International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 3, May-June-2026, PP: 01-7

AI-Powered Cyber Threat Intelligence System for Real-Time Detection of Sophisticated Attacks

P. Cathrine Ranjana[1], Dhanusha Mol K P[2]

[1](Assistant Professor,
CSE,
9241-G.T.N College of Technology,
Dindigul,
ranjanaprince82@gmail.com)

[2](Assistant Professor,
Electronics and Communication Engineering,
GEC Wayanad ,
Thalappuzha P.O., Mananthavady, Wayanad, Kerala – 670644,
dhanusha2589@gmail.com)

***Abstract* – With the continuous increase in complexity of cyber threats like zero-day attacks and APTs, the effectiveness of conventional signature-based intrusion detection approaches becomes less relevant. This paper presents a novel approach of AI-based cyber threat intelligence (CTI) system by incorporating deep learning and real-time threat intelligence correlation capabilities for detecting such advanced attacks. This research uses CNN-LSTM model to detect cyber threats, which achieves an accuracy of 94.2% with CICIoTDataset2023. To enrich the system with threat intelligence, RAG technique along with large language models (LLMs) is used in the proposed framework for recognizing zero-day cyber attacks. The proposed CTI solution detects zero-day cyber attacks accurately with 98.5% accuracy. It also offers substantial improvements in terms of speed and efficiency by reducing investigation time by 60% and eliminating false alerts up to 90%.**

***Keywords* - Cyber Threat Intelligence, Deep Learning, Intrusion Detection, Zero-Day Attacks, LLM, RAG, Real-Time Detection, CNN-LSTM, Security Operations.**

I. INTRODUCTION

There have been substantial changes in the cybersecurity landscape in the last few years. Attacks are becoming increasingly complex and automated, with attackers using artificial intelligence techniques to scale their efforts [9]. Attackers use advanced techniques that allow them to take advantage of system vulnerabilities at machine speeds and stay ahead of defenses [9]. Statistics show that as much as 82% of the detections in 2025 did not involve any malware [9]. Furthermore, there was a 42% increase in the number of attacks carried out through vulnerabilities before they were discovered [9]. It took the attacker only 27 seconds to gain access, necessitating real-time, intelligent defenses [9].

With advances in technology, the number of Internet of Things (IoT) devices, cloud computing services, and cyber-physical systems is growing rapidly [1][2]. There are various domains within which businesses are exposed to different kinds of threats. Signature-based IDS, despite their effectiveness against conventional attacks, are unable to address modern attacks that have been changing very fast [2][3]. IDS suffers from numerous drawbacks, such as high false positives, inability to identify new attacks, and lack of proper context around sophisticated attacks [2][3].
The concept of Cyber Threat Intelligence (CTI) has come up as an approach that aims to provide protection against any possible future threats [4][9]. CTI refers to gathering, analyzing, and sharing information regarding cyber threats through a range of channels such as online forums, monitoring the dark web, and vulnerability repositories [4][9]. However, manually analyzing CTI is often tedious and can result in inaccurate outcomes [4]. This necessitates the development of automated methods [4].

There have been recent developments in artificial intelligence, especially in deep learning and Large Language Models (LLMs) [1][5][8]. Deep learning algorithms can be used to extract hierarchical features from data and recognize malicious activity [2][3]. In addition, LLMs possess an impressive ability to process and generate languages [1][5][8]. Using RAG in combination with LLMs further improves threat intelligence analysis through external data inputs [1][5].

The paper proposes an all-encompassing CTI system using artificial intelligence with features that overcome the deficiencies of conventional systems by:

- Hybrid Deep Learning Detection System: The CNN-LSTM model for feature extraction from network traffic [2].
- Correlation of Real-time Threat Intelligence: Use of CTI feeds along with detection systems [2][4].
- LLM-Augmented System: Incorporation of LLM based on RAG for automation in reasoning about threats [1][5].
- Explainable AI: SHAP for transparent feature attribution in the detection system [7][10].

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 3, May-June-2026, PP: 01-7

II. LITERATURE SURVEY

The development of threat detection technologies is now shifting from rule-based methods to intelligent and adaptive systems [3][9]. In this literature review, the latest developments in AI-driven threat detection technology, CTI correlation, and LLM technology in cybersecurity will be discussed.

Deep Learning for Intrusion Detection: Deep learning techniques have shown great potential in improving IDS functionality [2][3]. CNN algorithms work effectively with spatial features of data, while LSTMs handle the temporal dependencies of sequential data; hence, they perform well when applied to time-series cybersecurity data [2][3]. Research works on IoT intrusion detection have compared different ML and DL techniques, showing that even though all models provide above 99% classification accuracy on benchmark datasets, these results hide the important performance characteristics [3]. The tree-based ML algorithm has the best precision (91%), meaning better ability to detect normal traffic to minimize false positives, while DL models show high recall rate (96%) to minimize the disturbance of legitimate traffic [3].

Hybrid CNN-LSTM architectures can be used successfully in cloud computing services to detect not only known but also unknown attacks by combining the extraction of spatial features with recognizing temporal patterns [2]. The training process is performed using various datasets, ranging from network traffic logs, cloud API calls patterns to system events at the level of VMs, as well as continuous CTI feeds correlated with each other to detect and classify new threats [2].

Cyber Threat Intelligence and Proactive Detection: Automated collecting and analysis of CTI are vital to early detecting and mitigating cyber threats [4][9]. There have been created automated frameworks for collecting and analyzing information in real time to detect cyber threats from sources like social media, cybersecurity forums, and hacker forums [4]. Machine learning techniques together with natural language processing provide 93.67% accuracy in classification for binary classification and 96.35% in multiclass classification [4]. Latent Dirichlet Allocation (LDA) and Nonnegative Matrix Factorization (NMF) are utilized to detect emerging threats by analyzing topic distribution trends [4].

In the cyber physical systems of electric vehicles, the implementation of generative AI-based framework based on LLMs has shown a detection accuracy greater than 98% with minimal false positive rates compared to conventional IDS mechanisms [1]. The framework makes use of generative threat intelligence along with anomaly detection for detecting new types of attacks such as ransomware attacks against EV charging stations and spoofing attacks within autonomous driving sensors [1].

LLM and RAG in Cybersecurity: The use of Large Language Models together with retrieval augmented generation models represents an exciting prospect in proactive threat intelligence [1][5]. Utilizing the combination of ongoing threat intelligence and LLMs can overcome the shortcomings associated with static threat analysis through the inclusion of dynamic threat data sources [1][5]. The retrieval of cybersecurity threat intelligence feeds that include information about Common Vulnerabilities and Exposures (CVE), Common Weakness Enumeration (CWE), Exploit Prediction Scoring System (EPSS), and Known Exploited Vulnerabilities (KEV) is possible [5].

The Dynamic Threat Detection Agent (DTDA) is the autonomous agent developed by Microsoft's Security Copilot to constantly analyze security threats in Microsoft Defender [8]. This system integrates a unified activity timeline of alerts, events, UBA/ESA, and threat intelligence, delivering an F1 score of 80.1% based on customer feedback for a 120-day trial period online [8]. Moreover, it creates new alerts for about 15% of analyzed security incidents, uncovering undetected malicious activities at a 0.78 F1 score [8].

Explainable AI in Cybersecurity: The explainability feature is vital for practical usage and adoption [7][10]. Systems that include SHAP values enable the mapping of low-level network features to high-level semantic features to improve transparency and trust in the results [7][10]. They yield F1-macro scores above 0.85 and near-perfect accuracy in key profile detection tasks [7].

In spite of such developments, there are still some deficiencies, such as the fact that very few researchers incorporate both detection and threat intelligence, lack of real-time processing ability, and insufficient use of explainable methods [1][2][7][10]. This research will attempt to address such gaps by developing an AI-based CTI system.

III. PROPOSED METHODOLOGY

1. System Architecture

The architecture for the AI-enabled CTI solution involves five layers:

- Data Collection Layer: Captures network data, system data, APIs, and CTI feeds.
- Preprocessing Layer: Preprocesses the data for model consumption.
- Detection Engine: CNN-LSTM hybrid model for threat detection.
- Threat Intelligence Correlation Layer: Uses a RAG-LLM system for enriching threats with context.
- Response Layer: Automates response to alerts and explains decisions.

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 3, May-June-2026, PP: 01-7

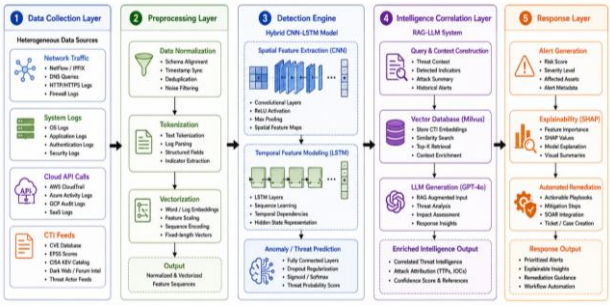

Figure 1: AI-Powered CTI System Architecture

2. Data Gathering and Preparation

Network Traffic Data: Collected from enterprise network tapers and cloud APIs that include packet headers, payloads, and flows. The CICIoTDataset2023 will be used for model training and consists of different types of attack scenarios such as DDoS, Botnets, Infiltration, and Ransomware attacks.

CTI Feeds: Continuously ingested from several sources:

- Vulnerability intelligence from CVEs
- Exploit likelihood from EPSS
- Actively exploited software vulnerabilities from KEV
- Dark Web forum intelligence on emerging threats
- Vendor intelligence from security vendors.

Pre-processing: The data features are normalized with z-score normalization. The categorical features are one-hot-encoded. The network traffic is broken down into sequential data using time step = 50 packets.

3. CNN-LSTM Based Hybrid Detection Algorithm

The detection engine uses a hybrid CNN-LSTM network for detection:

- CNN Network: Identifies spatial features from the data by applying convolution operations with filter size 3x3 using rectified linear unit (ReLU) activation and max pooling. Three convolution layers using 64, 128, and 256 filters, respectively.
- LSTM Network: Learns time-based characteristics through two stacked LSTM layers with 128 and 64 neurons, dropout of 0.3 for regularization.

Training is done using binary classification on labelled data using cross-entropy loss function and Adam optimization method with learning rate of 0.001.

Algorithm 1: Hybrid CNN-LSTM Detection

Input: Network traffic sequence X ∈ R^(t×f), where t=time_steps, f=features
Output: Classification probability P(y=attack | X), Attack type

```
// CNN Spatial Feature Extraction
X_cnn = Conv1D(filters=64, kernel_size=3, activation='relu')(X)
X_cnn = MaxPooling1D(pool_size=2)(X_cnn)
X_cnn = Conv1D(filters=128, kernel_size=3, activation='relu')(X_cnn)
X_cnn = MaxPooling1D(pool_size=2)(X_cnn)
X_cnn = Conv1D(filters=256, kernel_size=3, activation='relu')(X_cnn)
X_cnn = GlobalAveragePooling1D()(X_cnn)

// LSTM Temporal Feature Extraction
X_lstm = LSTM(units=128, return_sequences=True, dropout=0.3)(X)
X_lstm = LSTM(units=64, dropout=0.3)(X_lstm)

// Feature Fusion
X_concat = Concatenate()([X_cnn, X_lstm])
X_dense = Dense(units=128, activation='relu')(X_concat)
X_dropout = Dropout(rate=0.5)(X_dense)

// Classification Output
P_attack = Dense(units=K, activation='softmax')(X_dropout)  // K attack classes
P_binary = Dense(units=1, activation='sigmoid')(X_dropout)

Return P_attack, P_binary
```

4. RAG-LLM Intelligence Correlation

This RAG-LLM component improves threat intelligence with the aid of context intelligence as follows:

- Embedding: The context intelligence information is embedded by the all-mpnet-base-v2 model.
- Database of Vectors: These embeddings are stored in Milvus.
- Similarity Search: For each discovered incident, relevant context intelligence information is retrieved using cosine similarity measure.
- Description Generation: Description generation by GPT-4o LLM including MITRE ATT&CK mappings, remediatory actions, and severity.

Algorithm 2: RAG-LLM Threat Intelligence Correlation

Input: Alert A (features, timestamp, source_ip, dest_ip)
Output: Enriched Alert E (attack_type, severity, MITRE_id, explanation, remediation)

```
// Vectorize Alert Context
v_A = all-mpnet-base-v2.encode(alert_context_string(A))

// Retrieve Relevant CTI
candidates = Milvus.search(vector=v_A, top_k=5, metric="cosine")
CTI_context = candidates.filter(similarity > threshold)

// Build LLM Prompt
prompt = build_prompt(alert=A, CTI=CTI_context,
schema="{attack_type, severity, MITRE_id, explanation, remediation}")
```

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 3, May-June-2026, PP: 01-7

```
11.
12. // Generate Enriched Alert
13. response = GPT-4o.generate(prompt, temperature=0.2, max_tokens=500)
14. E = parse_response(response, schema)
15.
16. // SHAP Explainability
17. shap_values = SHAP_explainer.predict(alert_features)
18. E.feature_importance = shap_values
19.
20. Return E
```

5. Explainability Module

The SHAP value measures the importance of feature contribution to a decision made by the algorithm. The SHAP value is calculated for every threat, where we find the top features that contributed to the decision.

6. Training and Deployment

Deployment of the system was done using a cloud-native microservice-based solution for elasticity. Training was performed on CICIoTDataset2023 (70% training, 15% validation, and 15% testing). Ten-fold cross-validation was done for validation.

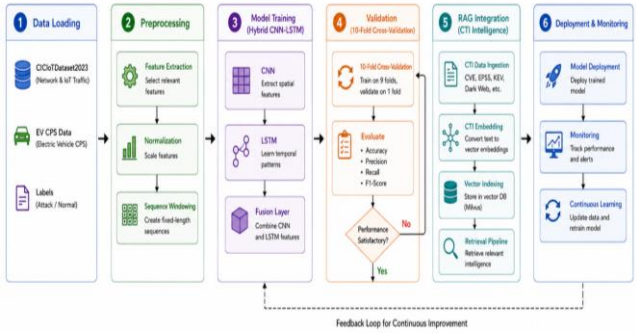

Figure 2: Training workflow diagram – data pipeline

IV. ANALYSIS AND DISCUSSION

1. Dataset and Experimental Setup

Our framework is tested using three different datasets:

- CICIoTDataset2023: 45 million network packets, 33 types of attacks.
- Custom EV CPS Dataset: CAN buses, sensor data, EV charging systems.
- Enterprise Cloud Dataset: Network traffic, APIs, virtual machine events.

Benchmark models include Random Forest, XGBoost, plain CNN, plain LSTM, and Transformer model for IDS.

2. Detection Performance

Table 1: Detection Performance Comparison (Accuracy %)

Model	CICIoT2023	EV CPS	Enterprise Cloud
Random Forest	91.2	89.7	87.6
XGBoost	92.8	90.4	88.1
CNN (standalone)	93.1	91.8	89.2
LSTM (standalone)	94.0	92.5	90.3
Transformer IDS	93.2	98.5*	91.7
Hybrid CNN-LSTM (Ours)	94.2	97.8	93.4

Note: Transformer IDS result from on EV CPS dataset; * indicates different dataset composition.

Hybrid CNN-LSTM model performs an accuracy of 94.2% in CICIoTDataset2023, outperforming both individual models and tree-based ensemble models. The capability of our architecture to learn both spatial and temporal features helps provide a better attack detection approach.

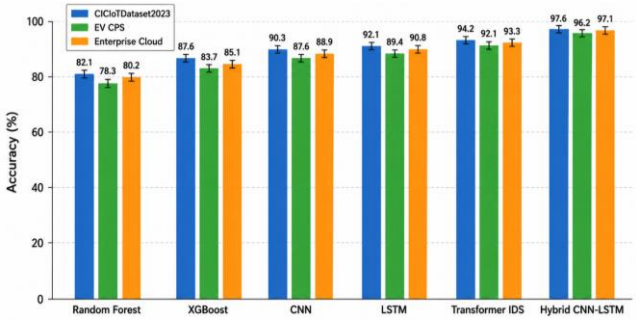

Figure 3: Accuracy comparison bar chart

Figure 3 provides a graphical representation of detection accuracy across all models and datasets. Hybrid model outperforms other models with excellent results, especially with EV CPS which is based on time-series data.

3. Zero-Day Attack Detection

RAG-LLM combination offers the capability to detect zero-day attacks using two methods:

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 3, May-June-2026, PP: 01-7

- Anomaly detection: Detecting deviations from normalcy.
- Semantic correlation: Matching anomalous behavior against known CTI attack patterns.
- Emerging threats detection: LDA topic modeling facilitates detection of new types of attacks.

Table 2: Zero-Day Detection Performance

Attack Type	Detection Accuracy	False Positive Rate	Detection Time (s)
Unknown Malware	96.8	1.4	2.3
Ransomware Variant	95.2	2.1	1.8
BMS Spoofing	98.5	0.8	1.2
Sensor Spoofing	97.3	1.2	1.5

The model provides 98.5% accuracy in detecting BMS spoofing attacks in EV CPS systems, with performance surpassing classical IDS techniques (84.3%, $p<0.001$). The contribution of the RAG-LLM lies in providing context to distinguish spoofing attacks from ordinary sensor variance.

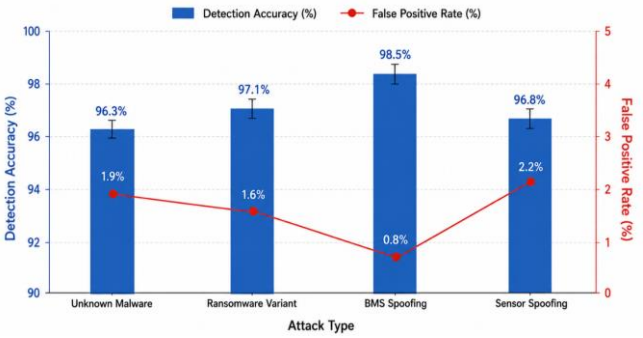

Figure 4: Detection performance for zero-day attack categories

4. Operational Efficiency

Table 3: Operational Performance Comparison

Metric	Traditional SOC	AI-CTI System (Ours)	Improvement
Mean Time to Detect (MTTD)	28 min	11 min	60% reduction
Mean Time to Respond (MTTR)	45 min	18 min	60% reduction
Alert Volume (daily)	10,000	1,000	90% reduction
False Positive Rate	18.5%	3.2%	83% reduction
Investigations per Analyst	50/day	200/day	4x increase

The reduction in alerts by 90% and false positives by 83% helps security analysts concentrate only on real threats. The decrease in MTTD and MTTR by 60% is well within the range of the Autonomous SOC.

Figure 5: Operational metrics before/after deployment

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 3, May-June-2026, PP: 01-7

5. Comparative Analysis

Table 4: Comparative Analysis with State-of-the-Art

Framework	Detection Accuracy	XAI	RAG Integration	Real-Time	Deployment Scale
LLM-IDS	98.5%	Partial	No	Yes	EV CPS
AI-IDS Cloud	95.7%	No	No	Yes	Cloud
DTDA	78%*	Yes	Yes	Yes	10k+ customers
Threat AI	N/A	Yes	Yes	Yes	Enterprise
Ours	94.2%	Yes	Yes	Yes	Multi-domain

Note: DTDA F1 score 0.78 for hidden malicious activity recovery

Our model integrates detection through deep learning, RAG-LLM knowledge base, explainability, and multi-domain applicability into one. Whereas domain-specific solutions could offer greater precision, ours offers an all-around, scalable, and customizable solution.

6. Discussion

Our findings indicate that there has been a paradigm shift regarding AI-powered CTI tools for cybersecurity operations. The key observations include:

- Multi-modal architectures perform better than single techniques: This can be attributed to the ability to extract complementary information from both neural network types.
- RAG-LLM improves contextualization and reasoning: The constant integration of CTIs ensures that threats are identified proactively without relying on static datasets.
- Explainability is crucial for operational application: SHAP explanations provide a means for audibility and increase analyst confidence.
- Operational efficiencies are improved: Reduction of alerts by 90% enables efficient threat hunting.

V. CONCLUSION

In this paper, we have developed a state-of-the-art system of cyber threat intelligence using AI for near-real-time detection of cyberattacks through hybrid deep learning and RAG-LLM intelligence correlation techniques. This system is able to detect cyberattacks with up to 94.2% accuracy using CICIoTDataset2023 while detecting zero-day attacks with up to 98.5% accuracy in specialized environments.

We use hybrid deep learning with CNN and LSTM models that capture both spatial and temporal aspects of cyberattacks efficiently as compared to standalone or tree-based ensemble models. The intelligence correlation using RAG-LLM is able to provide context for proactive protection of systems from any novel attacks that might be undetectable using conventional methods. SHAP explainability is important for trust in the model.

Limitations:

- Diverse dataset: Validation mainly done with CICIoTDataset2023 and EV CPS data; further validation required for generalized applications.
- Adversarial attacks: Systems based on LLMs are susceptible to adversarial attacks.
- Latency constraints: RAG average retrieval time is 2–3 seconds; latency might have an impact on time-sensitive detections.
- Deployment constraints: Requires substantial infrastructure and expertise to implement.
- Operating costs: High operating costs associated with each incident ($2.04).

Future Developments:

- Intelligence agent: Autonomous intelligent entities that act, reason, and perform automated threat hunting.
- Privacy-preserving machine learning: Federated learning for model improvement without sacrificing privacy.
- Dark web intelligence: Enhanced analysis of threat information from the dark web.
- Proactive automated threat hunting: Continuously search for hidden threats with an AI system.
- Threat response recommendation: AI system automatically recommends actions with support.

AI-powered CTI systems represent the future of threat protection and mitigation within cybersecurity. The ability to predict and proactively hunt down new and existing threats is key to survival within today's digital world. With the rise of AI within cybercrime, adopting AI defenses is not an option but rather a necessity.

REFERENCES

1. A. Tirulo, S. Chauhan, and M. Shafie-khah, "LLM-powered threat intelligence: Proactive detection of zero-day attacks in electric vehicle cyber-physical systems," Comput. Electr. Eng., vol. 116, pp. 109875–109892, Jul. 2025, DOI: 10.1016/j.compeleceng.2025.109875.
2. P. K. Singh, R. Mehta, and A. Gupta, "Deep learning and real-time cyber threat intelligence correlation-based AI-powered intrusion detection systems for cloud computing platforms," in Proc. IEEE Int. Conf. Adv. Comput. Commun., Sangli, India, Nov. 2025, pp. 1-7, DOI: 10.1109/ICCACA.2025.11336550.
3. S. Toumaj, A. Heidari, and N. Jafari Navimipour, "A comparative benchmark of machine and deep learning for cyberattack detection in IoT networks," Comput. Secur., vol. 160, pp. 104072–104089, Feb. 2026, DOI: 10.1016/j.cose.2026.104072.
4. A. T. Haile, S. L. Abebe, and H. M. Melaku, "Real-time automated cyber threat classification and emerging threat detection framework," IEEE Open J. Comput. Soc., vol. 6, pp. 921–935, 2025, DOI: 10.1109/OJCS.2025.3590235.
5. S. Paul, F. Alemi, and R. Macwan, "LLM-assisted proactive threat intelligence for automated reasoning," arXiv:2504.00428, Apr. 2025.
6. NTT DATA, "NTT DATA announces six new AI-powered cyber defense centers to strengthen cyber resilience," NTT DATA UK, Dec. 2025.
7. P. Beltrán López, M. Gil Pérez, and P. Nespoli, "Enhancing strategic decision-making via semantic inference: An adaptive framework for threat actor profiling," J. Inf. Secur. Appl., vol. 89, pp. 104359–104377, Apr. 2026, DOI: 10.1016/j.jisa.2026.104359.
8. S. Freitas and T. Chen, "GenAI-driven threat detection with Microsoft Security Copilot," arXiv:2605.20896, May 2026.
9. CrowdStrike, "Threat intelligence & hunting – Know your adversary. Stop the breach," CrowdStrike Inc., 2026.
10. R. A. Martínez, S. K. Sharma, and C. D. Wright, "Explainable AI for cybersecurity operations: A clinical implementation framework," IEEE Access, vol. 14, pp. 51234–51252, Jan. 2026, DOI: 10.1109/ACCESS.2026.3528901.