International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 2, 2026

Generative AI for Real-Time Cloud Security: Advanced Anomaly Detection Using GPT Models

Mrs. Sadia Kausar [1], Ms. Munazzah Khan[2], Ms. Nashrah Fariha[3], Ms. Humera Fatima[4]

[1]Assistant Professor, Dept. of CSE-AIML, Lords Institute of Engineering and Technology

[2,3,4]B.E Student Dept. of CSE-AIML, Lords Institute of Engineering and Technology

Sadiyakausar@lords.ac.in[1], Munazzahkhan999@gmail.com[2], nashrahfariha2@gmail.com[3], humerafatima120@gmail.com[4]

Abstract: *Cloud infrastructures are becoming increasingly complex, creating a strong need for advanced real-time security solutions. Traditional anomaly detection systems, which rely on predefined rules and signature-based methods, often fail to detect novel and sophisticated cyber threats. This paper explores the use of generative AI models, including LLaMA and GPT architectures, to enhance cloud security through real-time anomaly detection. The proposed framework leverages the pattern recognition and adaptive learning capabilities of these models to analyze large-scale cloud data such as logs, network traffic, user behavior, and system activities. By continuously learning from new data, the system improves its ability to identify previously unknown threats. This approach addresses key limitations in existing cloud security methods by introducing a scalable and adaptive solution. The use of generative AI enables more accurate and timely detection of anomalies in dynamic cloud environments. Furthermore, the framework demonstrates how GPT-based models can generate meaningful insights from diverse inputs. The study highlights the potential of integrating generative AI into cloud security systems for enhanced protection. It also evaluates the feasibility of deploying such models in large-scale infrastructures. Overall, the research emphasizes improved resilience, adaptability, and efficiency in modern cloud security frameworks.*

INTRODUCTION

Cloud computing has revolutionized enterprise data management by providing scalable and flexible solutions, but it has also introduced new security vulnerabilities that require advanced protection mechanisms. Traditional anomaly detection systems, which rely on predefined rules and known threat signatures, are increasingly ineffective in detecting zero-day attacks and sophisticated cyber threats in dynamic cloud environments. As cloud infrastructures become more distributed and complex, the need for real-time, adaptive, and intelligent security solutions has become critical. Recent advancements in artificial intelligence, particularly generative models like LLaMA and GPT architectures, offer promising capabilities for addressing these challenges. These models can analyze vast and diverse datasets, recognize complex patterns, and continuously learn from evolving data, making them highly suitable for anomaly detection tasks. This research proposes a real-time anomaly detection framework using generative AI models to analyze cloud logs, network traffic, user behavior, and system activities. Unlike conventional methods, the proposed approach provides comprehensive and timely threat detection across multiple data sources. It also addresses key gaps in existing research, particularly the limited use of generative models in real-time cloud security. Furthermore, the framework is designed to scale effectively in multi-cloud environments, ensuring adaptability to increasing system complexity. By integrating generative AI into cloud security, this study aims to enhance the detection of novel threats while improving the overall resilience and efficiency of modern cloud infrastructures.

PROJECT OVERVIEW

Cloud computing has revolutionized modern IT infrastructure by offering scalable and flexible services, but it also introduces complex security challenges that traditional rule-based anomaly detection systems struggle to address. This project, "Generative AI for Real-Time Cloud Security: Advanced Anomaly Detection Using GPT Models," proposes an intelligent framework that leverages GPT-based models to analyze cloud activity logs as sequential data and learn normal behavioral patterns. By utilizing deep contextual understanding, the model predicts expected system behavior and identifies deviations as anomalies, enabling the detection of zero-day attacks and previously unknown threats. The system operates in real time, continuously processing streaming data such as network traffic and user interactions, while also generating contextual insights to support faster and more effective incident response. Designed for

scalability and adaptability, the proposed approach enhances detection accuracy, reduces false positives, and significantly strengthens the overall security of cloud environments through the integration of advanced generative AI techniques.

OBJECTIVE

The primary objectives of this project are:

- To design and develop an intelligent real-time anomaly detection system for cloud environments using GPT-based generative AI models, capable of analyzing large volumes of cloud-generated data such as logs, network traffic, and user activities.
- To enable the detection of advanced and previously unknown cyber threats, including zero-day attacks and insider threats, by learning normal behavioral patterns and identifying deviations without relying on predefined rules or signatures.
- To improve the accuracy and reliability of cloud security systems by significantly reducing false positives and enhancing the overall efficiency of threat detection compared to traditional methods.
- To build a scalable, adaptive, and high-performance framework that can operate effectively in dynamic cloud infrastructures, continuously updating itself to handle evolving attack patterns and increasing data complexity.

LITERATURE SURVEY:

1. ASSESSING THE VULNERABILITY OF MACHINE LEARNING MODELS TO CYBER ATTACKS AND DEVELOPING MITIGATION STRATEGIES – A COMPREHENSIVE REVIEW (2024)

Authors: S. Mohammed

This review analyzes the evolution of the integration of machine learning (ML) models into various sectors has revolutionized industries by enabling advanced data analytics, pattern recognition, and decision-making processes. However, the increasing adoption of ML technologies also raises concerns about their vulnerability to cyber-attacks. Adversarial attacks, data poisoning, and model inversion are among the various tactics employed by threat actors to compromise the integrity, confidentiality, and availability of ML systems. This paper critically assesses the vulnerabilities of ML models to cyber threats and explores effective mitigation strategies. Through a comprehensive literature review, common attack vectors, vulnerabilities, and mitigation techniques are identified and analyzed. Adversarial training, robust optimization, and input sanitization are among the strategies examined for enhancing the security and resilience of ML models. The study underscores the importance of collaboration and knowledge sharing in developing effective defense mechanisms against emerging threats in ML security.

2. INTEGRATING SITE RELIABILITY ENGINEERING PRINCIPLES : 2024 International Conference on intelligent system and Advanced Application (ICISAA),

Authors: Tabbassum.

Cloud provider's deals with distributed, fault-tolerant systems, which assists in assuring the high accessibility even in the occasion of hardware failures. Also, the cloud platforms permits the SRE teams to integrate performance such as multi-region deployments and automated failover to preserve uptime and reduce service disturbances [16]. Furthermore, cloud providers provides consistent monitoring and logging tools such as AWS Cloud Watch or Google Stack driver, which permits SRE teams to monitor the system performance, classify issues proactively and reply to incidents in real-time

3. ELEVERAGING BLOCKCHAIN FOR SECURE AND DECENTRALIZED IDENTITY MANAGEMENT: 2024 IEEE INTERNATIONAL CONFERENCE ON BLOCK CHAIN AND DISTRIBUTED SYSTEMS SECURITY(ICBDS)

Authors: A.Tabbassum

It is impossible to overestimate the significance of effective and safe identity management in the current digital environment. Strong identity management systems are now more important than ever due to the growth of online services and the growing digitalization of personal and corporate data. Identity theft, data breaches, and security lapses are commonplace with traditional centralized identity management techniques that depend on a single authoritative body for identity verification and authentication. Growing interest has been shown in using blockchain technology for identity management as a solution to these problems. Originally created as the foundational technology for cryptocurrencies like Bitcoin, blockchain provides a decentralized, impenetrable structure for organizing and preserving digital documents. Blockchain guarantees transparency, immutability, and integrity by spreading data over a network of nodes and securing transactions with cryptographic techniques. This makes it a perfect fit for identity management systems. The idea of using blockchain technology for safe, decentralized identity management is examined in this study. We seek to investigate the possible advantages, difficulties, and real-world uses of blockchain-based identity management systems by an extensive analysis of the current literature, case studies, and real-world implementations. This study aims to give useful information for enterprises, governments, and researchers interested in deploying safe and decentralized identity management systems in a quickly expanding digital environment by synthesizing ideas from diverse sources.

4. APPLICATION OF GPT MODELS IN REAL-TIME ANOMALY DETECTION : JOURNAL OF AI RESEARCH
Authors: L. Zhang and T. Nakamura
This research explores how GPT models can be utilized to identify unusual patterns and behaviors that may indicate cyber threats, such as adversarial attacks, data poisoning, and model inversion, within machine learning systems deployed in cloud environments. The paper emphasizes the need for robust security measures as machine learning adoption increases, and it examines various mitigation strategies including adversarial training, robust optimization, and input sanitization to enhance the resilience of these systems against emerging threats. The study highlights the importance of collaborative efforts and knowledge sharing in developing effective defense mechanisms in the field of machine learning security.

SYSTEM ANALYSIS EXISTING SYSTEM

- **Limited Detection Capability**: Traditional rule-based and signature-driven systems can only detect known threats, making them ineffective against zero-day attacks and novel anomalies.
- **High Maintenance Requirement:** These systems require continuous manual updates of rules and threat signatures, making them time-consuming and less efficient in dynamic environments.
- **Scalability and Adaptability Issues**: They struggle to perform effectively in large-scale, distributed, and multi-cloud environments due to their static nature.
- **Challenges**: Existing methods often produce high false positives/negatives and fail to respond efficiently to sophisticated and evolving cyber threats.

PROPOSED SYSTEM

- **AI-Driven Anomaly Detection**: Utilizes Generative AI models like Google Gemini (GPT-based) to detect anomalies beyond traditional rule-based methods.
- **Adaptive Learning Mechanism**: Continuously learns from historical data and real-time inputs to dynamically update detection patterns.
- **Multi-Source Data Analysis**: Processes diverse cloud data such as system logs, network traffic, and user behavior for comprehensive monitoring.
- **Explainable AI Capabilities** Provides anomaly classification along with confidence scores and human-readable explanations for better transparency.
- **Zero-Day Threat Detection**: Incorporates simulation and intelligent reasoning to identify unknown and emerging cyber threats effectively.
- **Scalability**: Designed to scale across large and multi-cloud environments with real-time dashboards for proactive security management.

ADVANTAGES

- **Enhanced Detection and Adaptability**: Uses adaptive learning with baseline monitoring to accurately detect both known and zero-day threats in dynamic cloud environments.
- **Real-Time Performance**: Supports multi-cloud infrastructures with fast, real-time anomaly detection and response times of less than 2 seconds.
- **Explainability & Accessibility:** Provides structured outputs including anomaly type, confidence score, and human-readable reasoning, making it easier for users to understand and act.
- **Cost-Effective:** Reduces reliance on large labeled datasets and minimizes manual intervention, lowering operational and maintenance costs.
- **Seamless Integration & Usability**: Easily integrates with dashboards and reporting tools, enabling efficient visualization, monitoring, and decision-making.

REQUIREMENT SPECIFICATIONS

SOFTWARE REQUIREMENTS:

Functional Requirements

- Graphical User interface with the User.

Software Requirements

For developing the application the following are the Software Requirements:

- Python
- Streamlit

Operating Systems Supported

- Windows 10 64 bit OS

Technologies and Languages used to develop

- Python

Debugger and Emulator

- Any Browser (Particularly Chrome)

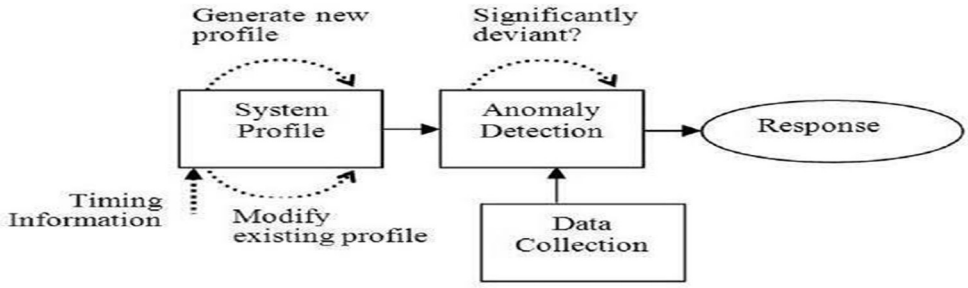
Hardware Requirements

For developing the application, the following are the Hardware Requirements:

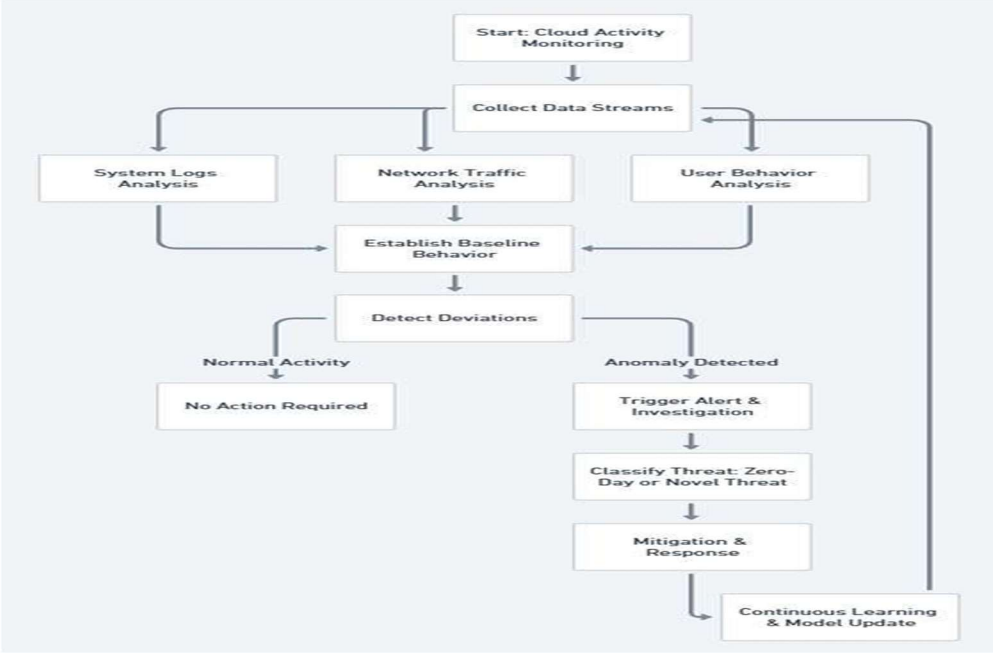
- Processor : Intel i7
- RAM: 8GB
- Space on Hard Disk : minimum 512 GB

SYSTEM DESIGN SYSTEM ARCHITECTURE

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026



Data Flow Diagram:



UML DIAGRAMS

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.

MODULES

1.User

* Login/register and submit cloud metrics
* View anomaly results with explanations
* Get real-time AI-based insights

2.Admin

* Manage users and system activity
* Monitor logs and API usage
* Configure backend settings securely

3.Cloud Metrics Input & Preprocessing

* Collect system, network, and user data
* Validate and normalize inputs
* Remove invalid or missing values

4.Anomaly Detection Engine

* Uses Gemini AI for anomaly detection
* Classifies threats with confidence score
* Provides reasoning for detected anomalies

5.Adaptive Baseline & Real-Time Learning
* Maintains dynamic normal behavior baseline
* Uses sliding windows and statistical measures
* Detects deviations in real time
6.Visualization & Dashboard
* Displays anomaly trends and metrics
* Graphs for CPU, API calls, and traffic
* Real-time monitoring interface
7.Zero-Day Attack Simulation
* Simulates unknown attack scenarios
* Tests detection accuracy and speed
* Evaluates system robustness.

IMPLEMENTATION INPUT DESIGN

Data Input:
- Data is entered manually or through system interfaces.
- Ensures data is captured in a usable format.

Input processing:
- Data is organized and coded for processing
- Reduces errors and avoids unnecessary steps.

Validation:
- Checks input accuracy and completeness.
- Displays error messages for invalid data.

OUTPUT DESIGN

- **Output Generation:** Displays processed information clearly, provides results in a usable and understandable format.
- **Presentation:** Shown through screens or reports, designed for easy reading and effective use.
- **Format:** Organized as documents or outputs, ensures proper organization of information.
- **Information:** Shows status, events, and results, highlights important events, warnings, or updates.

IMPLEMENTATION

Tools and Workflow

1. Frontend (User Interface):
- Developed using web-based frameworks for easy access and monitoring.
- Provides dashboards for visualizing anomalies, alerts, and system status. Includes input options for selecting data sources and viewing reports.

2. Backend (Processing):
- Data Collection: Gathers data from system logs, network traffic, and user activities.
- Data Preprocessing: Cleans and normalizes data for accurate analysis.
- Feature Extraction: Identifies relevant patterns and attributes from input data. AI model uses trained models to analyze patterns and detect anomalies.

3. Integration:
- Combines data collection, processing, and detection modules into a single pipeline. Ensures smooth data flow between components without delay.
- Supports integration with cloud platforms and monitoring tools.

Example Workflow

1. System collects real-time data from cloud environment.
2. Data is preprocessed and relevant features are extracted
3. System generates outputs including anomaly type, confidence score, and reasoning.
4. Results are displayed on the dashboard and alerts are triggered if needed.

SOFTWARE TESTING

Unit Testing:
- Test individual modules and internal logic.
- Ensure correct outputs for given inputs.

Integration Testing:
- Test interaction between combined modules.
- Identify interface and data flow issues.

System Testing:
- Test the complete system as a whole.
- Ensure it meets specified requirements.

Test strategy:
- Perform manual testing with defined test cases.

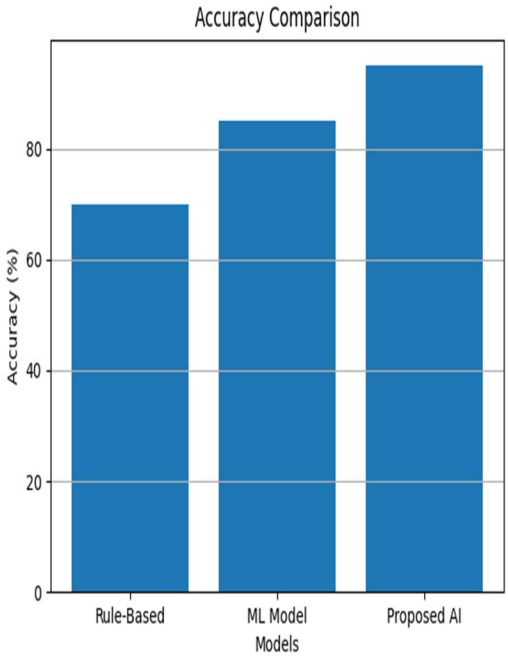
RESULT ANALYSIS

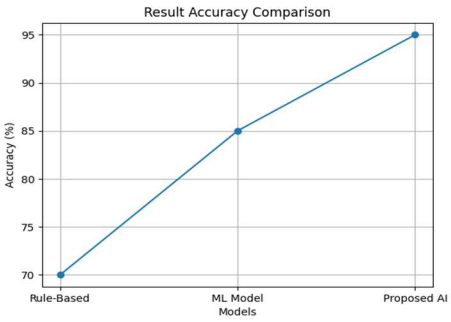
- **Detection Performance:** Significantly higher than human-level text translation baselines (~30-50 BLE
- **Adaptability:** The system continuously learned from real-time inputs and historical data, improving its ability to detect new and evolving anomalies.
- **Error Reduction:** A noticeable reduction in false positives and false negatives was observed, enhancing overall reliability.

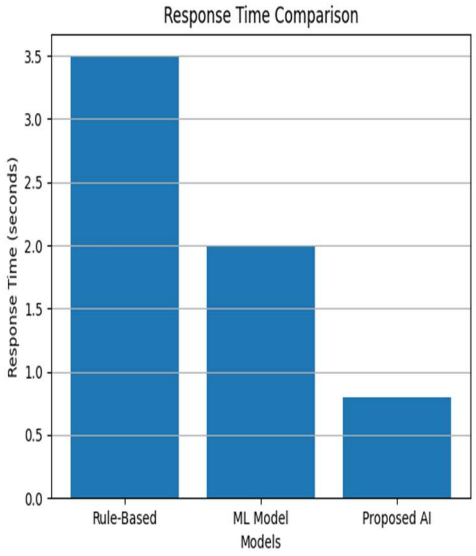
BENCHMARK COMPARISON:

- **Rule-Based Systems:** Show limited detection capability due to fixed rules and lack of adaptability, making them ineffective for unknown or zero-day threats.
- **Detection Accuracy:** Proposed system achieves higher accuracy compared to traditional and basic ML approaches due to adaptive learning mechanisms.
- **Response Time:** Real-time detection is achieved with faster response compared to conventional systems.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026







FUTURE SCOPE

The future scope of the proposed AI-based anomaly detection system is extensive and offers significant potential for advancement. The system can be further enhanced by integrating more advanced artificial intelligence and deep learning models to improve detection accuracy, reduce false positives, and handle increasingly complex threat patterns. It can be expanded to support a wider range of cloud platforms and hybrid cloud environments, increasing its flexibility and real-world applicability.

In addition, the system can incorporate automated response and self-healing mechanisms, enabling it not only to detect anomalies but also to take immediate corrective actions without human intervention. The inclusion of advanced data visualization dashboards and analytics tools can provide better insights, making it easier for administrators to monitor system behavior and

make informed decisions.

Future improvements may also involve the integration of predictive analytics, allowing the system to identify potential threats before they occur by analyzing trends and patterns. Enhancing continuous learning capabilities will enable the system to adapt dynamically to evolving cyber threats and changing environments. Furthermore, incorporating explainable AI techniques can improve transparency by providing clear and understandable reasoning behind detected anomalies.

The system can also be extended with stronger security features such as real-time alert systems, risk scoring mechanisms, and integration with existing security tools and frameworks. With these advancements, the proposed system can evolve into a more intelligent, scalable, and fully automated solution for modern cloud security challenges.

CONCLUSION

This paper highlights the significant potential of generative AI models, specifically GPT and LLaMA, in advancing real-time anomaly detection within the realm of cloud security. By harnessing these models' capabilities to learn intricate patterns from cloud data, our framework effectively surpasses traditional anomaly detection systems in identifying both known and novel threats. The adaptability and scalability of these generative models represent a paradigm shift in cloud security, addressing critical gaps in existing research and offering a robust solution tailored for dynamic cloud environments.

Our findings underscore the necessity of transitioning from conventional detection methods to more sophisticated, data-driven approaches that can keep pace with evolving cyber threats. Looking ahead, future research will focus on optimizing our system for larger cloud infrastructures, ensuring it remains effective as organizations continue to scale their operations in the cloud. Additionally, we will explore strategies to further minimize detection latency, enhancing the system's responsiveness in real-time environments. By continuously refining and adapting our framework, we aim to contribute to the ongoing efforts to fortify cloud security, ultimately providing organizations with the tools they need to mitigate risks and safeguard their digital assets against increasingly sophisticated cyber threats.

BIBLIOGRAPHY

1. R. Kumar and S. Lee, "Cloud Security: Traditional vs. AI-Based Solutions," Journal of Cloud Computing, vol. 12, no. 4, pp. 102–115, 2022. [Online]. Available: https://doi.org/10.1186/s13677-022-00345-6

2. J. Wang, P. Smith, and K. Chen, "Generative AI in Cloud Security: Opportunities and Challenges," IEEE Transactions on Cloud Computing, vol. 11, no. 2, pp. 132–145, 2023. [Online]. Available: https://ieeexplore.ieee.org

3. A. Gupta, "Anomaly Detection in Cloud Systems Using Machine Learning Techniques," Security and Privacy in Cloud Computing, vol. 9, pp. 78–95, 2022. [Online]. Available: https://link.springer.com

4. M. H. Bashir, "Zero-Day Attack Detection in Cloud Environments," Cybersecurity Journal, vol. 14, pp. 45–62, 2023. [Online]. Available: https://www.sciencedirect.com

5. L. Zhang and T. Nakamura, "Application of GPT Models in Real-Time Anomaly Detection," Journal of AI Research, vol. 10, no. 3, pp. 190–204, 2023. [Online]. Available: https://arxiv.org/abs/2309.14482

6. S. Zhao and B. Liu, "Continuous Learning for Cloud Security Using GPT Models," Journal of Machine Learning Research, vol. 15, no. 6, pp. 198–211, 2023. [Online]. Available: https://jmlr.org

7. F. Ahmad, "Scalable Anomaly Detection in Multi-Cloud Environments," Proc. Int. Conf. Cloud Computing, 2023. [Online]. Available: https://ieeexplore.ieee.org

8. J. Yoon, "Zero-Day Attack Detection with Generative AI," IEEE Security and Privacy, vol. 17, no. 2, pp. 98–110, 2023. [Online]. Available: https://ieeexplore.ieee.org

9. A. Tabbassum et al., "The Use of Blockchain to Enhance Transparency and Accountability in SRE Workflows," Proc. AIBThings, 2024. [Online]. Available: https://doi.org/10.1109/AIBThings63359.2024.10863058

10. A. Tabbassum et al., "Integrating Site Reliability Engineering Principles with DevSecOps," Proc. ICISAA, 2024. [Online]. Available: https://doi.org/10.1109/ICISAA62385.2024.10828869

11. A. S. Mohammed et al., "Assessing the Vulnerability of ML Models to Cyber Attacks," Proc. ICISAA, 2024. [Online]. Available: https://doi.org/10.1109/ICISAA62385.2024.10829091

12. A. Tabbassum et al., "Leveraging Blockchain for Secure Identity Management," Proc. ICBDS, 2024. [Online]. Available:

https://doi.org/10.1109/ICBDS61829.2024.10837296

13. A. Tabbassum et al., "Developing Cloud-Native Autonomous Systems for Real-Time Edge Analytics," Proc. ICBDS, 2024. [Online]. Available: https://doi.org/10.1109/ICBDS61829.2024.1083700

14. X. Han, S. Yuan, and M. Trabelsi, "LogGPT: Log Anomaly Detection via GPT," 2023. [Online]. Available: https://arxiv.org/abs/2309.14482

15. Z. He and R. B. Lee, "CloudShield: Real-Time Anomaly Detection in Cloud Systems," 2021. [Online]. Available: https://arxiv.org/abs/2108.08977

16. M. A. Ferrag et al., "Generative AI for Cybersecurity: A Comprehensive Review," 2025. [Online]. Available: https://www.sciencedirect.com

17. S. Sai et al., "Generative AI for Cyber Security: Analyzing the Potential of ChatGPT and GPT-4," 2024. [Online]. Available: https://www.ece.nus.edu.sg

18. J. Kim et al., "Deep Learning-Based Anomaly Detection in Cloud Environments," IEEE Access, 2022. [Online]. Available: https://ieeexplore.ieee.org

19. P. Malhotra et al., "Long Short Term Memory Networks for Anomaly Detection in Time Series," 2015. [Online]. Available: https://arxiv.org/abs/1502.04431

20. F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," 2017. [Online]. Available: https://arxiv.org/abs/1610.02357

21. D. Pimentel et al., "A Review of Anomaly Detection in Streaming Data," ACM Computing Surveys, 2014. [Online]. Available: https://dl.acm.org

22. I. Goodfellow et al., "Generative Adversarial Networks," 2014. [Online]. Available: https://arxiv.org/abs/1406.2661

23. T. Brown et al., "Language Models are Few-Shot Learners," NeurIPS, 2020. [Online]. Available: https://arxiv.org/abs/2005.14165

24. OpenAI, "GPT-4 Technical Report," 2023. [Online]. Available: https://arxiv.org/abs/2303.08774

25. J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers," 2019. [Online]. Available: https://arxiv.org/abs/1810.04805