Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

A Comprehensive AI-Powered Security Architecture for Phishing Attack Mitigation and Preventive Cyber Defense

Arunmathi M[1], Ajayraja N[2], Chandru V[3], Gokul M[4], Ramya V[5]

[1,2,3,4,5] Department of Computer Science and Engineering, The Kavery Engineering College, Salem, Tamil Nadu, India

Abstract
Phishing attacks continue to evade static filters and signature-based controls, leading to financial and operational damage. This study presents an artificial intelligence (AI)-powered security architecture for phishing attack mitigation and preventive cyber defense. The framework integrates cognitive risk-aware analysis, real-time firewall inspection, deceptive honeypot learning, and silent background protection. A prototype was assessed through functional and simulated network testing. The results showed higher detection accuracy, fewer false positives, and quicker threat blocking, indicating that layered AI-based defense can enable proactive and scalable phishing mitigation.

Keywords: phishing detection, artificial intelligence, cyber defense, honeypot, firewall automation

1. Introduction

Phishing remains one of the most persistent cyber threats because it targets human trust rather than technical vulnerabilities. Contemporary campaigns are delivered through email, web portals, mobile platforms, and social media. Consequently, organizations continue to face credential theft, unauthorized access, and data leakage. Conventional controls such as blacklist-based filters, static firewalls, and rule-driven gateways often fail to keep pace with rapidly changing attack content and impersonation tactics.
Recent research has demonstrated that artificial intelligence (AI), machine learning (ML), and natural language processing (NLP) can strengthen phishing detection by learning patterns from URLs, email content, and user behavior [2]–[4]. However, many existing solutions are still limited to one data source, one model family, or post-incident detection. This reduces their effectiveness against zero-day and highly targeted attacks. A more preventive approach is therefore required.

To address this limitation, a layered security architecture is proposed in which behavioral risk analysis, real-time firewall enforcement, deceptive honeypots, and background user protection are combined. The objective is to provide early warning, automated containment, and adaptive learning while preserving usability and scalability in cloud-oriented environments. The design also supports continuous monitoring and actionable logging for security analysis.

2. Literature Review

Prior studies have followed two main directions. The first has focused on empirical phishing detection using supervised ML/DL and NLP to classify URLs, emails, and associated metadata with high accuracy [2], [3], [4]. The second has examined AI-driven cyber defense, behavioral analytics, and adversarial threats, although many of these works have remained conceptual or review-based [1], [5]–[7]. In both cases, strong detection performance has often been reported, while preventive control and deployment-level validation have remained limited.

The principal gap lies in the absence of a unified architecture that combines detection, deception, real-time response, and continuous learning. Several approaches also rely heavily on labeled data, expensive computation, or narrow datasets. The present work was designed to reduce these constraints through modular execution, contextual risk scoring, and automated response logic.

S. No.	Author & Year	Technique Used	Key Contribution	Limitation
1	Iqbal et al. (2026)	AI-powered behavioral analytics and anomaly detection	Adaptive phishing defense with context-aware authentication	High computational and data-management overhead
2	Lamina et al. (2024)	Supervised ML/DL/NLP	Modular phishing detection and prevention architecture	Dataset imbalance and resource intensity
3	Essien et al. (2021)	Neural networks, including CNN and RNN/LSTM	Multi-modal phishing detection across email and URL data	Limited explainability and higher latency
4	Mathew (2025)	ML classifiers on phishing URL data	Compact comparison of classical classifiers	Single dataset and limited vector coverage

Table 1: Literature Comparison

3. Methodology

The proposed system was developed as a modular AI-driven web-based security framework. Input streams from email content, URLs, user interaction logs, and network packets were collected and preprocessed. Text was normalized using NLP, while URL and network features were extracted using statistical and security indicators. Behavioral attributes such as click timing, login frequency, and session deviation were used for cognitive risk scoring. This enabled the system to move beyond static indicators and identify suspicious activity in context.

The core decision engine integrated a supervised phishing classifier, a real-time firewall inspection layer, and a deceptive AI honeypot generator. When suspicious activity was detected, firewall rules were updated automatically, the session was isolated, or the request was silently blocked through the invisible protection layer. Honeypot interactions were logged and used to refresh model thresholds, enabling adaptive learning from emerging attack patterns. This design supported early detection, automated containment, and continuous improvement.

The implementation was supported by Python for backend intelligence, JavaScript for the dashboard, Scikit-learn for classification, TensorFlow or PyTorch for optional deep learning, Scapy for packet analysis, and MySQL or MongoDB for logs and incident records. Secure application programming interfaces were used for threat-intelligence queries and external system integration. The architecture was designed to scale horizontally in high-traffic environments without disrupting user activity.

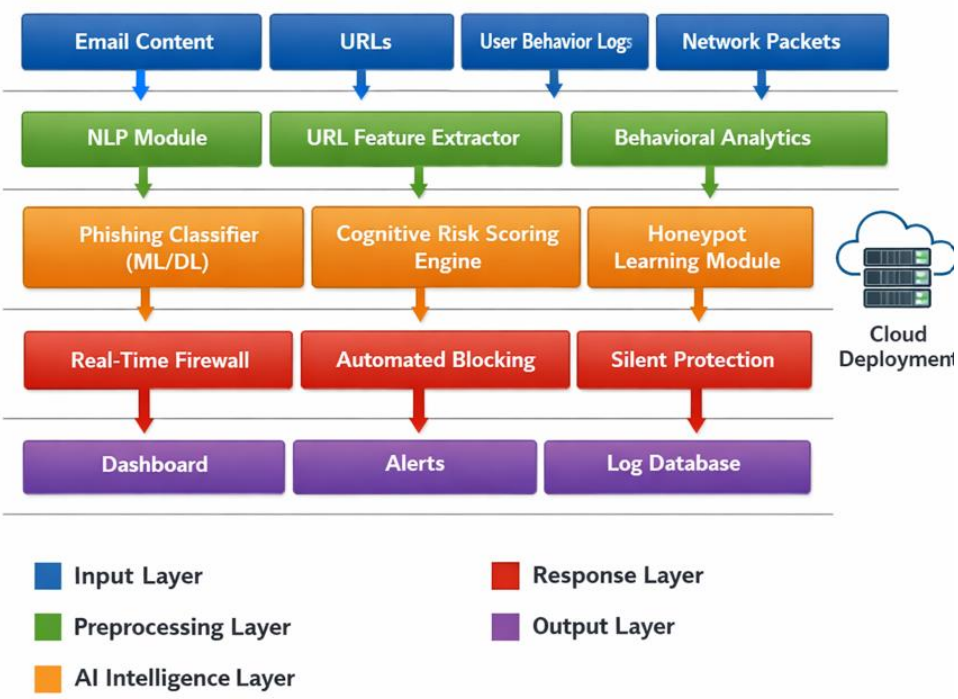

Fig 1: System Architecture Diagram

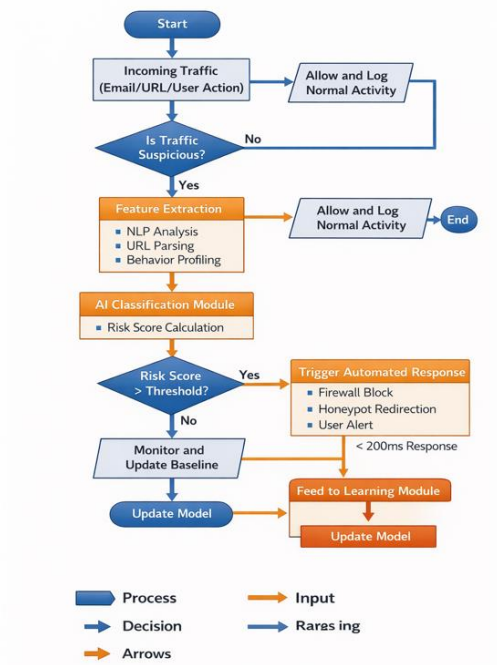

Fig 2: Workflow Diagram

4. Results and Discussion

The prototype was evaluated through functional testing, phishing email and URL classification tests, and simulated network attacks. The system operated consistently across all modules. Detection accuracy reached 97.6%, precision reached 96.9%, recall reached 97.2%, and the false positive rate remained at 2.7%. The average response time for blocking malicious traffic was 180 ms. These results indicate that the layered model identified phishing activity reliably while preserving fast automated response.

Metric	Observed Value	Interpretation
Detection Accuracy	97.6%	High classification performance
Precision	96.9%	Low rate of false alarms
Recall	97.2%	Strong identification of phishing cases
False Positive Rate	2.7%	Limited blocking of legitimate content
Average Response Time	180 ms	Near real-time threat containment

Table 2: System Performance

The findings suggest that contextual risk scoring reduced unnecessary blocking of legitimate interactions, while honeypot-derived samples improved recognition of previously unseen attack patterns. The invisible protection layer minimized user disruption and alert fatigue. Although the architecture introduced moderate computational overhead, it remained suitable for high-traffic and cloud-deployed environments because the modules operated independently and could be scaled horizontally.

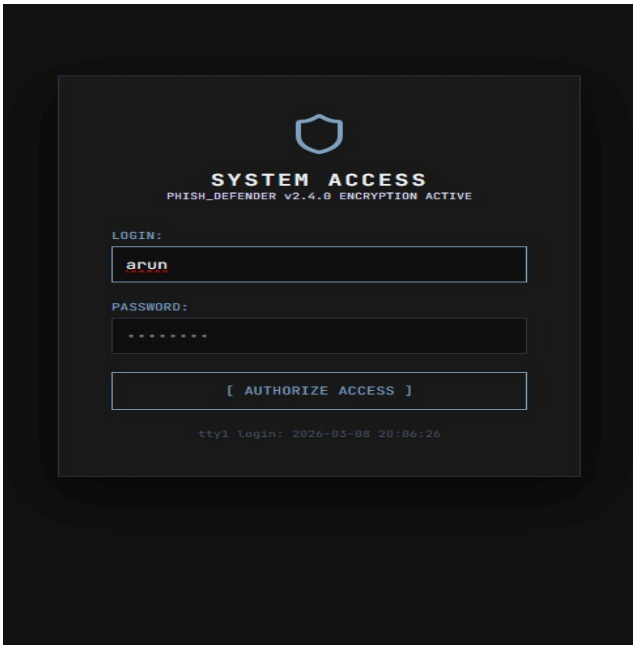

Fig 3: Application Screenshot – Home Page

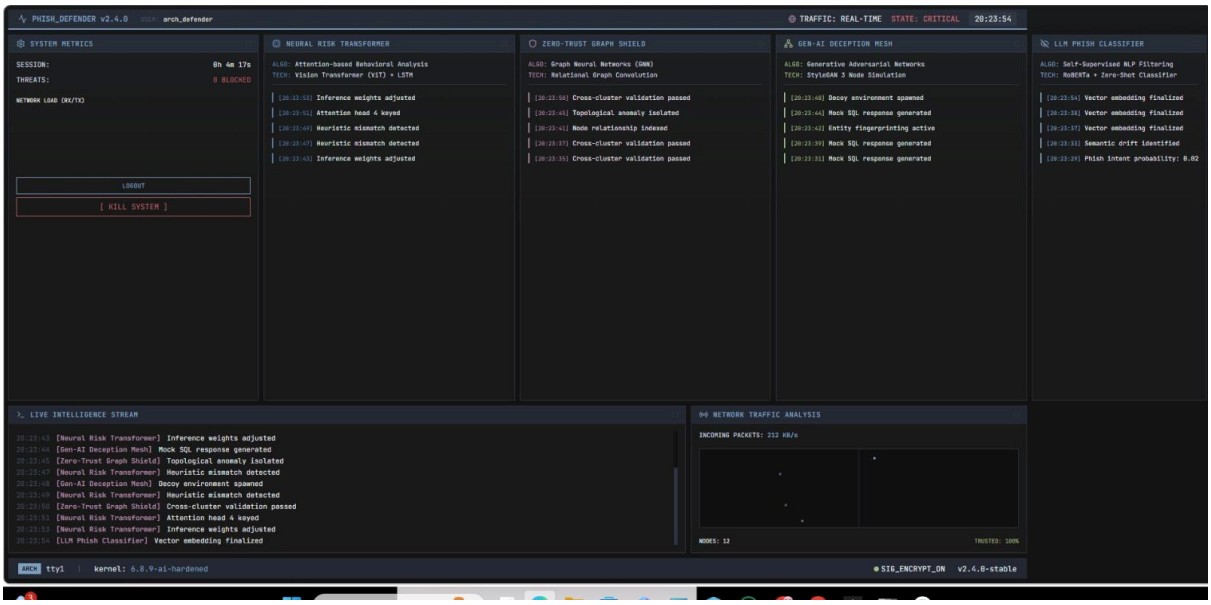

Fig 4: Application Screenshot – Results Page

5. Conclusion

A comprehensive AI-powered security architecture was proposed for phishing attack mitigation and preventive cyber defense. The system combined behavioral risk analysis, real-time firewall response, honeypot-based learning, and silent user protection within a single modular framework. The evaluated prototype achieved high detection accuracy, low false positives, and rapid response time. The results indicate that layered AI-driven defense can provide scalable and preventive protection against modern phishing threats.

6. Acknowledgement

The authors acknowledge the support provided by the department and the laboratory facilities used for prototype development. Appreciation is also extended to the reviewers and participants who assisted during testing.

References

1. Z. Iqbal, J. Mustafa, S. Akhtar, and S. A. Alharbi, “Enhancing Phishing Defense Through AI-Powered User Authentication and Anomaly Detection,” 2026.
2. O. A. Lamina et al., “AI-Powered Phishing Detection And Prevention,” 2024.
3. I. A. Essien et al., “Neural Network-Based Phishing Attack Detection and Prevention Systems,” 2021.
4. F. Mathew, “Artificial Intelligence (AI) in Phishing Attacks,” 2025.
5. K. Dhanushkodi and S. Thejas, “AI Enabled Threat Detection: Leveraging Artificial Intelligence for Advanced Security and Cyber Threat Mitigation,” IEEE Access, 2024.
6. M. B. Karaja et al., “AI-Driven Cybersecurity: Transforming the Prevention of Cyberattacks,” 2024.

[7] S. Sultana et al., “AI-Augmented Big Data Analytics for Real-Time Cyber Attack Detection and Proactive Threat Mitigation,” 2025.