Effectiveness of Generative Artificial Intelligence (GenAI) for Cybersecurity Threat Detection

An independent scholarly project presented in partial fulfilment of the requirements for the degree of

Master of Information Technology

Supervisor Name: **Dr. Md. Akbar Hossain**

Student Name: **Janakee M. Patabadige**

Eastern Institute of Technology

Auckland, New Zealand

2025

Abstract

The increase in frequency and complexity of cyberthreats poses significant risks to individuals, organizations, and governments. Cyberattacks, such as phishing, injection attacks, malware, data breaches, and other malicious activities, lead to serious economic and social impacts. Traditional threat detection techniques, including rule-based systems, machine learning, and deep learning models, often face challenges in addressing these evolving attacks. In this context, the adoption of Generative Artificial Intelligence (GenAI) in cybersecurity threat detection offers a transformative approach to enhance defense mechanisms. This systematic literature review (SLR) critically evaluates the effectiveness of GenAI techniques, such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs) and Transformer-based large language models (LLMs), in strengthening threat detection. Following PRISMA guidelines, this study synthesizes findings from 48 peer-reviewed studies published between 2021 and 2025 that focus on applications, performance impacts, challenges, and ethical considerations. This study identifies three core applications including synthetic data generation, anomaly detection, and adversarial example generation. Performance evaluations reveal that GenAI outperforms traditional detection techniques, achieving higher accuracy, precision and F1-score, particularly against zero-day attacks. Moreover, GenAI-based detection systems reduce false positives and negatives, enhancing system efficiency and reliability. Key datasets such as NSL-KDD, and UNSW-NB15 support these improvements. However, several challenges remain, including adversarial vulnerabilities, computational demands, and ethical concerns resulting from the dual-use nature of GenAI. These findings emphasize the need for robust threat detection strategies and ethical guidelines. Theoretically, this study provides a structured framework for understanding the effectiveness of GenAI in the field of cybersecurity. Practically, it provides insights for cybersecurity professionals to strengthen their defense systems. Future research should explore integrating GenAI with emerging technologies, implementing lightweight models, and enhancing interpretability to ensure scalable, secure and responsible systems deployments. This SLR offers practical insights for cybersecurity professionals and a theoretical framework to guide future research in GenAI-based threat detection.

Keywords: Generative Artificial Intelligence, Generative AI, GenAI, Cybersecurity, Threat Detection, Cyberattack, Generative Adversarial Networks, GANs, Variational Autoencoder, VAE, Large Language Model, LLM

Table of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

Acronyms/ Abbreviation	Definition
AE	Autoencoder
AI	Artificial Intelligence
AML	Adversarial Machine Learning
APTs	Advanced Persistent Threats
AR	Autoregression
ARGAN	Adversarially Robust Generative Adversarial Networks
AUC	Area Under the Curve
BBPE	Byte-Pair-Encoding
CDAAE	Conditional Denoising Adversarial Autoencoder
CDAAE-KNN	Conditional Denoising Adversarial Autoencoder with KNN
cGAN	Conditional Generative Adversarial Networks
CNN	Convolutional Neural Network
CVAE	Conditional Variational Autoencoders
DDGAN	Denoising Diffusion Generative Adversarial Networks
DDoS	distributed denial-of-services
DL	Deep Learning
DNN	Deep Neural Network
FNR	False Negative Rate
FPR	False Positive Rate
GAN	Generative Adversarial Networks
GenAI	Generative Artificial Intelligence
Generative AI	Generative Artificial Intelligence
GRU	Gate Recurrent Unit (GRU)
IoT	Internet of Things
KNN	K-Nearest Neighbour
LLM	Large Language Models
LSTM	Long Short-Term Memory
MCC	Matthew's correlation coefficient
ML	Machine Learning
MLP	Multiplayer Perception
NB	Naive Bayes
PPFLE	privacy-preserving encoding technique
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RF	Random Forest
RNN	Recurrent Neural Network
RSS	Range-Based Sequential System
SIEM	Security Information and Event Management
SLR	Systematic Literature Review

SME	Small and Medium Enterprises
SMOTE	Synthetic Minority Oversampling Technique
SOC	Security Operations Center
SVM	Support Vector Machine
TPR	True Positive Rate
VAE	Variational Autoencoder
WGAN	Wasserstein Generative Adversarial Networks
WGEN-GP	Wasserstein Generative Adversarial Networks with Gradient Penalty

1. Introduction

1.1. Chapter Overview

Cybersecurity threat detection is a critical component of information security, focusing on identifying and mitigating evolving attacks targeting systems, networks, and data. This chapter begins with the background and context of current cybersecurity threats and applications of Generative Artificial Intelligence (GenAI). Later, it presents the motivation for the study, outlines the research objectives, and defines key research questions. The significance of this study from both theoretical and practical contexts is highlighted. Finally, this chapter presents the report structure of the subsequent sections.

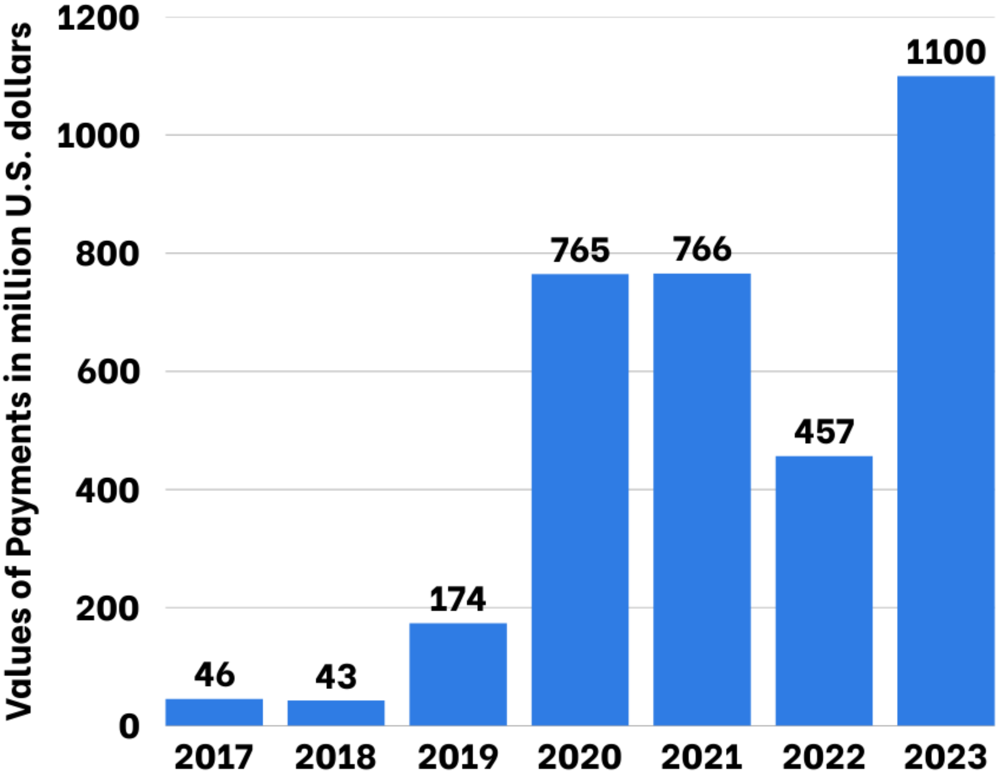
1.2. Background

In the current digital era, most individuals, organizations, and governments increasingly depend on digital systems. Cybersecurity threat detection has become a critical component of protecting these digital systems from malicious attacks, such as malware, phishing, ransomware, distributed denial-of-service (DDoS) attacks, and advanced persistent threats (APTs). With the increasing dependence on cloud platforms and the Internet of Things (IoT), organizations are further exposed to larger attack surfaces (Kasri et al., 2025). According to the report by Cybersecurity Ventures (2023), global cybercrime costs are projected to reach $10.5 trillion annually by 2025. Furthermore, the New Zealand Cyber Security Center reported 7122 incidents between 2023 and 2024 that caused NZD 21.6 million in financial losses (NCSC, 2025). Figure 1, based on Statista (2024) shows the increasing ransom payments from 2017 to 2023 highlighting the growing impact of cyberattacks. These growing cyberthreats emphasize the need for robust and adaptive cybersecurity strategies.

Traditional methods mainly depend on rule-based and signature-based systems (Demirbaga, 2024; Senthilkumar et al., 2024). Signature-based methods accurately detect known threats with a short processing time for previously known patterns (Ren et al., 2023; Vu et al., 2023). However, these methods struggle to detect unknown and zero-day attacks. Similarly, rule-based methods utilize predefined rules to identify suspicious activities but face challenges with false alerts. Traditional Machine Learning (ML) models like Support Vector Machine (SVM) and Random Forest (RF) models have achieved some improvements compared to conventional methods. However, the effectiveness of these models depends on high quality, large datasets, which are limited in cybersecurity due to data sensitivity and rare attack frequency (Alo et al., 2024).

Figure 1

Global Ransomware Payments 2017- 2023 (in million U.S. dollars)

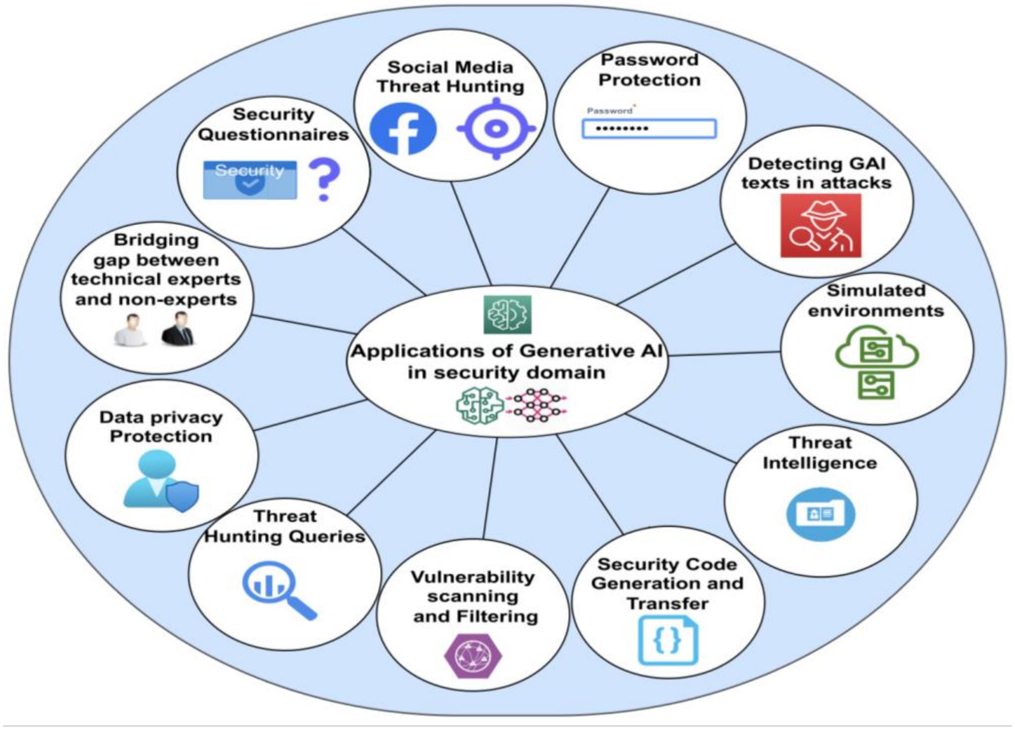


Note. From Statista (2024).

GenAI, a subset of Artificial Intelligence (AI), that provides innovative solutions to the challenges of traditional detection methods. According to Celik and Eltawil (2024) and Demirbaga (2024) GenAI models generate new data, such as text, images, audio, and video, based on learned patterns. Common GenAI models include Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Transformer-based models, and Diffusion Models (Sai et al., 2024). GenAI enhances cybersecurity through various applications, such as generating synthetic data (Aceto et al., 2024; Dina et al., 2022), detecting anomalies (Ferrag et al., 2023; Senthilkumar et al., 2024), and simulating threat environments (Mari et al., 2023; Sai et al., 2024), to proactively identify unknown threats (refer to Figure 2). However, it faces significant challenges including high computational demand, the risk of biased outputs, and the potential misuse of generated content (Alo et al., 2024; Sai et al., 2024). This study evaluates GenAI's effectiveness in cybersecurity threat detection across dynamic environments.

Figure 2

Applications of GenAI in Cybersecurity


Social Media Threat Hunting
Password Protection
Password
••••••••
Security Questionnaires
Security
?
Detecting GAI texts in attacks
Bridging gap between technical experts and non-experts
Simulated environments
Applications of Generative AI in security domain
Data privacy Protection
Threat Intelligence
Threat Hunting Queries
Vulnerability scanning and Filtering
Security Code Generation and Transfer
{}

Note. Retrieved from Sai et al. (2024).

1.3. Research Motivation

Traditional and early ML-based detection approaches struggle with the rising frequency and sophistication of cyberthreats. Recent studies by Chiriac et al., (2025) and Hamouda et al., (2024) highlight GenAI as an innovative solution due to its capabilities in generating synthetic data, modeling complex patterns, and simulating attacks. However, its effectiveness in cyberthreat detection needs to be further investigation.

Table 1 provides an overview of the existing literature in this area and highlights the unique contributions of this research. Most literature provides valuable insights into the use of GenAI in areas such as attack obfuscation and synthetic data generation. However, these studies do not specifically focus on threat detection in the cybersecurity domain, performance evaluation or discuss practical limitations. Therefore, this study aims to fill this gap by conducting research specifically focused on the applications, performance impacts, and associated challenges of GenAI in cyberthreat detection.

Table 1

Overview of Literature Surveys on GenAI for the Cybersecurity Threat Detection Context with Comparison to Our Work

Article	Focus	Cybersecurity Scope	Discussed GenAI Techniques	Timeframe covered
Hasanov et al., (2024)	Systematic Review of LLM applications in cybersecurity	LLM in offensive and defensive cybersecurity, ethical and governance regulations of LLMs	LLMs, AI	2018-2024
Goyal & Mahmoud, (2024)	Synthetic data generation using GenAI across domain	Not focused on cybersecurity, primary focus on synthetic data generation frameworks	GANs, VAEs, LLMs and other models	2014-2024
Cappolino et al., (2025)	Impact of GenAI on Cybersecurity	Network and Web security (defensive and offensive uses)	GANs, Conditional GANs (cGANs)	2020-2024
Dunmore et al., (2023)	Survey on applications of GAN in intrusion detection systems (IDS)	Intrusion detection (IoT, network, mobile, wireless, sensor, autonomous vehicle)	GANs and its variations	2014-2023
Vu et al., (2024)	Survey on GenAI in mobile and wireless networking	Wireless security: intrusion detection, jamming attacks, data obscurity	GANs, VAEs, Diffusion models, Transformer, Meta-learning, Multi-task GAN	2014-2024
Our work	**Evaluate the effectiveness of GenAI in cybersecurity threat detection**	**Specifically, cybersecurity threat detection: applications, performance impact, limitations**	**GANs, VAEs, Transformer-based Models, Hybrid models**	**2021- 2025**

1.4. Research Objective/aim

The primary objective of this research is to investigate the effectiveness of GenAI in enhancing cybersecurity threat detection. This includes identifying and analyzing the key applications of GenAI techniques in detecting and mitigating cyberthreats. Furthermore, this study aims to evaluate the impact of using GenAI on the performance of threat detection systems. Additionally, this study explores the challenges and ethical concerns associated with using GenAI. Overall, these objectives aim to provide a comprehensive understanding of the effectiveness of GenAI in the modern cybersecurity field.

1.5. Research Questions

This study aims to critically evaluate the effectiveness of GenAI techniques in enhancing cybersecurity threat detection. To achieve this aim, the study is guided by a main research question (RQ1), and three sub-questions are proposed to support RQ1.

Main Research Question:

RQ1: How effective are the applications of GenAI techniques in enhancing cybersecurity threat detection?

Sub Research Questions (SRQ):

SRQ1: What are the primary applications of GenAI in cybersecurity threat detection?

- Explores key GenAI applications, including synthetic data generation, anomaly detection, and adversarial training, in cybersecurity threat detection. It focuses on how these techniques support threat simulation, data augmentation, and system enhancement in cyber defense.

SRQ2: How do the applications of GenAI techniques affect the performance of threat detection systems?

- Examines how GenAI improves detection accuracy, reduces false positives, and enhances system reliability in threat detection systems.

SRQ3: What are the potential challenges and ethical concerns associated with using GenAI techniques in cybersecurity threat detection?

- Investigates potential limitations and drawbacks associated with using GenAI, such as resource requirements, data dependencies, adversarial attacks, and ethical concerns.

1.6. Research Significance (Theoretical and Practical)

From a theoretical perspective, this research contributes to cybersecurity by critically evaluating the effectiveness of GenAI models in enhancing threat detection. Generative models, such as GANs, VAEs and LLMs, have demonstrated various applications in industries like healthcare, tourism, finance, and hospitality (Dwivedi et al., 2024). For instance, GenAI models like ChatGPT are integrated into customer services to enable real-time interaction, assist with travel planning, and provide personalized recommendations (Sai et al., 2024). However, as mentioned in the research motivation section, the application of GenAI in cyberthreat detection has not been comprehensively examined. This study addresses this gap by analyzing how these models identify cybersecurity threats in dynamic digital environments. Furthermore, through an empirical evaluation, it emphasizes the strengths of these models, like anomaly detection, and limitations including interpretability and scalability issues. This study establishes a theoretical foundation for future research by explaining how GenAI enhances cyberthreat detection.

Practically, this study provides valuable insights for cybersecurity professionals, organizations, and policymakers. By identifying GenAI applications, like synthetic data generation and simulating adversarial attacks, this research provides techniques to enhance their defense mechanisms. These findings support the development of robust threat detection systems against sophisticated threats, like zero-day attacks and APTs. Moreover, this study analyzes challenges, such as data quality and computational demands. By providing a balanced overview, this study guides organizations in making informed decisions about adopting GenAI solutions (Alo et al., 2024). Additionally, improved GenAI-based threat detection will reduce financial losses, secure sensitive information, and overall improve national security.

1.7. Structure of the Report

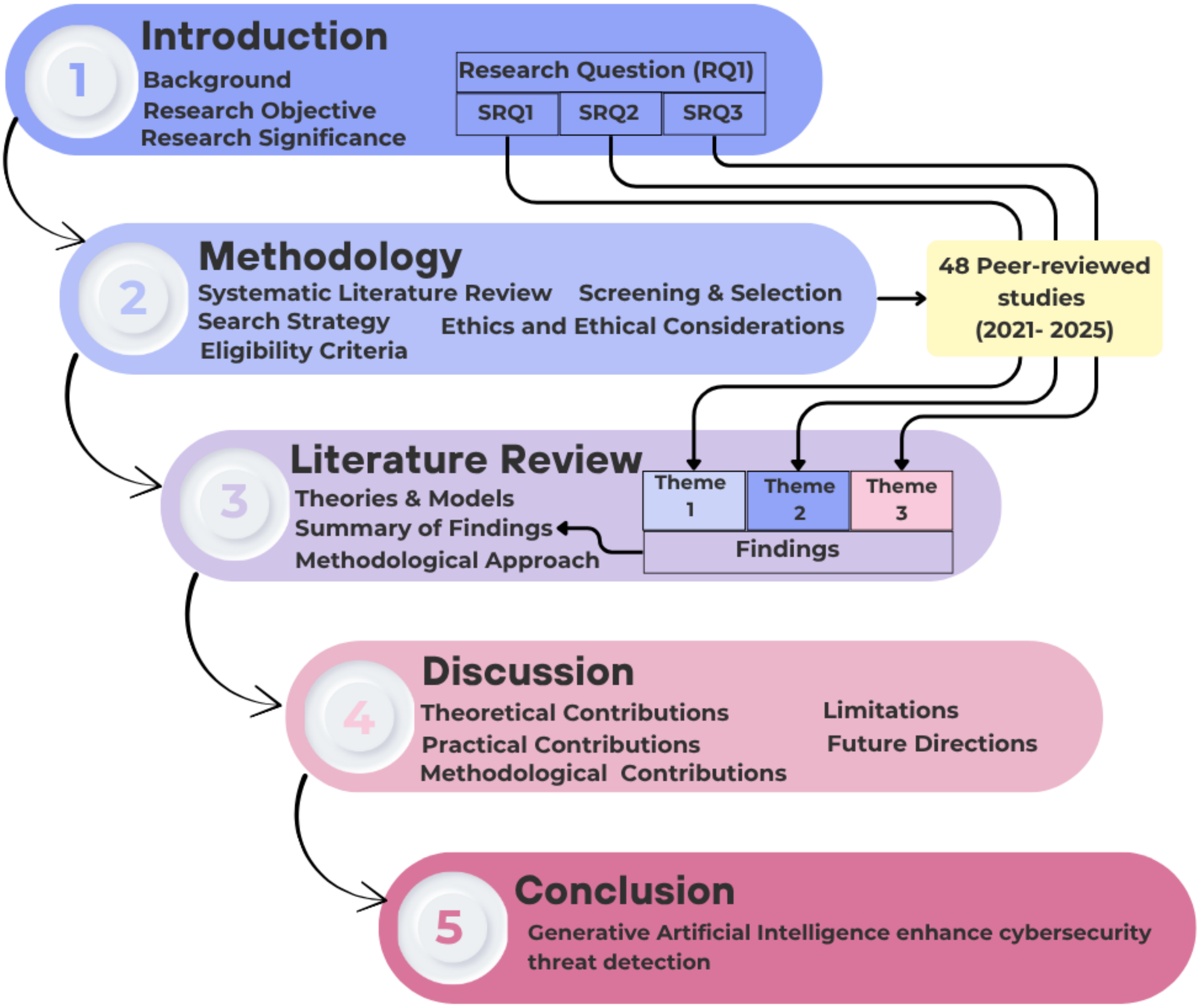
The structure of this report has been designed to provide a comprehensive evaluation of the effectiveness of GenAI in cybersecurity threat detection (see Figure 3). **Chapter 2** describes the research methodology, including the research approach, data collection, and data analysis. The literature review in **Chapter 3** reviews existing research on GenAI in cybersecurity threat detection. In **Chapter 4**, the discussion interprets the findings of the literature review and explores future directions. Finally, **Chapter 5** summarizes the key insights and highlights its contributions to the cybersecurity field.

1.8. Chapter Summary

This chapter has provided the foundation for this research by explaining the challenges in threat detection and the potential of GenAI to address these challenges. It has described the motivation for this study, its objectives, and key research questions, and provided a clear structure for the subsequent chapters. The next chapter will discuss the research methodology, including details of the research approach and techniques used to conduct this study.

Figure 3

Structure of the Report



2. Methodology

2.1. Chapter Overview

This chapter describes the methodology used to investigate the effectiveness of GenAI in enhancing cyberthreat detection. It begins with the justification for selecting the Systematic Literature Review (SLR) as the research methodology and defines key terms relevant to the study. Subsequently, it provides details of the search strategy, eligibility criteria, and the screening and selection process applied to identify relevant literature. Furthermore, ethical considerations specific to this study are discussed. This chapter concludes with a summary of the methodology used.

2.2. Systematic Literature Review

This study utilizes SLR as the research method to investigate the effectiveness of GenAI in cybersecurity threat detection. The SLR approach was chosen due to its systematic, transparent, and replicable nature (Mohamed Shaffril et al., 2021; Snyder, 2019). It supports the identification and mapping of various GenAI techniques in this domain. The SLR follows predefined protocols, unlike traditional or narrative reviews, which often lack methodological transparency and introduce bias due to undefined search strategies or selection criteria (Williams et al., 2021). This enhances reliability and repeatability of the study, making it suitable for evaluating emerging, interdisciplinary fields like GenAI and cybersecurity. By providing a thorough and objective process for identifying, screening, and synthesizing literature (Snyder, 2019; Williams et al., 2021), the SLR helps to understand the current state of knowledge, research gaps, and future directions.

Alternative research methodologies, such as qualitative and quantitative research methods, were considered. However, these approaches were found unsuitable for the objectives of this study. Each research methodology has its own strengths and limitations, and the selection depends on the study's context, research goals, and key questions. Quantitative methods focus on numerical data and statistical analysis, which may limit the exploration of the emerging nature of this field. Qualitative methods emphasize subjective insights but lack the structured synthesis needed for a comprehensive review of technical applications. Therefore, the SLR is the most suitable methodology, providing a strong, transparent, and replicable framework to advance research and practice in GenAI for cybersecurity threat detection.

2.3. Key Terms

The key terms and related sub-concepts were identified through the literature review process. These terms represent the core themes and scope of this study and provide the foundation for the search strategy. Table 2 presents the identified key terms and related keywords.

Table 2

Key Terms and Keywords

Key Terms	Related Keywords
GenAI	• GANs • VAEs • Transformer-based Models ↳ LLMs ↳ GPT ↳ BERT
Cybersecurity Threat Detection	• Synthetic Data • Anomaly detection • Adversarial example • Phishing Attacks • Intrusion Detection • Malware Detection
Performance Evaluation	• Accuracy • Precision • F1-score • Recall

2.4. Search Strategy

This study utilizes a systematic search method to retrieve a wide range of academic literature relevant to the topic. It systematically searches academic databases, such as IEEE Xplore, ACM digital Library, ScienceDirect, Google Scholar, and SpringerLink. These databases were selected to ensure the inclusion of peer-reviewed articles published in reputable journals and conferences from 2021 to 2025.

To search the databases, this study used search strings like "GenAI for cybersecurity threat detection" and "GANs for cyberthreat detection". Additionally, the previously defined key terms and keywords were strategically combined using Boolean operators, such as "AND", "OR", and "NOT", to develop effective search queries, such as:

- ("Generative AI" OR "GenAI") AND ("Cybersecurity" OR "Threat detection")
- ("Generative AI" OR "GANs") AND ("Cybersecurity" OR "Threat detection")
- ("GenAI" OR "VAEs") AND ("Cybersecurity" OR "Threat detection")
- ("GenAI" OR "GANs") AND Cybersecurity AND Anomaly detection
- ("Adversarial examples" OR "Generative models") AND "Threat detection"
- Generative AI AND Cybersecurity AND ("Synthetic data" OR "Anomaly detection")
- LLMs AND ("Cybersecurity" OR "Malware detection")

The search process was conducted by examining titles, keywords, and abstracts to ensure the relevance of the selected publications. Additionally, this study used backward and forward citation searching to identify key publications.

2.5. Eligibility Criteria

Following the search process, this study applies predefined inclusion and exclusion criteria to filter the retrieved articles. These eligibility criteria ensure the credibility and relevance of the reviewed articles (refer Table 3).

Table 3

Inclusion/Exclusion Criteria

Code	Inclusion Criteria
IC1	Articles that discuss the applications of GenAI in cybersecurity threat detection
IC2	Peer-reviewed articles, including journal papers, conferences proceedings, and industry reports
IC3	Research that includes empirical findings, theoretical contributions, or methodological approaches aligned with the research objectives
IC4	Articles published from 2021 to 2025
IC5	Articles written and published in English
IC6	Full-text availability is required
Code	**Exclusion Criteria**
EC1	Articles that do not primarily focus on the applications of GenAI in cybersecurity threat detection.
EC2	Non-peer-reviewed articles, posters, books, theses etc.
EC3	Systematic Literature Reviews and survey articles.
EC4	Articles that published prior to 2021
EC5	Articles not published in English

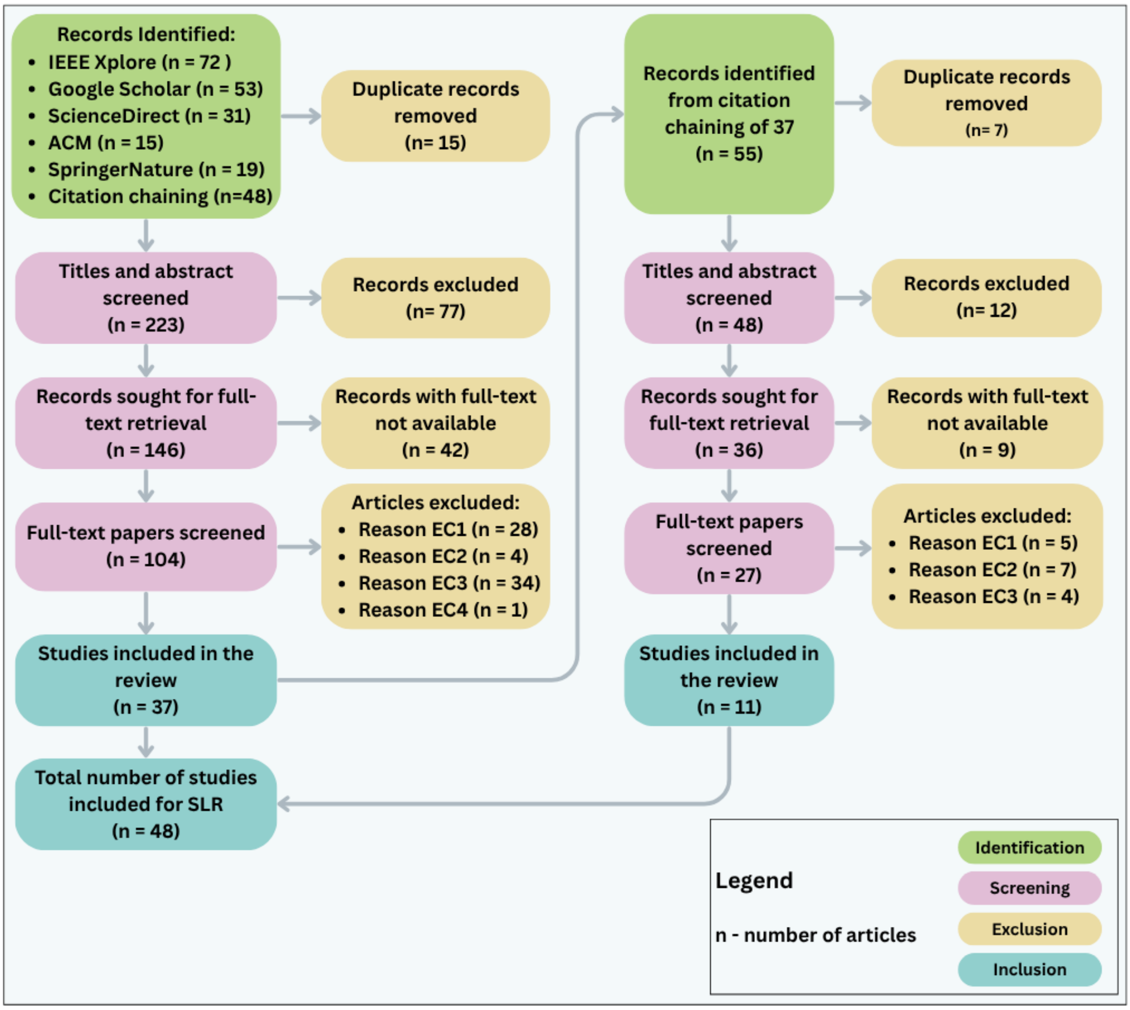
2.6. Screening and Selection

The process of screening and selecting articles follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines to ensure transparency and replicability (O'Dea et al., 2021). As shown in Figure 4, it involves four stages including identification, screening, eligibility, and inclusion. Initially, 238 articles were identified through citation chaining (backward and forward) and databases searches using defined keywords and search queries. After removing 15 duplicates, 223 articles remained for title and abstract screening.

During the screening phase, 77 articles were excluded for not aligning with the research topic or eligibility criteria, leaving 146 articles for full-text review. Of these, 42 were excluded due to the unavailability of full texts, leaving 104. In the eligibility phase, 67 articles were excluded for not meeting the eligibility criteria, resulting in 37 retained articles.

Figure 4

PRISMA Flowchart of Collection Procedure


Records Identified:
IEEE Xplore (n = 72)
Google Scholar (n = 53)
ScienceDirect (n = 31)
ACM (n = 15)
SpringerNature (n = 19)
Citation chaining (n=48)
Duplicate records removed (n= 15)
Titles and abstract screened (n = 223)
Records excluded (n= 77)
Records sought for full-text retrieval (n = 146)
Records with full-text not available (n = 42)
Full-text papers screened (n = 104)
Articles excluded:
Reason EC1 (n = 28)
Reason EC2 (n = 4)
Reason EC3 (n = 34)
Reason EC4 (n = 1)
Studies included in the review (n = 37)
Total number of studies included for SLR (n = 48)
Records identified from citation chaining of 37 (n = 55)
Duplicate records removed (n= 7)
Titles and abstract screened (n = 48)
Records excluded (n= 12)
Records sought for full-text retrieval (n = 36)
Records with full-text not available (n = 9)
Full-text papers screened (n = 27)
Articles excluded:
Reason EC1 (n = 5)
Reason EC2 (n = 7)
Reason EC3 (n = 4)
Studies included in the review (n = 11)
Legend
n - number of articles
Identification
Screening
Exclusion
Inclusion

Note. Adapted from O'Dea et al., (2021), The flowchart was modified to represent SLR process of this study.

Finally, a descriptive analysis was performed on 48 selected studies to investigate their distribution and characteristics. Figure 5 presents the yearly distribution of publications, indicating a growing academic interest in cybersecurity. Figure 6 illustrates the adoption of generative models, where GANs are more most frequently used, with a balanced representation across VAEs, LLMs, and hybrid models.

Figure 5

Distribution of Selected Publications per Year

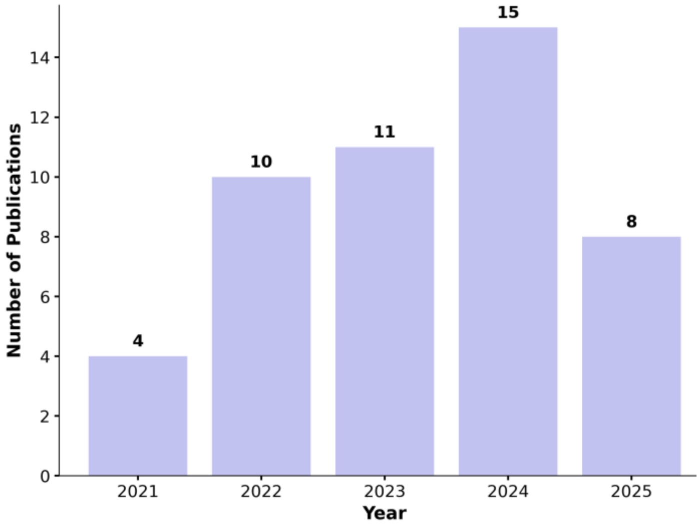


Figure 6

Distribution of Articles Across Generative Models

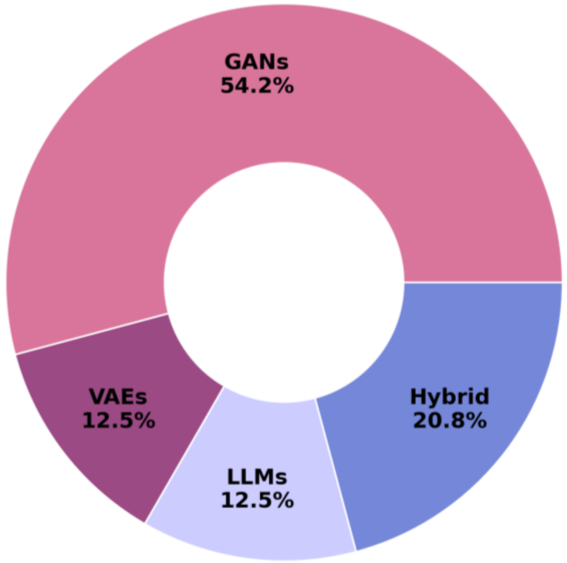


Figure 7 provides a multi-level breakdown of sources, categorizing studies by publication type (journal, conference) and their respective databases. Figure 8 classifies journal articles based on their impact factor range, with the majority falling between 3.1 and 4.0.

Figure 7

Distribution of Studies by Source and Publication Type

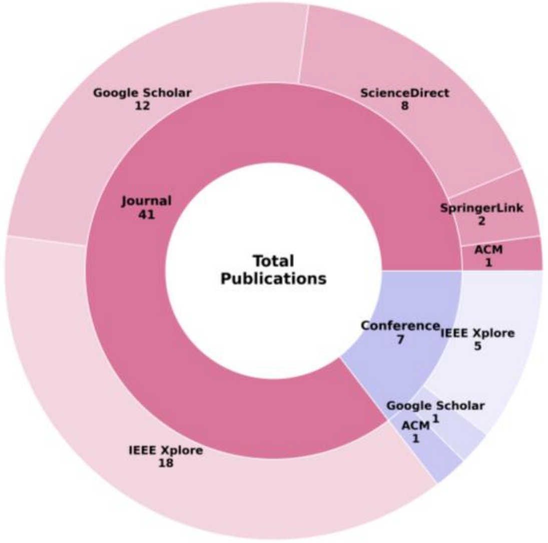


Figure 8

Distribution of Impact Factors

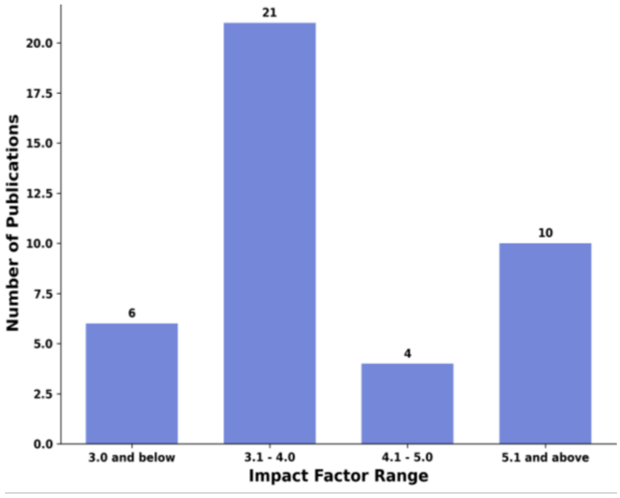
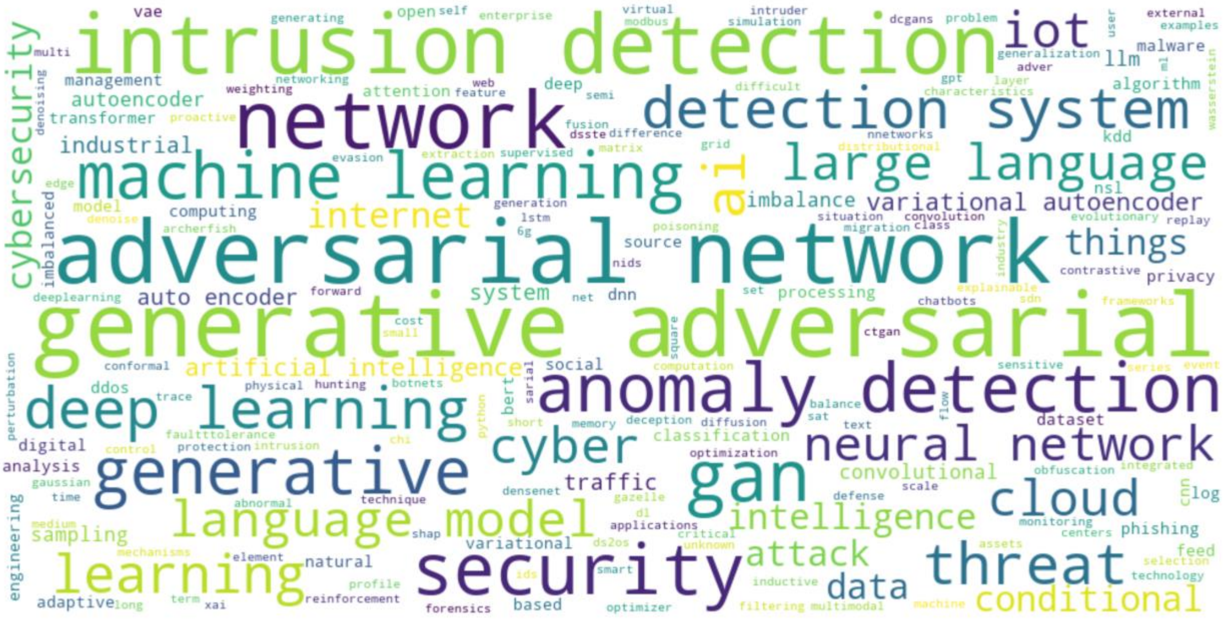


Figure 9 illustrates a graphical representation of keywords and terms extracted from the selected 48 studies, reflecting core themes in GenAI for cybersecurity threat detection.

Figure 9

Keyword Cloud Representing Keywords and Terms from Selected Studies



2.7. Ethics and Ethical Considerations

Ethical considerations refer to the principles and standards that researchers should follow when conducting research in a responsible manner (Goodwin et al., 2020). This study is entirely based on secondary data from peer-reviewed scholarly literature. It does not involve primary data collection, human participants, or personal information. Therefore, ethical concerns, such as consent, privacy, or participant protection, are not applicable to this study. All selected articles are properly cited in the reference section to ensure academic integrity and avoid plagiarism. Similarly, the article selection process was conducted objectively using predefined eligibility criteria without any bias to specific authors, institutions, or publishers. An Ethical Consideration Form was submitted to the Eastern Institute of Technology and formally approved by confirming adherence to the institution's ethical standards.

Moreover, this study aims to evaluate the effectiveness of GenAI in cybersecurity threat detection. GenAI-based activities, such as synthetic data generation and adversarial example creation, can be misused to bypass defense mechanisms, posing significant ethical and security risks. Therefore, researchers must handle these technologies according to the ethical principles. Due to this sensitive nature of cybersecurity, ethical considerations are essential to ensure the transparency and credibility. Overall, this study has been conducted ethically, adhering to institutional guidelines and scholarly standards.

2.8. Chapter Summary

This chapter has presented the research methodology used to conduct this SLR. It has justified the choice of the SLR approach compared to other methods. Furthermore, this chapter has defined key terms, described the search strategy, established eligibility criteria, and explained the screening and selection process. Ethics and ethical considerations have been discussed to ensure compliance with academic and institutional standards. The next chapter will comprehensively review existing research on GenAI in cyberthreat detection.

3. Literature Review

3.1. Chapter Overview

This chapter reviews literature on GenAI in cybersecurity threat detection by addressing key research questions. It evaluates the methodology, findings, significance, and relevance of three landmark studies. Subsequently, it defines key concepts and examines theoretical frameworks from prior research. It critically analyzes selected literature through key themes and summarizes key findings. The final sections analyze the methodological approaches and identify research gaps in the reviewed literature.

3.2. Synthesis of Landmarked Studies

3.2.1. Landmark Study 1: An Enhanced AI-Based Network Intrusion Detection System Using Generative Adversarial Networks

Citation: Park et al. (2023)

Objective: This study developed an AI-based network intrusion detection systems (NIDS) that addresses data imbalances in cybersecurity threat detection. It utilizes GANs, specifically a Boundary Equilibrium GAN (BEGAN) model, to generate synthetic data for minority attack classes and Autoencoder (AE) for feature extraction. The study aims to improve detection accuracy for rare threats that traditional AI methods struggle to identify due to insufficient malicious data.

Methodology: This research used a quantitative approach consisting of four stages.

1. **Preprocessing:** The stage includes outlier removal, one-hot encoding, and min-max normalization.
2. **Generative model training:** The BEGAN was trained to generate plausible synthetic data for minority classes using reconstruction error and Wasserstein distance.
3. **AE training:** Features were extracted by reducing dataset dimensionality
4. **Predictive model training:** Trained three deep learning (DL) classifiers (Deep Neural Network (DNN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM)) using the augmented dataset.

The proposed model was tested using four diverse datasets, including NSL-KDD, UNSW-NB15, IoT-23, and a real-world dataset collected from a large enterprise system. The model's performance was evaluated using accuracy, precision, recall, and F1-score.

Findings: The proposed GenAI-based NIDS outperformed baseline DL models, especially in detecting minority attacks like Remote-to-Local (R2L) and Probe. The proposed G-DNN$_{AE}$ and G-CNN$_{AE}$ models achieved accuracy up to 93.2% on the NSL-KDD dataset. Furthermore, moderate performance improvements were observed on the UNSW-NB15 and IoT-23 datasets. It demonstrated the efficiency of the proposed models in both traditional and distributed environments. Additionally, results on the real-world dataset validated the practical applicability of this model in mitigating data imbalance (see Appendix A for a key performance summary)

Significance: This article is peer-reviewed and was published in 2023 in the reputable IEEE Internet of Things Journal (Q1 ranked, impact factor 8.2). Furthermore, the proposed framework enhanced AI-based NIDS by addressing data imbalance using GANs and an AE and offered a scalable, real-world solution.

Relevance: This study is highly relevant to our research as it applies GANs to mitigate the data imbalance challenge in cyberthreat detection. It focuses on identifying rare cyberattacks by generating plausible synthetic data. Furthermore, these experimental results on diverse datasets provide valuable insights into the effectiveness of GenAI in distributed environments.

3.2.2. Landmark Study 2: Machine Learning with Variational AutoEncoder for Imbalanced Datasets in Intrusion Detection

Citation: Lin et al. (2022)

Objective: This study aims to address key challenges in IDS in complex networks by developing a hybrid model combining a VAE and Multilayer Perception (MLP). The targeted challenges include detecting massive attack variants, handling imbalanced datasets, and optimizing data segmentation. Furthermore, it focuses on enhancing detection accuracy and recall for zero-day and new attack variants in a heterogeneous environment.

Methodology: This study employs a quantitative approach using a VAE to generate synthetic data and MLP for supervised classification. The methodology consists of three phases, namely preprocessing, model training, and evaluation. The study pre-processed data from system logs (HDFS dataset) and network traffic (TTP dataset) using a range-based sequential system (RSS) algorithm to identify the optimal sequence length. Subsequently, the VAE generated synthetic samples for minority classes to address data imbalance. The augmented dataset was used to train the MLP for behavior classification. The performance of this model was evaluated using F1-score, precision, and recall.

Findings: The proposed model achieved significant improvements, with higher recall and F1-score, especially in data imbalance situations (legitimate to attack data ratios: 22:1(HDFS), up to 146:1(TTP)). It achieved a recall of 45%~61% on the imbalanced dataset and 70%~97% on the balanced dataset generated by the VAE. The balanced dataset improved the F1-score by up to 35% and recall by 27% outperforming state-of-the-art IDS solutions. However, system logs showed higher false negatives due to limited features (the grey area effect). This model outperformed the MLP (0%) in detecting attack variants while identifying 100% of Group 1 and 82% of Group2.

Significance: This peer-reviewed article was published in IEEE Access (Q1 rank, impact factor 3.4). Additionally, this study contributes to the IDS through a novel combination of VAE-MLP models and the RSS algorithm. This approach effectively addresses IDS challenges, outperforming traditional approaches.

Relevance: This study is highly relevant to our research as it demonstrates how GenAI with an MLP enhances cybersecurity threat detection by addressing critical challenges in IDS. Furthermore, this study focuses on detecting zero-day attacks and variants in heterogeneous environments, which mitigates modern cybersecurity challenges. Additionally, the experimental results of this study provide insights into the practical applications of GenAI in IDS.

3.2.3. Landmark Study 3: Adversarial Deep Learning Approach Detection and Defense against DDoS Attacks in SDN Environments

Citation: Novaes et al., (2021)

Objective: This study aims to introduce a novel anomaly detection system using a GAN framework to detect and defend against DDoS attacks in Software-Defined Networking (SDN) environments. It focuses on enhancing detection accuracy and resilience against adversarial cybersecurity threats. This system aims to overcome the vulnerability of DNNs using adversarial training to enhance network security and real-time threat detection.

Methodology: This study used a quantitative research methodology and utilized a GAN-based adversarial DL approach with four modules, namely data collection, data processing, anomaly detection, and mitigation. It collected network traffic every second using the OpenFlow protocol and captured IP flow features, like bits, packets, and entropy metrics. These features were processed to extract characteristics for anomaly detection. The GAN with a generator and discriminator, was trained to detect adversarial examples. This system was evaluated using the Mininet emulator and the CICDDoS2019 dataset. It analyzed system performance using the F1-score, accuracy, precision and recall and compared the GAN approach against CNNs, LSTMs and MLPs.

Findings: The GAN-based system achieved high performance with an accuracy of 99.78%, precision of 99.76%, recall of 99.99% and an F1-score of 99.87% in the emulated SDN scenario by outperforming CNNs, LSTMs, and MLPs. On the CICDDoS2019 dataset, it achieved significant recall and F1-scores, demonstrating robustness of proposed model against various DDoS attacks. Additionally, the proposed mitigation module effectively reduced anomalous traffic and restored normal network behavior.

Significance: This article was published in a peer-reviewed, reputable journal, Future Generation Computer Systems (Q1 ranked, impact factor 8.8). This study highlights the efficacy of a GenAI model, especially a GAN, in enhancing resilience to DDoS attacks in SDN environments.

Relevance: This research is highly relevant to our study as it demonstrates the potential of GANs to improve detection accuracy and robustness against adversarial DDoS attacks. This study provides a practical framework for real-world threat detection, which can directly apply to modern network security challenges.

3.3. Definitions/Concepts and Terms

This section explains core concepts and terms relevant to the study of GenAI in cybersecurity threat detection to provide a clear foundation for the literature review (refer to Table 4).

Table 4

Definition/Concepts and Terms

Term	Definition
GenAI	GenAI is a subset of AI that is capable of generating new data or content, such as text, images, audio, or code (Vadisetty & Polamarasetti, 2024). These models learn patterns from an existing dataset and use that knowledge to produce new outputs.
Generative Adversarial Networks (GANs)	A GAN is a type of GenAI model consisting of two neural networks, namely a generator that produces synthetic data and a discriminator that evaluates the authenticity of the data (Nadella et al., 2025). Applications of GAN includes image processing, video generation, and data augmentation. (Alo et al., 2024; Nadella et al., 2025).
Variational Autoencoders (VAEs)	A VAE is a GenAI model that learns latent representations for effective data reconstruction and denoising (Li et al., 2024; Nadella et al., 2025). VAEs are used to model normal behavior and detect anomalies by identifying deviations from normal patterns (Pandian, 2024)
Large Language Models (LLMs)	LLMs are transformer-based models that are trained on large volumes of data to understand and generate realistic data. Common LLMs include the OpenAI GPT series, Google PaLM, and Meta's LLaMA.
Cybersecurity Threat Detection	Cybersecurity threat detection is the process of identifying and mitigating potential security breaches, such as malware, phishing, and unauthorized access, using various tools and techniques.
Synthetic Data Generation	The process of generating artificial data that simulates the characteristics of real-world data. In cybersecurity, synthetic data generation by GenAI models, especially for rare and minority attack addresses data imbalance in training datasets (Park et al., 2023).
Anomaly Detection	Anomaly detection is a technique used to identify abnormal or unusual patterns that indicate cyberattacks or intrusions (Nadella et al., 2025). Generative models can learn the distribution of normal traffic and flag suspicious activities (Pandian, 2024).

3.4. Theories and Models for Prior Researchers

This section explores the theoretical frameworks and models applied in prior research on GenAI in cybersecurity threat detection.

a) **Game Theory:** Game theory is a mathematical framework for modeling strategic interactions between rational decision makers, such as attackers and defenders in cybersecurity (He et al., 2025). It supports the identification of optimal defense strategies. GenAI enhances this by simulating attacker strategies and enabling defenders to optimize their responses.

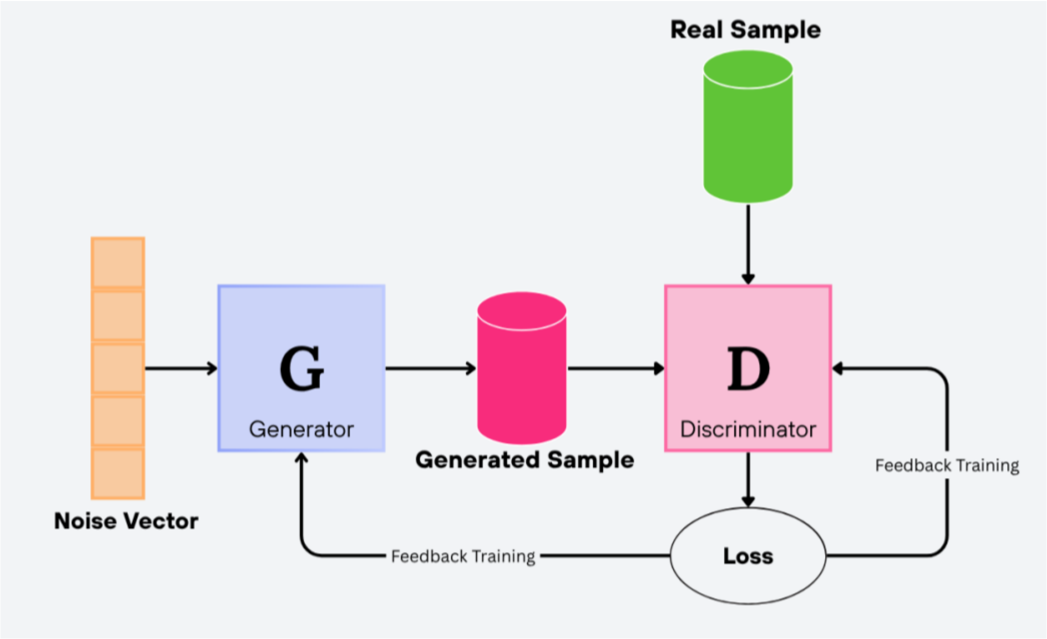
b) **Anomaly Detection Theory:** Anomaly detection theory focuses on identifying deviations from normal data distribution to signal cyberthreats (Li et al., 2023; Pandian, 2024). GenAI applies this theory by creating realistic synthetic datasets to detect anomalies, like unusual network traffic and phishing attacks.

c) **Simulation and Modeling Theory:** This theory focuses on generating simulated environments to replicate real-world scenarios, that enables effective training and testing of systems. GenAI applies it to simulate cyberattacks, enhancing the resilience of defense mechanisms against real cyberthreats.

d) **Generative Adversarial Networks:** GANs are the most widely used GenAI technique, introduced by Goodfellow et al. in 2014, for applications like image synthesis, text generation, and recently, cybersecurity (Chiriac et al., 2025; Nadella et al., 2025). The basic GAN architecture consists of a generator that creates synthetic data samples from random noise (Nadella et al., 2025) and a discriminator that evaluates each sample to distinguish between real and artificial data samples (refer to Figure 10) (Alo et al., 2024; Shieh et al., 2022).

Figure 10

The Basic GAN Architecture

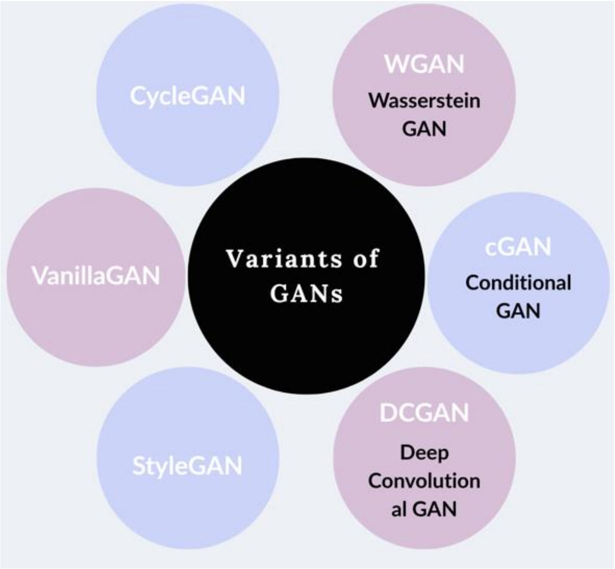


Note. Retrieved from Shieh et al. (2022).

This adversarial minimax game, based on game theory (Cherqi et al., 2023; Park et al., 2023), iterates until the generator produces data that closely matches real data, and the discriminator is unable to differentiate. Several GAN variants (refer to Figure 11) have been developed to address the specific challenges.

Figure 11

Variants of GANs



e) **Variational Autoencoders:** VAEs were introduced by Kingma and Welling in 2013 (Lin et al., 2022). A VAE is an improved version of a traditional AE that learns data distributions using probabilistic modeling (Nadella et al., 2025; Pandian, 2024). These authors mention that the VAE architecture consists of an encoder that converts input data into a set of latent variables and a decoder that reconstructs the output, enabling the learning of complex patterns. Applications of VAEs include image generation (Pandian, 2024), natural language processing (Vadisetty & Polamarasetti, 2024) and anomaly detection (Ren et al., 2023). In cybersecurity, VAEs simulate normal traffic patterns to identify deviations as potential attacks.

f) Transformer-based Models and LLMs: Transformer-based models were introduced by Vaswani et al. in 2017 and utilizes a self-attention mechanism to effectively capture contextual relationships in sequential data (Ferrag et al., 2023; Shafee et al., 2025). The standard transformer model is built with an encoder and a decoder. Senevirathne et al. (2024) highlight that this architecture led to the development of major LLMs, including GPT (Generative Pre-trained Transformers), BERT (Bidirectional Encoder Representation from Transformers). These LLMs excel in understanding and generating human language, which is highly effective for tasks such as machine translation, summarization, and text generation. In cybersecurity, LLMs enhance defensive applications, including phishing email detection (Zhang et al., 2025), and log anomaly detection (Senevirathne et al., 2024).

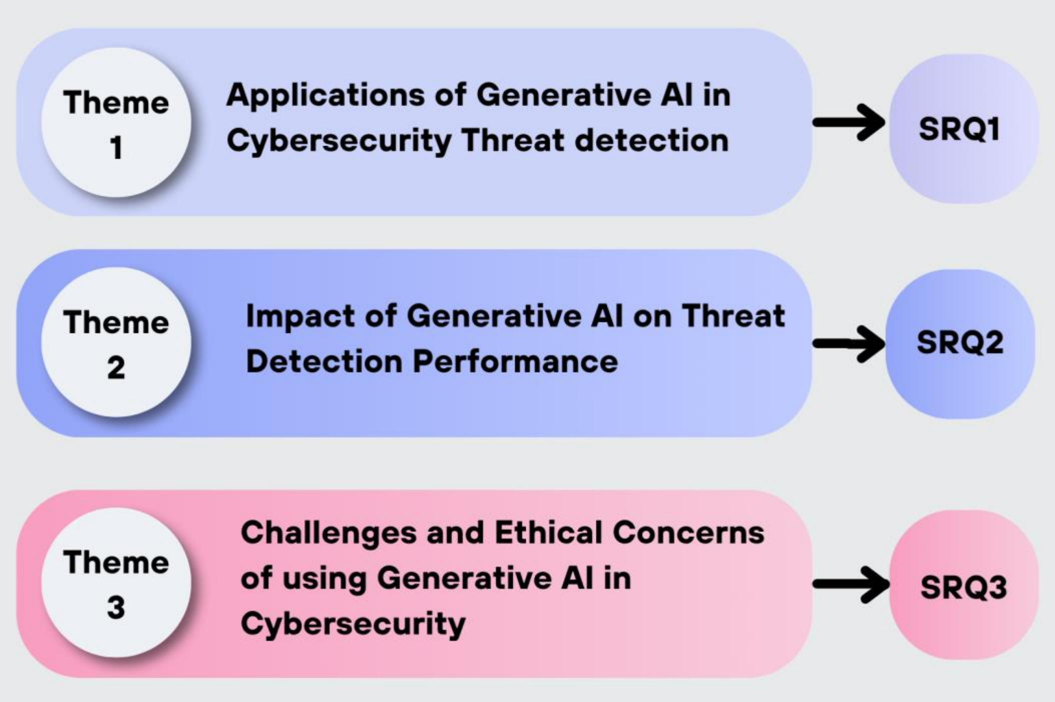
g) Hybrid Models: Hybrid models combine GANs, VAEs, Transformers, ML and DL techniques to improve cybersecurity threat detection. Recent studies have introduced hybrid models, namely ConGAN-BERT (Cherqi et al., 2023), CIDF-VAW-GAN-GOA (Senthilkumar et al., 2024), and CDAAE (Conditional Denoising Adversarial Autoencoder) (Vu et al., 2023), to enhance detection accuracy for cyberattacks.

3.5. Findings from the Previous Literature and Recurring Themes

This section explores key findings of the scholarly literature on the effectiveness of GenAI in cybersecurity threat detection. It is organized into three themes that align with SRQs and their objectives (see Figure 12)

Figure 12

Thematic Framework Mapping Research Themes to SRQs


Theme 1
Applications of Generative AI in Cybersecurity Threat detection
SRQ1
Theme 2
Impact of Generative AI on Threat Detection Performance
SRQ2
Theme 3
Challenges and Ethical Concerns of using Generative AI in Cybersecurity
SRQ3

3.5.1. Applications of GenAI in Cybersecurity Threat Detection

GenAI has become a transformative tool in cybersecurity threat detection, addressing limitations of traditional methods. Advanced GenAI technologies, such as GANs, VAEs, transformer-based models, and hybrid approaches, improve the detection of threats like malware, phishing attacks, network intrusions, and abnormal traffic patterns. This theme explores three key applications, including synthetic data generation, anomaly detection, and adversarial example generation, by answering SRQ1.

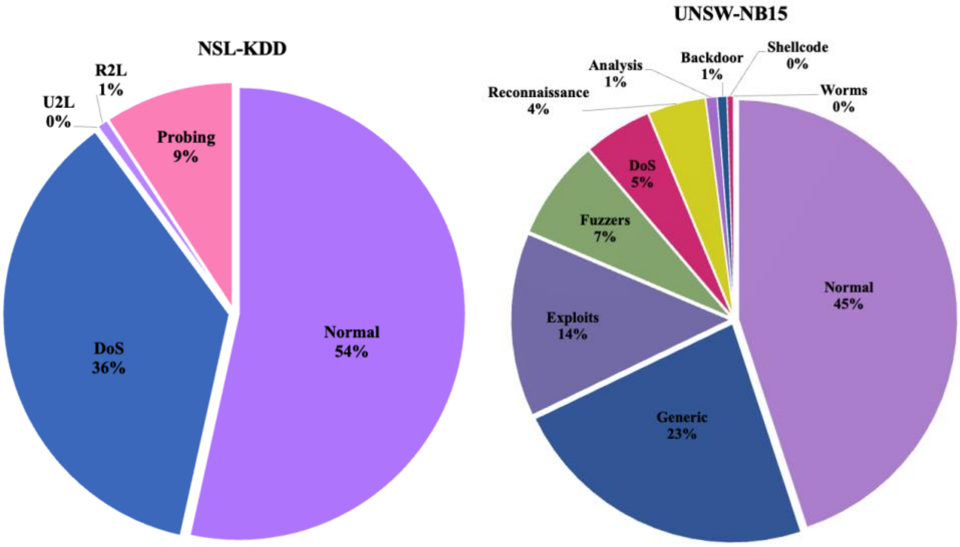
3.5.1.1. Synthetic Data Generation

A key application of GenAI is the generation of synthetic data samples to train detection models. Modern threat detection systems require large, high-quality, labelled, and diverse data (Cherqi et al., 2023;

Hamouda et al., 2024). However, real-world cybersecurity data is often scarce, imbalanced, or restricted due to privacy concerns (see Figure 13). GANs and other generative models address this challenge by producing realistic and diverse synthetic data that mimics real-world threats.

Figure 13

Class Distribution of NSL-KDD & UNSW-NB15



Note. Adapted from Vu et al. (2023).

Most studies in the literature have widely used GANs (Alabrah, 2022; Alo et al., 2024; Benaddi et al., 2022; Moti et al., 2021) and cGANs (Ahsan et al., 2022; Dina et al., 2022; Ullah & Mahmoud, 2021; Yang et al., 2022) to generate balanced synthetic datasets that enable detection systems to learn a broader range of examples. Both GANs and cGANs aim to enhance the performance of threat detection. These methods differ mainly in their architecture, such as the inclusion of class labels in cGANs to control output generation. Some studies utilized advanced GAN variants to create synthetic data. For example, Cao et al. (2024) introduced a network intrusion detection (NID) approach using Denoising Diffusion GAN (DDGAN) combined with a Multi-scale CNN (DDGAN-MCNN) to generate synthetic data for minority classes like "web", "bot", "infiltration", and "heartbleed". Similarly, the study by Bao et al. (2025) utilized a WGAN with Gradient Penalty (WGAN-GP) to generate synthetic malware samples.

Additionally, some studies utilized other GenAI models like VAEs, Adversarial Autoencoders (AAEs), and hybrid models to create realistic data samples. Wasswa et al. (2023) used VAEs to enhance IoT-botnet detection, mitigating high-dimensional, imbalanced data. The study by Vu et al. (2023) proposed a CDAAE and a hybrid model combining CDAAE with the K-Nearest Neighbor (KNN) algorithm. These models were designed to generate synthetic malicious samples in cloud environments, aiming to address unknown attacks and enhance the detection of cyberattacks. Vadisetty and Polamarasetti (2024) introduced a GAN and Transformer-based model to generate synthetic data to train cyber threat hunting (CTH) systems in 6G-enabled IoT networks. These generated samples were used to augment imbalanced training datasets, addressing the challenges of data scarcity and imbalance.

3.5.1.2. Anomaly Detection

Anomaly detection is crucial for identifying cyberattacks, such as intrusions, deception attacks, and malware. GenAI models excel in anomaly detection by learning normal system behavior and flagging deviations as potential threats. Several studies have highlighted the effectiveness of VAE-based approaches in anomaly detection. For instance, Cai and Koutsoukos (2023) proposed a model combining a VAE and a Recurrent Neural Network (RNN) to detect anomalies in cyber-physical systems. This approach encodes sensor data into a latent space and uses RNNs to forecast future inputs. Similarly, Pandian (2024) introduced an innovative application combining VAEs with a CNN for real-time network traffic monitoring to adapt to evolving threats.

Furthermore, research by Lin et al. (2022) integrated a VAE and MLP for unsupervised anomaly detection in IDSs, addressing the imbalanced data problem, numerous attack variants, and data segmentation. For anomaly detection in industrial control systems, Ren et al. (2023) developed a lightweight unsupervised intrusion detection model named LVA-SP using a VAE model. This approach utilized a gate recurrent unit (GRU) and autoregression (AR) modules to maintain accuracy and computational efficiency. Beyond VAEs, some studies applied other frameworks, including GANs and LLMs. Senthilkumar et al. (2024) proposed the CIDF-VAWGAN-GOA model, combining VAE and a WGAN, to distinguish normal and anomalous cloud traffic patterns. The study by Qu et al. (2024) introduced MFGAN (Multimodal Fusion GAN), an innovative multimodal framework with an attention-based AE and GAN to enhance anomaly detection in industrial systems.

3.5.1.3. Adversarial Example Generation

Another innovative application of GenAI is adversarial example generation to test and improve model resilience against evasion techniques. Numerous studies discussed adversarial example generation across various scenarios. Zhou et al. (2022) developed an innovative detection framework using GANs and evolutionary computations to learn the features of unseen attacks, detecting unknown threats in IDS by strengthening IDS capabilities. Furthermore, the work by Liu et al. (2023) introduced a GAN based method called AIGAN (anomaly-based intrusion using GAN) that generates adversarial examples for poisoning attacks on IoT-based NIDS. The authors highlighted the potential of GANs in adversarial training by evaluating their effectiveness against various ML classifiers for threat detection.

Similarly, Mari et al. (2023) utilized a GAN to create adversarial network traffic that bypasses the ML-based IDS. By using these adversarial examples for training, the detection capabilities of ML-based IDS were significantly enhanced, especially against new or modified attacks. However, Choi et al. (2022) noted that GAN-based models may misclassify legitimate inputs. Therefore, they proposed an Adversarially Robust GAN (ARGAN), a defense mechanism that enhances DNN robustness against adversarial examples using a two-step transformation architecture. Shieh et al. (2022) introduced the Symmetric Defense GAN (SDGAN) to detect adversarial DDoS attacks. This model achieved higher performance, outperforming traditional ML models like RF, KNN, and SVM.

In addition to GANs and their variants, some studies have explored other GenAI techniques, such as VAE, and hybrid models. The article by Siniosoglou et al. (2021) presented an IDS called "MENSA" that utilizes an AE-GAN architecture to generate adversarial examples, enhancing cyberattack detection in smart grid environments. Furthermore, Cherqi et al. (2023) introduced ConGAN-BERT, an enhanced semi-supervised GAN-BERT framework with contrastive learning to improve cyberthreat identification in open-source intelligence feeds.

Table 5

Summary of Reviewed Literature Based on Primary GenAI Application

Application	Primary focus of Literature Findings
Synthetic Data Generation	(Aceto et al., 2024; Ahsan et al., 2022; Alabrah, 2022; Alo et al., 2024; Bao et al., 2025; Benaddi et al., 2022; Cao et al., 2024; Chiriac et al., 2025; Constantin et al., 2024; Dina et al., 2022; Hamouda et al., 2024; Kotb et al., 2025; Li et al., 2024; Moti et al., 2021; Nadella et al., 2025; Park et al., 2023; Saikam & Ch, 2024; Shafee et al., 2025; Ullah & Mahmoud, 2021; Vadisetty & Polamarasetti, 2024; Vu et al., 2023; Wasswa et al., 2023; Yang et al., 2022)
Anomaly Detection	(Abdalgawad et al., 2022; Cai & Koutsoukos, 2023; Chernyshev et al., 2023; Demirbaga, 2024a; Ferrag et al., 2023, 2024; Li et al., 2023; Lin et al., 2022; Pandian, 2024; Qi et al., 2024; Ren et al., 2023; Senevirathne et al., 2024; Senthilkumar et al., 2024; Zeng et al., 2025)
Adversarial Example Generation	(Cherqi et al., 2023; Choi et al., 2022; Coppolino et al., 2025; Liu et al., 2023; Mari et al., 2023; Novaes et al., 2021; Shieh et al., 2022; Siniosoglou et al., 2021; Zhou et al., 2022)

This study identifies synthetic data generation, anomaly detection, and adversarial example generation as key applications of GenAI in cybersecurity threat detection. In addition, a few studies, such as those by Heiding et al. (2024) and Zhang et al. (2025) utilized GenAI to detect phishing attacks. Table 5 summarizes the reviewed literature by GenAI applications.

3.5.2. Evaluating the Effectiveness of GenAI in Enhancing Threat Detection Performance

The growing complexity and frequency of cyberthreats require advanced threat detection systems that are capable of detecting sophisticated attacks with high accuracy and reliability. Theme 2 addresses SRQ2 by examining GenAI's impact across two key areas, including improving detection accuracy, and reducing false positives.

3.5.2.1. Enhancement of Cyberthreat Detection Accuracy

GenAI models significantly enhance cybersecurity threat detection, outperforming traditional and ML methods in detecting cyberthreats. By generating more diverse training datasets, synthetic data enable detection models to identify both known and unknown attacks. Most studies used standard performance metrics to evaluate these performance enhancements. These metrics provide a quantitative basis for evaluating how models distinguish between normal activity and malicious attacks.

Standard Metrics

The most commonly used primary metrics in studies are accuracy, F1-score, precision, and recall (refer to Table 6).

- **Accuracy:** Measures the ratio of correctly identified threats to total instances, measuring the overall correctness of the classification (Alo et al., 2024; Ferrag et al., 2023). Out of 48 studies, 30 used accuracy to evaluate the performance of their experiments. However, Abdalgawad et al. (2022) and Wasswa et al. (2023) emphasized that accuracy is less suitable for imbalanced datasets as it can mislead the experimental results
- **Precision:** Measures the ratio of correctly classified attack samples to predicted attacks (Ferrag et al., 2023) indicating a lower false positive rate with higher values (Ren et al., 2023).
- **Recall:** Recall, also known as sensitivity(Moti et al., 2021) or detection rate (Saikam & Ch, 2024), measures the fraction of actual attacks correctly identified (Ferrag et al., 2023).

- **F1-Score:** The harmonic mean of precision and recall (Ferrag et al., 2023; Zeng et al., 2025). Of 48 studies, 38 used the F1-score as a key metric. Several studies (Ren et al., 2023; Vu et al., 2023; Zhang et al., 2025) demonstrated that the F1-score is a comprehensive performance metric with high robustness to class imbalance problems.

Additionally, various metrics, like True Positive Rate (TPR), False Positive Rate (FPR), False Negative Rate (FNR), Area Under the Curve (AUC), execution time, and Matthews correlation coefficient (MCC), were used to evaluate model performance (Kotb et al., 2025; Shieh et al., 2022).

Table 6

Research Method and Key Performance Metrics Used in Reviewed Literature

Article	**Performance Metrics**							**Research Method**
	F1-Score	**Accuracy**	**Precision**	**Recall**	**TPR**	**FPR**	**FNR**	
Abdalgawad et al., 2022	✓							**Quantitative**
Aceto et al., 2024	✓							
Ahsan et al., 2022	✓	✓	✓	✓				
Alabrah, 2022	✓	✓	✓					
Bao et al., 2025	✓							
Benaddi et al., 2022	✓	✓	✓	✓				
Cai & Koutsoukos, 2023						✓	✓	
Cao et al., 2024	✓	✓	✓	✓				
Chernyshev et al., 2023	✓		✓	✓				
Cherqi et al., 2023	✓	✓	✓					
Chiriac et al., 2025	✓	✓	✓	✓				
Choi et al., 2022;		✓						
Constantin et al., 2024		✓	✓	✓				
Coppolino et al., 2025	✓	✓	✓	✓		✓		
Demirbaga, 2024	✓	✓	✓	✓				
Dina et al., 2022	✓	✓						
Ferrag et al., 2023	✓	✓	✓	✓				
Ferrag et al., 2024	✓		✓	✓				
Hamouda et al., 2024		✓	✓			✓	✓	

Article	Performance Metrics							Research Method
	F1-Score	Accuracy	Precision	Recall	TPR	FPR	FNR	
Kotb et al., 2025	✓	✓	✓	✓	✓	✓		Quantitative
Li et al., 2023	✓		✓	✓				
Li et al., 2024	✓	✓	✓	✓				
Lin et al., 2022	✓		✓	✓				
Liu et al., 2023	✓							
Mari et al., 2023	✓	✓	✓	✓				
Moti et al., 2021	✓	✓	✓	✓				
Nadella et al., 2025	✓	✓	✓	✓				
Novaes et al., 2021	✓	✓	✓	✓				
Park et al., 2023	✓	✓		✓				
Qi et al., 2024	✓	✓	✓	✓				
Ren et al., 2023	✓	✓	✓	✓				
Saikam & Ch, 2024	✓	✓	✓	✓		✓		
Senevirathne et al., 2024						✓		
Senthilkumar et al., 2024	✓	✓	✓	✓		✓	✓	
Shafee et al., 2025	✓		✓	✓				
Shieh et al., 2022					✓	✓		
Siniosoglou et al., 2021	✓	✓			✓	✓		
Ullah & Mahmoud, 2021	✓	✓	✓	✓				
Vadisetty & Polamarasetti, 2024		✓				✓	✓	
Vu et al., 2023	✓		✓	✓		✓		
Wasswa et al., 2023	✓		✓	✓				
Yang et al., 2022	✓	✓	✓	✓				
Zeng et al., 2025	✓		✓	✓				
Zhang et al., 2025	✓	✓	✓	✓				
Zhou et al., 2022	✓		✓	✓				
Alo et al., 2024		✓	✓	✓	✓	✓		Mixed
Heiding et al., 2024		✓						
Pandian, 2024			✓	✓				

Note. Used - ✓, Not Used - empty

Datasets

Similarly, the effectiveness of GenAI in threat detection greatly depends on the quality and characteristics of training and evaluation datasets. Variations in dataset size, diversity, and authenticity directly affect performance metrics (Cao et al., 2024; Mari et al., 2023). As shown in Figure 14, of 48 selected studies, 43 unique datasets were identified. The most frequently used datasets are described in Table 7.

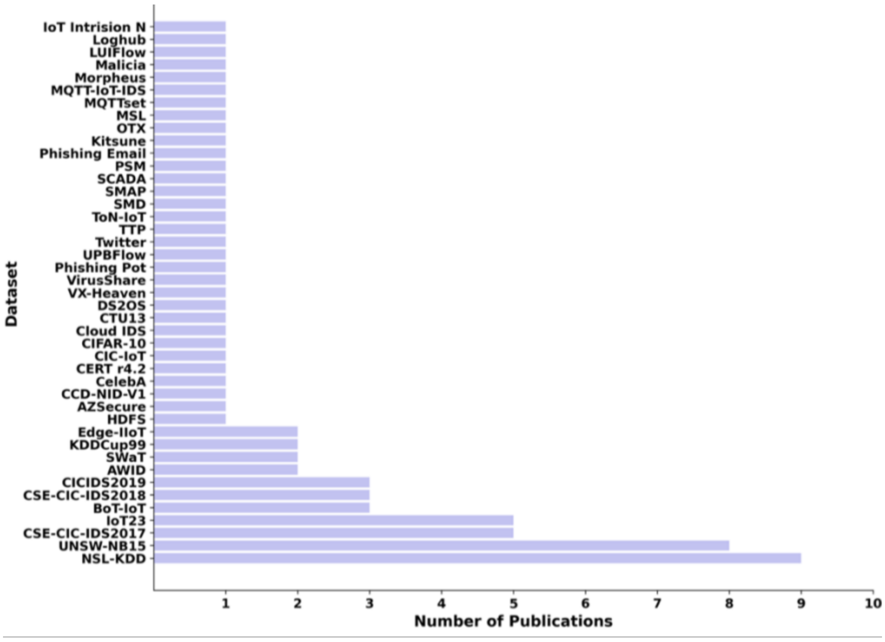
Table 7

Key Datasets Used in Studies

Dataset	Features	Attack Types	Description
NSL-KDD	41	DoS, Probe, R2L, U24	Benchmark dataset for intrusion detection (refined KDDcup99) (Mari et al., 2023; Vu et al., 2023)
UNSW-NB15	49	Fuzzers, Analysis, Backdoors (9 Types)	Normal and synthetic attack traffic (Dina et al., 2022; Yang et al., 2022)
CSE-CIC-IDS2017	76	15 different cyberattacks	Covers complex network scenarios (Cao et al., 2024)
IoT23	21	7 malware, 3 benign captures	IoT network traffic (Aceto et al., 2024)

Figure 14

Frequency of Dataset Usage



Additionally, a few studies used real-world data as part of their experiments. For instance, Park et al. (2023) collected real network flow data with raw security events from a large enterprise system. Similarly, Demirbaga (2024) collected raw log and performance data from Hadoop clusters using "SmartMonit".

Performance Enhancement

GenAI models enhance detection accuracy compared to traditional methods, especially identifying zero-day attacks. Selected studies reported significant performance improvements across various metrics and datasets using diverse GenAI approaches. GAN-based adversarial training improved DDoS detection in software-defined networking environments (Novaes et al., 2021). The study by Zhou et al. (2022) introduced an intrusion detection method (IDM-GE) that combines a GAN and an evolutionary algorithm to identify unknown threats. It achieved over 90% accuracy and recall on the CSE-CIC-IDS2018 dataset compared to baseline methods like oversampling, and SMOTE.

Furthermore, VAE-based models demonstrated substantial improvements in cyberthreat detection. For example, Lin et al. (2022) introduced a VAE-MLP framework that enhanced the F1-score by up to 35% and recall by 27% on the HDFS dataset, outperforming state-of-the-art IDS solutions. Aceto et al. (2024) showed that a NIDS trained on Conditional Variational Autoencoder-generated (CVAE) synthetic data had minimal F1-score loss compared to training on real-data. Additionally, hybrid model, including ConGAN-

BERT (Cherqi et al., 2023), CIDF-VAWGAN-GOA (Senthilkumar et al., 2024), and MENSA (Siniosoglou et al., 2021), demonstrated GenAI capabilities in cybersecurity threat detection. For instance, a VAE-WGAN-based intrusion detection method achieved 83.45% accuracy and 83.69% F1-score on the NSL-KDD dataset and over 98.9% on the AWID dataset, outperforming traditional ML, DL, and deep reinforcement learning models (Li et al., 2024). These performance improvements validate GenAI's effectiveness in handling data imbalances and enhancing cybersecurity threat detection.

3.5.2.2. Reduction of False Alarm Rates using GenAI Models

Traditional cybersecurity tools often struggle with false positives (classifying normal activity as attacks) and false negatives (failing to detect actual attacks). GenAI contributes to reducing these errors by refining detection models and improving discrimination. The GAN-based NIDS proposed by Alo et al. (2024) achieved a significantly lower FPR of 2.4% outperforming SVM (5.1%) and RF (4.7%). This reduction in false positives improved efficiency and system reliability. Shieh et al. (2022) demonstrated that the SDGAN model outperformed traditional models (RF, KNN, SVM, Naive Bayes (NB)) by achieving a TPR of 85.7% on the NSL-KDD dataset and 87.2% on the CIC-IDS2018 dataset. It maintained a 70.9% TPR compared to RF(9.4%), achieving higher detection performance for unknown adversarial attacks. The work by Hamouda et al. (2024) introduced the FedGENID framework, which improved zero-day attack detection by minimizing benign traffic misclassification and achieving lower FPR and FNR for the "Normal" class.

The study by Cai and Koutsoukos (2023) proposed a VAE-based approach for detecting deception attacks on a real-world dataset, which outperformed other methods by reducing both FPR and FNR to less than 10%. Similarly, Vadisetty and Polamarasetti (2024) proposed a GAN-VAE-based NIDS model and achieved a 10% increase in detection accuracy, a 6% reduction in FPR, and a 7% reduction in FNR compared to traditional methods.

Additionally, GenAI-based models enhance the efficiency of threat detection systems by reducing detection time (Nadella et al., 2025) and optimizing resource utilization (Senthilkumar et al., 2024). However, as discussed in the following section, their success depends on addressing challenges, including high computational demands and extended training times.

3.5.3. Challenges and Ethical Concerns of GenAI in Cybersecurity Threat Detection

As mentioned in Themes 1 and 2, GenAI significantly enhances cybersecurity threat detection by generating synthetic data, simulating attacks, and improving model robustness. However, its practical deployment faces significant challenges. This section explores three critical challenges, with a focus on answering SRQ3.

3.5.3.1. Adversarial Vulnerabilities and Dual-Use Risk

As discussed previously, GenAI models have become a powerful defensive approach for cybersecurity. However, they are often vulnerable to adversarial attacks (Ferrag et al., 2023). As defined by Choi et al. (2022), adversarial examples refer to input instances with small, intentional modifications designed to mislead threat detection mechanisms. Attackers may utilize these techniques to generate false data that mimics legitimate inputs (Hamouda et al., 2024), leading to misclassification or bypassing GenAI-driven detection systems (Alo et al., 2024; Coppolino et al., 2025; Hamouda et al., 2024). Shieh et al. (2022) highlighted that cybercriminals use adversarial ML (AML) techniques to identify weaknesses in detection systems and mislead them with adversarial attacks. Similarly, Liu et al. (2023) described the concept of transferability, which attackers use to create adversarial inputs across systems without detailed knowledge of the target systems.

Furthermore, the dual-use nature of GenAI, including both offensive and defensive capabilities, raises significant risks. Attackers can exploit defensive tools like GANs to create sophisticated threats, like polymorphic malware, phishing emails or synthetic traffic to mask attacks (Vadisetty & Polamarasetti, 2024). Coppolino et al. (2025) demonstrated how cGANs were used as an attack obfuscation strategy to conceal real attacks from IDS, emphasizing GenAI's offensive capabilities.

3.5.3.2. Computational Resource Constraints and Scalability

According to most studies, GenAI models, including GANs (Alo et al., 2024; Benaddi et al., 2022; Coppolino et al., 2025) and transformer-based architectures (Ferrag et al., 2023, 2024), requires significant computational resources for training and deployment. As mentioned by Ferrag et al. (2023), LLMs like GPT-3 and GPT-4 require extensive computational power to train and operate. Similarly, GAN-augmented detection systems require substantial computational overhead and require prolonged training times (Alo et al., 2024; Moti et al., 2021; Shafee et al., 2025). For example, Bao et al. (2025) emphasized that converting malware files into an image format using models like GANs and VAEs is computationally expensive and time consuming.

Additionally, these systems require high-performance GPUs and larger amounts of memory, which lead to cost and feasibility challenges for many organizations (Nadella et al., 2025; Saikam & Ch, 2024; Shafee et al., 2025). For instance, Alo et al. (2024) reported that a GAN-based NIDS required significantly more computational resources, including up to 10.5 hours training time, 85% GPU utilization, 18.7 GB of memory. This approach significantly exceeded traditional models, such as SVM (0.5 hours) and RF (1.2 hours), in training time and had considerably higher GPU and memory usage. However, Cherqi et al. (2023) noted that, despite high computational costs, the improved accuracy justifies the investment for organizations aiming to enhance cybersecurity. Furthermore, these computational overheads create scalability issues in real-world settings. For this reason, deployment becomes challenging in distributed networks or resource constraint environments, like IoT, small and medium enterprises (SMEs) (Ferrag et al., 2023). Expert interviews in Alo et al. (2024) supported this concern, as 11 out of 15 experts highlighted scalability as a major challenge.

3.5.3.3. Ethical and Privacy Concerns

Ethical and privacy concerns are a major issue with advancement in GenAI models. Their dual-use nature can be utilized to produce sophisticated attacks, such as phishing emails and polymorphic malware (Vadisetty & Polamarasetti, 2024). For instance, research by Heiding et al. (2024) demonstrated how LLMs could generate phishing emails by rephrasing prompts (e.g., from ‘‘phishing email’’ to ‘‘informative email’), bypassing ethical restrictions. Detection systems struggle to differentiate the malicious intent from legitimate use. This highlights the difficulty of controlling the malicious use of GenAI.

Additionally, there are significant privacy risks associated with synthetic data generation (Aceto et al., 2024; Hamouda et al., 2024). Several researchers have proposed privacy-preserving techniques to mitigate these privacy risks. For example, Aceto et al. (2024) proposed a method using a CVAE with non-reversible binning to prevent data leakage. Similarly, Ferrag et al. (2024) introduced "SecurityBERT", using a privacy-preserving encoding technique (PPFLE) and Byte-Pair-Encoding (BBPE) tokenization to ensure the privacy of network data.

In summary, GenAI demonstrates significant enhancements in cybersecurity threat detection. However, its effectiveness is constrained by adversarial attacks, ethical and privacy risks, high resource demands, and scalability issues. Additionally, data quality, integration, and interpretability issues limit the real-world deployment of GenAI-based detection mechanisms.

3.6. Summary of Findings from Section 3.5

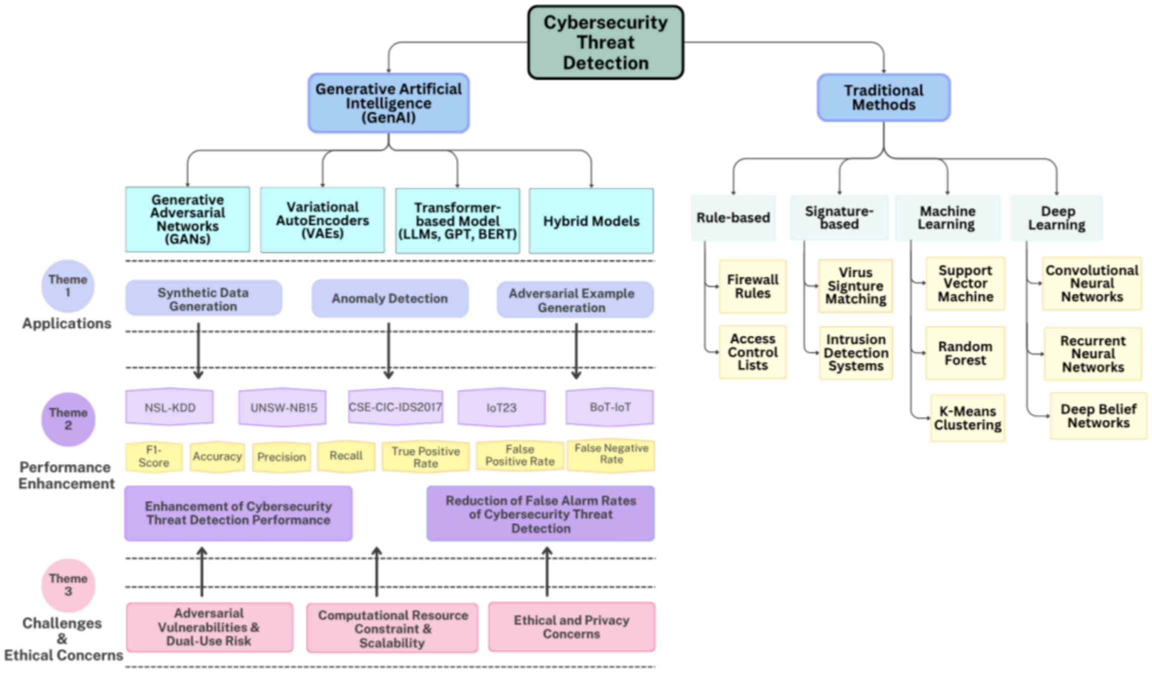
Section 3.5 examines the effectiveness of GenAI in cybersecurity threat detection using peer-reviewed scholarly articles. The findings are organized into three key themes: applications of GenAI in cyberthreat detection, measuring the effectiveness of GenAI in performance enhancement, challenges and ethical concerns associated with use of GenAI in threat detection as represented in Figure 15. These themes address the research objectives of identifying key applications, evaluating performance impact, and exploring drawbacks and risks.

- **Applications of GenAI:** This study highlights three primary applications of GenAI techniques, including GANs, VAEs, and LLMs and hybrid models. The first application is synthetic data generation, which addresses the data scarcity and imbalance by producing realistic synthetic data samples for training detection models. The second application is anomaly detection, which uses VAEs and hybrid models to identify deviations in network traffic or system behavior. The third application is adversarial example generation, which enhances system resilience by simulating sophisticated attacks. Furthermore, some studies explored the use of GenAI in detecting phishing attacks and analyzing malware. These applications emphasize the effectiveness of GenAI models in detecting cyberattacks.
- **Evaluating impact of GenAI on threat detection performance:** GenAI significantly enhances cybersecurity threat detection performance, outperforming traditional detection methods. It improves threat detection across multiple datasets and performance metrics. Key performance metrics, such as accuracy, precision, recall, and F1-score are used to measure model performance. Of the 48 reviewed studies, 38 used the F1-score because of its reliability and robustness in handling imbalanced datasets. Commonly used datasets include NSL-KDD, UNSW-NB15,CICIDS2017 and IoT23. Furthermore, GenAI improves detection accuracy by generating synthetic data to augment limited or imbalanced real-world datasets. Notable improvements include a 35% increase in the F1-score achieved by (Lin et al., 2022) and over 90% accuracy for GAN-based intrusion detection (Zhou et al., 2022). Similarly, GenAI reduces false positives by refining detection models and distinguishing benign from malicious activities.

- **Challenges and Ethical Considerations:** In contrast, GenAI faces significant challenges in the cybersecurity threat detection. One major issue is adversarial vulnerabilities, which allow attackers to exploit manipulated inputs or advanced threats like polymorphic malware. This highlights the dual-use nature of GenAI techniques in threat detection. Furthermore, computational resource demands, such as high GPU utilization and prolonged training times, limit the scalability of models, especially in real-time applications and resource-constrained environments. Additionally, ethical and privacy concerns arise from the potential misuse of GenAI to create realistic attacks. These challenges emphasize the need for a robust GenAI-based defense approach and ethical guidelines.

Figure 15

Overview of Key Findings of This SLR



Cybersecurity Threat Detection
Generative Artificial Intelligence (GenAI)
Traditional Methods
Generative Adversarial Networks (GANs)
Variational AutoEncoders (VAEs)
Transformer-based Model (LLMs, GPT, BERT)
Hybrid Models
Rule-based
Signature-based
Machine Learning
Deep Learning
Firewall Rules
Access Control Lists
Virus Signture Matching
Intrusion Detection Systems
Support Vector Machine
Random Forest
K-Means Clustering
Convolutional Neural Networks
Recurrent Neural Networks
Deep Belief Networks
Theme 1
Applications
Synthetic Data Generation
Anomaly Detection
Adversarial Example Generation
Theme 2
Performance Enhancement
NSL-KDD
UNSW-NB15
CSE-CIC-IDS2017
IoT23
BoT-IoT
F1-Score
Accuracy
Precision
Recall
True Positive Rate
False Positive Rate
False Negative Rate
Enhancement of Cybersecurity Threat Detection Performance
Reduction of False Alarm Rates of Cybersecurity Threat Detection
Theme 3
Challenges & Ethical Concerns
Adversarial Vulnerabilities & Dual-Use Risk
Computational Resource Constraint & Scalability
Ethical and Privacy Concerns

3.7. Methodological Approaches used in Previous Studies

This section critically evaluates the methodologies of 48 peer-reviewed studies on GenAI for cybersecurity threat detection, focusing on research methods, sampling, data collection, and analysis.

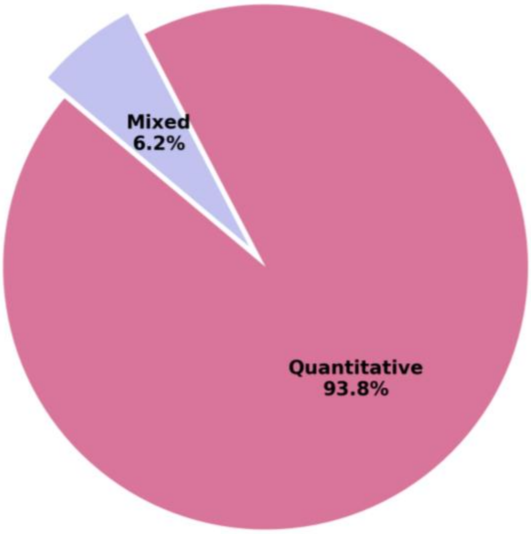
3.7.1. Research Methods

The reviewed studies primarily utilized quantitative experimental evaluations to assess proposed models, and techniques in a cybersecurity context. These evaluations often compared GenAI-based models with state-of-the-art techniques. This SLR analyzed 48 studies employing a quantitative or mixed method approach (see Table 6). Of these, 93.8% (45 studies) used quantitative methods, using standard metrics and statistical analysis to evaluate model performance (Alabrah, 2022; Ferrag et al., 2023, 2024; Vu et al., 2023) (see Figure 16). This quantitative approach provides empirical evidence for technical validity and the reproducibility of results.

Conversely, the remaining 6.2% (3 studies) used a mixed-method approach, combining quantitative results with qualitative insights gained from experiments, practitioner interviews, and discussions of practical limitations (Alo et al., 2024; Heiding et al., 2024) (refer to Table 6). For instance, Alo et al. (2024) combined empirical evaluations of a GAN-based NIDS with expert interviews, providing a balanced perspective on research findings.

Figure 16

Methodological Distribution of Reviewed Studies



3.7.2. Sampling Methods

Sampling is the technique of selecting a subset from a population to analyze (Bryman & Bell, 2011, p.172). Most reviewed studies used non-probabilistic sampling methods, including purposive (Bao et al., 2025; Coppolino et al., 2025; Shafee et al., 2025) and convenience sampling (Constantin et al., 2024; Heiding et al., 2024). Purposive sampling targets relevant datasets or specific subjects to align with research objectives. For instance, Hamouda et al. (2024) purposively chose the Edge-IIoTset dataset for its IoT relevance, and Zhang et al. (2025) selected 40 emails from four categories for phishing detection. Some studies used convenience sampling, for example Heiding et al. (2024) recruited 112 university students via flyers and emails based on availability. Conversely, some studies used probabilistic methods, like random sampling (Li et al., 2023; Yang et al., 2022; Zhou et al., 2022). For instance, Zeng et al. (2025) randomly sampled 40% - 80% of normal data for training, and (Choi et al., 2022) selected 1000 images per class from the CIFAR-10 dataset.

3.7.3. Data Collection Approaches

Data collection is the systematic process of gathering data to gain insights into a specific research objective (Taherdoost, 2021). In most reviewed studies that employed quantitative methods, researchers primarily utilized pre-existing, publicly available benchmark datasets or generated synthetic data using GenAI models (Benaddi et al., 2022; Li et al., 2023; Wasswa et al., 2023). Theme 2 discussed the frequently used datasets by in the reviewed articles. Additionally, a few studies focused on real-world data collection. For instance, Park et al. (2023) combined data from three public datasets with real-world enterprise network flow and security event data, logged and labeled by Security Operations Center (SOC) analysts over five months. Furthermore, mixed method studies combined statistical data with qualitative approaches, like semi-structured interviews and surveys. Alo et al. (2024) used existing datasets and semi-structured expert interviews regarding the deployment of a GAN-enhanced NIDS. Heiding et al. (2024) collected click-through rates from phishing emails and qualitative survey data from university students.

3.7.4. Data Analysis

Data analysis is the systematic process used by researchers to evaluate collected data (Bryman & Bell, 2011, p.144). The reviewed quantitative studies utilized statistical techniques like descriptive, inferential, and advanced ML and DL methods. Most of these studies used performance metrics for ML and DL models, as discussed in Theme 2 (Ahsan et al., 2022; Li et al., 2023; Nadella et al., 2025; Novaes et al., 2021; Yang et al., 2022). Additionally, some studies used descriptive analysis (Constantin et al., 2024) and inferential statistics (Coppolino et al., 2025). Conversely, mixed method studies combined quantitative and qualitative analysis approaches, like thematic and content analysis (Alo et al., 2024; Heiding et al., 2024). For instance, Alo et al. (2024) combined quantitative techniques, like GAN-based model training with performance metrics and comparative benchmarking, with qualitative thematic analysis of expert interviews.

3.8. Research Gaps in Previous Literature

This study identified several gaps in the reviewed studies on GenAI for cybersecurity threat detection. Most studies focused on specific cyberthreats, like malware, anomaly, and intrusion detection, through methods such as generating synthetic malware samples (Bao et al., 2025; Moti et al., 2021) or detecting anomalies (Benaddi et al., 2022; Li et al., 2023). However, critical threats like phishing, DDoS attacks, and insider threats remain underexplored. To address this gap, researchers should evaluate the effectiveness of GenAI against diverse threats, including social engineering attacks, APT, and phishing attacks.

Scalability and real-time applicability are other underexplored challenges in the reviewed studies. Computationally intensive models, like GANs, and Transformers, limit real-time threat detection, especially in resource-constraint environments like IoT networks (Ferrag et al., 2023) and SMEs (Zhang et al., 2025). Some studies proposed lightweight models (Ren et al., 2023), but the trade-off between computational efficiency and detection accuracy has not been thoroughly explored. Moreover, the integration of GenAI into existing cybersecurity systems, like Security Information and Event Management (SIEM) systems, is rarely addressed. Additionally, most studies conducted their experiments using benchmarked datasets in controlled, simulated environments rather than real-world scenarios, which limits insights into practical performance.

Another research gap is the limited discussion of the ethical and privacy implications of GenAI in cybersecurity. Synthetic data generated by these models could be misused in cyberattacks, like phishing or deepfakes, and poses privacy risks by replicating sensitive information. However, only a few studies proposed a robust ethical framework or privacy-preserving techniques to adapt GenAI for cybersecurity. Therefore, future research should prioritize developing ethical guidelines and privacy protection techniques to address these gaps.

3.9. Chapter Summary

This chapter provided a SLR on GenAI in cybersecurity threat detection analyzing three landmark studies, defining key concepts and theories. Furthermore, by analyzing 48 studies, it covered applications, performance impact, and challenges, and ethical concerns. Additionally, methodological approaches were reviewed, and research gaps were identified. The next chapter will highlight the research's contributions, identify limitations, and propose future directions.

4. Discussion

4.1. Chapter Overview

This chapter discusses the contributions of this research to the field of cybersecurity by focusing on GenAI for threat detection. It highlights the theoretical, practical, and methodological contributions of this SLR. Additionally, it addresses the limitations of this study and proposes future research directions.

4.2. Contribution of the Research

This study makes theoretical, practical, and methodological advancements that improve the efficacy of GenAI approaches in cybersecurity threat detection.

4.2.1. Theoretical Contribution

This SLR significantly enhances the theoretical understanding of GenAI's effectiveness in cybersecurity threat detection by critically evaluating 48 peer-reviewed studies. It provides a clear, structured framework by categorizing key applications, such as synthetic data generation, anomaly detection, and adversarial example generation. This approach enhances the conceptual understanding of how GenAI models, such as GANs, VAE, and LLMs, address cybersecurity challenges. By evaluating the performance of these models across different datasets (NSL-KDD, CICIDS2017) and performance metrics (accuracy, F1-score), this SLR provides insights into their strengths, like improved detection of zero-day attacks, and limitations, like vulnerability to adversarial inputs. This comparative analysis contributes to theoretical discussions on model selection and contextual factors that influence performance.

Furthermore, it identifies critical challenges, including adversarial vulnerabilities, high resource demands, and ethical and privacy concerns, emphasizing the need for robust frameworks, such as hybrid models or lightweight architectures, to enhance scalability and resilience. Similarly, it recognizes research gaps, including underexplored threats (phishing, insider attacks), real-world scalability issues, and integration with existing systems, providing a roadmap for future theoretical advancements. These insights guide researchers toward developing adaptive, ethically responsible models to strengthen the application of GenAI in cyberthreat detection.

4.2.2. Practical Contribution

This SLR offers practical insights for cybersecurity professionals, organizations, and policymakers. It synthesizes key findings from the literature to enhance cybersecurity practices, improve defense mechanisms, and guide decision making. For instance, organizations with sensitive data, like those in finance and healthcare, can use GenAI models, such as GANs, VAEs, and DDGAN-MCNN, to generate synthetic data, addressing data scarcity and imbalance to train robust detection systems. Cybersecurity professionals can apply these insights to implement GenAI solutions that improve the detection of rare and zero-day attacks, like APT or polymorphic malware, reducing vulnerabilities in organizations.

Furthermore, this study evaluates the impact of GenAI on threat detection using key performance metrics and demonstrates significant improvements over traditional methods. Cybersecurity teams can use these evidence-based insights to benchmark and refine detection systems, improving operational efficiency and response times. This SLR highlights GenAI's computational demands and scalability challenges and suggests lightweight solutions, like LVA-SP (Ren et al., 2023) for constrained environments. This information support organizations in making decisions about adopting GenAI-based solutions. The study's findings on ethical and privacy risks associated with GenAI, guide organizations and policy makers in implementing solutions that ensure compliance with ethical standards and data protection regulations.

4.2.3. Methodological Contribution

This SLR on the effectiveness of GenAI in cybersecurity threat detection significantly enhances the methodological approach. Following PRISMA guidelines, it ensures transparency and reproducibility. Using a systematic and structured approach, it synthesizes diverse literature and provides a methodological framework to guide future research. Its comprehensive search strategy across databases like IEEE Xplore, and ACM, focusing on peer-reviewed Q1 and Q2 journals and conference proceedings published between 2021 and 2025, reflects the latest developments in the field. Defining eligibility criteria (IC1, IC2, EC3, EC4) ensures the quality, and relevance of studies, while a multistage screening process, including citation chaining, improves comprehensiveness.

Similarly, this SLR focus on the recent literature reflects the rapid evolution of GenAI in cybersecurity. Its structured thematic approach analyzes GenAI applications, performance metrics, and challenges ,addressing research objectives. By excluding other SLRs (EC3), this systematic approach offers a credible and updated analysis that supports researchers and practitioners in enhancing threat detection approaches.

4.3. Limitations

This study provides a comprehensive SLR but has several limitations that may affect its scope, generalizability, and depth due to its methodology and focus. A key limitation is that it synthesizes existing research without conducting primary experiments or collecting new data, which limits the verification of findings. The literature search was limited by the specific academic databases and the exclusion of non-peer-reviewed sources (IC2), and non-English articles (IC5), potentially missing relevant research. Additionally, reliance on high quality peer-reviewed articles may introduce bias and limit generalizability.

Furthermore, its reliance on synthetic data generated by GenAI models introduces potential biases and may not fully represent real-world cyberthreats accurately. By focusing on major GenAI models, this study may have missed insights from emerging methods, like diffusion models which may limit its technical scope. These constraints reduce the generalizability of the findings and suggest future research directions, including broader literature coverage, inclusion of diverse sources, and empirical validation of GenAI in cybersecurity threat detection.

4.4. Future Directions

Future research should explore integrating GenAI with cutting-edge technologies, such as blockchain, to enhance secure data sharing and the reliability of threat intelligence. Similarly, the use of quantum computing may improve data processing speed, enabling faster real-time threat detection in cybersecurity (Vadisetty & Polamarasetti, 2024). Researchers should further investigate innovative hybrid models to effectively respond to emerging threats, like zero-day attacks and APTs (Saikam & Ch, 2024). Additionally, future work should focus on optimizing GenAI algorithms for improved performance in large-scale environments and developing efficient, lightweight models that require less computational power. Another area for future research is enhancing GenAI interpretability by utilizing techniques like explainable AI (Demirbaga, 2024; Kotb et al., 2025). This would enable clearer decision-making and build trust in cybersecurity systems. These future research directions address current limitations and enhance the GenAI's role in adaptive and secure threat detection in cybersecurity.

4.5. Chapter Summary

This chapter summarizes the findings of the reviewed literature, highlighting the contributions of this study, identifying its limitations, and proposing future directions. The next chapter will conclude the study by covering the research purpose, identified gaps, methodology, key findings, limitations, and future directions.

5. Conclusion

This study critically evaluated the effectiveness of GenAI techniques in enhancing cybersecurity threat detection, guided by the main research question and three sub-questions. It employed a SLR approach, adhering to PRISMA guidelines, ensuring transparency and reproducibility. This SLR synthesized findings from 48 peer-reviewed articles published between 2021 and 2025. The research aimed to identify key applications, evaluate performance impacts, and explore challenges associated with GenAI in this domain.

This SLR identified three core applications of GenAI in cybersecurity threat detection, namely synthetic data generation, anomaly detection, and adversarial example generation. Synthetic data generation, primarily driven by models like GANs, VAEs and LLMs, addresses critical challenges, such as data scarcity and class imbalances. These models create realistic datasets that enhance the training of detection systems. This is particularly important for enhancing the detection of rare and zero-day attacks, as demonstrated by recent studies showing that synthetic data significantly enhances detection accuracy. Additionally, of the 48 reviewed studies, 23 primarily focused on synthetic data generation, and of those, 70% used GANs and their variants, like cGANs, and WGAN. Anomaly detection leverages generative models to learn patterns in normal behaviour and identify deviations as potential threats. In particular, VAEs and their variants (LVA-SP, VAE-CNN) showed effectiveness in real-time monitoring, emphasizing higher performance in detecting anomalies. Furthermore, to strengthen system resilience, adversarial example generation is used to simulate sophisticated attacks, using GenAI to mimic real-world threat scenarios. In addition, a few studies discussed the use of GenAI for detecting phishing attacks, malware analysis and DDoS detection. These applications highlight GenAI's capacity to mitigate challenges in traditional threat detection methods.

GenAI models outperform traditional and early ML methods by leveraging diverse generative synthetic data to accurately identify both known and novel attacks. Most reviewed studies utilized standard evaluation metrics, such as accuracy, precision, recall, and the F1-score, to quantify performance. In particularly, the F1-score proved effective for addressing class imbalance in threat detection. Additional metrics, like TPR, FPR, FNR, AUC and MCC, were utilized to provide a comprehensive evaluation of model performance. The efficacy of GenAI-based detection systems depends heavily on the quality and diversity of the datasets used for training and evaluation. Benchmarked datasets, like NSL-KDD, UNSW-NB15, and CICIDS2017, provide diverse, realistic scenarios for training and testing models. Empirical results from reviewed studies demonstrated the potential of GenAI models to detect sophisticated attacks. Furthermore, GenAI reduces false positives and negatives, enhancing detection system reliability. Several studies reported significantly lower false alarm rates compared to traditional models, like SVM and RF. These improvements emphasize the ability of GenAI to enhance both reliability and accuracy in cybersecurity threat detection systems.

Despite these advancements in GenAI, several challenges remain in the field of cybersecurity threat detection. A major challenge is its vulnerability to adversarial attacks. Attackers can produce sophisticated inputs designed to bypass detection, including phishing campaigns and polymorphic malware. This dual-use nature of GenAI, working as both a defensive and an offensive tool, emphasizes the need for a robust defence mechanism. Furthermore, this study found that the high computational demands of these models, such as the need for powerful GPUs and prolonged training times, limit their scalability in resource constrained environments, like IoT networks and SMEs. In addition to technical limitations, ethical and privacy concerns result from the misuse of synthetic data, which can create highly realistic attack simulations and potentially expose sensitive information. The findings of this SLR emphasize the need for balanced strategies that optimize the benefits of GenAI while minimize its challenges.

This study proposed several future research directions to support its continued evolution. Exploring the integration of GenAI with technologies like blockchain could enhance secure data sharing and address privacy concerns. Similarly, leveraging quantum computing could accelerate processing times for real-time threat detection, which is currently limited by computational demands. Furthermore, developing hybrid models that combine multiple GenAI techniques could enhance defences against novel and sophisticated threats. Another future direction is optimizing GenAI algorithms for efficiency and developing lightweight models to improve scalability. Similarly, improving the interpretability of GenAI through explainable AI techniques can enhance trust and adoption in real-world applications.

Moreover, this SLR offers several practical insights for cybersecurity practitioners. Organizations can utilize synthetic data generation to train robust models without compromising sensitive data. Anomaly detection and adversarial example generation facilitate proactive threat detection and enhance system resilience against sophisticated attacks. Furthermore, policymakers can use these findings to develop regulations that ensure ethical AI deployment. These contributions support bridge the gap between theoretical knowledge and operational requirements by providing guidance for improving cybersecurity defences.

In conclusion, GenAI offers significant potential for enhancing cybersecurity threat detections by providing innovative solutions to complex challenges. This SLR analysed its applications, performance impacts, and limitations, providing a foundation for future research and practice to develop adaptive, secure, and ethical cybersecurity systems.

References

Abdalgawad, N., Sajun, A., Kaddoura, Y., Zualkernan, I. A., & Aloul, F. (2022). Generative deep learning to detect cyberattacks for the IoT-23 dataset. *IEEE Access*, *10*, 6430–6441. https://doi.org/10.1109/ACCESS.2021.3140015

Aceto, G., Giampaolo, F., Guida, C., Izzo, S., Pescapè, A., Piccialli, F., & Prezioso, E. (2024). Synthetic and privacy-preserving traffic trace generation using generative AI models for training Network Intrusion Detection Systems. *Journal of Network and Computer Applications*, *229*, 103926. https://doi.org/10.1016/j.jnca.2024.103926

Ahsan, R., Shi, W., Ma, X., & Lee Croft, W. (2022). A comparative analysis of CGAN-based oversampling for anomaly detection. *IET Cyber-Physical Systems: Theory & Applications*, *7*(1), 40–50. https://doi.org/10.1049/cps2.12019

Alabrah, A. (2022). A novel study: GAN-based minority class balancing and machine-learning-based network intruder detection using chi-square feature selection. *Applied Sciences*, *12*(22), 11662. https://doi.org/10.3390/app122211662

Alo, S. O., Jamil, A. S., Hussein, M. J., Al-Dulaimi, M. K. H., Taha, S. W., & Khlaponina, A. (2024). Automated Detection of Cybersecurity Threats Using Generative Adversarial Networks (GANs). *2024 36th Conference of Open Innovations Association (FRUCT)*, 566–577. https://doi.org/10.23919/FRUCT64283.2024.10749874

Bao, T., Trousil, K., Duy Tran, Q., Di Troia, F., & Park, Y. (2025). Generating Synthetic Malware Samples Using Generative AI. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3556704

Benaddi, H., Jouhari, M., Ibrahimi, K., Ben Othman, J., & Amhoud, E. M. (2022). Anomaly detection in industrial IoT using distributional reinforcement learning and generative adversarial networks. *Sensors*, *22*(21), 8085. https://doi.org/10.3390/s22218085

Bryman, A., & Bell, E. (2011). *Business Research Methods* (Third Edition). OXFORD University Press Inc.

Cai, F., & Koutsoukos, X. (2023). Real-time detection of deception attacks in cyber-physical systems. *International Journal of Information Security*, *22*(5), 1099–1114. https://doi.org/10.1007/s10207-023-00677-z

Cao, Y., Dong, F., Xin, Z., Wei, Z., & Han, L. (2024). Denoising Diffusion Generative Adversarial Network Integrating Multi-Scale CNN. *2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS)*, 145–150. https://doi.org/10.1109/DDCLS61622.2024.10606860

Celik, A., & Eltawil, A. M. (2024). At the dawn of generative AI era: A tutorial-cum-survey on new frontiers in 6G wireless intelligence. *IEEE Open Journal of the Communications Society*, *5*, 2433–2489. https://doi.org/10.1109/OJCOMS.2024.3362271

Chernyshev, M., Baig, Z., & Doss, R. R. M. (2023). Towards Large Language Model (LLM) Forensics Using LLM-based Invocation Log Analysis. *Proceedings of the 1st ACM Workshop on Large AI Systems and Models with Privacy and Safety Analysis*, 89–96. https://doi.org/10.1145/3689217.3690616

Cherqi, O., Moukafih, Y., Ghogho, M., & Benbrahim, H. (2023). Enhancing cyber threat identification in open-source intelligence feeds through an improved semi-supervised generative adversarial learning approach with contrastive learning. *IEEE Access*, *11*, 84440–84452. https://doi.org/10.1109/ACCESS.2023.3299604

Chiriac, B. N., Anton, F. D., Ioniță, A. D., & Vasilică, B. V. (2025). A Modular AI-Driven Intrusion Detection System for Network Traffic Monitoring in Industry 4.0, Using Nvidia Morpheus and Generative Adversarial Networks. *Sensors*, *25*(1). https://doi.org/10.3390/s25010130

Choi, S.-H., Shin, J.-M., Liu, P., & Choi, Y.-H. (2022). ARGAN: Adversarially robust generative adversarial networks for deep neural networks against adversarial examples. *IEEE Access*, *10*, 33602–33615. https://doi.org/10.1109/ACCESS.2022.3160283

Constantin, M. G., Stanciu, D. C., Stefan, L. D., Dogariu, M., Mihailescu, D., Ciobanu, G., Bergeron, M., Liu, W., Belov, K., Radu, O., & Ionescu, B. (2024). Exploring Generative Adversarial Networks for Augmenting Network Intrusion Detection Tasks. *ACM Transactions on Multimedia Computing, Communications and Applications*, *21*(1). https://doi.org/10.1145/3689636

Coppolino, L., D'Antonio, S., Mazzeo, G., & Uccello, F. (2025). The good, the bad, and the algorithm: The impact of generative AI on cybersecurity. *Neurocomputing*, *623*. https://doi.org/10.1016/j.neucom.2025.129406

Cybersecurity Ventures. (2023). *Cybercrime to cost the world $9.5 trillion USD annually by 2024*. https://cybersecurityventures.com/cybercrime-to-cost-the-world-9-trillion-annually-in-2024/

Demirbaga, U. (2024). Advancing anomaly detection in cloud environments with cutting-edge generative AI for expert systems. *Expert Systems*, *42*(2). https://doi.org/10.1111/exsy.13722

Dina, A. S., Siddique, A. B., & Manivannan, D. (2022). Effect of balancing data using synthetic data on the performance of machine learning classifiers for intrusion detection in computer networks. *IEEE Access*, *10*, 96731–96747. https://doi.org/10.1109/ACCESS.2022.3205337

Dwivedi, Y. K., Pandey, N., Currie, W., & Micu, A. (2024). Leveraging ChatGPT and other generative artificial intelligence (AI)-based applications in the hospitality and tourism industry: practices, challenges and research agenda. *International Journal of Contemporary Hospitality Management*, *36*(1), 1–12. https://doi.org/10.1108/IJCHM-05-2023-0686

Ferrag, M. A., Debbah, M., & Al-Hawawreh, M. (2023). Generative AI for cyber threat-hunting in 6G-enabled IoT networks. *2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW)*, 16–25. https://doi.org/10.1109/CCGridW59191.2023.00018

Ferrag, M. A., Ndhlovu, M., Tihanyi, N., Cordeiro, L. C., Debbah, M., Lestable, T., & Thandi, N. S. (2024). Revolutionizing cyber threat detection with large language models: A privacy-preserving BERT-based lightweight model for IoT/IIoT devices. *IEEE Access*, *12*, 23733–23750. https://doi.org/10.1109/ACCESS.2024.3363469

Goodwin, D., Mays, N., & Pope, C. (2020). Ethical issues in qualitative research. *Qualitative Research in Health Care*, 27–41.

Goyal, M., & Mahmoud, Q. H. (2024). A systematic review of synthetic data generation techniques using generative AI. In *Electronics* (Vol. 13, Issue 17). https://doi.org/10.3390/electronics13173509

Hamouda, D., Ferrag, M. A., Benhamida, N., Seridi, H., & Ghanem, M. C. (2024). Revolutionizing intrusion detection in industrial IoT with distributed learning and deep generative techniques. *Internet of Things*, *26*. https://doi.org/10.1016/j.iot.2024.101149

Hasanov, I., Virtanen, S., Hakkala, A., & Isoaho, J. (2024). Application of Large Language Models in Cybersecurity: A Systematic Literature Review. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3505983

He, L., Sun, G., Niyato, D., Du, H., Mei, F., Kang, J., Debbah, M., & Han, Z. (2025). Generative AI for game theory-based mobile networking. *IEEE Wireless Communications*, *32*(1), 122–130. https://doi.org/10.1109/MWC.007.2400133

Heiding, F., Schneier, B., Vishwanath, A., Bernstein, J., & Park, P. S. (2024). Devising and detecting phishing emails using large language models. *IEEE Access*, *12*, 42131–42146. https://doi.org/10.1109/ACCESS.2024.3375882

Kasri, W., Himeur, Y., Alkhazaleh, H. A., Tarapiah, S., Atalla, S., Mansoor, W., & Al-Ahmad, H. (2025). From Vulnerability to Defense: The Role of Large Language Models in Enhancing Cybersecurity. *Computation*, *13*(2). https://doi.org/10.3390/computation13020030

Kotb, H. M., Gaber, T., AlJanah, S., Zawbaa, H. M., & Alkhathami, M. (2025). A novel deep synthesis-based insider intrusion detection (DS-IID) model for malicious insiders and AI-generated threats. *Scientific Reports*, *15*(1), 207. https://doi.org/10.1038/s41598-024-84673-w

Li, Z., Chen, S., Dai, H., Xu, D., Chu, C.-K., & Xiao, B. (2023). Abnormal traffic detection: Traffic feature extraction and DAE-GAN with efficient data augmentation. *IEEE Transactions on Reliability*, *72*(2), 498–510. https://doi.org/10.1109/TR.2022.3204349

Li, Z., Huang, C., & Qiu, W. (2024). An intrusion detection method combining variational auto-encoder and generative adversarial networks. *Computer Networks*, *253*. https://doi.org/10.1016/j.comnet.2024.110724

Lin, Y.-D., Liu, Z.-Q., Hwang, R.-H., Nguyen, V.-L., Lin, P.-C., & Lai, Y.-C. (2022). Machine learning with variational autoencoder for imbalanced datasets in intrusion detection. *IEEE Access*, *10*, 15247–15260. https://doi.org/10.1109/ACCESS.2022.3149295

Liu, Z., Hu, J., Liu, Y., Roy, K., Yuan, X., & Xu, J. (2023). Anomaly-based intrusion on IoT networks using AIGAN-a generative adversarial network. *IEEE Access*, *11*, 91116–91132. https://doi.org/10.1109/ACCESS.2023.3307463

Mari, A.-G., Zinca, D., & Dobrota, V. (2023). Development of a machine-learning intrusion detection system and testing of its performance using a generative adversarial network. *Sensors*, *23*(3), 1315. https://doi.org/10.3390/s23031315

Mohamed Shaffril, H. A., Samsuddin, S. F., & Abu Samah, A. (2021). The ABC of systematic literature review: the basic methodological guidance for beginners. *Quality and Quantity*, *55*(4), 1319–1346. https://doi.org/10.1007/s11135-020-01059-6

Moti, Z., Hashemi, S., Karimipour, H., Dehghantanha, A., Jahromi, A. N., Abdi, L., & Alavi, F. (2021). Generative adversarial network to detect unseen internet of things malware. *Ad Hoc Networks*, *122*, 102591. https://doi.org/10.1016/j.adhoc.2021.102591

Nadella, G. S., Addula, S. R., Yadulla, A. R., Sajja, G. S., Meesala, M., Maturi, M. H., Meduri, K., & Gonaygunta, H. (2025). Generative AI-Enhanced Cybersecurity Framework for Enterprise Data Privacy Management. *Computers*, *14*(2), 55. https://doi.org/10.3390/computers14020055

NCSC. (2025). *Cyber Threat Report 2023/24 released | National Cyber Security Centre*. https://www.ncsc.govt.nz/news/cyber-threat-report-2024

Novaes, M. P., Carvalho, L. F., Lloret, J., & Proença, M. L. (2021). Adversarial Deep Learning approach detection and defense against DDoS attacks in SDN environments. *Future Generation Computer Systems*, *125*, 156–167. https://doi.org/10.1016/j.future.2021.06.047

O'Dea, R. E., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Gurevitch, J., Page, M. J., Stewart, G., Moher, D., & Nakagawa, S. (2021). Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: A PRISMA extension. *Biological Reviews*, *96*(5), 1695–1722. https://doi.org/10.1111/brv.12721

Pandian, A. P. D. (2024). Variational Autoencoders using Convolutional neural network for highly advanced cyber threats. *In 2024 IEEE Integrated STEM Education Conference (ISEC)*, 01–06. https://doi.org/10.1109/ISEC61299.2024.10664944

Park, C., Lee, J., Kim, Y., Park, J. G., Kim, H., & Hong, D. (2023). An enhanced AI-based network intrusion detection system using generative adversarial networks. *IEEE Internet of Things Journal*, *10*(3), 2330–2345. https://doi.org/10.1109/JIOT.2022.3211346

Qi, S., Chen, J., Chen, P., Wen, P., Niu, X., & Xu, L. (2024). An efficient GAN-based predictive framework for multivariate time series anomaly prediction in cloud data centers. *Journal of Supercomputing*, *80*(1), 1268–1293. https://doi.org/10.1007/s11227-023-05534-3

Qu, X., Liu, Z., Wu, C. Q., Hou, A., Yin, X., & Chen, Z. (2024). MFGAN: Multimodal fusion for industrial anomaly detection using attention-based autoencoder and generative adversarial network. *Sensors*, *24*(2), 637. https://doi.org/10.3390/s24020637

Ren, Y., Feng, K., Hu, F., Chen, L., & Chen, Y. (2023). A lightweight unsupervised intrusion detection model based on variational Auto-Encoder. *Sensors (Basel, Switzerland)*, *23*(20). https://doi.org/10.3390/s23208407

Sai, S., Yashvardhan, U., Chamola, V., & Sikdar, B. (2024). Generative ai for cyber security: Analyzing the potential of chatgpt, dall-e and other models for enhancing the security space. *IEEE Access*, *12*, 53497–53516. https://doi.org/10.1109/ACCESS.2024.3385107

Saikam, J., & Ch, K. (2024). EESNN: Hybrid Deep Learning Empowered Spatial–Temporal Features for Network Intrusion Detection System. *IEEE Access*, *12*, 15930–15945. https://doi.org/10.1109/ACCESS.2024.3350197

Senevirathne, P., Cooray, S., Dinal Herath, J., & Fernando, D. (2024). Virtual Machine Proactive Fault Tolerance using Log-based Anomaly Detection. *IEEE Access*, *12*, 178951–178970. https://doi.org/10.1109/ACCESS.2024.3506833

Senthilkumar, G., Tamilarasi, K., & Periasamy, J. K. (2024). Cloud intrusion detection framework using variational auto encoder Wasserstein generative adversarial network optimized with archerfish hunting optimization algorithm. *Wireless Networks*, *30*(3), 1383–1400. https://doi.org/10.1007/s11276-023-03571-7

Shafee, S., Bessani, A., & Ferreira, P. M. (2025). Evaluation of LLM-based chatbots for OSINT-based Cyber Threat Awareness. *Expert Systems with Applications*, *261*, 125509. https://doi.org/10.1016/j.eswa.2024.125509

Shieh, C. S., Nguyen, T. T., Lin, W. W., Lai, W. K., Horng, M. F., & Miu, D. (2022). Detection of adversarial ddos attacks using symmetric defense generative adversarial networks. *Electronics*, *11*(13). https://doi.org/10.3390/electronics11131977

Siniosoglou, I., Radoglou-Grammatikis, P., Efstathopoulos, G., Fouliras, P., & Sarigiannidis, P. (2021). A unified deep learning anomaly detection and classification approach for smart grid environments. *IEEE Transactions on Network and Service Management*, *18*(2), 1137–1151. https://doi.org/10.1109/TNSM.2021.3078381

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Statista. (2024). *Total annual amount of money received by ransomware actors worldwide from 2017 to 2023*. https://www.statista.com/statistics/1410498/ransomware-revenue-annual/

Taherdoost, H. (2021). Data collection methods and tools for research; a step-by-step guide to choose data collection technique for academic and business research projects. *International Journal of Academic Research in Management (IJARM)*, *2021*(1), 10–38. https://hal.science/hal-03741847

Ullah, I., & Mahmoud, Q. H. (2021). A framework for anomaly detection in IoT networks using conditional generative adversarial networks. *IEEE Access*, *9*, 165907–165931. https://doi.org/10.1109/ACCESS.2021.3132127

Vadisetty, R., & Polamarasetti, A. (2024). Generative AI for Cyber Threat Simulation and Defense. *In 2024 12th International Conference on Control, Mechatronics and Automation (ICCMA)*, 272–279. https://doi.org/10.1109/ICCMA63715.2024.10843938

Vu, L., Nguyen, Q. U., Nguyen, D. N., Hoang, D. T., & Dutkiewicz, E. (2023). Deep generative learning models for cloud intrusion detection systems. *IEEE Transactions on Cybernetics*, *53*(1), 565–577. https://doi.org/10.1109/TCYB.2022.3163811

Wasswa, H., Lynar, T., & Abbass, H. (2023). Enhancing IoT-botnet detection using variational auto-encoder and cost-sensitive learning: A deep learning approach for imbalanced datasets. *In 2023 IEEE Region 10 Symposium (TENSYMP)*, 1–6. https://doi.org/10.1109/TENSYMP55890.2023.10223613

Williams, R. I., Clark, L. A., Clark, W. R., & Raffo, D. M. (2021). Re-examining systematic literature review in management research: Additional benefits and execution protocols. *European Management Journal*, *39*(4), 521–533. https://doi.org/10.1016/j.emj.2020.09.007

Yang, Y., Yao, C., Yang, J., & Yin, K. (2022). A network security situation element extraction method based on conditional generative adversarial network and transformer. *IEEE Access*, *10*, 107416–107430. https://doi.org/10.1109/ACCESS.2022.3212751

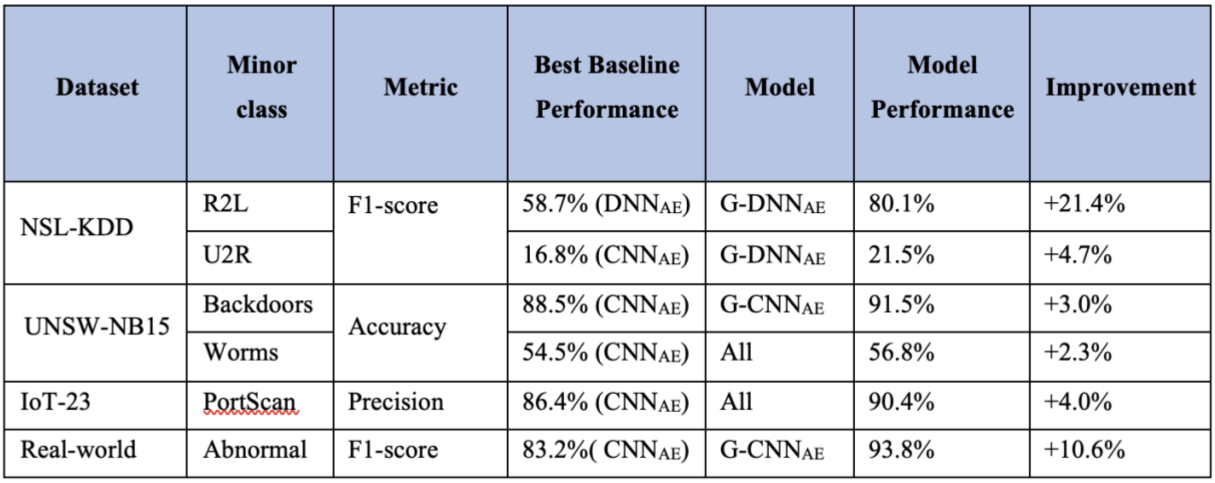
Zeng, X., Zhuo, Y., Liao, T., & Guo, J. (2025). Cloud-GAN: Cloud generation adversarial networks for anomaly detection. *Pattern Recognition*, *157*. https://doi.org/10.1016/j.patcog.2024.110866

Zhang, J., Wu, P., London, J., & Tenney, D. (2025). Benchmarking and evaluating large language models in phishing detection for small and midsize enterprises: A comprehensive analysis. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3540075

Zhou, J., Wu, Z., Xue, Y., Li, M., & Zhou, D. (2022). Network unknown-threat detection based on a generative adversarial network and evolutionary algorithm. *International Journal of Intelligent Systems*, *37*(7), 4307–4328. https://doi.org/10.1002/int.22766

Appendix A: Key Findings of Landmark Study 1

Table A1. Key findings of Landmark Study1

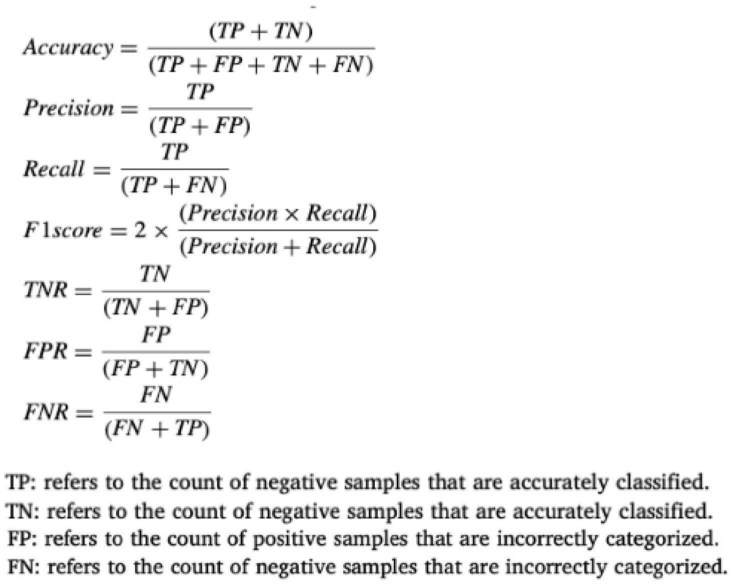


Dataset	Minor class	Metric	Best Baseline Performance	Model	Model Performance	Improvement
NSL-KDD	R2L	F1-score	58.7% (DNN$_{AE}$)	G-DNN$_{AE}$	80.1%	+21.4%
	U2R		16.8% (CNN$_{AE}$)	G-DNN$_{AE}$	21.5%	+4.7%
UNSW-NB15	Backdoors	Accuracy	88.5% (CNN$_{AE}$)	G-CNN$_{AE}$	91.5%	+3.0%
	Worms		54.5% (CNN$_{AE}$)	All	56.8%	+2.3%
IoT-23	PortScan	Precision	86.4% (CNN$_{AE}$)	All	90.4%	+4.0%
Real-world	Abnormal	F1-score	83.2%(CNN$_{AE}$)	G-CNN$_{AE}$	93.8%	+10.6%

Appendix B: Performance Metrics Formulas

Figure B1

Performance Metrics Formulas



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

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

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

$$F1score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$

$$TNR = \frac{TN}{(TN + FP)}$$

$$FPR = \frac{FP}{(FP + TN)}$$

$$FNR = \frac{FN}{(FN + TP)}$$

TP: refers to the count of negative samples that are accurately classified.
TN: refers to the count of negative samples that are accurately classified.
FP: refers to the count of positive samples that are incorrectly categorized.
FN: refers to the count of negative samples that are incorrectly categorized.

Note. Adapted from Hamouda et al. (2024) and Ullah and Mahmoud (2021).

Appendix C: Annotated Bibliography

Article 1: Advancing anomaly detection in cloud environments with cutting-edge Generative AI for expert systems

Citation:

Demirbaga, U. (2024). Advancing anomaly detection in cloud environments with cutting-edge Generative AI for expert systems. *Expert Systems*. https://doi.org/10.1111/exsy.13722

Background: This article addresses the challenges of anomaly detection in cloud computing environments with a dynamic, distributed and data-intensive nature. Traditional anomaly detection techniques such as statistical, rule-based, machine learning and clustering approach struggle to ensure cloud security. In the cloud systems, anomalies occur due to inconsistency, incompatibility, dynamic resource allocation and distributed architecture. GenAI especially GANs supports to generate of synthetic data and improves model training in complex environments. Therefore, this study introduces CloudGEN, a new approach using GANs and Convolutional Neural Networks (CNNs) to enhance anomaly detection accuracy and adaptability in cloud environments.

Significance of the Article: The article is highly significant to my study as it introduces CloudGEN a new system that integrates GANs, CNNs and SHAP (Shapley Additive exPlanations) techniques for anomaly detection in cloud environments. This study investigates how GenAI techniques identify security threats such as abnormal behaviors and resource misuse within dynamic and large-scale cloud environments. This study provides the most important methodological and empirical insights that can be effectively applied to improve the cloud-based threat detection approach. Furthermore, this article emphasizes that CloudGEN improves accuracy and flexibility which are key critical factors to build strong security systems. Similarly, the integration of SHAP enhances the explainability of systems by improving trust and understandability of AI decisions. Additionally, this study used unsupervised learning and transfer learning to improve the detection of unseen and evolving threats.

Research Methodology: This research follows a quantitative and conceptual research methodology. In conceptual approach presents the CloudGEN system architecture explaining three components Monitoring module, GAN model and CNN-based anomaly detection system. The monitoring module collects raw log and performance data from big data clusters like Hadoop in real time using SmartMonit. The YARN and Sigar API are utilized to collect metrics such as execution time, job status, CPU/memory utilization and network traffic. Next, collected real data is transmitted through RabbitMQ to store in a time-series database called InfluxDB. Data preprocessing techniques were applied after collecting this real-world dataset. This data preprocessing stage includes data wrangling to clean inconsistencies of raw data, feature extraction to identify patterns from cleaned data and SHAP-based feature selection to select the most critical attributes.

The second phase, the generative model utilizes GANs to generate synthetic data that can simulate the statistical properties and distribution of real-world anomalies. This generated synthetic is important to overcome challenges like scarcity and imbalance of labeled threat data in cloud environments. This hybrid dataset enhances model training and improves the model's generalizability. CloudGEN incorporates SHAP to analyze feature importance by enhancing the explainability and interpretability of the generated data. Finally, the CNN-based anomaly detection system is trained using both real data collected from cloud-based big data clusters and synthetic data generated by the GAN model to detect anomalies. In the quantitative approach, the trained CNN model is compared against other traditional models using evolution metrics such as accuracy, precision, recall and F1-score.

Key Findings: The study demonstrates that the CloudGEN approach significantly outperforms traditional anomaly detection methods in cloud environments. This new CloudGEN approach achieved 99.7% accuracy in anomaly detection exceeding the performance of models like CNN-only (97.07%), Ensemble Learning (97.06%), Isolation Forest (IF) (92.4%) and K-Nearest Neighbors (KNN) (90.7%). Overall, CloudGEN showed approximately 11% improvement across key performance metrics such as accuracy, precision, recall, and F1-Score. The use of GAN-generated synthetic data is the key factor in the performance boost because it closely simulates real-world anomalies and significantly enhances the model's generalizability. Similarly, the model achieved low false positive and false negative rates which the highlights robustness of this approach in large-scale and complex cloud environments.

Another major finding of this study is the integration of SHAP which improves model explainability by identifying and ranking the most important features. This study has identified critical factors to detect anomalies such as data locality, network faults, CPU usage and memory usage in cloud-based large-scale systems. Overall, the CloudGEN approach is a highly effective and transparent solution for threat detection in dynamic cloud environments.

Strengths and Weaknesses:

Strengths: Introducing CloudGEN, this study presents several significant strengths that contribute to improving anomaly detection in cloud platforms. The most of strength of this study is that CloudGEN demonstrates high performance achieving 99.7% accuracy by integrating GNNs, CNNs, and SHAP techniques. The use of real-time data enhances the credibility of this study and the open-access dataset on GitHub supports reproducibility. The integration of SHAP significantly improves the transparency of the model while improving the user's trust. Furthermore, the unsupervised generative modeling addresses the challenges of limited labeled anomaly data which is a common issue in cloud environments.

Weaknesses: Although, this study has certain limitations in addition to its strengths. This model has a complex architecture that combines GNNs, CNNs, and SHAP. This can lead to significant challenges in implementation, tunning and maintenance across diverse cloud settings. Furthermore, this study focuses on specific metrics and platforms such as resource utilization in Hadoop. This focus can limit the generalizability of this model in other cloud services. In addition, the real time responsiveness of the model is not fully addressed in this study can be considered another limitation.

Article 2: Enhancing Cyber Threat Identification in Open-Source Intelligence Feeds Through an Improved Semi-Supervised Generative Adversarial Learning Approach with Contrastive Learning

Citation:

Cherqi, O., Moukafih, Y., Ghogho, M., & Benbrahim, H. (2023). Enhancing cyber threat identification in open-source intelligence feeds through an improved semi-supervised generative adversarial learning approach with contrastive learning. *IEEE Access*, *11*, 84440–84452. https://doi.org/10.1109/ACCESS.2023.3299604

Background: This article focuses on the growing importance of Cyber Threat Intelligence (CTI) due to the increase of cybersecurity threats. This article explores the use of Open-source Threat Intelligence Feeds (OTIFs) to effectively improve cybersecurity in organizations. Traditional supervised learning methods have faced challenges due to these methods require large volumes of high-quality labeled data. This article introduces ConGAN-BERT, a new and improved version of the GEN-BERT model which effectively utilizes both labeled and unlabeled data for automated threat detection. The ConGAN-BERT approach used contrastive learning with semi-supervised techniques to improve the accuracy and robustness of threat classification.

Significance of the Article: This article is directly significant to my study as it introduces a GenAI-based solution called ConGAN-BERT (GANs +BERT + Contrastive Learning) for threat detection using minimum labeled data. This approach directly aligns with my research domain on GenAI as it uses a semi-supervised GAN-based model. It generated high-quality annotated data from limited labeled data and a large volume of unlabeled cyber threat intelligence. Furthermore, this proposed semi-supervised model reduced depends on labeled data which is a common challenge in cloud environments. By testing on diverse datasets including dark web forums and Open Threat Exchange (OTX), this article highlights the applicability of generative models to threat detection in real-world environments.

Research Methodology: This article followed a quantitative experiment research approach by introducing a novel approach to threat detection. This study utilized diverse OTIFs from various sources such as OTX, Exploit Database, Global Database of Events, Language and Tone (GDELT) and Dark Web markets data from the AZSecure Hacker Assets Portal. These datasets cover a wide range of cyber threat types, platforms, periods. For instance, OTX data was collected from 2016 to 2021, and Dark Web data covers 2003 to 2016. To ensure the data quality and consistency, preprocessing techniques such as tokenization, character normalization, and UTF-8 encoding were applied.

The model training began with a pre-trained BERT model and finetuned using a task-specific muti-layer perception (MLP). This developed model architecture included a discriminator and generator with GELU activation functions. This model trained over five epochs using a high-performance system (NVIDIA TITAN Xp GPU, Intel 9 processor and 128 GB RAM). In the evaluation stage, the ConGAN-BERT model was compared with BERT and GEN-BERT using metrics like F1-score, accuracy and precision. This testing was conducted with different amounts of labeled data (0.1%- 2%) and unlabeled data (2000 to 15000 samples) to evaluate performance. Additionally, evaluates the impact of hyperparameters such as batch size, λ and temperature τ on the performance of the ConGAN-BERT framework using grid search.

Key Findings: The primary key finding of this article is introducing ConGAN-BERT, a new semi-supervised learning model to improve threat detection. It combines GAN-BERT with contrastive learning to improve threat detection addressing challenges of limited labeled data and overlapping terminology across threat classes. This model demonstrated significant performance improvement over BERT and GAN-BERT across multiple databases. This ConGAN-BERT model achieved 71.44% accuracy on the Exploit dataset with only 0.1% labeled data (10 samples). This was 12.53 points higher than BERT (58.91%) and 7.8 points higher than GAN-BERT (63.67%).

Furthermore, this proposed model improved the F1-score by 3-12% across different datasets. Similarly, ConGAN-BERT achieved 26.79% accuracy on the GDELT dataset outperforming both BERT (16.00%) and GEN-BERT (20.16%). Although GEN-BERT slightly outperformed ConGAN-BERT in certain points (0.6% and 1% of labeled examples). Overall, the ConGAN-BERT model performs well when minimal labeled data setups are the model improves when more labeled data is added. The hard negative sampling strategy further improved accuracy on noisy, crowd-sourced datasets like OTX. For example, accuracy on

OTX increases from 55.10% to 56.47% at 0.1% labeled data. The optimal results were achieved with a batch size of 32, λ set to 0.006 and temperature set to 2.5. Furthermore, ConGAN-BERT performed well on dark web data highlighting it strong scalability and generalizability in the real-world applications.

Strengths and Weaknesses:

Strengths: The main strength of this study is the introduction of the ConGAN-BERT method combining contrastive learning with GAN-BERT. This approach addresses the challenge of overlapping threat types and limited labeled data in cybersecurity. The study used a diverse set of real-world datasets with different characteristics ensuring comprehensive testing results across accuracy, F1-score and precision. This innovative approach significantly strengthens threat detection by reducing the need for human supervision. This study provides a clear research methodology with data collection, preprocessing and training details offering transparency and reproducibility of the findings.

Weaknesses: This study has several limitations such as limited generalizability, and sensitivity to hyperparameters. The major weakness of the study is its dependency on factors such as the quality and diversity of annotated data, the complexity of the classification tasks and the selection of appropriate hyperparameters. The authors highlighted the need for further evaluations of limitations and applicability of the ConGAN-BERT framework as these factors affect its generalizability. Furthermore, this model requires high computational resources that may limit the scalability.

Article 3: Benchmarking and Evaluating Large Language Models in Phishing Detection for Small and Midsize Enterprises: A Comprehensive Analysis

Citation:

Zhang, J., Wu, P., London, J., & Tenney, D. (2025). Benchmarking and evaluating large language models in phishing detection for small and midsize enterprises: A comprehensive analysis. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3540075

Background: The article discusses the increasing cyber threats of phishing attacks such as spear phishing and business email compromise (BEC) which target human vulnerabilities and bypass traditional detection systems. This study emphasizes the urgent need of advanced, cost-effective phishing detection mechanisms specially developed for small and medium-size enterprises (SMEs) which often lack of financial and technical resources. This study evaluates the effectiveness of large language models (LLMs) to detect phishing especially in SMEs to address real-world challenges like financial and operational damages.

Significance of the Article: This research is highly significant to my study as it directly examines the application of LLMs, a type of GenAI to detect phishing attacks which is a major threat for SMEs. This study highlights how LLM models such as Llama-3-8b-instruct, identify complex phishing attacks that commonly bypass traditional security measures. It closely aligns with my research objectives by presenting that GenAI can function not only as a threat but also as a robust defensive mechanism. Furthermore, it emphasizes the ability of LLMs to provide scalable, cost-effective solutions without fine-tuning for real-time threat detection. In addition, this approach provides detailed reports explaining threats so that users can understand the risks.

Research Methodology: This study used a comprehensive benchmarking methodology to evaluate the effectiveness of LLMs in detecting phishing emails especially in the context of SMEs. Initially, this research used publicly available datasets from sources such as Google Scholar, Kaggle and GitHub to collect 160 emails categorized into four groups namely fraud, false positives (legitimate emails), phishing and commercial plan. Additionally, 20 custom-generated emails by LLMs were added to increase the dataset's diversity. Afterward, the final balanced dataset was created with 40 emails and evenly divided into 4 categories: human-legitimate, AI-generated legitimate, human-phishing, and AI-generated phishing.

In this research, twelve LLMs including four proprietary models (GPT-4o, GPT-3.5) and eight open-source models (Llama-3-8b-Instruct, OpenHermes-2.5-) were selected based on the ranking in public leaderboard. This study used models with default settings without fine-tuning to simulate realistic and accessible situation for SMEs. Finally, model outputs are manually evaluated and scored using classification matrices such as accuracy, F1-Score, Geometric Mean, balanced detection rate and Matthew's Correlation Coefficient. Overall, this research used experimental methodology based on standardized inputs and quantitative performance metrics to evaluate the feasibility of LLMs in phishing detection.

Key Findings: The major finding of the research is that LLMs effectively detect phishing attacks within SMEs, especially with limited cybersecurity infrastructure. This research found that Llama-3-8b-instruct, a lightweight open-source model outperformed large models including LIama-3-70b and OpenAI model GPT-4o. Among benchmarked 12 LLMs, the Llama-3-8b-instruct model achieved the highest accuracy (97.50%) and F1-score (97.56%) in identifying both phishing and legitimate mail showing its adaptability and cost-effectiveness in SMEs. Furthermore, proprietary models like GPT-3.5 and Claude 3.5 Sonnet perform more competitively than older models such as OpenHermes-2.5.

Another significant finding of this study is that base models with default parameters significantly performed in phishing detection tasks. Therefore, this approach reduced the barrier for SMEs to use these base models as these do not require expensive fine-tuning or additional domain-specific training. Moreover, these results discovered that model size does not directly correlate with detection performance. For example, the Llama-3-8b-instruct model outperformed the Llama-3-70b-instruct and GPT-3.5. performed over GPT-4o (GPT-4o-2024-05-13). Furthermore, this study provides practical recommendations such as deployment strategies and consideration of human factors to improve real-world usability. Overall, this study mentioned GPT-3.5.(proprietary) and Llama-3-8b-instruct (open-source) as top models for phishing detection due to their affordable, reliable and well-supported to SMEs. However, due to privacy concerns of OpenAI, this study recommended self-hosting Llama-3-8b-instruct for handling sensitive data within SMEs.

Strengths and Weaknesses

Strengths: The article provides several strengths that make it a valuable resource for cybersecurity research. The key strength of this research is the practical focus on cybersecurity solutions for SMEs which have limited computational resources and finance investments. Another strength is the strong benchmarking of 12 LLMs using high-quality, balanced datasets of AI and human-generated phishing and legitimate emails. This approach provides a clear, practical comparison for SMEs to decide on the use of GenAI for threat detection. In addition, this research offers practical recommendations to enhance applicability especially for organizations with limited budgets and resources.

Weaknesses: A weakness in the study is small the sample size of 40 emails in only English. This limited dataset is unable to fully cover the diversity and complexity of real-world phishing attacks which potentially limit the generalizability of results. Additionally, this research depended on base models without fine-tuning to specific tasks, which limits accuracy for complex threat detections. Another limitation is that the study used the Llama-3-8b-instruct model both as a sample generator and as one of the benchmarked models which leads to biased results.

Article 4: Deep Generative Learning Models for Cloud Intrusion Detection Systems

Citation:

Vu, L., Nguyen, Q. U., Nguyen, D. N., Hoang, D. T., & Dutkiewicz, E. (2023). Deep generative learning models for cloud intrusion detection systems. *IEEE Transactions on Cybernetics*, *53*(1), 565–577. https://doi.org/10.1109/TCYB.2022.3163811

Background: Cloud systems are vulnerable to many types of cyberattacks due to their rapid growth and complex architecture. Traditional IDSs struggle with imbalanced datasets due to a lack of sufficient malicious activity compared to normal behaviors. This leads to biased prediction of novel or rare threats.

This article introduces two generative learning models namely Conditional Denoising Adversarial Autoencoder (CDAAE) and a hybrid of CDAAE with the KNN algorithm (CDAEE-KNN). These models are designed to generate realistic, class-specific malicious samples that can used to augment training data for IDS. This approach enhances the training dataset with synthetic malicious samples and improves IDS accuracy across diverse threats.

Significance of the Article: This article highly relevant resource to my research because it examines the use of generative models especially adversarial AE techniques, a type of GenAI to improve intrusion detection in cloud environments. By introducing two novel deep generative models namely CDAAE and CDAEE-KNN, this article addresses the challenge of imbalanced malicious data. These models generate realistic malicious attack samples and significantly improve the detection accuracy of cloud IDS across diverse attack types. This study supports my research objective by presenting the practical application of a GenAI-driven threat detection approach and emphasizing the potential of generative models to outperform traditional methods.

Research Methodology: This study used a quantitative experimental research methodology to evaluate the efficiency of GenAI models for enhancing IDSs. In the model development phase, this study introduced CDAAE and CDAEE-KNN models. The CDAAE model generates malicious data for any specific attacks on the cloud systems by combining Denoising Autoencoder (DAE) and Adversarial Network techniques. The CDAEE-KNN is a hybrid model integrating KNN to filter and identify generated samples that are near decision boundaries.

This study used six diverse IDS datasets to evaluate the effectiveness of the proposed models. These datasets include a cloud IDS dataset, two network IDS datasets (NSL KDD, UNSW NB15) and three malware datasets from the CTU-13 dataset system. The generated synthetic malicious samples from models combined with these original datasets to form augment datasets. Using these augmented datasets trained the three classifiers namely Support Vector Machine (SVM), Decision Tree (DT) and Random Forest (RF). The performance of the proposed models was measured using F1-Score, AUC and GEO score and Gaussian Parzen window distribution used to evaluate the quality of generated synthetic data. Finally, this study conducted a comprehensive analysis to evaluate the effectiveness of these models against traditional sampling methods (SMOTE-SVM, BalanceCascade) and generative models (ACGAN, CVAE, CAAE).

Key Findings: The most significant contribution of this study is development of two innovative generative models namely CDAAE and CDAEE-KNN that generate synthetic data for minor classes to address the class imbalance in the intrusion detection dataset. These models significantly outperformed traditional methods and other generative methods across multiple datasets. The CDAAE and CDAEE-KNN achieved higher F1-scores, AUC and GEO scores across a range of cloud and network intrusion datasets. Moreover, CDAEE-KNN was highly effective as it uses KNN to generate samples near decision boundaries which enhanced classifier accuracy, generalizability and reduced misclassifications.

Additionally, Parzen window-based log-likelihood experimental results showed that synthetic samples generated by CDAAE more closely aligned with the original data distribution compared to other models. Classifiers like SVM, DT and RF showed significant improvement in AUC and GEO scores when trained on data augmented by these models. For instance, on the NSL-KDD dataset, SVM, DT and RF classifiers achieved GEO scores of 0.760, 0.672 and 0.715 respectively from 0 when trained using CDAEE-KNN augmented dataset and their AUC scores increased by 15-30%. These results showed that classifiers perform better when trained on datasets augmented with the proposed model. Overall, these key findings emphasized that CDAAE and CDAEE-KNN models generate quality synthetic data by reducing imbalance and improving cloud IDS accuracy.

Strengths and Weaknesses:

Strengths: The major strength of this article is the development of two innovative GenAI models to address the class imbalance in cloud IDS. These models effectively produce synthetic malicious samples and significantly improve the performance of classifiers. The utilization of six benchmark datasets including cloud, network and malware types increased the robustness and generalizability of the research's results. Furthermore, the authors provide a quantitative evaluation comparing these proposed methods with traditional and generative methods across diverse metrics. This comprehensive analysis increased the scalability and reproducibility of this research approach.

Weaknesses: The one limitation of this study is the higher computational cost during the training and generating phase. This could limit the applicability for small or resource-constraint organizations and limit the scalability and efficiency in fast-evolving cloud environments. Another limitation of this study is that evaluation was limited to traditional classifiers without considering modern deep learning models like CNNs, and RNNs. Furthermore, this study did not analyze the attack interpretability that focuses on understanding the nature of the generated synthetic attacks. Another weakness in this study is that it did not evaluate the effectiveness of these models in detecting entirely new attack types.

Article 5: Automated Detection of Cybersecurity Threats Using Generative Adversarial Networks (GANs).

Citation:

Alo, S. O., Jamil, A. S., Hussein, M. J., Al-Dulaimi, M. K. H., Taha, S. W., & Khlaponina, A. (2024). Automated Detection of Cybersecurity Threats Using Generative Adversarial Networks (GANs). *2024 36th Conference of Open Innovations Association (FRUCT)*, 566–577. https://doi.org/10.23919/FRUCT64283.2024.10749874

Background: Traditional network intrusion detection systems (NIDS) are heavily dependent on rule-based and signature-based detection methods and these methods struggle to detect unknown, new threats effectively. Furthermore, they struggle with large-scale data and lack of available data for specific cyberattacks such as endpoint attacks and zero-day vulnerabilities that impact the model's accuracy. Therefore, advanced GenAI techniques such as GANs are required to identify both known and zero-day threats in NIDS. This study explores the application of GANs to enhance NIDS for cybersecurity threat detection. GANs technique can generate synthetic data that can used to enhance the performance of NIDS in identifying new attacks.

Significance of the Article: This article provides strong significance to my research as it explores the integration of GANs into NIDS to enhance threat identification. By investigating both traditional and advanced attacks including obfuscated and adversarial threats, this article directly supports my research focus on using GenAI for sophisticated threat detection. Furthermore, this article explores the effectiveness of GAN-boosted NIDS to detect various emerging cyber-attacks across different environments by generating high-quality synthetic data. It presents the GAN-based NIDS systems significantly improve detection accuracy and reduce the false positive rates compared to traditional models like SVM, and RF.

Research Methodology: The research methodology of this article is a comprehensive mixed-method approach, and it is organized a with set of subsections. In the quantitative approach, this research utilized two larger cybersecurity datasets UNSW-NB15 and CICIDS2017 which includes various types of cyberattacks such as denial of service (DoS), brute force, SQL, injection and advanced persistent threats (APTs). The preprocessing step included data cleaning (remove corrupted and irrelevant data entries), feature selection(choose 40 relevant features like source IP and packet size) and data augmentation to balance attack classes. A GAN model is developed with generator (G) and discriminator (D) to generate high-quality synthetic data using Stochastic Gradient Descent(GSD) with the Adam optimizer. The experimental design compared GAN-enhanced NIDS model with the traditional models (SVM, RF) which were trained without synthetic data. The performance of this model was evaluated using key metrics such as detection accuracy, false positive rate (FPR), precision and recall. This study further validated the GAN-based NIDS model using 5-fold cross-validation to evaluate generalization and adversarial testing (FGSM, BIM) to evaluate robustness.

Similarly, this research gathered qualitative insights from 15 cybersecurity experts from different industries including telecommunications, financial services and critical infrastructure. This study conducted semi-structured interviews with open-ended questions to explore the practical deployment challenges of GAN-based models.

Key Findings: The key finding of this article is that the GAN-enhanced NIDS model achieved better performance than traditional ML models in cybersecurity threat detection. Table 1 summarizes the key findings of the quantitative approach of this article.

Summary of Key findings of Article

Performance Matrix	Key Findings
Detection Accuracy	GAN-enhanced NIDS achieved 95.8% accuracy which is better than SVM(89.6%) and RF(91.2%)
False Positive Rate	False positive rate reduced to 2.4% significantly lower than SVM(5.1%) and RF(4.7%)
Performance on Novel attacks	Detected novel threats with 88.2% accuracy and 5.3% FRP addressing traditional models limitations.
Performance of Obfuscated Attacks	Maintained 83.4% accuracy for obfuscated attacks improving generalization capability.
Robustness Against Adversarial Attacks	GAN-enhanced NIDS detection accuracy of attacks; Fast Gradient Sign Method (FGSM) -82.7%, Basic Iterative Method (BMI) -76.8%, Projected Gradient Descent (PGD)-71.9%, Carlini & Wagner - 69.2%
Computational Challenges	Higher resource computation of GAN-based NIDS: 10.5 hour training time, 85% GOU utilization and 18.7GB memory usage compared to traditional model.

Note. Adapted from Alo et al., (2024).

These key findings summarized in Table 1, demonstrate the GAN-based NIDS outperformed traditional ML models by achieving higher accuracy, reducing FPRs and strong resilience against novel obfuscated and adversarial attacks. Additionally, expert's opinions of this qualitative approach highlighted concerns about deployment feasibility, scalability issues, ethical considerations and integration challenges of GAN-enhanced NIDS.

Strengths and Weaknesses:

Strengths: The key strength of this research is the innovative application of GANs to improve cybersecurity especially in NIDS. This study utilized two well-known datasets in cybersecurity research ensuring research's reproducibility. This GAN-enhanced NIDS achieves higher detection accuracy (95.8%) and lower false positive rates (2.4%) making it a more reliable solution for threat detection. Furthermore, this model achieved 88.2% accuracy in the detection of novel obfuscated attacks highlighting its generalizability in IDS. Another strength is that the study uses a comprehensive research methodology by combining quantitative experiments and qualitative insights. The expert interviews strengthen the study by providing practical insights into real-world deployments.

Weaknesses: The most significant challenge of this study is high computational demand which may limit the use of this model especially for organizations with limited computational resources. This GAN-enhanced model requires 10.5 hours of training, 85% GPU utilization and 18.76 GB memory usage. Furthermore, this study presents the effectiveness of the GAN-based model, but its low interpretability reduces trust and usability in cybersecurity applications. Another limitation is the integration challenges of GAN-based NIDS into existing security infrastructures. Deploying GAN-based NIDS into existing security systems requires specialized knowledge. Additionally, this article does not provide detailed discussions on ethical concerns such as misuse of synthetic data.