ISSN: 3049-3277

International Journal of Integrative Studies (IJIS)

Journal homepage: www.ijis.co.in

RESEARCH ARTICLE

AI Security Framework: A Real-Time Threat Detection and Compliance Monitoring System

[1]K. Roshan, [2]T. Laya, [3]K. Saketh, [4]M. Kaushik, [5]Potharaju Chandra Mounika

[2]Department of CSE (CS),Vignana Bharathi Institute of Technology,Hyderabad,Telangana, India.
Email: tlaya504@gmail.com.
[3]Department of CSE (CS),Vignana Bharathi Institute of Technology,Hyderabad,Telangana, India.
Email: sakethshetty1234@gmail.com.
[4]Department of CSE (CS),Vignana Bharathi Institute of Technology,Hyderabad,Telangana, India.
[5]Assistant Professor, Department of CSE (CS),Vignana Bharathi Institute of Technology,Hyderabad, Telangana, India.
Email: mounika803@gmail.com

ABSTRACT

This paper presents an AI Security Framework designed to provide a practical and effective solution for securing AI-based systems. The proposed framework focuses on protecting sensitive data, securing AI models, and ensuring compliance with privacy regulations. It includes key components such as secure data handling, threat detection, access control, continuous monitoring, and real-time alerts. The system also supports safe integration of LLMs while reducing risks such as adversarial attacks and unauthorised access.

The main goal of this framework is to offer a simple, scalable, and reliable approach to AI security. It helps organisations use AI technologies safely while maintaining strong protection against modern cyber threats.

Keywords: *AI Security, Large Language Models, Threat Detection, Data Protection, Cybersecurity, anomaly detection, risk assessment.*

INTRODUCTION

Artificial intelligence (AI) and large language (LLMs) are transforming modern digital systems by enabling automation, intelligent decision-making, and improved user interaction. Organisations across various domains are adopting these technologies to enhance productivity and drive innovation. However, along with these benefits, AI systems also introduce significant security challenges that cannot be ignored. AI-based systems rely heavily on data and models, making them vulnerable to threats such as data leakage, model misuse, adversarial attacks, and unauthorised access. Traditional cybersecurity solutions are not fully effective in addressing these challenges because AI systems have unique characteristics, including dynamic learning behaviour and complex data processing. As a result, there is a need for a specialised security approach that protects both data and AI models.

[1]Department of CSE (CS), Vignana Bharathi Institute of TechnologyHyderabad, Telangana, India.
Email: roshankappala@gmail.com.

Corresponding Author: K. Roshan,
Department of CSE (CS),Vignana Bharathi Institute of Technology Hyderabad, Telangana, India.
Email: roshankappala@gmail.com

DOI: https://doi.org/10.63856/ijis/v2i4/00034

How to cite this article: Roshan, K., (2026). AI Security Framework: A Real-Time Threat Detection and Compliance Monitoring System. *International Journal of Integrative Studies*, 2(4), 54-58.

Source of support: Nil

Conflict of interest: None.
Received: 26//03/2026 **Revised:** 30/03/2026 **Accepted:** 02/04/2026
Published: 22/04/2026

Several existing frameworks, such as the NIST AI Risk Management Framework, Google Secure AI Framework (SAIF), and Microsoft security guidelines, provide structured methods for securing AI systems. Still, they are often complex and difficult to implement in real-world environments. This creates a gap between theoretical guidelines and practical implementation. To address this gap, this paper proposes an AI Security Framework that provides a simple, scalable, and effective solution for securing AI systems. The main goal of the project is to design a unified security framework that ensures secure data handling, real-time threat detection, access control, continuous monitoring, and compliance with privacy regulations. The framework also supports the safe integration of LLMs while reducing risks related to adversarial attacks and misuse, thereby enabling organisations to adopt AI technologies securely and efficiently.

Related Work

Security in artificial intelligence (AI) and large language model (LLM) systems has been widely studied in recent years. Many research works focus on using machine learning and deep learning techniques for threat detection and intrusion detection. These methods help in identifying unusual patterns and improving system security. In addition, frameworks such as the NIST AI Risk Management Framework, Google Secure AI Framework (SAIF), and Microsoft security guidelines provide structured approaches for managing AI-related risks, including data protection, governance, and model security.

However, most existing solutions are complex, resource-intensive, and difficult to implement in real-world environments. Many approaches address only specific issues such as anomaly detection or access control, rather than providing a complete solution. They also lack real-time monitoring and integrated security mechanisms. Therefore, there is a need for a simple and unified AI security framework that combines data protection, threat detection,

access control, and compliance in a single system, which is addressed in the proposed work.

System Architecture And Design

The proposed AI Security Framework is designed as a structured and sequential system to support secure and efficient data processing. The architecture starts with multiple input sources such as user inputs, system logs, network activity, and application data. This information is passed to the data handling and preprocessing module, where it is collected, filtered, validated, and prepared for further processing. After preprocessing, the data is stored in a secure storage system that preserves confidentiality and integrity through controlled access and protection mechanisms. The stored data is then sent to the AI/LLM analysis engine, where large language models are applied to identify patterns, detect anomalies, and support intelligent decision-making processes.

Back-End

- **Data Processing and Storage:**

The system collects, validates, and stores data securely with controlled access to ensure confidentiality and integrity

- **AI Analysis and Threat Detection:**

AI/LLM models analyse data to identify patterns

- , detect anomalies, and recognise potential security threats.
- Access Control, Monitoring, and Compliance:

The system enforces authorised access, continuously monitors activities, and ensures compliance with security policies and regulations.

Front-End

- **User Interface and Dashboard:**

Provides an interactive dashboard that displays system status, security insights, and real-time updates.

- **Visualisation and Alerts:**

Presents detected threats, alerts, and analysis results using clear visual elements to help users quickly understand system conditions.

- **User Interaction and Control:**

Allows users to monitor activities, view reports,and take necessary actions based on system recommendations.

Methodlogy

The framework follows a structured approach to ensure secure data processing and effective threat management. The methodology combines data management, AI-driven analysis, security controls, and continuous monitoring to detect and mitigate potential risks in real time. Each stage of the process is designed to preserve data integrity, enforce secure access, and support reliable decision-making within the systems.

- **Data Collection and Preprocessing:**

The system gathers data from user inputs, logsand network activity, then cleans and validates it for accurate analysis.

- **Secure Storage and AI Analysis:**

The processed data is stored securely and analysed using AI/LLM models to identify patterns and detect anomalies.

- **Threat Detection and Monitoring:**

The system identifies potential threats, assesses risks, and continuously monitors system activities in real time.

- **Access Control and Reporting:**

The framework enforces authorised access, ensures compliance with security policies, and generates alerts, dashboards, and reports for user action.•

- **AI/LLM-Based Analysis:**

The stored data is analyzed using AI and large language models to identify patterns, detect anomalies, and understand system behaviour.

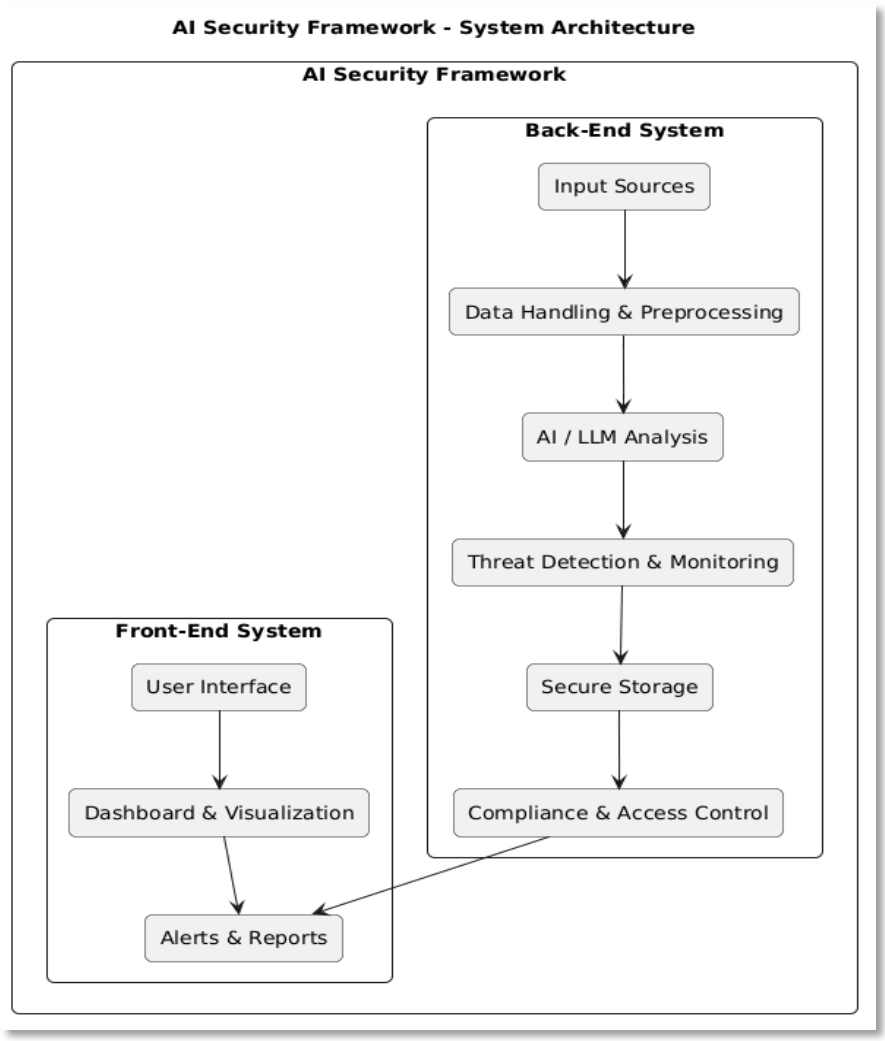

Fig.1 System Architecture

Implementation Details and Results

The implementation of the proposed AI Security Framework focuses on developing a secure, scalable, and efficient system that integrates data processing, AI-based analysis, and security mechanisms. The system is implemented as a web-based platform that enables user interaction, real-time monitoring, and risk evaluation. It combines structured data handling, intelligent analysis, and compliance checking to provide a complete security solution for AI-based systems.

A. Ai-Based Risk Evaluation and Analysis:

The implementation of the proposed AI Security Framework focuses on analysing system inputs and evaluating security risks using AI-driven techniques. The system processes user-provided data, system configurations, and activity inputs to assess the overall security posture. Large Language Models (LLMs) and analytical logic are used to identify patterns, detect inconsistencies, and highlight potential vulnerabilities within the system.

The framework generates a quantitative risk score (e.g., 75%) that represents the current security level based on evaluated parameters. This score is derived from multiple factors such as data protection practices, access control mechanisms, and compliance status. The system also provides insights into identified risks and suggests improvements, enabling users to enhance system security effectively.

B. Compliance Evaluation and Remediation Logic:

The framework incorporates a compliance evaluation mechanism to ensure adherence to security standards and data protection regulations, such as DPDP guidelines. The system checks whether the provided inputs meet required security practices, including encryption, secure storage, and controlled access. Any deviations from standard practices are identified and highlighted as compliance gaps.

In addition to evaluation, the system implements remediation logic by generating recommendations for improving security. These include suggestions such as enabling encryption, strengthening access controls, and improving data handling practices. This approach helps users not only identify issues but also understand the necessary steps to mitigate risks and maintain compliance.

C. System Deployment and Output Visualisation:

The proposed system is implemented as a web-based application, enabling easy access and interaction through a user interface. The framework integrates

front-end components for user interaction and back-end modules for processing, analysis, and security evaluation. The deployment ensures scalability and supports real-time monitoring of system activities.

The results are presented through an interactive dashboard that displays risk scores, compliance status, and security insights in a clear and structured format. Visual elements such as indicators, checklists, and summaries help users quickly understand the system's security condition. The implementation demonstrates that the framework effectively combines AI-based analysis with security mechanisms to provide accurate risk evaluation and support informed decision-making.

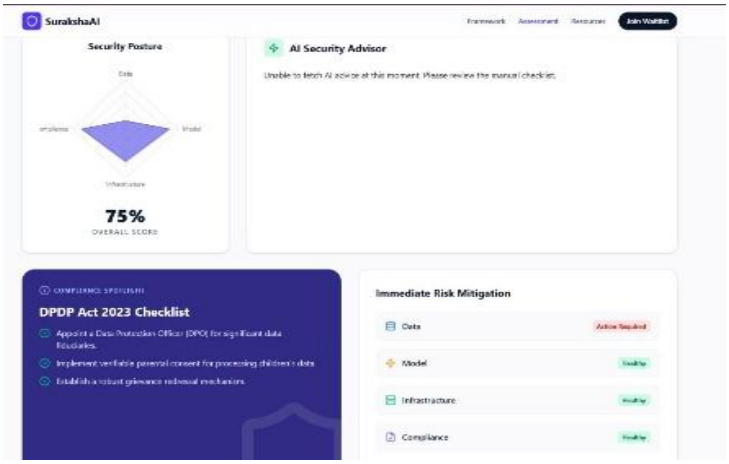

Fig .2 Framework Dashboard

D. Real-Time Monitoring and Alert Generation:

The implemented framework supports continuous monitoring of system activities to ensure timely detection of security risks. The system tracks user inputs, data flow, and processing behaviour in real time, enabling early identification of suspicious or abnormal patternsIn addition to monitoring, the system generates alerts and notifications based on detected risks and anomalies. These alerts are integrated into the dashboard, providing users with instant updates on security status. The alert mechanism helps in prioritising critical issues and enables a quick response to potential threats. The results demonstrate that the framework effectively maintains continuous surveillance and improves overall system security through proactive monitoring and alert generation.

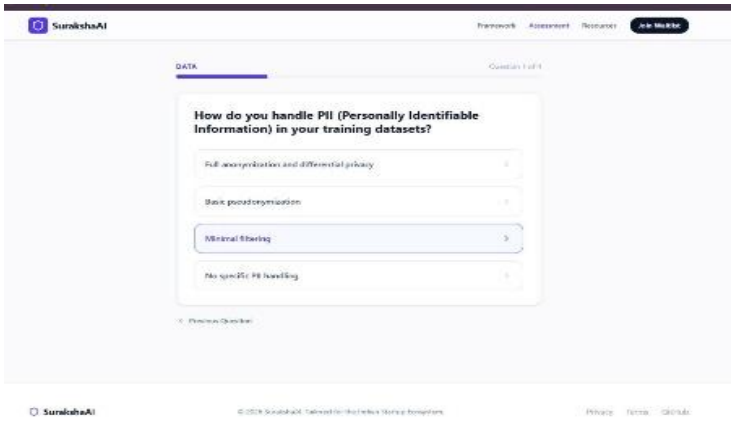

Fig .3 Framework Insights

Conclusion

The proposed AI Security Framework provides an effective solution for securing AI-based systems by integrating data processing, AI-driven analysis, and security mechanisms into a unified architecture. The system successfully evaluates security posture, identifies potential risks, and ensures compliance with data protection standards. By utilising structured data handling and AI/LLM-based analysis, the framework is able to detect vulnerabilities and generate meaningful insights for improving system security.

The implementation results demonstrate that the framework can accurately assess risk levels, generate security scores, and provide clear visual outputs through dashboards and alerts. The inclusion of real-time monitoring and compliance evaluation enhances the reliability and responsiveness of the system. Overall, the proposed framework offers a scalable and efficient approach for protecting AI systems and supports informed decision-making for maintaining secure and compliant environments.

Referances

1. A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, “Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications,” IEEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2347–2376, 2015.
2. NIST, “Artificial Intelligence Risk Management Framework (AI RMF 1.0),” National Institute of Standards and Technology, 2023.
3. Google, “Secure AI Framework (SAIF),” Google Research, 2023.
4. Microsoft, “Responsible AI Standard and Security Guidelines,” Microsoft Corporation, 2022.
5. Databricks, “AI Security Framework 2.0,” Databricks Documentation, 2023.
6. OWASP Foundation, “OWASP Top 10 for Large Language Model Applications,” 2023.
7. K. Kim and J. Park, “Deep Learning for Network Threat Detection,” IEEE Transactions on Network Security, 2019.
8. P. Sharma and R. Singh, “AI for Cyber Security: Challenges and Opportunities,” ACM International Conference on Security, 2020.
9. R. Kumar and M. Patel, “Predictive Threat Analysis Using AI,” International Journal of Cybersecurity, 2023.
10. I. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and Harnessing Adversarial Examples,” International Conference on Learning Representations (ICLR), 2015