mitigate threats without data loss.

The shift of AI to being an autonomous, anticipatory security framework rather than an AI-based reactive augmentation tool is a paradigm shift in AI. To explain this shift and define the variety of GenAI uses, Table (11) below compares reactive and proactive applications, showing the goals, model design, results, and effect of operation

Table (11): Reactive vs. Proactive GenAI Applications

Aspect	Reactive GenAI	Proactive GenAI
Primary Objective	Threat detection and classification post-occurrence	Threat anticipation, prevention, and autonomous mitigation
Typical Technologies	GAN, VAE, diffusion models for detection and anomaly synthesis [34], [35], [37], [71], [72],[112]	LLM, hybrid generative-discriminative pipelines, autonomous agents [66], [82], [83]
Operational Impact	Improved incident response time and accuracy	Reduced attack surfaces, automated patching, dynamic adaptation
Example Applications	Intrusion detection, malware classification, phishing detection [39], [42], [62], [63]	Vulnerability generation/scoring, attack simulation, adaptive defense [66], [81], [84], [86]
Data Requirements	Historical attack data and network logs	Synthetic adversarial environments, real-time learning feedback

6. Challenges and Future Directions

The course toward the application of GenAI to the area of cybersecurity is full of major challenges that outline clear, concentrated opportunities for further investigation. These limitations found in the existing scholarship directly require and activate the strategic agenda of the discipline.

6.1 Hardware Requirements and Computational Costs

The hardware requirement is not consistently reported in the existing literature. The data in Table 12 were taken out of 25 sources. The findings below illustrate some key trends despite some shortcomings in the reporting of the data.

Table (12): Hardware Requirements and Computational Costs of Selected GenAI Models

GenAI Model	Model/ Variant	Ref.	CPU	GPU	Training Time	Inference Latency
GAN	GAN (GPIDS)	[40]	Intel i5 (12 cores)	RTX 3060	~2-4 sec per attack	Not reported
	GAN (FEDGAN-IDS)	[42]	Intel i7-10750H	RTX 2060 (4GB)	20 global epochs	Not reported
	GAN (SGAN-IDS)	[110]	Intel i9-7900X	Not specified	10,000 epochs	Not reported
	GAN (EAC-GAN)	[62]	Intel i5-12600KF	GTX 4060Ti	10,000 epochs	Not reported
	GAN (MCOGAN)	[77]	Intel Xeon Gold 6148	TITAN XP	Not specified	Not reported
	GAN (R-GAN)	[71]	AMD Ryzen 5 5600X	Not specified	10,000 epochs	Not reported
	GAN (Cloud-GAN)	[123]	Not specified	Tesla V100	1,000 epochs	Not reported
	GAN (T-GAN)	[72]	Intel Xeon Platinum 8352V	Dual v-GPU (32GB)	100 epochs	Not reported
	Side Channel GAN	[96]	Arduino Mega / ATmega2560 MCU	None	Not reported	Not reported
	GAN-AIIPot	[45]	Not specified (Google Cloud)	Not specified	2.5 hours	Not reported
	GAN (Energy Systems)	[126]	Intel Xeon E5-2603 v3	Not specified	35 epochs	Not reported
VAE	VAE (CVAE)	[79]	2x Intel Xeon Gold 6230	4x RTX 2080 Ti	Not specified	Not reported
	VAE	[91]	Not specified	Tesla V100	268 seconds	Not reported
LLM	LLM (SecurityBERT)	[58]	Intel Xeon @ 2.20GHz	A100 (40GB)	1 hour 47 min	<0.15 sec
	LLM (SmartGuard)	[82]	Intel i7-10700K	RTX 3090 (24GB)	Not specified	Not reported
	LLM (CodeLlama-7b)	[84]	Intel Xeon Silver 4210	Tesla V100-32GB	3 epochs (LoRA)	Not reported
	LLM-APTDS	[49]	Intel Core2 Duo T7700	VMware virtual GPU	2-4 days	Not reported
	LLM (IoE)	[114]	Apple M4 Max (16-core)	Apple M4 Max (40-core)	Not reported	0.03s/flow
	LLM+m	[48]	Apple M4 Max (16-core)	Apple M4 Max (40-core)	Not specified	Not reported
Diffusion Models	Diffusion (LW-Diff)	[66]	Intel i7-14700K	RTX 4070	Not specified	Not reported
	Diffusion (NT-DDPM)	[98]	Intel Xeon Platinum 8168	4x A40	20 epochs	High latency
	Diffusion (Anonymous)	[57]	AMD Ryzen 9 7950X3D	RTX 4070 Ti	Inference only	5.53 sec
	Diffusion (Privacy-Diff)	[120]	AMD Ryzen 9 7950X3D	RTX 4080 SUPER	Inference only	13.94 sec
Hybrid	Hybrid (VAE-WGAN)	[50]	Intel i7-9570H	GTX 1660Ti	VAE: 6,000 cycles; LSTM: 2,000 cycles	Not reported
	Hybrid (VAE+Diffusion)	[75]	Not specified	RTX 4090	VAE: 5,000 epochs; Diffusion: 5,000 epochs	Not reported

Training time and inference latency are not mentioned in most studies. Of the ones that do, GANs frequently require thousands of epochs, LLMs can run on high-end GPUs and can run at low latency (less than 0.15 sec), and diffusion models can have high inference latency (5-14 seconds), so they are hard to deploy in real-time.

6.2 Fundamental Challenges and Corresponding Research Opportunities

The mixture of operational and technical issues hinders the implementation of GenAI. To overcome such limitations, specific studies are required that go beyond the area of improving the work of the models and introduce improvements to their computational efficiency, robustness, and full compatibility in real-world conditions.

- **Bringing the Compute-Hungry GenAI to the Edge Efficiently**
  - Challenge: GenAI models highly consume computational resources, making the training stages costly and making real-time inference on edge devices difficult.
  - Research Opportunity: The greatest need is to create lightweight GenAI structures using approaches like pruning and quantization and simultaneously come up with training algorithms with increased stability and efficiency, especially when operating in an Internet of Things (IoT) or edge scenario.
- **Data Problems and The Movement towards Modernization**
  - Threat: GenAI systems rely on outdated datasets, ineffective quality synthetic data of very rare threats, and phenomena like model collapse.
  - Research Opportunity: The expertise of these shortcomings is the development of modern, realistic data warehouse systems that can replace outdated standards, as well as the development of improved evaluation measures that could be used to determine the quality of synthetic data.
- **Generalization Gaps and Model Instability**
  - Problem: Unstable training systems, and the models arising from them cannot transfer to new domains or result in changes in threat space without retraining and thus cannot effectively cope with concept drift.
  - Research Opportunity: The shortcoming prompts the creation of cross-domain adaptation methods, long-term learning systems, and investigation of hybrid models based on the combination of GANs and VAEs. These are to enhance stability and better generalization.
- **Essentially, there are security, ethical, and operational risks**

  - Issue: GenAI is a technology with dual-use potential, and its models can be manipulated adversarial. The implementation of this technology in current work tools is a complicated procedure. These systems are opaque, which undermines the confidence of an analyst.
  - Research Opportunity: These issues highlight the importance of implementing different privacy mechanisms and secure multi-party computation in order to protect the confidentiality of data and architectural solutions to create more reliable and confidence-based systems.
- **Lack of Standardized Hardware Reporting**
  - Challenge: From Table 12, just 25 out of 99 articles mention hardware configurations. The training duration is not reported or is expressed in terms of different measures, like epochs, cycles, or seconds. The inference latency, an essential indicator for real-time implementations, is measured in just two papers. However, the area lacks standard methodologies to evaluate system efficiency and feasibility simultaneously.
  - Research Opportunity: There is an urgent need for the establishment of a minimum standard set of metrics in terms of hardware requirements to facilitate a comparison of different GenAI security systems and realistic deployment assessment. Such metrics should include: CPU/GPU model and memory; training time in hours/seconds; and inference latency per data sample. Furthermore, there is an urgent need for robust benchmarking to match practical implementation scenarios (edge devices, network constraints, adversarial environments).

6.3 Strategic Future Directions

The aforementioned research opportunities enable a strategic advancement of isolated tools into the amalgamation of systems that partake in the proactive development. The field objectives to be used are:

- **Relocating defenses and deploying lingering defenses with greater autonomy:** Supporting real-time threat simulation, automated response, and steady evolution of defense mechanisms with autonomous cyber defense. As an example, recent works on GAN and diffusion models prove the ability to produce real-world attack settings, hence warning about adaptive defensive procedures.
- **Privacy-Preserving Collaborative Defense:** To obtain the benefits of data sharing securely for threat intelligence by promoting the adoption of federated learning methods and differential privacy techniques that may be inspired by new findings in privacy-preserving GAN and VAE. These designs solve privacy-utility tradeoffs, and other important processes in sharing information on threats, which multi-party cooperation needs.
- **Availability of AI Security Software:** Democratize high-end defense solutions by offering cloud-based GenAI services (AI-as-a-Service) to facilitate synthetic data creation and threat analysis that relies on the current models of malware detection and vulnerability detection that take advantage of LLM and diffusion algorithms.
- **Explainable AI Operations (XAI-Ops):** Trying to instill explain-ability techniques (SHAP, LIME, attention heatmaps, etc.) into AI-based security processes to make them more trustworthy, interpretable, and viewable. Such uses as vulnerability detection, vulnerability scoring, and intrusion detection are examples of parallel critiques that stress the challenge of the need to have clear adversarial insights.
- **Generating frauds, malware, and intrusion attempts:** Integrating multi-modal data synthesis (e.g., GAN, VAE, diffusion models) with interpretable machine learning and formal privacy guarantees to simulate, detect and respond to complex and evolving threats, including phishing, malware, and obstructions, among others.
- **Accentuate Scalability and Real-Time Deployment:** Generally, make efficient, low-latency-based models, such as any of the variants of lightweight diffusion or GAN, and resource-aware models that can execute at the edge or in a decentralized setup, informed by current research on edge-compatible lightweight diffusion models.

It will be essential to overcome these obstacles by tackling the challenges with the use of multidisciplinary approaches that will help achieve maximum benefits of GenAI as a disruptive tool in proactive, resilient cybersecurity protection. In addition, the rapid change in GenAI makes it important to have a dynamic and living review framework that is constantly updated with new findings and new technology so as to remain relevant and guide researchers and practitioners to work within this rapidly evolving field.

7. Conclusions

The review summarizes 99 peer-reviewed articles (2021-2026) on generative artificial intelligence (GenAI) applications in cyber defense, including generative adversarial networks (GANs), variational autoencoders (VAE), large language models (LLM), and diffusion models in seven application areas: intrusion detection, malware analysis, privacy preservation, threat intelligence, anomaly detection, vulnerability detection, and phishing detection.

The results show that GenAI allows making a paradigm shift to proactive security and improving the threat detection, data augmentation and the automated response considerably. GANs can generate synthetic data with high-fidelity, enhancing the accuracy of detection to over 98.0% in a majority of intrusion detection and phishing tasks, but are unstable in training and have high computational cost. VAE provide more stable anomaly detection and feature extraction that are privacy-preserving, but with lower reconstruction fidelity. LLM offer better semantic perception to vulnerability detection and threat intelligence, but need extensive computing resources and are subject to hallucination. Diffusion models achieve high-quality synthetic data with stable training, yet due to its high inference latency, prevents real-time deployment on resource-constrained devices.

The result of this synthesis is a persistent performance-efficacy trade-off: high-accuracy models (diffusion, LLM) require large amounts of computer resources, whereas lightweight models (VAE, distilled LLM) can make compromises on accuracy to be run in real-time. This is the trade-off that is the key strategic issue of practitioners who choose GenAI models in particular operation situations.

There are major gaps in the literature. First, the majority of research is based on old data sets (e.g., NSL-KDD, KDDCUP99), which restricts their application in the real world. Second, edge or IoT hardware validation is not commonly done in real-time, even when there is high reported accuracy. Third, the field lacks standardized benchmarks and evaluation metrics, preventing direct cross-study comparison. Fourth, instability of models (especially mode collapse in GAN) and researchers consistently underreport computational expenses.

Resting on these results, we suggest a research strategy roadmap that focuses on autonomous cyber defense systems that can simulate and respond to threats in real-time; privacy-preserving federated learning models that can support secure cross-organizational sharing of threat intelligence; explainable AI operations (XAI-Ops), which can be integrated into security processes with techniques like SHAP and LIME; and lightweight, edge-deployable models. Limitations of this review include its database scope (Scopus, ScienceDirect, Springer), which may exclude relevant work from other repositories, and its temporal focus (2021–2026), which reflects the rapid evolution of the field.

Acknowledgement: The authors acknowledge the College of Computer Science and Mathematics of the University of Mosul that offered the required facilities.

Conflict of Interest: The authors declare that there are no conflicts of interest associated with this research project. We have no financial or personal relationships that could potentially bias our work or influence the interpretation of the results.

References

[1] V. T. Truong, L. B. Dang, and L. B. Le, "Attacks and defenses for generative diffusion models: A comprehensive survey," *ACM Comput. Surv.*, vol. 57, no. 8, Art. no. 216, p. 216, 2025, doi: 10.1145/3721479.

[2] W. Ibrar, Author, Author, Author, Author, and Author, "Generative AI: A double-edged sword in the cyber threat landscape," *Artif. Intell. Rev.*, vol. 58, p. 285, 2025, doi: 10.1007/s10462-025-11285-9.

[3] S. Ayyaz and S. M. Malik, "A comprehensive study of generative adversarial networks (GAN) and generative pre-trained transformers (GPT) in cybersecurity," in *Proc. Int. Conf. Intell. Comput. Data Sci. (ICDS)*, 2024, pp. 1–8, doi: 10.1109/ICDS62089.2024.10756505.

[4] J. L. L. Delgado and J. A. L. Ramos, "A comprehensive survey on generative AI solutions in IoT security," *Electronics*, vol. 13, no. 24, p. 4965, 2024, doi: 10.3390/electronics13244965.

[5] F. Li, I. Udoidiok, and J. Zhang, "A Comprehensive Study on Lightweight Convolution Techniques for Malicious Traffic Synthesis in Diffusion Models," in *IEEE Access*, vol. 13, 2025, pp. 88306–88317.

[6] W. Kasri, Author, Author, Author, Author, and Author, "From vulnerability to defense: The role of large language models in enhancing cybersecurity," *Computation*, vol. 13, no. 2, p. 30, 2025, doi: 10.3390/computation13020030.

[7] M. Uddin, Author, Author, Author, Author, and Author, "Generative AI revolution in cybersecurity: A comprehensive review of threat intelligence and operations," *Artif. Intell. Rev.*, vol. 58, p. 236, 2025, doi: 10.1007/s10462-025-11219-5.

[8] R. Vadisetty and A. Polamarasetti, "Generative AI for cyber threat simulation and defense," in *Proc. 12th Int. Conf. Control, Mechatronics Autom. (ICCMA)*, 2024, pp. 272–277, doi: 10.1109/ICCMA63715.2024.10843938.

[9] P. Radanliev, O. Santos, and U. D. Ani, "Generative AI cybersecurity and resilience," *Front. Artif. Intell.*, vol. 8, p. 1568360, 2025, doi: 10.3389/frai.2025.1568360.

[10] M. Gupta, C. Akiri, K. Aryal, E. Parker, and L. Praharaj, "From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy," *IEEE Access*, vol. 11, pp. 80218–80245, 2023, doi: 10.1109/ACCESS.2023.3300381.

[11] M. Alanezi and R. M. A. Al-Azzawi, "AI-Powered Cyber Threats: A Systematic Review," *Mesopotamian J. Cybersecurity*, vol. 4, no. 3, pp. 166–188, 2024, doi: 10.58496/MJCS/2024/021.

[12] L. O. Ofusori, T. Bokaba, and S. Mhlongo, "Artificial intelligence in cybersecurity: A comprehensive review and future direction," *Applied Artificial Intelligence*, vol. 38, no. 1, p. 2439609, 2024, doi: 10.1080/08839514.2024.2439609.

[13] M. Ring, S. Wunderlich, D. Scheuring, D. Lands, and A. Hotho, "A survey of network-based intrusion detection data sets," *Computers & Security*, vol. 86, pp. 147–167, 2019, doi: 10.1016/j.cose.2019.06.005.

[14] I. J. Goodfellow, Author, Author, Author, Author, and Author, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, vol. 27, pp. 2672–2680, doi: 10.5555/2969033.2969125.

[15] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. Int. Conf. Learn. Represent (ICLR)*, 2014, doi: 10.48550/arXiv.1312.6114.

[16] L. Xu, Author, Author, Author, Author, and Author, "G-VAE: Variational autoencoder-based adversarial attacks and defenses in industrial control systems," *Computers and Electrical Engineering*, vol. 124, p. 110290, 2025, doi: 10.1016/j.compeleceng.2025.110290.

[17] A. Vaswani, Author, Author, Author, Author, and Author, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, doi: 10.5555/3295222.3295349.

[18] F. Chiarello, Author, Author, Author, Author, and Author, "Future applications of generative large language models: A data-driven case study on ChatGPT," *Technovation*, vol. 133, 2024, doi: 10.1016/j.technovation.2024.103002.

[19] H. Pearce, Author, Author, Author, Author, and Author, "Examining zero-shot vulnerability repair with large language models," in *Proc. 44th Int. Conf. Softw. Eng. (ICSE)*, 2022, pp. 1490–1501, doi: 10.1109/SP46215.2023.10179420.

[20] J. Sohl-Dickstein, Author, Author, Author, Author, and Author, "Deep unsupervised learning using nonequilibrium thermodynamics," ed, 2015.

[21] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 8780–8794, 2021, doi: 10.5555/3540261.3540933.

[22] T. Karras, M. Aittala, S. Laine, and T. Aila, "Elucidating the design space of diffusion-based generative models," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 26565–26577, 2022, doi: 10.5555/3600270.3602196.

[23] M. Al-Ajlan and M. Ykhlef, "A review of generative adversarial networks for intrusion detection systems: Advances, challenges, and future directions," *Comput., Mater. Contin.*, vol. 81, no. 2, pp. 2053–2095, 2024, doi: 10.32604/cmc.2024.055891.

[24] S. R. Sindiramutty, Author, Author, Author, Author, and Author, "Generative AI in network security and intrusion detection," in *Advanced Applications of Generative AI and Natural Language Processing Models*, *IGI Global*, 2025, pp. 77–123.

[25] Z. Deng, Author, Author, Author, Author, and Author, "Generative AI in intrusion detection systems for Internet of Things: A systematic literature review," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 4689–4715, 2025, doi: 10.1109/OJCOMS.2025.3573194.

[26] M. Andreoni, Author, Author, Author, Author, and Author, "Enhancing autonomous system security and resilience with generative AI: A comprehensive survey," *IEEE Access*, vol. 12, pp. 109470–109495, 2024, doi: 10.1109/ACCESS.2024.3439363.

[27] O. Emehin, Author, Author, Author, Author, and Author, "Generative AI in forensic data analysis: Opportunities and ethical implications for cloud-based investigations," *Int. J. Res. Publ. Rev.*, vol. 5, no. 10, pp. 2941–2957, 2024, doi: 10.55248/gengpi.5.1024.2904.

[28] M. A. Ferrag, Author, Author, Author, Author, and Author, "Generative AI in cybersecurity: A comprehensive review of LLM applications and vulnerabilities," *Internet Things Cyber-Phys. Syst.*, vol. 5, pp. 1–46, 2025, doi: 10.1016/j.iotcps.2025.01.001.

[29] S. Sai, Author, Author, Author, Author, and Author, "Generative AI for cyber security: Analyzing the potential of ChatGPT, DALL-E, and other models for enhancing the security space," *IEEE Access*, vol. 12, pp. 53497–53522, 2024, doi: 10.1109/ACCESS.2024.3385107.

[30] H. Alqahtani and G. Kumar, "A comprehensive review of generative AI techniques and their impact on cybersecurity," *Soft Comput.*, vol. 29, pp. 4945–4982, 2025, doi: 10.1007/s00500-025-10702-z.

[31] S. L. Mirtaheri, N. Movahed, R. Shahbazian, V. Pascucci, and A. Pugliese, "Cybersecurity in the age of generative AI: A systematic taxonomy of AI-powered vulnerability assessment and risk management," *Future Generation Computer Systems*, vol. 175, p. 108107, 2026/02/01/ 2026, doi: https://doi.org/10.1016/j.future.2025.108107.

[32] E. Hilario, Author, Author, Author, Author, and Author, "Generative AI for pentesting: The good, the bad, the ugly," *Int. J. Inf. Secur.*, vol. 23, pp. 2075–2097, 2024, doi: 10.1007/s10207-024-00835-x.

[33] K. Curran, Author, Author, Author, Author, and Author, "The role of generative AI in cyber security," *Metaverse*, vol. 5, no. 2, p. 2796, 2024, doi: 10.54517/m.v5i2.2796.

[34] G. Andresini, Author, Author, Author, Author, and Author, "GAN augmentation to deal with imbalance in imaging-based intrusion detection," *Future Gener. Comput. Syst.*, vol. 123, pp. 108–127, 2021, doi: 10.1016/j.future.2021.04.017.

[35] X. Zhao, K. W. Fok, and V. L. L. Thing, "Enhancing network intrusion detection performance using generative adversarial networks," *Computers & Security*, vol. 145, p. 104005, 2024, doi: 10.1016/j.cose.2024.104005.

[36] H. Zouhri and A. Idri, "A novel CTGAN-ENN hybrid approach to enhance the performance and interpretability of machine learning black-box models in intrusion detection and IoT," *Future Gener. Comput. Syst.*, vol. 173, p. 107882, 2025, doi: 10.1016/j.future.2025.107882.

[37] M. Arafah, I. Phillips, A. Adnane, M. Alauthman, and N. Aslam, "Anomaly-based network intrusion detection using denoising autoencoder and Wasserstein GAN synthetic attacks," *Applied Soft Computing*, vol. 168, p. 112455, 2024, doi: 10.1016/j.asoc.2024.112455.

[38] M. S. Alshehri, Author, Author, Author, Author, and Author, "A hybrid Wasserstein GAN and autoencoder model for robust intrusion detection in IoT," *Computer Modeling in Engineering & Sciences*, 2025, doi: 10.32604/cmes.2025.064874.

[39] S. Li, Author, Author, Author, Author, and Author, "HDA-IDS: A Hybrid DoS Attacks Intrusion Detection System for IoT by using semi-supervised CL-GAN," *Expert Systems With Applications*, vol. 238, p. 122198, 2024, doi: 10.1016/j.eswa.2023.122198.

[40] J. Qin, Author, Author, Author, Author, and Author, "GPIDS: GAN assisted contextual pattern-aware intrusion detection system for IVN," *IEEE Trans. Veh. Technol.*, vol. 73, no. 9, 2024, doi: 10.1109/TVT.2024.3383449.

[41] J. D. Yoo, H. Kim, and H. K. Kim, "GUIDE: GAN-based UAV IDS Enhancement," *Comput. Secur.*, vol. 147, p. 104073, 2024, doi: 10.1016/j.cose.2024.104073.

[42] A. Tabassum, Author, Author, Author, Author, and Author, "FEDGAN-IDS: Privacy-preserving IDS using GAN and Federated Learning," *Comput. Commun.*, vol. 192, pp. 299–310, 2022, doi: 10.1016/j.comcom.2022.06.015.

[43] F. Cirillo and C. Esposito, "Intrusion detection using quantum generative adversarial networks: a federated approach with noisy simulators," in *IET Space and Communications Conference 2025*, 2025, pp. 31–35, doi: 10.1049/icp.2025.2226.

[44] M. Sekhar, V. Alukapelly, T. Neelima, R. Kotoju, and V. Veerabhadram, "Generative Adversarial Networks for Cyber Threat Simulation and Defence Strategies," *Journal of Theoretical and Applied Information Technology*, vol. 103, no. 4, pp. 1388–1400, 2025.

[45] V. S. Mfogo, A. Zemkoho, L. Njilla, M. J. Nkenlifack, and C. A. Kamhoua, "GAN-AIIPot: GAN-Based Cyber Deception for Probing Attacks on IoT Devices," in *IEEE Transactions on Network and Service Management*, vol. 23, 2026, pp. 417–431.

[46] F. S. Melícias, T. F. R. Ribeiro, C. Rabadão, L. Santos, and R. L. D. C. Costa, "GPT and Interpolation-Based Data Augmentation for Multiclass Intrusion Detection in IIoT," in *IEEE Access*, vol. 12, 2024, pp. 17945–17965.

[47] P. Fernández-Saura, J. Bernal-Bernabé, and A. Skarmeta-Gómez, "Enhancing federated intrusion detection through LLM-Driven alert enrichment and collaborative threat information sharing," *Future Gener. Comput. Syst.*, vol. 178, 2026, doi: 10.1016/j.future.2025.108319.

[48] R. H. Hwang, Y. H. Hsiao, Y. D. Lin, and Y. C. Lai, "LLM+m: Dual-model chatGPT-based product training and testing with adversarial attack and defense," *Future Generation Computer Systems*, vol. 179, p. 108328, 2026, doi: 10.1016/j.future.2025.108328.

[49] L. Yang, A. Ye, Y. Liu, W. Lu, and C. Huang, "LLM-APTDS: A high-precision advanced persistent threat detection system for imbalanced data based on large language models with strong interpretability," *Future Generation Computer Systems*, vol. 178, p. 108315, 2026, doi: 10.1016/j.future.2025.108315.

[50] Z. Li, C. Huang, and W. Qiu, "An intrusion detection method combining variational auto-encoder and generative adversarial networks," *Computer Networks*, vol. 253, p. 110724, 2024, doi: 10.1016/j.comnet.2024.110724.

[51] Y. Wan, Y. Qu, L. Gao, and Y. Xiang, "Differentially Privacy-Preserving Federated Learning Using Wasserstein Generative Adversarial Network," in *2021 IEEE Symposium on Computers and Communications (ISCC)*, 2021, pp. 1–6, doi: 10.1109/ISCC53001.2021.9631541.

[52] A. Kotal and A. Joshi, "Differentially Private Synthetic Data Generation Using Context-Aware GANs," in *2024 IEEE International Conference on Big Data (BigData)*, 2024, pp. 6289–6297, doi: 10.1109/BigData62323.2024.10826047.

[53] S. Han, Author, Author, Author, Author, and Author, "Fed-GAN: Federated generative adversarial network with privacy-preserving for cross-device scenarios," *IEEE Trans. Dependable Secure Comput.*, 2025, doi: 10.1109/TDSC.2025.3567564.

[54] A. Torfi, E. A. Fox, and C. K. Reddy, "Differentially private synthetic medical data generation using convolutional GANs," *Inf. Sci.*, vol. 586, pp. 485–500, 2021, doi: 10.1016/j.ins.2021.12.018.

[55] H. Montenegro and J. S. Cardoso, "Anonymizing medical case-based explanations through disentanglement," *Medical Image Analysis*, vol. 95, p. 103209, 2024, doi: 10.1016/j.media.2024.103209.

[56] G. Aceto, Author, Author, Author, Author, and Author, "Synthetic and privacy-preserving traffic trace generation using generative AI models for training network intrusion detection systems," *J. Netw. Comput. Appl.*, vol. 229, p. 103926, 2024, doi: 10.1016/j.jnca.2024.103926.

[57] P. C. Hsu, Z. Yu, N. Ghafoori, S. Mise, and H. Miyaji, "Anonymous-Diffusion: Blockchain-Based Privacy-Preserving Stable Diffusion," in *2025 1st International Conference on Consumer Technology (ICCT-Pacific)*, 2025, pp. 1–4, doi: 10.1109/ICCT-Pacific63901.2025.11012859.

[58] M. A. Ferrag, Author, Author, Author, Author, and Author, "Revolutionizing Cyber Threat Detection With Large Language Models: A Privacy-Preserving BERT-Based Lightweight Model for IoT/IIoT Devices," in *IEEE Access*, vol. 12, 2024, pp. 23733–23750.

[59] D. Rodriguez, I. Yang, J. M. D. Alamo, and N. Sadeh, "Large language models: A new approach for privacy policy analysis at scale," *Computing*, vol. 106, pp. 3879–3903, 2024, doi: 10.1007/s00607-024-01331-9.

[60] H. Montenegro, W. Silva, and J. S. Cardoso, "Privacy-Preserving Generative Adversarial Network for Case-Based Explainability in Medical Image Analysis," in *IEEE Access*, vol. 9, 2021, pp. 148037–148047.

[61] V. R. Saddi, S. K. Gopal, A. S. Mohammed, S. Dhanasekaran, and M. S. Naruka, "Examine the role of generative AI in enhancing threat intelligence and cyber security measures," in *2024 2nd International Conference on Disruptive Technologies (ICDT)*, 2024, pp. 537–541, doi: 10.1109/ICDT61202.2024.00089.

[62] L. Zhi and W. Wang, "Research on modeling of the imbalanced fraudulent transaction detection problem based on embedding-aware conditional GAN," *Big Data Research*, vol. 41, p. 100557, 2025, doi: 10.1016/j.bdr.2025.100557.

[63] P. Ghadekar, R. Paimode, S. Pandav, P. Patange, A. Pardeshi, and P. Patil, "GAN AI for Predictive Threat Detection with Explainable Risk Insights," in *2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI)*, 2025, pp. 1446–1452, doi: 10.1109/ICCSAI64074.2025.11063972.

[64] A. p. Durai pandian, "Variational Autoencoders using Convolutional neural network for highly advanced cyber threats," in *2024 IEEE Integrated STEM Education Conference (ISEC)*, 2024, pp. 1–6, doi: 10.1109/ISEC61299.2024.10664944.

[65] S. Jalalipour and B. Rekabdar, "Noisy-Defense Variational Auto-Encoder (ND-VAE): An Adversarial Defense Framework to Eliminate Adversarial Attacks," in *2023 Fifth International Conference on Transdisciplinary AI (TransAI)*, 2023, pp. 50–57, doi: 10.1109/TransAI60598.2023.00018.

[66] F. Li, H. Wu, and J. Zhang, "Lightweight Diffusion Model for Synthesizing Malicious Network Traffic," in *NAECON 2024 - IEEE National Aerospace and Electronics Conference*, 2024, pp. 409–413, doi: 10.1109/NAECON61878.2024.10670640.

[67] M. M. Helal, M. Abu-Elkheir, and M. Mashaly, "A Generative AI Framework for Cloud Security: Automated Attack Simulation and Threat Detection," in *2025 International Conference on Innovation in Artificial Intelligence and Internet of Things (AIIT), Jeddah*, 2025, pp. 1–8, doi: 10.1109/AIIT63112.2025.11082823.

[68] M. M. Yamin, E. Hashmi, M. Ullah, and B. Katt, "Applications of LLMs for generating cybersecurity exercise scenarios," *IEEE Access*, vol. 12, pp. 143806–143819, 2024, doi: 10.1109/ACCESS.2024.3468914.

[69] R. Honnalli and J. Farooq, "LLM-Powered Agentic AI Approach to Securing EV Charging Systems Against Cyber Threats," in *2025 IEEE 26th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2025, pp. 266–274, doi: 10.1109/WoWMoM65615.2025.00053.

[70] A. Patel, R. C. Sachan, H. Ragothaman, A. Sheth, P. Pandey, and S. K. Udayakumar, "Leveraging Generative AI for Proactive Cybersecurity Threat Detection in Cloud Environments," in *2025 8th International Conference on Information and Computer Technologies (ICICT)*, 2025, pp. 80–85, doi: 10.1109/ICICT64582.2025.00019.

[71] J. Lee, D. Jung, J. Moon, and S. Rho, "Advanced R-GAN: Generating anomaly data for improved detection in imbalanced datasets using regularized generative adversarial networks," *Alexandria Engineering Journal*, vol. 111, pp. 491–510, 2025, doi: 10.1016/j.aej.2024.10.084.

[72] W. Yin, C. Wang, and Y. Qin, "T-GAN: Transformer-based Generative Adversarial Network for Network Traffic Anomaly Detection," in *Proc. 2025 2nd Int. Conf. Comput. Netw. Cloud Comput. (CNCC)*, 2025, pp. 64–69, doi: 10.1145/3744451.3744461.

[73] R. Changala, S. Kayalvili, M. Farooq, L. M. Rao, V. S. Rao, and S. Muthuperumal, "Using Generative Adversarial Networks for Anomaly Detection in Network Traffic: Advancements in AI Cybersecurity," in *2024 International Conference on Data Science and Network Security (ICDSNS)*, 2024, pp. 1–6, doi: 10.1109/ICDSNS62112.2024.10690857.

[74] C. K. Akiri, K. Jayabalan, J. Lopes, S. A. Kareem, and A. Tabbassum, "Generative AI for Real-Time Cloud Security: Advanced Anomaly Detection Using GPT Models," in *2025 IEEE Conference on Computer Applications (ICCA), Yangon, Myanmar*, 2025, pp. 1–6, doi: 10.1109/ICCA65395.2025.11011269.

[75] L. Feng, J. Yu, S. Li, and S. Jiu, "VAE-SMOTE Augmented Diffusion for Anomaly Detection," in *Proc. 2025 2nd Int. Conf. Generative Artif. Intell. Inf. Secur*, 2025, pp. 163–170, doi: 10.1145/3728725.3728751.

[76] Y. Zhao, F. Ullah, C. Chen, M. Amoon, and S. Kumari, "Efficient malware detection using hybrid approach of transfer learning and generative adversarial examples with image representation," *Expert Systems*, vol. 42, no. 5, p. 13693, 2025, doi: 10.1111/exsy.13693.

[77] F. B. Khan, Author, Author, Author, Author, and Author, "Design and performance analysis of an anti-malware system based on generative adversarial network framework," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3358454.

[78] H. H. Reshi and K. Singh, "Enhancing Malware Detection using Deep Learning Approach," in *2024 International Conference on Automation and Computation (AUTOCOM), Dehradun, India*, 2024, pp. 497–501, doi: 10.1109/AUTOCOM60220.2024.10486116.

[79] Y. Ban, J. H. Yi, and H. Cho, "Augmenting Android malware using conditional variational autoencoder for the malware family classification," *Computer Systems Science & Engineering*, vol. 46, no. 2, 2023, doi: 10.32604/csse.2023.036555.

[80] D. Devadiga, Author, Author, Author, Author, and Author, "GLEAM: GAN and LLM for Evasive Adversarial Malware," in *2023 14th International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island*, 2023, pp. 53–58, doi: 10.1109/ICTC58733.2023.10393706.

[81] J. Haurogné, N. Basheer, and S. Islam, "Vulnerability detection using BERT based LLM model with transparency obligation practice towards trustworthy AI," *Machine Learning with Applications*, vol. 18, p. 100598, 2024, doi: 10.1016/j.mlwa.2024.100598.

[82] H. Ding, Y. Liu, X. Piao, H. Song, and Z. Ji, "SmartGuard: An LLM-enhanced framework for smart contract vulnerability detection," *Expert Systems With Applications*, vol. 269, p. 126479, 2025, doi: 10.1016/j.eswa.2025.126479.

[83] G. Lu, X. Ju, X. Chen, W. Pei, and Z. Cai, "GRACE: Empowering LLM-based software vulnerability detection with graph structure and in-context learning," *The Journal of Systems & Software*, vol. 212, p. 112031, 2024, doi: 10.1016/j.jss.2024.112031.

[84] Y. Guo, C. Patsakis, Q. Hu, Q. Tang, and F. Casino, "Outside the Comfort Zone: Analysing LLM Capabilities in Software Vulnerability Detection," in *Proc. 29th Eur. Symp. Res. Comput. Secur. (ESORICS)*, 2024, pp. 271–289, doi: 10.1007/978-3-031-70879-4_14.

[85] S. L. Mirtaheri and A. Pugliese, "Leveraging Generative AI to Enhance Automated Vulnerability Scoring," in *2024 IEEE Conference on Dependable, Autonomic and Secure Computing (DASC)*, 2024, doi: 10.1109/DASC64200.2024.00014.

[86] V. Akuthota, R. Kasula, S. T. Sumona, M. Mohiuddin, M. T. Reza, and M. M. Rahman, "Vulnerability Detection and Monitoring Using LLM," in *2023 IEEE 9th International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, 2023, pp. 309–314, doi: 10.1109/WIECON-ECE60392.2023.10456393.

[87] F. Z. Qachfar, R. M. Verma, and A. Mukherjee, "Leveraging Synthetic Data and PU Learning For Phishing Email Detection," in *Proc. Twelveth ACM Conf. Data Appl. Secur. Privacy*, 2022, doi: 10.1145/3508398.3511524.

[88] L. Jovanovic, Author, Author, Author, Author, and Author, "Generative Adversarial Networks for Synthetic Training Data Replacement in Phishing Email Detection Using Natural Language Processing," *Smart Data Intelligence*, 2024, doi: 10.1007/978-981-97-3191-6_46.

[89] M. Rathakrishnan, S. Gayan, S. Edirisinghe, and H. Inaltekin, "A Multi-Model Framework for Synthesizing High-Fidelity Network Intrusion Data Using Generative AI," in *2025 5th International Conference on Advanced Research in Computing (ICARC), Belihuloya, Sri Lanka*, 2025, pp. 1–6, doi: 10.1109/ICARC64760.2025.10963129.

[90] A. Rashidi, M. Salehi, and S. Najari, "CGANS: a code-based GAN for spam detection in social media," *Social Network Analysis and Mining*, vol. 14, p. 218, 2024, doi: 10.1007/s13278-024-01379-7.

[91] M. K. Prabakaran, P. M. Sundaram, and A. D. Chandrasekar, "An enhanced deep learning-based phishing detection mechanism to effectively identify malicious URLs using variational autoencoders," *IET Information Security*, vol. 17, no. 3, pp. 423–440, 2023, doi: 10.1049/ise2.12106.

[92] C. Thanomwong, K. Puangnak, N. Rachsiriwatcharabul, M. Tiawongsuwan, and K. Puangnak, "Applying Generative AI for Fraud and Cybercrime Prevention in Thailand," in *2025 11th International Conference on Engineering, Applied Sciences, and Technology (ICEAST), Thailand*, 2025, pp. 65–68, doi: 10.1109/ICEAST64767.2025.11088187.

[93] A. B. Beydemir, U. Sezgin, U. Doğan, B. E. Aşıklar, F. A. Yerlikaya, and Ş. Bahtiyar, "A Dynamically Selected GPT Model for Phishing Detection," in *2024 14th International Conference on Advanced Computer Information Technologies (ACIT), Ceske Budejovice, Czech Republic*, 2024, pp. 481–484, doi: 10.1109/ACIT62333.2024.10712553.

[94] R. Nair, F. Abbasi, and S. Pervez, "PhishEmailLLM: A meta model approach to detect phishing emails by leveraging LLMs and machine learning models," *Australasian Computer Science Week (ACSW)*, 2025, doi: 10.1145/3727166.3727169.

[95] A. Alqarni, Author, Author, Author, Author, and Author, "How Generative AI Transforms Spam Detection," in *Tech Fusion in Business and Society*, 2025, p. 233.

[96] K. A. Vedros, C. Kolias, D. Barbara, and R. C. Ivans, "From Code to EM Signals: A Generative Approach to Side Channel Analysis-based Anomaly Detection," in *Proc. 19th Int. Conf. Availability, Rel. Secur. (ARES)*, 2024, doi: 10.1145/3664476.3664520.

[97] R. Talibzade, F. Bergadano, and I. Drago, "On LLM Embeddings for Vulnerability Management," in *2025 9th Network Traffic Measurement and Analysis Conference (TMA)*, 2025, pp. 1–4, doi: 10.23919/TMA66427.2025.11097007.

[98] S. Cai, Author, Author, Author, Author, and Author, "DDP-DAR: Network intrusion detection based on denoising diffusion probabilistic model and dual-attention residual network," *Neural Netw.*, vol. 184, p. 107064, 2025, doi: 10.1016/j.neunet.2024.107064.

[99] S. Gul, S. Arshad, S. M. U. Saeed, A. Akram, and M. A. Azam, "WGAN-DL-IDS: An Efficient Framework for Intrusion Detection System Using WGAN, Random Forest, and Deep Learning Approaches," *Computers*, vol. 14, no. 4, pp. 1–21, 2024, doi: 10.3390/computers14010004.

[100] C. Park, J. Lee, Y. Kim, J. G. Park, H. Kim, and D. Hong, "An Enhanced AI-Based Network Intrusion Detection System Using Generative Adversarial Networks," *IEEE Internet of Things Journal*, vol. 10, no. 3, pp. 2330–2345, 2023, doi: 10.1109/JIOT.2022.3211346.

[101] Z. Ullah, Author, Author, Author, Author, and Author, "Interpretable and Adaptive GAN-BiLSTM Approach for Cyber Threat Detection in IoMT-based Healthcare 5.0," *IEEE Journal of Biomedical and Health Informatics*, 2025, doi: 10.1109/JBHI.2025.3573097.

[102] Y. Feng, Y. Si, J. Zhang, Z. Cai, and H. Zhao, "SA-WGAN based data enhancement method for industrial Internet intrusion detection," *Computational Materials and Continua*, vol. 84, no. 3, pp. 4432–4448, 2025, doi: 10.32604/cmc.2025.064696.

[103] S. S. Khatami, M. Shoeibi, A. E. oskouei, D. Martín, and M. K. Dashliboroun, "5DGWO-GAN: A novel five-dimensional gray wolf optimizer for generative adversarial network-enabled intrusion detection in IoT systems," *Computers, Materials & Continua*, vol. 82, no. 1, pp. 882–910, 2025, doi: 10.32604/cmc.2024.059999.

[104] M. Jamoos, A. M. Mora, M. Alkhanafseh, and O. Surakhi, "A new data-balancing approach based on generative adversarial network for network intrusion detection system," *Electronics*, vol. 12, no. 12, p. 2851, 2023, doi: 10.3390/electronics12132851.

[105] M. S. Siddique, Author, Author, Author, Author, and Author, "An intelligent intrusion detection system for cyber-physical systems using GAN-LSTM networks," *Franklin Open*, vol. 11, p. 100281, 2025, doi: 10.1016/j.fraope.2025.100281.

[106] X. Wang, Y. Xu, Y. Xu, Z. Wang, and Y. Wu, "Intrusion Detection System for In-Vehicle CAN-FD Bus ID Based on GAN Model," *IEEE Access*, vol. 12, pp. 82402–82412, 2024, doi: 10.1109/ACCESS.2024.3412933.

[107] H. N. Nguyen, T. Lan-Phan, and C. J. Song, "Generative Adversarial Network-Based Network Intrusion Detection System for Supervisory Control and Data Acquisition System," in *2024 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, 2024, pp. 1–3, doi: 10.1109/ICCE-Asia63397.2024.10773791.

[108] A. Rahman, Author, Author, Author, Author, and Author, "SYN-GAN: A robust intrusion detection system using GAN-based synthetic data for IoT security," *Internet Things*, vol. 26, p. 101212, 2024, doi: 10.1016/j.iot.2024.101212.

[109] U. Kumaran, S. Thangam, T. V. N. Prabhakar, J. Selvaganesan, and H. N. Vishwas, "Adversarial defense: A GAN-IF based cyber-security model for intrusion detection in software piracy," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 14, no. 4, pp. 96–114, 2023, doi: 10.58346/JOWUA.2023.I4.008.

[110] S. Aldhaheri and A. Alhuzali, "SGAN-IDS: Self-attention-based generative adversarial network against intrusion detection systems," *Sensors*, vol. 23, no. 18, p. 7796, 2023, doi: 10.3390/s23187796.

[111] M. A. Ferrag, D. Hamouda, M. Debbah, L. Maglaras, and A. Lakas, "Generative Adversarial Networks-Driven Cyber Threat Intelligence Detection Framework for Securing Internet of Things," in *2023 19th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT), Cyprus*, 2023, pp. 196–200, doi: 10.1109/DCOSS-IoT58021.2023.00042.

[112] S. Koparde, S. Kothari, A. Pawar, P. Vyas, S. Sawant, and P. Tank, "Intrusion detection system (IDS) using generative adversarial networks (GAN) & XGBoost (XGB)," in *Proc. Int. Conf. Appl. Mach. Intell. Data Analytics (ICAMIDA)*, 2025, pp. 1–6, doi: 10.1109/ICAMIDA64673.2025.11208916.

[113] K. Indhumathi, G. Pandiselvi, P. J, N. M. F, P. E, and A. B, "Utilizing Generative Adversarial Networks for Synthetic Data Generation in Cybersecurity Applications," in *2025 2nd Int Conf Intelligent Algorithms for Computational Intelligence Systems (IACIS)*, 2025, pp. 1–7, doi: 10.1109/IACIS65746.2025.11211266.

[114] M. N. Halgamuge, Author, Author, Author, Author, and Author, "LLM-Driven Adaptive Security for the Internet of Energy (IoE)," *IEEE Network*, vol. 40, no. 1, pp. 35–43, 2026, doi: 10.1109/MNET.2025.3620562.
[115] A. Appenzeller, M. Leitner, P. Philipp, E. Krempel, and J. Beyerer, "Privacy and utility of private synthetic data for medical data analyses," *Applied Sciences*, vol. 12, no. 23, p. 12320, 2022, doi: 10.3390/app122312320.
[116] S. Imtiaz, M. Arsalan, V. Vlassov, and R. Sadre, "Synthetic and Private Smart Health Care Data Generation using GAN," in *2021 International Conference on Computer Communications and Networks (ICCCN)*, 2021, pp. 1–7, doi: 10.1109/ICCCN52240.2021.9522203.
[117] S. Hu, Author, Author, Author, Author, and Author, "Protecting Facial Privacy: Generating Adversarial Identity Masks via Style-robust Makeup Transfer," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 14994–15003, doi: 10.1109/CVPR52688.2022.01459.
[118] S. Ho, Author, Author, Author, Author, and Author, "DP-GAN: Differentially private consecutive data publishing using generative adversarial nets," *Journal of Network and Computer Applications*, vol. 185, p. 103066, 2021, doi: 10.1016/j.jnca.2021.103066.
[119] Y. Li, Q. Tan, and B. S. Shin, "CryptoGAN: Privacy-Preserving Federated Generative Adversarial Networks With Homomorphic Encryption in Healthcare Systems," *IEEE Transactions on Computational Social Systems*, vol. 12, no. 6, pp. 5330–5341, 2025, doi: 10.1109/TCSS.2025.3570990.
[120] P. C. Hsu, Z. Yu, S. Mise, and H. Miyaji, "Privacy-Diffusion: Privacy-preserving stable diffusion without FHE and differential privacy," *IEEE Access*, vol. 13, pp. 75914–75203, 2025, doi: 10.1109/ACCESS.2025.3562563.
[121] S. Misbah, Author, Author, Author, Author, and Author, "Generative AI-Driven Smart Contract Optimization for Secure and Scalable Smart City Services," *Smart Cities*, vol. 8, no. 4, p. 118, 2025, doi: 10.3390/smartcities8040118.
[122] S. Yin, X. Zhang, and L. Zuo, "Defending against adversarial attacks using spherical sampling-based variational auto-encoder," *Neurocomputing*, vol. 478, pp. 1–10, 2022, doi: 10.1016/j.neucom.2021.12.080.
[123] X. Zeng, Y. Zhuo, T. Liao, and J. Guo, "Cloud-GAN: Cloud generation adversarial networks for anomaly detection," *Pattern Recognit.*, vol. 157, p. 110866, 2025, doi: 10.1016/j.patcog.2024.110866.
[124] F. Xiao, J. Zhou, K. Han, H. Hu, and J. Fan, "Unsupervised anomaly detection using inverse generative adversarial networks," *Inf. Sci.*, vol. 689, p. 121435, 2025, doi: 10.1016/j.ins.2024.121435.
[125] A. Baral, B. K. Paikaray, and S. Baral, "Enhancing cyber threat detection with generative adversarial networks and anomaly detection techniques," in *Proc. 2nd Int. Conf. Circuits, Power Intell. Syst. (CCPIS)*, 2025, pp. 1–6, doi: 10.1109/CCPIS65231.2025.11234235.
[126] S. Badakhshan and J. Zhang, "Generative AI-Enhanced Real-Time Anomaly Detection in Integrated Energy Systems," in *IEEE Transactions on Smart Grid*, 2025.
[127] A. J, V. Ebenezer, A. J. Isaac, J. Marshell, P. Pradeepa, and V. Naveen, "Adversarial Attacks on Generative AI Anomaly Detection in the Quantum Era," in *2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2023, pp. 1833–1840, doi: 10.1109/ICECA58529.2023.10395092.
[128] J. Lee, Author, Author, Author, Author, and Author, "Anomaly Detection Using Generative Language Models and Deep Feature-Based Time Series Similarity," in *IEEE Access*, vol. 13, 2025, pp. 157147–157159.
[129] A. Choi, A. Jiang, S. Jumani, D. Luong, and F. D. Troia, "Synthetic Malware Using Deep Variational Autoencoders and Generative Adversarial Networks," *EAI Endorsed Transactions on Internet of Things*, vol. 10, 2024, doi: 10.4108/eetiot.6566.
[130] T. Taylor and A. Eleyan, "Using variational autoencoders to increase the performance of malware classification," in *2021 International Symposium on Networks, Computers and Communications (ISNCC)*, 2021, pp. 1–8, doi: 10.1109/ISNCC52172.2021.9615643.
[131] T. Bao, K. Trousil, Q. D. Tran, F. D. Troia, and Y. Park, "Generating synthetic malware samples using generative AI," *IEEE Access*, vol. 13, pp. 59725–59736, 2025, doi: 10.1109/ACCESS.2025.3556704.