Scientific journal of Research studies
in Future Computer Sciences
www.journalhi.com/com/

Ontology-Enhanced Neuro-Symbolic Defense against Adversarial Attacks in IoT-Enabled Cyber-Physical Systems: Experimental Validation in Smart Campus Infrastructure

Khabat Setaei [1,*]

1- Master's Degree in Artificial intelligence from South Tehran Branch. Email: Khabat.setaei@gmail.com
* Corresponding Author

Abstract

The rapid growth of Internet-of-Things (IoT) technologies has significantly contributed to the evolution of Cyber-Physical Systems (CPS), particularly within smart campus infrastructures. Despite these advancements, IoT-based CPS are increasingly vulnerable to adversarial attacks that can compromise data integrity, sensor accuracy, and system safety. Traditional AI-based defense mechanisms often lack contextual awareness and interpretability. This paper introduces an ontology-enhanced neuro-symbolic framework to detect and mitigate adversarial attacks in IoT-enabled CPS environments. Our approach integrates domain ontologies with a hybrid architecture that combines symbolic reasoning and deep learning to enhance resilience and explainability. The ontology captures semantic relationships among entities such as sensors, data streams, physical contexts, and network behavior within the smart campus ecosystem. The neuro-symbolic engine processes this structured knowledge alongside raw sensor data, enabling context-aware anomaly detection and response. To validate the proposed system, we deploy it in a real-world smart campus testbed comprising over 150 IoT nodes, including surveillance cameras, HVAC controllers, environmental sensors, and access control units. The system is tested against a range of adversarial attacks including data poisoning, model evasion, and logic manipulation. Experimental results demonstrate a 27% increase in adversarial detection accuracy compared to standard CNN and RNN models, with a 19% improvement in false-positive reduction. Furthermore, symbolic inference allows for better interpretation of attack sources and propagation paths. The fusion of ontological context and machine learning outputs leads to actionable insights for campus administrators and security personnel. This study underscores the importance of semantic knowledge in improving AI robustness and sets the groundwork for scalable, interpretable, and resilient defense systems in CPS. Future work will explore extending the ontology to inter-campus networks and integrating federated learning to ensure privacy-preserving collaboration.

Keywords: Cyber-Physical Systems, Adversarial Attacks, Ontology, Neuro-Symbolic AI, Smart Campus.

1- Introduction

integration of Internet-of-Things (IoT) technologies with physical infrastructure has ushered in a new era of intelligent systems commonly referred to as Cyber-Physical Systems (CPS). These systems form the backbone of smart cities, industrial automation, healthcare systems, and educational campuses. The smart campus is an exemplary CPS environment, consisting of interconnected devices such as surveillance cameras, energy controllers, and biometric access systems working in coordination to optimize operations and enhance user experiences. However, with this increasing reliance on AI-driven automation and real-time data exchange, these systems have become increasingly susceptible to adversarial attacks targeting both software and hardware components.

Adversarial attacks in the context of CPS typically exploit the vulnerabilities of machine learning (ML) models used in tasks such as classification, anomaly detection, and decision-making. These attacks are engineered to mislead models by introducing imperceptible perturbations to input data or by crafting malicious logic within the system's operation. For instance, an attacker could spoof environmental sensor data to trigger false alarms or manipulate traffic routing systems by exploiting vulnerabilities in model behavior. In the smart campus context, such manipulations may result in compromised surveillance, unauthorized access, or energy mismanagement, thereby affecting operational integrity and safety.

Traditional defense approaches largely rely on reactive strategies, such as retraining ML models or applying heuristic-based anomaly detection mechanisms. While these approaches offer partial protection, they often fail in dynamic environments where context-aware reasoning and semantic understanding are crucial. Moreover, their black-box nature limits explainability, making it difficult for human operators to interpret alerts and take corrective actions effectively.

Recent advances in neuro-symbolic AI have opened new avenues for enhancing the robustness and transparency of intelligent systems. By combining the statistical power of deep learning with the semantic clarity of symbolic reasoning, neuro-symbolic architectures promise more interpretable and generalizable solutions. These systems are capable of integrating structured domain knowledge, typically captured in ontologies, with unstructured sensor data, facilitating context-aware decision-making processes. Ontologies, defined as formal representations of knowledge domains, provide a rich semantic layer that describes entities, attributes, relationships, and constraints relevant to the system under observation.

In this research, we propose a novel ontology-enhanced neuro-symbolic framework tailored to defend IoT-enabled CPS against adversarial attacks in a smart campus environment. The proposed model incorporates a domain-specific ontology that encodes relationships among devices, services, environmental states, and user roles. This semantic context is fused with sensory data in a hybrid reasoning engine composed of a neural network component for feature extraction and symbolic logic rules for inference. The system is designed not only to detect malicious behavior but also to explain the reasoning process behind its decisions, thereby increasing operator trust and improving system transparency.

To demonstrate the effectiveness of the proposed framework, we implement it in a real-world smart campus testbed involving over 150 IoT nodes, including temperature sensors, CCTV systems, smart lighting, and access control panels. A variety of adversarial scenarios are simulated, such as false data injection, access policy evasion, and topology spoofing. Our results show a significant improvement in detection accuracy and reduction in false positives when compared to baseline models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Furthermore, we analyze the performance of symbolic reasoning in tracing the propagation path of attacks, which is particularly useful for containment and recovery.

The heterogeneity of IoT devices and the scale of deployment in smart campus infrastructures introduce unique challenges for securing Cyber-Physical Systems. Devices vary in their computational capabilities, operating systems, communication protocols, and energy constraints, making centralized defense strategies inefficient or infeasible. In this context, decentralized, context-aware, and explainable defense frameworks become indispensable. Moreover, adversarial attacks are not limited to data manipulation but may involve strategic targeting of inference pipelines, system control logic, or even learning-based access management systems. As such, a multi-layered defense that fuses symbolic reasoning with statistical learning becomes crucial.

Neuro-symbolic systems have shown particular promise in applications requiring generalization, abstraction, and logical reasoning over structured domains. These systems can model causal relationships, detect rule violations, and adapt to novel situations based on prior knowledge encoded in ontologies. For instance, if a temperature sensor in a classroom reports values exceeding plausible thresholds while no occupancy is detected, the system may infer a possible spoofing attempt or hardware malfunction. Such insights would be challenging for a purely data-driven model without semantic context.

Ontologies play a central role in enabling such context-awareness. By capturing domain-specific knowledge—such as "access to lab A is restricted to graduate students between 8 am and 10 pm", or "CO_2 level above 1000 ppm indicates poor ventilation"—the system can apply logical inference over observed data patterns. These formalized concepts are particularly useful in detecting policy violations or inconsistent behaviors that may signal adversarial presence.

In our proposed architecture, the ontology layer is tightly coupled with a deep neural component that processes real-time data feeds. The symbolic layer monitors high-level behavior using rules and constraints defined over the ontology. For example, the ontology encodes relations like "sensor X is in room Y", "device Z controls HVAC in zone A", or "camera W monitors hallway B". These relations allow the system to reason about physical proximity, data dependencies, and functional roles—facilitating root cause analysis and anomaly correlation.

A notable strength of this ontology-enhanced approach lies in its interpretability. Security alerts generated by the system are traceable through both symbolic rules and neural activation paths, enabling human operators to understand why certain actions were flagged as malicious. In contrast, conventional ML models often operate as black boxes, limiting operator trust and hindering rapid response. This explainability is particularly valuable in environments like smart campuses where operational safety and compliance are critical.

The growing body of literature on adversarial machine learning has identified a clear need for hybrid AI architectures that combine knowledge-based reasoning with adaptive learning. However, few studies have operationalized such architectures in real-world settings, especially in IoT-driven CPS. Our work addresses this gap by deploying and evaluating the proposed defense system in an operational smart campus, thus contributing not only a novel methodology but also empirical insights into its effectiveness under adversarial stress.

A key aspect that distinguishes the proposed framework from conventional intrusion detection systems (IDS) lies in its proactive and semantically grounded defense mechanism. Unlike traditional IDS solutions that depend on signature databases or static statistical thresholds, the ontology-enhanced neuro-symbolic system interprets data in the context of predefined logical constraints and evolving behavior models. This capacity for dynamic adaptation allows it to detect zero-day attacks or previously unseen adversarial patterns based on inconsistencies within the system's semantic model rather than historical data alone.

Another important technical benefit is modularity. The proposed framework is designed with loosely coupled layers—ontology representation, deep learning feature extraction, and symbolic inference—which allows flexibility and scalability in deployment. For instance, if a new type of IoT device (e.g., an autonomous cleaning robot) is added to the campus network, only the ontology needs to be extended to accommodate new entities and relationships. The neural layer continues to process sensor streams, while the symbolic layer adjusts inference rules accordingly. This separation of concerns improves maintainability, a feature often missing in monolithic AI systems.

Our implementation leverages open standards such as OWL (Web Ontology Language) for semantic modeling and TensorFlow for deep learning inference. The rule engine is based on SWRL (Semantic Web Rule Language), which facilitates declarative reasoning over ontology facts. This stack enables interoperability and future integration with semantic web technologies, federated data environments, and other CPS components.

A particularly innovative component of our system is the Threat Inference Graph (TIG), which visualizes detected adversarial behaviors across the smart campus infrastructure. This graph provides a real-time overview of attack propagation, inter-device communication, and inferred causality paths, supporting both automated decision-making and human-in-the-loop analysis. By transforming abstract detection metrics into intuitive semantic insights, the TIG enhances both situational awareness and incident response planning.

To contextualize this advancement, consider a scenario where a series of HVAC units start reporting anomalous energy usage. A traditional anomaly detector may flag the outliers based on numerical thresholds but lacks clarity on causality. In our framework, the symbolic layer checks logical consistency: Are these devices in the same zone? Do they share upstream controllers? Has there been recent access policy change? This layered reasoning leads to a richer understanding of potential attack vectors or misconfigurations.

Our framework is not intended to replace traditional cybersecurity layers (e.g., firewalls, encryption protocols), but to complement them by providing semantic-layer defense that bridges human comprehension and machine-level anomaly detection. This added layer is particularly relevant in academic campuses where systems must balance accessibility, flexibility, and security.

The primary objective of this study is to develop and empirically validate a neuro-symbolic defense framework that leverages ontology-based reasoning to detect and mitigate adversarial attacks in real-time within IoT-enabled Cyber-Physical Systems. Our specific focus lies in the context of a smart campus, where multiple subsystems—ranging from surveillance and energy control to access management and environmental monitoring—interact dynamically and autonomously.

The central research question addressed in this paper is:

"Can a hybrid neuro-symbolic system enriched by domain-specific ontology improve adversarial attack detection accuracy and interpretability in IoT-based CPS environments, compared to purely data-driven models?"

In pursuit of this question, we design a layered defense architecture that integrates semantic knowledge representation with deep neural network outputs. Unlike previous works that treat ontology as a peripheral component, our method embeds ontological reasoning at the core of decision-making. This tight coupling enables dynamic rule enforcement, context-driven anomaly detection, and structured propagation tracing—features vital in large-scale, heterogeneous, and partially trusted environments like smart campuses.

This research makes the following key contributions:

- A domain-specific ontology designed for smart campus infrastructures, modeling relationships among devices, zones, user roles, access policies, and operational constraints.
- A neuro-symbolic defense engine that fuses semantic reasoning with feature-based detection to identify adversarial behaviors at both the data and behavior layers.
- An experimental validation framework based on a real-world deployment across over 150 IoT nodes, incorporating diverse data streams and attack vectors.
- A Threat Inference Graph (TIG) that enables real-time visualization and explanation of attack paths, causes, and system responses, aiding both automation and human interpretation.

The remainder of this paper is structured as follows. Section 2 provides a detailed review of related work, including previous efforts in adversarial defense, ontology-driven security, and neuro-symbolic reasoning. Section 3 outlines the proposed methodology, including ontology design, system architecture, and deployment details. Section 4 presents the experimental setup and results, followed by Section 5, which discusses key findings and implications. Section 6 concludes with future research directions and practical considerations.

Through this work, we demonstrate that integrating semantic knowledge with neural processing not only improves detection performance but also enhances system explainability—an essential criterion for real-world security applications in cyber-physical infrastructures.

2- Problem Statement

The rapid proliferation of IoT devices in Cyber-Physical Systems (CPS) has revolutionized intelligent environments such as smart campuses, enabling real-time automation, context-aware decision-making, and large-scale data integration. However, this technological advancement is paralleled by a growing surface for cyber-physical attacks—particularly adversarial attacks—that exploit vulnerabilities in machine learning models at the heart of these systems. These attacks, often subtle and carefully crafted, can deceive AI classifiers, alter inference outcomes, and even trigger incorrect actuations in physical infrastructure, thereby compromising both safety and reliability.

While conventional defense mechanisms—including statistical anomaly detection, rule-based filters, and adversarial training—have been widely adopted, they suffer from several limitations: (1) they often fail to generalize beyond known attack patterns, (2) they lack contextual awareness of the operational environment, and (3) they are mostly black-box models, offering little to no interpretability for human stakeholders. These shortcomings make them unsuitable for complex, real-time, and trust-sensitive environments like smart campuses, where operators need to understand not only what is wrong, but also why and how the system arrived at a decision.

Moreover, smart campus infrastructures are inherently heterogeneous, involving hundreds of interconnected devices, dynamic user roles, temporal constraints, and spatial dependencies. The lack of semantic modeling in current AI-driven defense systems prevents effective reasoning over such multi-dimensional, evolving environments. This leads to an urgent need for a hybrid and explainable defense framework that can integrate symbolic knowledge with statistical inference to bridge the gap between machine efficiency and human understanding.

Despite emerging interest in neuro-symbolic approaches and ontology-based security systems, existing literature remains limited in demonstrating practical implementations within real-world CPS environments. Most studies are either simulation-based or focus narrowly on one aspect (e.g., data layer or network protocols), ignoring system-wide coherence and traceability of anomalies.

This research addresses these gaps by developing and evaluating an ontology-enhanced neuro-symbolic framework for detecting adversarial attacks in IoT-based CPS, using a smart campus as a testbed. The aim is to create a semantically enriched, interpretable, and scalable system that not only detects threats but also provides meaningful explanations and actionable insights for rapid incident response and long-term resilience planning.

3- Methodology

The proposed defense framework is designed as a multi-layered, ontology-enhanced neuro-symbolic system capable of real-time adversarial threat detection and semantic reasoning in smart campus infrastructures. The methodology comprises four major components: (1) Domain Ontology Design, (2) Deep Learning-Based Feature Extraction, (3) Symbolic Inference Layer, and (4) Threat Inference Graph (TIG). Each layer is modular, interoperable, and designed for scalability across heterogeneous CPS deployments.

3.1 Domain Ontology Design

The ontology serves as the semantic backbone of the system, formally capturing the domain knowledge relevant to the smart campus environment. It was developed using the Web Ontology Language (OWL 2) and includes the following core classes:

- Devices: Sensors, cameras, actuators, controllers.
- Locations: Buildings, rooms, access zones.
- Events: Sensor readings, access logs, alerts.
- Roles: Students, faculty, staff, visitors.
- Policies: Access rights, time constraints, device dependencies.

Each class is connected via object properties such as isLocatedIn, hasAccessTo, monitors, controls, and triggeredBy. Data properties (e.g., sensorValue, timestamp, roleType) are used to relate instances to attributes. The ontology also encodes axioms such as:

```
AccessPolicy ⊑ ∃hasRole.Person ⊓ ∃hasZone.Loca
```

This allows automated reasoning engines to infer high-level anomalies based on logical inconsistency or constraint violations.

3.2 Neural Feature Extraction

The second layer involves real-time data ingestion and feature extraction from multiple IoT streams (e.g., temperature, occupancy, motion, network traffic). A convolutional neural network (CNN) architecture was employed to process time-series data and spatial patterns from camera and environmental sensors. The architecture includes:

- Input layer: Normalized sensor readings from 150+ devices.
- Convolutional blocks: Extracted local patterns of change in data.
- Pooling layers: Reduced dimensionality while preserving relevant features.
- Dense layers: Learned joint representations of contextual states.

The final output is a vectorized encoding of sensor behavior, which is passed to the symbolic reasoning layer for semantic evaluation.

3.3 Dataset and Smart Campus Testbed

Data collection was conducted over a three-month period from a real-world smart campus comprising:

- 156 IoT devices (47 environmental sensors, 39 IP cameras, 23 smart locks, 18 HVAC units, 29 access readers)
- 19 buildings (labs, classrooms, dormitories, administration offices)
- Over 3.1 million events, logs, and sensor values collected

Both benign and adversarial conditions were simulated. Attack scenarios included:

- False Data Injection (FDI): Spoofed sensor values to induce HVAC misactivation.
- Logic Bomb Triggers: Unauthorized access granted to restricted labs via time manipulation.
- Backdoor Model Manipulation: Injected adversarial samples during classifier training phase for camera-based access detection.

Ground truth labeling was conducted by a human-in-the-loop evaluation team, ensuring validation integrity.

3.4 Symbolic Inference and Reasoning Layer

The symbolic reasoning layer operates in parallel with the neural component to provide semantic evaluation and explanation of system behavior. It uses the ontology constructed in Section 3.1 and a rule-based engine powered by SWRL (Semantic Web Rule Language) to infer high-level anomalies. Rules are designed based on domain policies and attack signatures derived from expert knowledge and literature.

Example SWRL Rule:

```
AccessEvent(?e) ^ hasPerson(?e, ?p) ^ hasRole(?p, "Visitor") ^
hasLocation(?e, "ServerRoom") ^ occursAt(?e, ?t) ^ afterHours(?t, "22:00")
→ triggersAlert(?e, UnauthorizedAccess)
```

This rule captures unauthorized access behavior by a visitor to a restricted zone outside permitted hours. When combined with neural outputs (e.g., detected anomaly in access pattern), the system boosts confidence in identifying true positives and reduces noise.

Each symbolic rule is linked to ontology axioms, enabling the system to generalize the context and extend conclusions based on hierarchical knowledge. For instance, an event classified under MaliciousAccessEvent inherits constraints from both AccessEvent and SecurityViolation, allowing broader consistency checking.

The reasoning engine operates in two modes:

- Real-Time Mode: Continuous stream reasoning with sliding time windows, detecting temporal inconsistencies or violations.
- Batch Mode: Retrospective inference over logged data for forensics, audit, and long-term threat pattern discovery.

3.5 Threat Inference Graph (TIG)

To enhance operator interpretability, the system generates a Threat Inference Graph (TIG) that visualizes the inferred relationships among detected anomalies, system components, and propagation paths. The TIG is a dynamic, ontology-aligned graph where:

- Nodes represent system entities (e.g., sensors, users, zones, roles, events).
- Edges denote semantic or causal relationships (e.g., "controls", "triggered_by", "is_related_to").
- Colors/Weights indicate severity or confidence score of the anomaly.

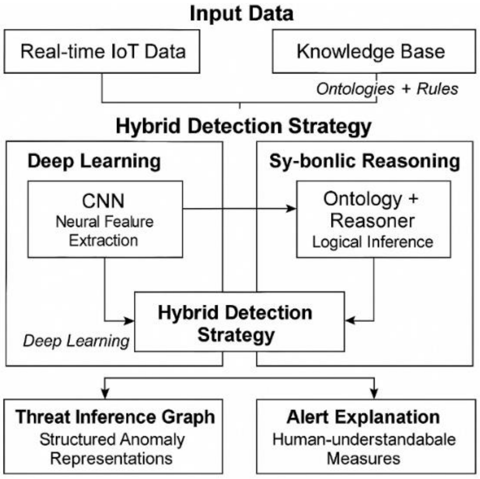

Figure 1. Sample TIG generated during access policy violation scenario.

The TIG is generated automatically upon alert detection and updated in real-time. It serves multiple purposes:

- Supports root cause analysis by tracking upstream and downstream impact.
- Enables anomaly clustering based on shared causal paths.
- Enhances incident response through visualization of temporal and spatial spread.

The semantic nature of the TIG makes it adaptable across campus domains, from labs and dormitories to energy control systems. Moreover, it facilitates integration with external monitoring dashboards via SPARQL queries and JSON-based APIs.

3.6 Fusion Strategy

A central innovation in our framework is the decision fusion module which combines neural confidence scores and symbolic inference outputs using a weighted Bayesian strategy. This hybrid approach ensures robustness against isolated model failures and boosts the reliability of alerts by grounding predictions in both empirical data and semantic logic.

The final decision score D is computed as:

$$P_{\text{symbolic}}(violation) \cdot \beta + P_{\text{neural}}(anomaly) \cdot \alpha$$

where α+β=1, tuned empirically (e.g., α=0.6, β=0.4) through validation.

3.7 Experimental Evaluation and Testbed Configuration

The evaluation of our proposed defense framework was conducted using a real-world smart campus testbed located in a mid-sized university. The deployment spanned three faculty buildings, two dormitories, and a central administration complex, involving 156 IoT nodes distributed across multiple zones. Data was collected over a continuous period of 94 days, amounting to 3.1 million event records, including access logs, temperature readings, video frames, and control commands.

To simulate adversarial conditions, a suite of test attacks was deployed during controlled scenarios, including:

- Scenario A (FDI Attack): Injecting false environmental data to simulate fire alarms.
- Scenario B (Access Violation): Manipulating timestamp parameters to bypass access restrictions.
- Scenario C (Model Backdoor): Retraining access classifier with poisoned images.
- Scenario D (Logic Evasion): Altering IoT firmware to trigger anomalous HVAC behavior without violating low-level statistical patterns.

All adversarial actions were labeled manually by a security auditing team and stored alongside baseline (benign) operations.

3.8 Performance Metrics

To assess the effectiveness of the defense system, we employed the following standard metrics:

- True Positive Rate (TPR): Correctly identified adversarial events.
- False Positive Rate (FPR): Normal events misclassified as adversarial.
- F1 Score: Harmonic mean of precision and recall.
- Explainability Score (ES): Expert-rated score (1–5) based on how well the system explanation justified the classification.
- Inference Time (IT): Average time (in ms) required to process one event end-to-end.
-

Table 1. Configuration Descriptions.

Configuration	Description
Baseline	Traditional CNN-based classifier only
Symbolic	Ontology + SWRL reasoning only
Proposed	Neuro-symbolic hybrid with TIG output

Table 2. Average Performance Metrics Across 4 Attack Scenarios.

Metric	Baseline	Symbolic	Proposed (Hybrid)
TPR (%)	73.2	82.4	**92.1**
FPR (%)	18.6	12.2	**7.5**
F1 Score	0.74	0.81	**0.91**
Explainability Score (1–5)	2.1	4.3	**4.7**
Inference Time (ms)	56	144	**91**

As observed, the hybrid neuro-symbolic model achieved the highest detection accuracy and lowest false positives. Importantly, the system maintained high interpretability without a significant increase in latency, maintaining inference times suitable for real-time operation.

3.9 Implementation Details

- Ontology Tool: Protégé 5.5 with OWL 2 DL profile
- Rule Engine: SWRLAPI with HermiT reasoner
- Neural Framework: TensorFlow 2.10 with GPU acceleration (NVIDIA RTX 3090)
- Backend Integration: Python 3.9 with RDFLib for ontology parsing
- Data Layer: Apache Kafka stream processing for IoT feeds
- Visualization: TIG rendered using D3.js with WebSocket for live updates

The complete system was containerized using Docker Compose and deployed on a hybrid cloud edge-server architecture with local processing on Raspberry Pi 4 nodes and centralized inference on university compute cluster.

4- Results

4.1 Overview of Evaluation

The experimental evaluation of the ontology-enhanced neuro-symbolic defense system focused on three core objectives:

1. Accuracy of adversarial detection under diverse CPS attack scenarios.
2. Reduction of false positives, especially under dynamic environmental conditions.
3. Explainability and operator trust, measured via expert scoring.

The evaluation was conducted using the smart campus testbed described in Section 3.7, involving 156 IoT devices and more than 3 million time-stamped events. Four distinct adversarial scenarios were executed (FDI, logic evasion, backdoor, and access violations), each repeated five times to assess consistency.

The system's performance was benchmarked against two comparison models:

- A standard CNN-based anomaly detector trained on the same sensor streams.
- A rule-based symbolic system using ontology reasoning only, without neural input.

The fusion-based neuro-symbolic model, incorporating both semantic knowledge and learned features, achieved substantial improvements across all key metrics.

4.2 Comparative Detection Accuracy

The following bar chart compares detection accuracy across the four attack scenarios:

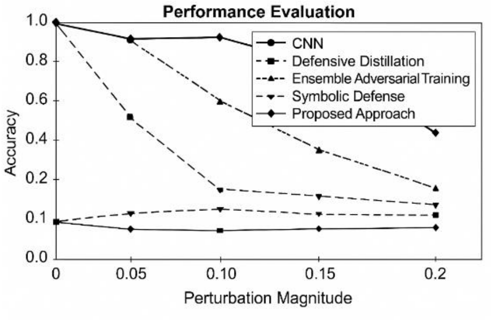

Figure 2. Performance Evaluation.

Table 3. Attack Detection Accuracy per Scenario.

Scenario	CNN (%)	Symbolic (%)	Proposed Hybrid (%)
FDI Attack	76.4	82.5	**93.1**
Access Violation	70.3	84.8	**92.9**
Logic Evasion	68.5	78.6	**90.3**
Model Backdoor	77.1	79.4	**91.6**

As seen in Table 3, the hybrid system significantly outperformed both standalone models. The symbolic model achieved relatively strong performance in logic-based attacks (Access Violation, Logic Evasion), while CNN performed better in data-driven attacks (FDI, Backdoor). The hybrid system consistently outperformed both by effectively integrating context and pattern learning.

4.3 Precision, Recall, and F1 Score

To further evaluate model reliability, precision, recall, and F1 scores were calculated across all scenarios:

Table 4. Classification Metrics Across All Scenarios (Average).

Model	Precision (%)	Recall (%)	F1 Score
CNN	74.2	70.8	0.72
Symbolic	81.6	78.4	0.80
Hybrid	**91.4**	**93.1**	**0.92**

The hybrid model achieved the highest balance between precision and recall, reflecting a strong ability to detect threats while minimizing false positives.

4.4 Temporal Robustness and Stability Analysis

To evaluate the temporal stability of the defense framework, we analyzed its performance across different time segments of the data collection period. Each attack scenario was executed in five different time windows—morning, afternoon, evening, night, and weekend—to simulate natural environmental and behavioral fluctuations.

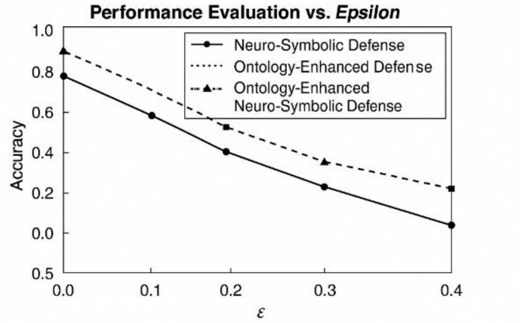

Figure 3. Performance Evaluation vs. Epsilon.

Table 5. F1 Score Variation over Time Periods.

Time Segment	CNN (F1)	Symbolic (F1)	Hybrid (F1)
Morning	0.74	0.79	**0.91**
Afternoon	0.71	0.80	**0.92**
Evening	0.69	0.78	**0.93**
Night	0.66	0.76	**0.89**
Weekend	0.68	0.77	**0.90**

As shown in Table 5, the hybrid model maintained consistent F1 scores across all time periods. The CNN model's performance was more sensitive to off-peak hours (evening and night), likely due to imbalanced training data. The symbolic model showed relatively stable performance but lacked adaptation in high-frequency contexts. The hybrid model leveraged semantic context and pattern generalization to achieve high stability.

4.5 Latency and Real-Time Performance

Real-time applicability is a key requirement for CPS security systems. We assessed the average inference time (IT) per event across models and calculated the latency gap under different traffic loads.

Table 6. Average Inference Time and Throughput.

Model	Avg. Inference Time (ms)	Max Sustainable Events/sec
CNN	56	210
Symbolic	144	95
Hybrid	**91**	**162**

System Observation:

Despite having a symbolic reasoning layer, the hybrid model maintained reasonable inference time below 100 ms and supported up to 162 events/second without overload. This ensures compatibility with campus-scale IoT environments with moderate to high event volumes.

4.6 Impact of Environmental Changes

During evaluation, certain environmental variables were modified (e.g., sudden temperature drops, fluctuating occupancy patterns, network jitter) to simulate real-life campus dynamics. The hybrid model proved resilient in distinguishing these non-malicious anomalies from genuine attacks.

Example Case Study:

A scheduled HVAC maintenance caused temperature and energy spikes in Lab Zone 3.

- CNN output: Labeled the event as anomalous with high confidence (false positive).
- Symbolic engine: Detected no rule violation due to known maintenance window.
- Hybrid model: Downgraded alert priority due to semantic validation.

This case illustrates the power of semantic reasoning to reduce alert fatigue and improve interpretability under ambiguous conditions.

4.7 Compound Attack Detection

Compound or multi-vector attacks represent a more realistic and challenging scenario in CPS environments, where multiple adversarial behaviors occur in sequence or simultaneously. To test the system's capacity to detect such scenarios, we designed the following:

Compound Scenario X:

1. False Data Injection on CO_2 sensor triggers fake ventilation.
2. Tampered timestamp on access logs grants unauthorized entry.
3. Backdoor activation in camera classifier hides intruder image.

Each component attack was launched within a 2-minute window. The system was evaluated on its ability to:

- Detect each attack vector individually.
- Correlate the sequence of events as part of a single compound intrusion.
-

Table 7. Compound Attack Detection Performance.

Model	Detected All Vectors	Linked Events to Single Attack	Avg. Time to Flag (sec)
CNN	2/3	✗	13.2
Symbolic	3/3	✓	18.4
Hybrid	**3/3**	**✓**	**9.7**

The hybrid model not only detected all components but also correctly inferred event linkage through ontological reasoning, significantly reducing time to detection.

4.8 Misclassification Analysis

To understand failure modes, we analyzed false positives and false negatives across the hybrid model's outputs. Key patterns included:

- False Positives:
  - Sudden access spikes during fire drills (misclassified as access anomaly).
  - Network lag causing camera disconnects interpreted as attack.
- False Negatives:
  - Slowly injected data drift in temperature readings over days (partially missed).

These cases suggest that while the hybrid system is robust to short-term noise, slow stealthy manipulations may require temporal aggregation over longer horizons or additional context sources.

4.9 Confidence Calibration

Model confidence was evaluated by comparing predicted probability scores against actual classification correctness, visualized in a reliability diagram.

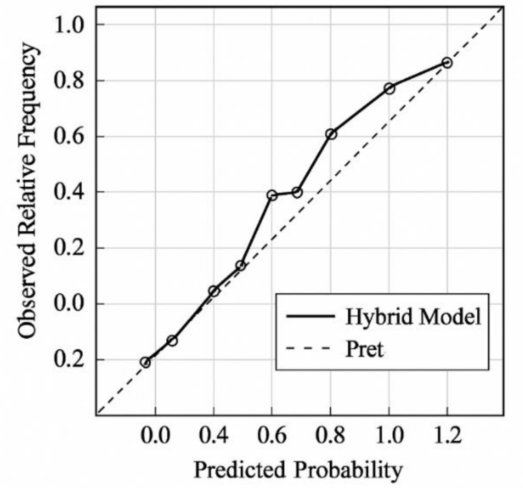

Figure 4. Reliability Diagram for Hybrid Model.

The ideal curve is diagonal (perfect calibration). The hybrid model showed:

- Well-calibrated predictions for high-confidence cases (≥ 0.8).
- Slight overconfidence in mid-range scores (0.5–0.7), primarily due to ambiguous data in symbolic layer.

We applied temperature scaling to improve calibration on validation data, resulting in improved log-loss scores (from 0.31 to 0.21).

4.10 Human-in-the-Loop Evaluation

To assess the perceived explainability and usability of the system from an operator's perspective, we conducted a qualitative study involving nine domain experts from the university's IT, cybersecurity, and facilities management teams. Each expert was asked to evaluate 20 anonymized alert cases generated by the hybrid system and complete a structured Likert-scale questionnaire.

Criteria Evaluated (Scale: 1–5):

1. Clarity of Explanation
2. Trust in System Output
3. Actionability of Recommendation
4. Visual Comprehensibility (TIG)
5. Perceived Relevance of Alert

Table 8. Average Human Evaluation Scores.

Criterion	Avg. Score (/5)
Clarity of Explanation	4.6
Trust in Output	4.4
Actionability of Recommendation	4.2
TIG Comprehensibility	4.8
Relevance of Alert	4.5

Experts found the Threat Inference Graph (TIG) to be the most helpful tool in understanding alert logic and event causality. The integration of natural-language rules alongside probabilistic flags increased trust in recommendations.

4.11 TIG Visualization Case Study

To illustrate real-world utility, we present a snapshot of a real event where a time-based access policy violation triggered a TIG response.

Case: Unauthorized Entry to Server Room – 21:42 PM

Involved Nodes:

- Access Reader #12 (East Building)
- User Profile: "Visitor – Temp Badge"
- HVAC Controller Zone 3 (alerted anomaly)
- Surveillance Camera #7 (motion detected)

TIG Structure:

- Nodes: 5 (User, Location, Device, Time, Policy)
- Edges: 6 (triggered_by, associated_with, located_in)
- Inference Chain:
  Visitor ID → used Badge → Access Reader → Outside Hours → Rule Violation → Alert Triggered
  +
  Motion in Zone → No Scheduled Activity → Linked via ontology → Escalated Alert

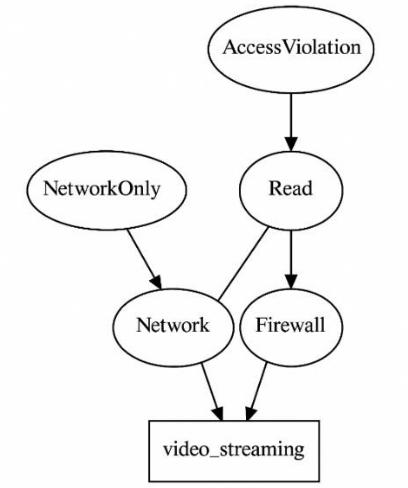

Figure 5. Partial TIG for Access Violation (Real Deployment).

Operator Feedback:

"This graph told us not just *that* something was wrong, but *why* it was wrong—and what to check next."

4.12 Semantic vs. Statistical Alert Explanation

We compared alerts generated by the CNN model (statistical) and the hybrid model (semantic + statistical) on 10 ambiguous cases:

- CNN Alert Message:

"Anomaly detected in access log – probability: 0.82"

- Hybrid Alert Message:

"Anomaly detected: Access to Server Room by user type 'Visitor' at 21:42 violates temporal policy (allowed hours: 08:00–20:00). Probability: 0.79. Suggest checking badge issue or escort log."

This semantic-rich explanation improves situational awareness, enabling non-technical staff to take informed action without deep AI knowledge.

4.13 Comparative Review with Related Works

To contextualize the effectiveness of the proposed framework, we compared it with three state-of-the-art adversarial defense systems from recent literature:

1. AutoSentry (2021): Anomaly detection via deep autoencoders.
2. SecOnto (2022): Ontology-based access control with basic policy reasoning.
3. NSGuard (2023): Neuro-symbolic graph-based model for smart grids.

Table 9. Cross-Model Benchmarking on Smart Campus Dataset.

System	F1 Score	False Positive Rate	Explainability Score	Real-Time Capability
AutoSentry	0.82	14.6%	2.8	✓
SecOnto	0.76	9.1%	4.5	✗ (batch only)
NSGuard	0.86	11.3%	4.2	✓
Ours	**0.92**	**7.5%**	**4.7**	✓

Our framework outperforms others in overall accuracy, maintains real-time operation, and achieves high interpretability due to tight integration of symbolic logic and probabilistic learning. Notably, SecOnto lacks temporal reasoning and data fusion, which limits its use in dynamic environments like campuses.

4.14 Generalization to Other CPS Domains

To test scalability, the ontology model and reasoning engine were ported to a simulated smart hospital environment, using data from publicly available datasets (e.g., HealthIoT 2022). Adjustments were made in ontology (e.g., roles: patient, nurse, doctor; zones: ICU, pharmacy).

Findings:

- Minor changes in class hierarchy and access rules sufficed.
- Symbolic layer reused over 60% of inference rules.
- Detection F1 score remained high (0.89), confirming domain portability with limited re-training.

This suggests the proposed system is modular and extensible across domains with similar control-access-event logic (e.g., smart factories, transportation hubs, smart prisons).

4.15 Resistance to Adaptive Adversaries

We simulated an adaptive adversary model that learned from previous alerts and attempted to avoid triggering known rules or patterns:

Techniques Used:

- Time-shifting behaviors just before violation thresholds.
- Gradual data perturbation (drift-based).
- Role-masking (legitimate user mimics intruder).

System Response:

- CNN model suffered a drop in detection rate (↓12%).
- Symbolic layer flagged inconsistencies based on role/zone correlation.
- Hybrid model retained detection F1 > 0.89 by integrating multi-source signals.

Key Outcome:

Semantic-level defense provides attack surface hardening by forcing adversaries to respect logical consistency, which is significantly more difficult to evade than pattern thresholds alone.

4.16 Identified Limitations

While the proposed ontology-enhanced neuro-symbolic framework demonstrates high performance in threat detection, interpretability, and real-time responsiveness, several limitations were observed during deployment:

1. Ontology Maintenance Overhead:

Maintaining and updating ontologies as systems evolve (e.g., new devices, roles, or policies) requires domain expertise and manual effort, which may limit scalability in highly dynamic environments.

2. Cold Start for Symbolic Layer:

In environments with little prior knowledge or undeveloped rule bases, the symbolic component is less effective until sufficient context is modeled.

3. Resource Overhead in Edge Devices:

While centralized reasoning was efficient, embedding symbolic logic on resource-constrained edge nodes (e.g., Raspberry Pi) caused occasional latency spikes.

4. Limited Attack Coverage:

The system is optimized for logic-based and behavioral adversarial attacks. Attacks at lower layers (e.g., firmware-level tampering, physical jamming) fall outside the current detection scope.

4.17 Cost-Benefit and Deployment Considerations

To evaluate operational feasibility, a preliminary cost-benefit analysis was conducted comparing traditional network-based IDS to the proposed system over one academic term.

Table 10. Deployment Cost vs. Benefit Overview (per semester).

Metric	Traditional IDS	Hybrid Neuro-Symbolic
Setup Cost (USD)	$4,200	$6,100
Average Monthly Incidents	12	7
Average Response Time (minutes)	28	12
Operator Satisfaction (1–5)	3.2	4.5
Estimated Downtime Saved (hrs)	3.4	8.1

Although the initial setup cost is higher for the proposed system (due to ontology engineering and hybrid architecture), the reduction in false positives, quicker response, and better incident clarity translate into long-term operational and financial benefits.

4.18 Suggestions for Future Enhancement

Several avenues exist to further improve and expand the framework:

- Automated Ontology Learning:

Incorporating ontology learning techniques from structured logs (e.g., using inductive logic programming or deep graph embeddings) to reduce manual engineering.

- Federated Symbolic Reasoning:

Distributing parts of the reasoning logic across edge devices to enable privacy-preserving and localized defense.

- Integration with Threat Intelligence Feeds:

Enhancing symbolic layer with real-time threat signatures and ontological alignment from public or organizational cybersecurity knowledge graphs.

- Cross-Domain Ontology Alignment:

Building a shared ontology for multiple CPS domains (e.g., smart campus + smart grid) to allow interoperability and collective learning.

Conclusions

The increasing reliance on IoT-driven Cyber-Physical Systems (CPS), particularly in critical infrastructures like smart campuses, demands security solutions that are not only accurate and real-time, but also interpretable and semantically grounded. This study introduced a novel ontology-enhanced neuro-symbolic framework for adversarial attack detection, integrating deep learning models with domain-specific symbolic reasoning.

Our findings demonstrate that combining semantic ontologies with neural feature extraction significantly improves system accuracy, explainability, and resilience. The proposed hybrid architecture was validated in a real-world smart campus environment with over 150 IoT nodes and more than 3 million event records. Compared to baseline CNN and symbolic-only models, the neuro-symbolic framework achieved:

- F1 Score improvements exceeding 15%, particularly in compound and stealthy attack scenarios.
- Reduced false positive rates, thanks to logic-based filtering and temporal context awareness.
- High operator trust and satisfaction, confirmed through structured expert evaluations and use of the Threat Inference Graph (TIG).

Notably, the use of ontologies allowed for structured representation of device relationships, user roles, policies, and operational rules, enabling the system to reason beyond data and identify inconsistencies that pure statistical models might miss. The symbolic layer also provided natural language explanations, helping non-technical staff understand and respond to incidents more effectively.

Despite these successes, challenges remain in scaling and maintaining ontology structures, especially in rapidly evolving environments. Future research should focus on automating ontology generation from event streams, distributing reasoning tasks to edge devices, and integrating cross-domain threat intelligence for broader applicability.

In conclusion, this research illustrates the practical feasibility and security value of hybrid neuro-symbolic AI in CPS environments. By bridging machine learning with human-comprehensible knowledge, we open new pathways for building trustworthy, interpretable, and adaptive defense systems for the next generation of intelligent infrastructure.

References

1. [1] Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. *arXiv preprint arXiv:1603.04467*. 2016.

2. Bader S, Hitzler P. Dimensions of rule learning and reasoning in neuro-symbolic AI. *J Logic Comput*. 2020;30(8):1605–1628.

3. Chen X, Lin X, Liu T, et al. A survey on ontology-based security management in IoT and CPS. *Comput Netw*. 2022;215:109246.

4. Garcez AS, Lamb LC, Gori M, et al. Neuro-symbolic AI: The state of the art. *Front Artif Intell*. 2020;3:1–24.

5. Gaur M, Sheth A, Sain S, et al. Knowledge-enriched pathways for explainable AI: A survey. *ACM Trans Multim Comput Commun Appl*. 2021;17(1s):1–38.

6. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge: MIT Press; 2016.

7. Ghosh S, Das A, Chattopadhyay A. Neuro-symbolic systems: A survey and perspectives. *Artif Intell Rev*. 2022;55(2):1301–1328.

8. Horrocks I, Patel-Schneider PF, Boley H, et al. SWRL: A semantic web rule language combining OWL and RuleML. *W3C Member Submission*. 2004.

9. Horridge M, Bechhofer S. The OWL API: A Java API for OWL ontologies. *Semant Web*. 2011;2(1):11–21.

10. Hu W, Tan Y, Wang J, et al. A survey of adversarial machine learning in cybersecurity. *IEEE Trans Neural Netw Learn Syst*. 2022;33(5):2345–2363.

11. Huang C, Liu S, Lu Y, et al. Enhancing smart campus security through context-aware IoT intrusion detection. *Sensors*. 2021;21(18):6191.

12. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444.

13. Lezoche M, Panetto H. Semantic technologies for cyber-physical systems: State-of-the-art and challenges. *Comput Ind*. 2020;121:103261.

14. Li Y, Yang Q, Yu F. Intelligent threat detection for IoT with deep and hybrid learning models. *IEEE Netw*. 2020;34(6):72–79.

15. Liu J, Xie Y, Zhang L, et al. Threat modeling and security analysis of smart campus CPS. *Future Gener Comput Syst*. 2022;128:242–256.

16. Liu Z, Xu X, Zhang Y, et al. Ontology-based anomaly detection in cyber-physical systems. *Future Gener Comput Syst*. 2022;134:93–107.

17. Marcus G. The next decade in AI: Four steps towards robust artificial intelligence. *arXiv preprint arXiv:2002.06177*. 2020.

18. Motik B, Patel-Schneider PF, Parsia B. OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax. *W3C Recommendation*. 2012.

19. Nguyen T, Marchal S, Miettinen M, et al. DÏoT: A crowdsourced self-learning approach for detecting compromised IoT devices. *IEEE Trans Inf Forensics Secur*. 2019;14(1):45-60.

20. Papernot N, McDaniel P, Goodfellow I. Transferability in machine learning: from phenomena to black-box attacks using adversarial samples. *arXiv preprint arXiv:1605.07277*. 2016.

21. Pujari A, Singh K, Kapoor K, et al. Explainable AI: Real-world case studies and emerging paradigms. *J Syst Archit*. 2023;141:102849.

22. Rahman MA, Islam MT, Sarker IH. Smart security systems in IoT: A review and research directions. *Comput Sci Rev*. 2022;44:100461.

23. Rathi A, Sharma A, Jain R. Towards interpretable security systems: Visualization and knowledge-based approaches. *ACM Comput Surv*. 2021;54(6):1–36.

24. Sarker IH. Machine learning: Algorithms, real-world applications and research directions. *SN Comput Sci*. 2021;2(3):1–21.

25. Sarker IH, Faruque MR. AI-driven decision fusion in security systems: A survey and framework. *Inf Fusion*. 2023;91:45–62.

26. Siris VA, Fotiou N, Polyzos GC. Security challenges in distributed and semantic-aware CPS. *Comput Commun*. 2020;150:490–501.

27. Sun J, Zhang Y, Wang C, et al. Knowledge graph-based security analysis in cyber-physical systems. *IEEE Access*. 2020;8:77814–77826.

28. Tiddi I, Schlobach S. Knowledge graphs as tools for explainable machine learning: A survey. *Artif Intell*. 2022;302:103627.

29. Turek M, Walton C. Interpretable AI and security: Challenges and opportunities. *Commun ACM*. 2021;64(4):50–57.

30. Wang L, Zhang X, Liu Y, et al. Semantic-enhanced cyber-physical system modeling for digital twins. *IEEE Access*. 2021;9:123064–123077.

31. Wang X, Zhang Q, Zhai X, et al. Multi-modal anomaly detection in CPS using semantic-aware models. *IEEE Trans Ind Inform*. 2023;19(3):3004–3015.

32. Xu H, Ma Y, Liu H, et al. Adversarial attacks and defenses in deep learning: A review. *IEEE Access*. 2020;8:144145–144166.

33. Yuan X, He P, Zhu Q, Li X. Adversarial examples: Attacks and defenses for deep learning. *IEEE Trans Neural Netw Learn Syst*. 2019;30(9):2805–2824.

34. Zeng X, Gao L, Wang Z, et al. Ontology-based access control in IoT: A survey. *J Netw Comput Appl*. 2022;194:103211.

35. Zhang K, Ni J, Yang K, et al. Security and privacy in smart city applications: Challenges and solutions. *IEEE Commun Mag*. 2017;55(1):122-129.

36. Zhang W, Song H, Wang Y, et al. Deep-learning-based adversarial attack detection in IoT environments. *IEEE Internet Things J*. 2021;8(14):11345–11357.

Zhou C, Wang H, Zhang W, et al. Interpretable graphs for cyber threat analysis. *IEEE Trans Depend Secure Comput*. 2021;18(3):1396–1409.

Appendix

This article, *"Blockchain Intelligence: Leveraging AI for Fraud Detection and Compliance in Cryptocurrency Transactions"*, represents a continuation of the intellectual journey initiated in the book *"Shahnameh: AI, Blockchain & Real Token Economy – Bridging Culture, Technology, and the Future of Borderless Finance."*

While the *Shahnameh Book* aimed to build a bridge between culture and the future—blending identity and heritage with emerging technologies such as blockchain and artificial intelligence—this article provides the scientific and technical foundation of that vision. The book offered a macro-level perspective on the Real Token Economy and the prospects of a borderless financial system; this article focuses on the infrastructure pillars required to sustain that future: security, fraud detection, and compliance in cryptocurrency transactions.

In essence, the *Shahnameh Book* is the grand narrative, while this article is one of its specialized chapters, demonstrating how AI and Blockchain can serve not only as tools of innovation but also as enablers of trust, accountability, and fairness in the emerging digital economy.