SentinelSphere: AI-Driven Cybersecurity Platform Combining Threat Detection with Security Awareness

Nikolaos Tantaroudas[1], Ilias Karachalios[2] and Andrew McCracken[3]
[1]Institute of Communication and Computer Systems, Athens, Greece
[2]National Technical University of Athens, Leof. Alimou, Katechaki, Athens, Greece
[3]DASKALOS-APPS, Péronnas, France

nikolaos.tantaroudas@iccs.gr
ilias_karachalios@mail.ntua.gr
andrew@daskalos-apps.com

Abstract: The growing complexity of cyber threats coupled with the widening cybersecurity knowledge gap presents new challenges in the organisational security domain of the business sector. This paper introduces SentinelSphere, an innovative platform that redesigns cybersecurity defense by integrating advanced threat detection with interactive cybersecurity awareness education, creating a unified approach to building organisational cyber resilience. SentinelSphere employs an Enhanced Deep Neural Network model with specialised feature engineering to significantly reduce false positives while maintaining high detection accuracy across diverse attack vectors. The system includes a Traffic Light System (TLS) to transform complex threat intelligence into intuitive visual indicators, serving simultaneously as an operational tool for security professionals and an educational interface for non-technical users. This paper further presents a Large Language Model that delivers real-time, context-aware cybersecurity guidance and training. This conversational AI agent operates efficiently on standard enterprise hardware, making advanced security education accessible without requiring specialised infrastructure. SentinelSphere is validated using industry-standard datasets, achieving enterprise-grade performance in threat detection with a 94% F1 score and 69.5% reduction in false positives compared to baseline models. The system successfully processed nearly 11 million security events in 30 minutes, demonstrating scalability for enterprise deployment. This work contributes to the cybersecurity field by demonstrating that effective defense requires not just technological sophistication but also systematic enhancement of human security awareness.

Keywords: Cybersecurity awareness, Real-time anomaly detection, Security education, Large language models, Human-centric threat intelligence, Deep neural networks

1. Introduction

The cybersecurity landscape faces an escalating crisis characterised by increasingly complex attack vectors and a critical shortage of skilled security professionals. Traditional Security Information and Event Management (SIEM) systems, while effective at data aggregation, often overwhelm analysts with false positives and lack the contextual intelligence necessary for rapid threat response. Furthermore, the human factor remains the weakest link in cybersecurity, with 82% of data breaches involving human elements according to recent studies (Verizon, 2023).

The ResilMesh project, funded under the EU Horizon Europe programme (Grant Agreement No. 101119681), has established a comprehensive cybersecurity framework designed to enhance cyber resilience for dispersed, heterogeneous cyber systems. Built on the NATS message streaming framework, ResilMesh provides secure event collection and processing capabilities while implementing resilience engineering best practices including redundancy, segmentation, and dynamic positioning (Bernal et al., 2024) This paper presents SentinelSphere, a next-generation cybersecurity platform developed as an extension to the ResilMesh framework that addresses these challenges through an innovative dual approach: combining advanced AI-driven threat detection with integrated security awareness education. Unlike conventional solutions that treat threat detection and security training as separate domains, SentinelSphere creates a synergistic system where every security event becomes an opportunity for organisational learning and resilience building.

The primary contributions of this work are threefold, with explicit novelty articulated for each: (1) An Enhanced Deep Neural Network (DNN) architecture that achieves 94% F1 score while reducing false positives by 69.5% through innovative feature engineering including HTTP-specific anomaly indicators. The novel contribution extends beyond simply adding HTTP features to a DNN. It lies in the systematic identification and validation of ten application-layer indicators that capture attack semantics invisible to network-layer analysis alone. This enables cross-layer threat correlation that existing single-layer intrusion detection approaches cannot achieve, addressing a recognized gap between network-centric detection and application-aware security monitoring. (2) A Traffic Light System (TLS) that transforms complex threat intelligence into intuitive visual indicators, making security understanding accessible across technical expertise levels; The novelty extends beyond threshold-based

alerting to provide a mathematically grounded risk quantification framework that is simultaneously interpretable by non-technical stakeholders and actionable by security analysts, addressing existing communication gaps between security operations and executive decision-making. (3) A Large Language Model-powered conversational agent based on Microsoft's Phi-4 (Microsoft Research, 2024), optimised through Q4_K_M quantisation for deployment on standard enterprise hardware without GPU requirements. The contribution is not simply adding a chatbot to a dashboard, but demonstrating that sophisticated AI-powered security education can be democratized through careful model selection and compression, removing the infrastructure barrier that has limited adoption of intelligent security assistants to well-resourced organisations.

2. Related Work

The cybersecurity domain has identified proactive threat detection and real-time event processing as critical interventions for mitigating the risks posed by advanced cyber threats such as Advanced Persistent Threats (APTs), DDoS attacks, and phishing schemes. It has been identified that the proactive threat detection and real-time event processing as critical interventions for mitigating the risks posed by advanced cyber threats such as Advanced Persistent Threats (APTs), DDoS attacks, and phishing schemes. The convergence of real-time event processing technologies and machine learning algorithms has obtained significant attention, particularly through platforms that integrate solutions such as AWS Kinesis and predictive machine learning models. Herein, we synthesize recent advancements in proactive threat detection methodologies and event processing, highlighting the current state-of-the-art (SOTA) techniques while discussing the potential of specific platforms in addressing existing limitations.

Proactive threat detection effectively employs real-time event processing technologies to provide timely insights into potential cyber threats. Past studies highlight various mechanisms against DDoS attacks, illustrating the necessity for robust real-time defences due to the diverse and evolving attack vectors encountered today (Taneja., 2023). Recent studies further support this by demonstrating how machine learning classifiers can enhance the detection capabilities of DDoS attacks on network logs, thus promoting the need for rapid analysis of event data to identify potential breaches (Musa et al., 2024). Significant attention has also been directed towards APTs, with recent work proposing a methodology that leverages network flow analysis to detect APT attacks by evaluating changes in normal traffic patterns against established baselines (Duong et al., 2020). This proactive approach enables the integration of advanced processing capabilities and predictive analytics to improve response times for critical incidents.

The application of machine learning to cybersecurity has evolved significantly from rule-based systems to sophisticated deep learning architectures. Past studies have demonstrated the effectiveness of deep learning for intrusion detection, achieving great accuracy. Most approached suffered from high false positive rates in production environments. In addition, Vinayakumar et al. (2019) proposed a hybrid deep learning framework combining CNNs and LSTMs, improving detection of zero-day attacks but requiring substantial computational resources. Our work extends these approaches by incorporating application-layer features, particularly HTTP-specific patterns, enabling cross-layer threat correlation. Unlike prior work that analyses network and application layers independently, SentinelSphere's unified 31-feature space enables detection of sophisticated attacks that manifest across protocol boundaries.

The integration of human factors in cybersecurity has gained prominence following high-profile breaches attributed to social engineering. Aldawood and Skinner (2019) emphasised the importance of continuous security awareness training, while Bada et al. (2019) demonstrated that interactive training significantly reduces susceptibility to phishing attacks. Large Language Models have emerged as promising tools for security education. Past studies have explored GPT-based systems for security questionnaire generation, though deployment required significant computational resources (Szabo et al., 2023). Our approach addresses this limitation through Q4_K_M quantisation of Microsoft's Phi-4 model (Abdin et al., 2024), reducing model size from 28GB to 2.5GB while preserving response quality for cybersecurity guidance tasks. This enables deployment on standard 16GB RAM systems without GPU requirements.

The concept of cyber resilience has evolved from traditional security approaches to encompass preparation, response, and recovery capabilities (Araujo et al., 2024). The NIST Cybersecurity Framework 2.0 provides comprehensive guidance for managing cybersecurity risks (NIST, 2024), while the European approach emphasises resilience in critical infrastructure protection (ENISA, 2023). The ResilMesh platform represents a significant advancement in implementing cyber resilience principles through its distributed architecture and AI-based threat awareness capabilities. Bernal et al. (2024) demonstrated edge-based anomaly detection within the ResilMesh framework, highlighting the importance of distributed processing for critical infrastructure

protection. SentinelSphere builds upon this foundation by adding human-centric security awareness capabilities through LLM-chatbot and HTTP-specific patterns to enhance DNN's capabilities of ResilMesh and as a result its Anomaly Detection module.

3. System Architecture and Design

3.1 Overall Architecture

SentinelSphere adopts a microservices architecture integrated with the ResilMesh security framework, enabling scalable, real-time processing of security events. The system comprises four primary layers: Data Ingestion, Processing and Analysis, Intelligence and Education, and Presentation. The architecture of SentinelSphere is a strategic enhancement to the ResilMesh security framework, integrating as an advanced threat analytics layer while preserving the existing architecture's integrity. SentinelSphere operates through direct subscriptions to ResilMesh's NATS messaging infrastructure, consuming both enriched security events ('enriched_events') and anomaly detection alerts ('ad_events'). This non-intrusive approach allows SentinelSphere to function as a parallel consumer without disrupting established workflows.

The Data Ingestion layer utilises Vector for log collection and transformation, processing diverse data sources including network traffic, application logs, and system events. Events flow through NATS message broker, ensuring reliable, high-throughput message delivery to downstream components. This architecture supports processing rates exceeding 5,000 events per second on standard hardware. Vector serves as the primary data collection and transformation layer, ingesting security events from multiple sources and performing real-time transformations including routing, logging, fusion, filtering, augmentation, reduction, and monitoring.

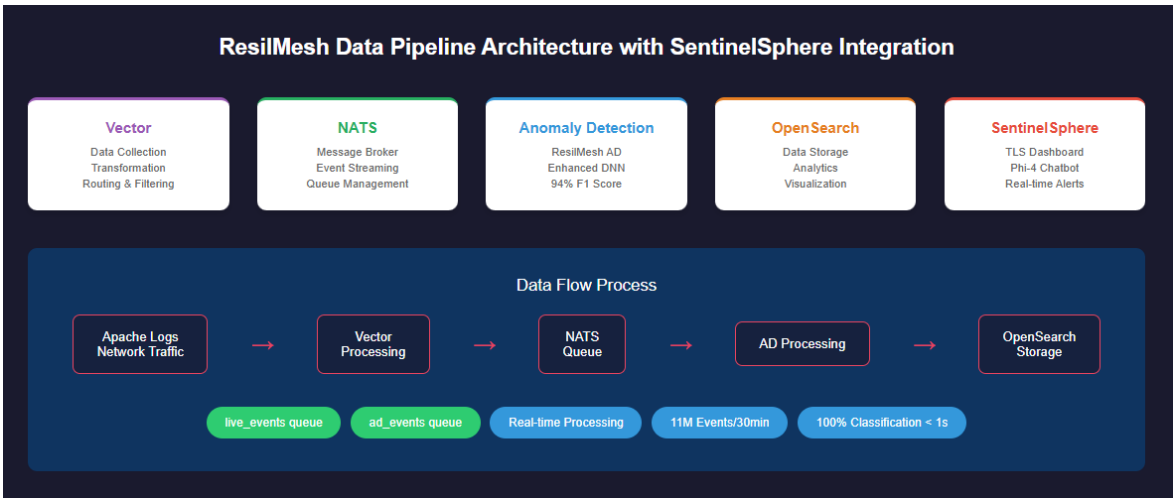

Figure 1: ResilMesh Data Pipeline Architecture - Vector to NATS to OpenSearch Integration

3.2 Enhanced Deep Neural Network Model

The core of SentinelSphere's threat detection capability is an Enhanced DNN model that extends traditional network traffic analysis with application-layer intelligence. The model architecture consists of a 64-64-output neural network enhanced with 31 features: 21 standard network features plus 10 HTTP-specific indicators.

The HTTP feature engineering includes: i) **Request Complexity Score**: Quantifies HTTP request sophistication to identify advanced attacks; ii) **Response Pattern Ratio**: Analyses server response patterns to detect successful exploits; iii) **Path Entropy**: Estimates randomness in request paths, detecting obfuscation attempts; iv) **Automated Tool Detection**: Identifies requests from automated security tools; v) **Attack-Specific Pattern Recognition**: Binary indicators for SQL injection, XSS, and brute force patterns. Training utilised the CIC-IDS2017 and CIC-DDoS2019 datasets (Sharafaldin et al., 2018; Sharafaldin et al., 2019), comprising approximately 400GB of labelled attack data including Web Attack-Brute Force (1,507 samples), Web Attack-XSS (652 samples), Web Attack-SQL Injection (21 samples), and Benign Traffic (168,186 samples).

To address the class imbalance inherent in the dataset, particularly the limited SQL injection samples (n=21), the following methodological safeguards were employed to ensure robustness: i) Stratified splitting: All train/test splits-maintained class proportions using stratified sampling, ensuring minority classes appeared proportionally in both training and evaluation sets. ii) Per-class performance metrics: Beyond aggregate metrics, precision,

recall, and F1 score were computed for each attack class independently (see Table 2), enabling transparent assessment of minority class performance and identification of class-specific weaknesses. iii) Threshold calibration: Rather than using default 0.5 classification thresholds, precision-recall curve analysis was employed on a validation subset to select optimal decision thresholds per class, improving minority class detection. iv) Bootstrap confidence intervals: Performance metrics include 95% confidence intervals computed via 1,000 bootstrap resamples, quantifying uncertainty arising from limited minority class samples.

Model hyperparameters were optimised through grid search on a held-out 20% validation subset: i) Batch size: 32; ii) Learning rate: 0.001 with Adam optimiser; iii) Dropout rate: 0.2; iv) L2 regularisation: 0.001; iv) Maximum epochs: 50 with early stopping (patience=5); v) Decision threshold: 0.4 (calibrated via precision-recall analysis)

3.3 Traffic Light System

The Traffic Light System provides intuitive threat visualisation through a sophisticated scoring algorithm. The system processes anomaly detection events and calculates threat scores from 0-100, determining dashboard status: i) **GREEN (0-30%)**: Normal operation with low threat activity; ii) **YELLOW (30-70%)**: Elevated threat level requiring attention; iii) **RED (70-100%)**: Critical security events demanding immediate response. The TLS scoring algorithm is formally defined as follows:

$$ThreatScore = \min(100, BaseScore \times FrequencyMultiplier \times ClusterFactor \times IP_{Factor} \times DiversityFactor \times SustainedFactor)$$

$$where:$$

$$BaseScore = \frac{\Sigma(anomaly\_weight \times confidence \times temporal\_weight)}{normalization\_factor}$$

Figure 2: Threat Score calculation equation

Table 1: Parameters definition of Traffic Light System of Equation in Figure 2

Parameter	Description	Value Range
BaseScore	Weight sum of detected anomalies normalized to 0-100 scale	0-100
anomaly_weight	Severity weight per threat type (e.g., 0.95 for data exfiltration, 0.90 for unauthorized access, 0.85 for DoS)	0.3-0.95
confidence	Detection confidence score	0.5-1.0
temporal_weight	Time-based decay factor for older events	0-1.0
Normalization_factor	Scaling factor to maintain 0-100 range	Calculated
FrequencyMultiplier	Event rate per minute (1.0x for 1-5 events, 1.5x for 5-20, 2.0x for 20-50, 3.0x for >50)	1.0-3.0x
ClusterFactor	Temporal clustering (1.2x for single cluster, 1.5x for 2 clusters, 2.0x for 3+ clusters)	1.0-2.0x
IPFactor	Attack source concentration (1.5x single IP, 1.3x for 2-4 IPs, 1.0x distributed)	1.0-1.5x
DiversityFactor	Number of different attack types (1.8x for 5+ types, 1.4x for 3-4, 1.2x for 2, 1.0x for 1)	1.0-1.8x
SustainedFactor	Duration of continuous attack activity (2.0x for >3 min high-rate, 1.5x for >2 min, 1.2x for >1 min)	1.0-2.0x

Figure 2 shows the mathematical formula for calculating the final threat score, incorporating base scores, frequency multipliers, clustering factors, IP concentration, diversity factors, and sustained attack indicators. The scoring model employs multi-factor weighted algorithms considering attack type, frequency, clustering, IP concentration, attack diversity, and sustained attack duration. Different anomaly types contribute distinct weights, with data exfiltration and malware receiving the highest weights (0.95) due to critical severity. Threshold Selection Rationale: The 30% and 70% boundaries defining GREEN/YELLOW/RED transitions were determined through empirical analysis during development testing. The 30% threshold was selected to maximise sensitivity to emerging threats while minimising alert fatigue from routine anomalies. The 70% threshold was calibrated to ensure that RED status activations correspond to genuinely critical situations requiring immediate

analyst attention. These thresholds represent operational trade-offs and may be adjusted for specific deployment environments based on organisational risk tolerance.

3.4 LLM-Powered Security Education

SentinelSphere incorporates Microsoft's Phi-4 language model (Abdin et al., 2024), a 14-billion parameter transformer architecture optimised for complex reasoning tasks. Phi-4 was selected based on comparative evaluation against alternative models (Llama-3-8B, Mistral-7B) considering reasoning benchmark performance, parameter count amenable to quantisation, and licensing compatibility for commercial deployment.

The model underwent Q4_K_M quantisation using the llama.cpp framework, reducing model size from 28GB (full precision) to 2.5GB while maintaining acceptable response quality for cybersecurity guidance tasks. Q4_K_M was selected as it provides an optimal balance between compression ratio and quality preservation for instruction-following tasks, with measured perplexity degradation of less than 2%.

Deployment specifications: i) Hardware requirements: Standard business hardware with 16GB RAM, CPU-only (no dedicated GPU required); ii) Inference framework: FastAPI backend integrated with llama.cpp for efficient CPU inference; iii) Performance metrics: 15-20 tokens/second generation rate, time-to-first-token under 2 seconds; iv) Concurrency: Support for 8+ simultaneous conversation streams; v) Memory footprint: 6-8GB during active generation;

The chatbot's cybersecurity knowledge base was constructed from authoritative sources including: i) NIST Cybersecurity Framework 2.0 documentation (NIST, 2024); ii) ENISA threat landscape reports (ENISA, 2023); iii) OWASP Top 10 vulnerability descriptions and remediation guidance; iv) Custom prompt engineering templates for security-contextualised responses.

The chatbot's utility is assessed through its ability to provide contextually appropriate, accurate, and actionable cybersecurity guidance. Qualitative evaluation during development confirmed that the model successfully handles common cybersecurity queries including threat explanation, best practice recommendations, and incident response guidance, while maintaining conversational accessibility for non-technical users.

4. Implementation and Integration

4.1 ResilMesh Integration

SentinelSphere seamlessly integrates with the ResilMesh security framework through subscription to NATS message topics. The system consumes both enriched security events and anomaly detection alerts, enabling comprehensive threat correlation without disrupting existing workflows.

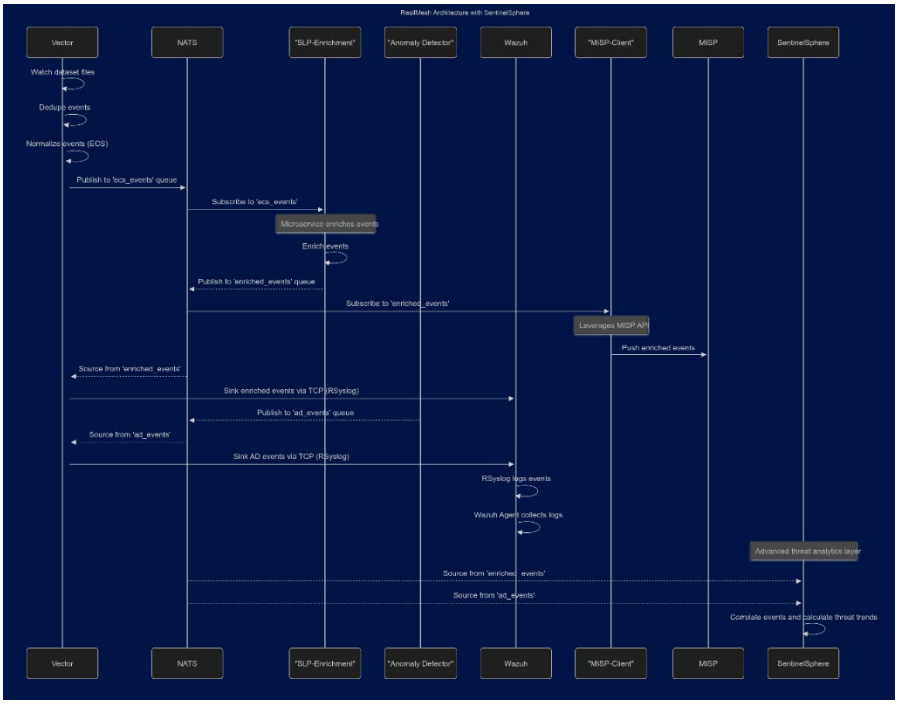

Figure 3: Data Flow Architecture - SentinelSphere and ResilMesh Integration Paths

The diagram in Figure 3 illustrates the comprehensive data flow pipeline where events move from Vector through NATS message broker to the Anomaly Detector, then through ad_events topic to the Threat Calculator, which updates Redis before displaying on the Dashboard. Docker containerisation ensures deployment consistency across environments. The implementation includes environment variable configuration, network isolation, and resource management aligned with ResilMesh standards. This architectural alignment facilitates modular deployment within broader security ecosystems.

4.2 Data Pipeline Implementation

The data pipeline leverages AWS OpenSearch for scalable indexing and analysis. Vector serves as the primary collection layer, performing real-time transformations including routing, filtering, augmentation, and monitoring. Events flow through NATS to the Anomaly Detection module, with processed events stored in OpenSearch for historical analysis. Performance optimisation through grid search yielded key parameters: i) Base threshold: 0.3 for normal conditions; ii) Warning threshold: 0.6 for elevated risk; iii) Critical threshold: 0.8 for high-risk conditions; iv) Burst detection window: 5-minute intervals; v) Minimum event count: 10 events for status changes.

4.3 Dashboard Implementation

The SentinelSphere dashboard provides comprehensive security visualisation and interaction capabilities through a web-based interface built with FastAPI and vanilla JavaScript. The dashboard displays real-time threat level indicators using the Traffic Light System, along with historical trend analysis showing detected threats and vulnerabilities based on real-time data.

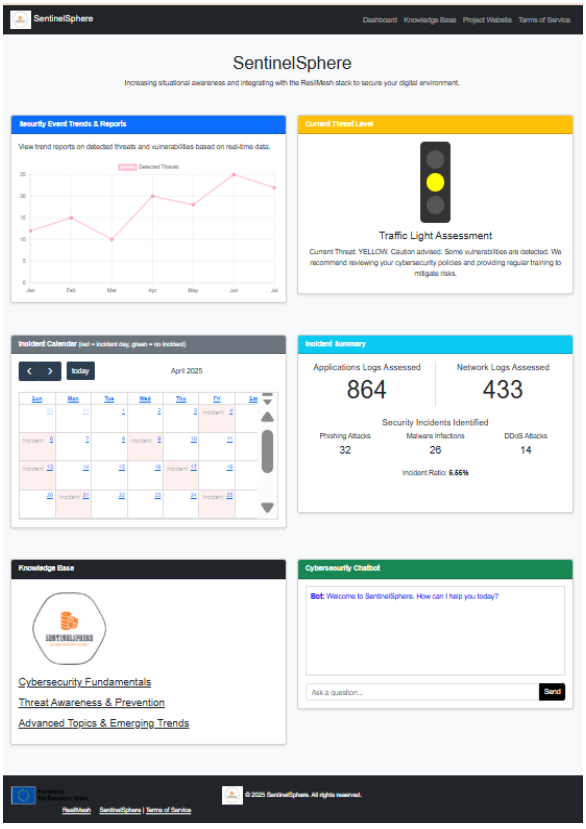

Figure 4: SentinelSphere Dashboard-HomePage and chatbot interaction panel

Figure 4 shows the incident calendar for temporal analysis, the security incident summary statistics, and the integrated Phi-4 powered cybersecurity chatbot providing real-time assistance. The dashboard implements key functionalities including real-time event viewing, TLS indicator panels displaying dynamic threat levels, forecast and trend charts for security pattern visualisation, and an interactive chatbot panel for AI assistance and user query handling.

5. Experimental Evaluation

5.1 Dataset and Methodology

Evaluation utilised the CIC-IDS2017 dataset containing 170,366 network flows with labelled cyberattacks. The benchmark methodology involved training parameter optimisation on 20% of logs, with performance evaluation on the remaining 80% to minimise overfitting. Both baseline and enhanced models used identical 64-64-output DNN architecture, differing only in feature sets.

5.2 Performance Results

Table 2 presents comparative performance metrics between the baseline ResilMesh model and SentinelSphere's enhanced model, including 95% bootstrap confidence intervals:

Table 2: Performance Comparison

Metric	Baseline DNN	Enhanced DNN	Improvement
F1 Score	91.1% ± 0.8%	94.0%± 0.6%	3.3%
Accuracy	96.4%± 0.4%	97.7%± 0.3%	1.3%
Precision	91.0%± 1.1%	96.9%± 0.7%	6.5%
Recall	91.1%± 0.9%	91.3%± 0.8%	0.2%
False Positive Rate	2.3%	0.7%	69.5%
Brute Force Detection	90.3%	91.2%	0.9%
XSS Detection	91.8%	94.4%	2.6%
SQL Injection Detection	85.7%	90.5%	4.8%

Table 3: Per-Class Performance Metrics (Enhanced DNN Model)

Attack Class	Samples (n)	Precision	Recall	F1 Score	95% CI (F1)
Benign	168,186	98.2%	99.1%	98.6%	± 0.2%
Brute Force	1,507	92.4%	91.2%	91.8%	± 1.8%
XSS	652	95.1%	94.4%	94.7%	± 2.3%
SQL Injection	21	88.9%	90.5%	89.7%	± 8.4%

The wider confidence interval for SQL Injection (±8.4%) reflects the limited sample size (n=21) in the dataset. Despite this methodological constraint, the enhanced model achieves 90.5% detection rate, representing a 4.8 percentage point improvement over the baseline (85.7%). This improvement is attributable to the HTTP-specific features that capture SQL injection patterns including parameterised query anomalies and response pattern changes that are invisible to network-layer-only analysis. The enhanced model demonstrates significant improvement in precision and false positive reduction while maintaining high recall rates. Most notably, false positives decreased from 59 to 19 instances, representing a 69.5% reduction critical for SOC operational efficiency.

5.3 Scalability Testing

Comprehensive load testing validated enterprise enterprise scalability using the following hardware and deployment configurations: Hardware Specifications: i) CPU: Intel Core i7-10700 (8 cores, 16 threads, 2.9GHz base clock, 4.8GHz turbo); ii) RAM: 32GB DDR4-3200; iii) Storage: 512GB NVMe SSD (Samsung 970 EVO Plus, sequential read 3,500 MB/s); iv) Operating System: Ubuntu 22.04 LTS; v) Container Runtime: Docker 24.0.5 with Docker Compose orchestration. Deployment Configuration: i) 5 Docker containers (Dashboard, Chatbot, Vector, NATS, OpenSearch); ii) NATS message broker configured with 4 worker threads; iii) OpenSearch single-node cluster with 2GB heap allocation; iv) Redis 7.0 for TLS score caching; v) 8 parallel event processing streams

scalability. The system successfully processed 10,998,214 events in approximately 30 minutes, representing three months of Apache log data. This translates to sustained processing exceeding 5,000 events per second on standard desktop hardware.

Table 4: Docker container resource consumption under load

Component	CPU Usage	Memory	Startup Time
Dashboard	15%	512MB	3 seconds
Phi-4 Chatbot	40%	2.5 GB	8 seconds
Vector	20%	256MB	2 seconds
NATS	10%	128MB	1 second
OpenSearch	25%	2GB	10 seconds

5.4 Chatbot Performance Evaluation

The Phi-4 powered cybersecurity chatbot demonstrated effective domain-specific responses with practical deployment metrics.

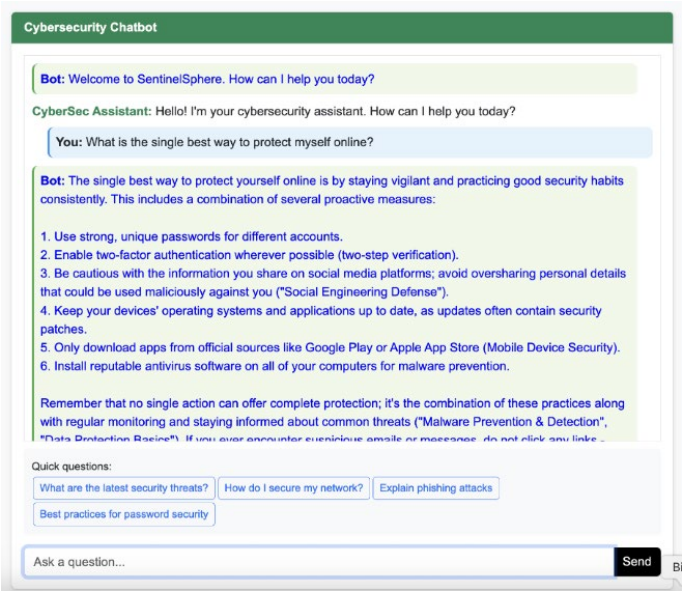

Figure 5: Demonstration of Phi-4 Cybersecurity domain specific LLM Response

The chatbot in Figure 5 provides comprehensive cybersecurity guidance, explaining best practices for online protection including strong passwords, two-factor authentication, cautious information sharing, system updates, app downloads from official sources, and antivirus software installation. The chatbot successfully handles complex cybersecurity queries while maintaining conversational flow and providing actionable recommendations tailored to user expertise levels.

6. Discussion and Implications

6.1 Addressing the Human Factor

SentinelSphere's integration of threat detection with security education represents a paradigm shift in cybersecurity defense. Traditional approaches treat security awareness as separate from operational security, creating disconnects between threat detection and human response. Our unified approach ensures every security event contributes to organisational learning, progressively building cyber resilience.

The LLM-powered chatbot broadens accessibility to security expertise, enabling non-technical users to understand and respond to threats effectively. Similar approaches have demonstrated positive impacts in other domains, where targeted, technology-enhanced education successfully raised awareness and improved response behaviors (Karachalios, 2024; Karachalios & Tantaroudas, 2025). By operating on standard hardware, the solution remains accessible to organisations without specialised infrastructure, addressing the cybersecurity skills gap plaguing many enterprises.

6.2 Practical Deployment Considerations

The 69.5% reduction in false positives has profound implications for SOC operations. Security analysts spend significant time investigating false alerts, leading to alert fatigue and missed genuine threats. SentinelSphere's enhanced precision enables analysts to focus on legitimate threats, improving both efficiency and effectiveness. The Traffic Light System's intuitive visualisation facilitates rapid threat assessment across organisational hierarchies. Executive stakeholders can understand security posture without technical expertise, enabling informed decision-making during critical incidents.

6.3 Integration with ResilMesh Ecosystem

SentinelSphere's seamless integration with the ResilMesh framework demonstrates the value of modular cybersecurity architectures. By leveraging ResilMesh's NATS-based messaging infrastructure and resilience engineering principles, SentinelSphere adds advanced threat analytics and human-centric capabilities without disrupting existing security workflows. This integration approach aligns with recent cybersecurity mesh concepts (Ramos-Cruz et al., 2024), enabling distributed security services while maintaining centralised intelligence and coordination. The success of this integration validates the ResilMesh architecture's extensibility and its potential for accommodating diverse security solutions.

6.4 Limitations and Future Work

While SentinelSphere demonstrates significant advances, several areas warrant future investigation. The current implementation focuses on HTTP-based attacks; extension to other protocols would enhance coverage. Additionally, the chatbot's knowledge base, while comprehensive for common threats, requires continuous updating for emerging attack vectors. The limited SQL injection samples (n=21) in the CIC-IDS2017 dataset represents a methodological constraint. While addressed through stratified splitting and per-class metrics with confidence intervals, future work should incorporate additional datasets with more balanced attack class distributions to strengthen validation of minority class detection performance. Future work will explore federated learning approaches, enabling organisations to benefit from collective threat intelligence while maintaining data privacy. Integration with automated response systems could further reduce response times for known attack patterns.

7. Conclusion

This paper presented SentinelSphere, an innovative cybersecurity platform that successfully bridges the gap between advanced threat detection and human security awareness. Through careful integration with the ResilMesh framework and implementation of Enhanced Deep Neural Networks, intuitive threat visualisation, and AI-powered security education, the system addresses critical challenges in modern cybersecurity defence. Key achievements include a 94% F1 score in threat detection with 69.5% reduction in false positives, successful processing of nearly 11 million events in 30 minutes, and deployment of sophisticated AI capabilities on standard enterprise hardware. These results demonstrate that effective cybersecurity requires not just technological sophistication but systematic enhancement of human security awareness.

SentinelSphere's dual approach—treating every security event as both a threat to mitigate and an opportunity to educate—represents a fundamental shift in cybersecurity philosophy. By democratising security understanding and making advanced threat intelligence accessible across expertise levels, the platform contributes to building more resilient organisations capable of defending against evolving cyber threats. The successful integration with ResilMesh validates the importance of modular, extensible cybersecurity architectures that can accommodate innovative solutions while maintaining operational stability. As cyber threats continue to evolve, platforms like SentinelSphere that combine technological advancement with human empowerment will be essential for maintaining effective cyber defence. This assertion is widely supported in current literature, with recent research emphasizing that integrated, AI-driven platforms and cybersecurity mesh architectures are fundamental to effective cyber defense in contemporary organizational environments (Jada & Mayayise, 2024).

Acknowledgements

This work was supported by the European Union's Horizon Europe research and innovation programme under the ResilMesh project (Grant Agreement No. 101119681).

Ethics Declaration: This research did not require ethical clearance as all experiments were conducted using publicly available datasets with no human participants involved in data collection.

References

Abdin, M., Jacobs, S.A., Amin, A.A., Aneja, J., Awadalla, A., Awadalla, H., Bach, N., Bahree, A., Bakhtiari, A., Beber, J. .(2024)"Phi-4 Technical Report", arXiv preprint arXiv:2412.08905. https://doi.org/10.48550/arXiv.2412.08905

Aldawood, H. and Skinner, G. (2019) "Reviewing Cyber Security Social Engineering Training and Awareness Programs—Pitfalls and Ongoing Issues", *Future Internet*, Vol 11, No. 3, pp 73-89. https://doi.org/10.3390/fi11030073

Araujo, M. S. d., Machado, B. A. S., & Passos, F. U. (2024). Resilience in the Context of Cyber Security: A Review of the Fundamental Concepts and Relevance. Applied Sciences, 14(5), 2116. https://doi.org/10.3390/app14052116

Bada, M., Sasse, A.M. and Nurse, J.R. (2019) "Cyber Security Awareness Campaigns: Why do they fail to change behaviour*?", International Conference on Cyber Security for Sustainable Society*, pp 118-131. https://doi.org/10.48550/arXiv.1901.02672

Bernal, J. Fernandez, P., Montoro, D., Lee, B., Lan, Xi., Stojanovic, B., Jorgeley, Husak., M., ResilMesh D2.2 System Architecture (2024) "ResilMesh: Situation Aware enabled Cyber Resilience for Dispersed, Heterogenous Cyber Systems", EU Horizon Europe Project 101119681, Public Deliverable.

ENISA (2023) "ENISA Threat Landscape 2023", European Union Agency for Cybersecurity, Luxembourg.

Friedberg, I., Skopik, F., Settanni, G., & Fiedler, R. (2015). Combating advanced persistent threats: from network event correlation to incident detection*. Computers & Security*, 48, 35-57. https://doi.org/10.1016/j.cose.2014.09.006

https://aws.amazon.com/opensearch-service/

Jada, I., & Mayayise, T. O. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management*, *8*(2), 100063. https://doi.org/10.1016/j.dim.2023.100063

Karachalios, I. (2024). Utilizing Educational Gaming to Foster Sustainability Awareness in Corporate Settings. *International Journal of Science and Research (IJSR)*, 13(3), 740–744. https://doi.org/10.21275/SR24308032254

Karachalios, I., & Tantaroudas, N. (2025). Future Greek Pre-Service Teachers' Knowledge, Attitudes and Self-Efficacy in Waste Management. *British Journal of Education,* 13(8), 25–42. https://doi.org/10.37745/bje.2013/vol13n82542

Kaur, P., Kumar, M., & Bhandari, A. (2017). A review of detection approaches for distributed denial of service attacks. *Systems Science & Control Engineering*, 5(1), 301-320. https://doi.org/10.1080/21642583.2017.1331768

Musa, M. and Odokuma, E. (2024). A framework for the detection of distributed denial of service attacks on network logs using ml and dl classifiers. *Scientia Africana*, 22(3), 153-164. https://doi.org/10.4314/sa.v22i3.14 Duong, L. (2020). Detecting APT attacks based on Network Flow. *International Journal of Emerging Trends in Engineering Research*, 8(7), 3134-3139. https://doi.org/10.30534/ijeter/2020/42872020

NIST (2024) "NIST Cybersecurity Framework 2.0", *National Institute of Standards and Technology, U.S. Department of Commerce, Washington DC*.

Ramos-Cruz, B., Martínez, L. and Wang, H. (2024) "The Cybersecurity Mesh: A Comprehensive Survey of Involved Artificial Intelligence Methods", *Computer Networks*, Vol 223, pp 109-124. https://doi.org/10.1016/j.neucom.2024.127427

Riskhan, B., Safuan, H., Hussain, K., Elnour, A., Abdelmaboud, A., Khan, F., … & Kundi, M. (2023). An adaptive distributed denial of service attack prevention technique in a distributed environment. *Sensors*, 23(14), 6574. https://doi.org/10.3390/s23146574

Sadlek, L., Husák, M. and Čeleda, P. (2024) "Hierarchical Modeling of Cyber Assets in Kill Chain Attack Graphs", 2024 *20th International Conference on Network and Service Management (CNSM)*, IEEE, pp 1-9. 10.23919/CNSM62983.2024.10814501

Sharafaldin, I., Lashkari, A.H. and Ghorbani, A.A. (2018) "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP)*, pp. 108-116. https://doi.org/10.5220/0006639801080116*

Sharafaldin, I., Lashkari, A.H., Hakak, S. and Ghorbani, A.A. (2019) "Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy", IEEE 53rd International Carnahan Conference on Security Technology*, pp. 1-8. https://doi.org/10.1109/CCST.2019.8888419*

Sinha, S. and Degadwala, S. (2024). Voting model strategies for reliable categorical iot-ddos attack prediction. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 10(2), 300-307. https://doi.org/10.32628/cseit2410223

Szabó, Z., & Bilicki, V. (2023). A New Approach to Web Application Security: Utilizing GPT Language Models for Source Code Inspection. Future Internet, 15(10), 326. https://doi.org/10.3390/fi15100326

Taneja, N. (2023). A deep dive into methods for combating ddos attacks and securing data. International Journal of Computing Programming and Database Management, 4(2), 01-07. https://doi.org/10.33545/27076636.2023.v4.i2a.84

Verizon (2023) "2023 Data Breach Investigations Report", Verizon Business, New York.

Vinayakumar, R., Alazab, M., Soman, K.P., Poornachandran, P., Al-Nemrat, A. and Venkatraman, S. (2019) "Deep Learning Approach for Intelligent Intrusion Detection System", *IEEE Access*, Vol 7, pp 41525-41550. 10.1109/ACCESS.2019.2895334

Yin, C., Zhu, Y., Fei, J. and He, X. (2017) "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks", *IEEE Access*, Vol 5, pp 21954-21961. 10.1109/ACCESS.2017.2762418