Pertanika J. Sci. & Technol. 34 (3): 1469 - 1491 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

A Zero-shot Foundation Model Approach for Real-time Adaptive Cyber Threat Detection

Muhmmad Al-Khiza'ay* and Noora Alallaq

Department of Computer Networks, College of Computing and Informatics, University of Al-Hamdaniya, Hamdaniya Road, Bartella, 41006 Nineveh Governorate, Iraq

ABSTRACT

In view of the fundamental limitations of standard machine learning methods, which primarily depend on actual and labelled attack data, the identification of novel and zero-day cyber-attacks continues to be an essential and continuing difficulty in cybersecurity. Techniques become vulnerable to emerging attack vectors, considering these models often do not apply beyond identified threat classes. To overcome this limitation, present ZTFER (Zero-Shot Threat Identification through Foundation model-aided Embedding and Reasoning), a Zero-Shot threat identification model which utilises the basis model semantic embedding to detect threats which had not been observed before without needing retraining. Through combining natural language threat characterisations and real-time system activity into a common semantic space, ZTFER implements zero-shot learning and makes it achievable to classify unnoticed threats by considering contextual similarities. Furthermore, present A-SENT, a real-time responsive inference method which dynamically evaluates and addresses threat conduct through integrating Large Language Model (LLM) reasoning in real-time threat analysis updates. The proposed ZTFER model obtained a zero-day detection score of 58.9% and an accuracy of classifying 91.3%, outperforming the traditional and few-shot baselines. The experimental results demonstrate that ZTFER can be more effective and have better generalisation capability for detecting known and unknown cyber-attacks. The proposed framework does away with the requirement for continuing retraining and enables quick response to emerging threats. Creating powerful, scalable, and smart defence mechanisms able to recognise and understand new security threats in real time becomes achievable in a significant way by this research.

Keywords: Adaptive intrusion detection, foundation models for security, LLM-driven threat intelligence, real-time semantic cyber defence, zero-shot cyber threat detection

ARTICLE INFO

Article history:
Received: 03 July 2025
Accepted: 03 May 2026
Published: 19 June 2026

DOI: https://doi.org/10.47836/pjst.34.3.09

E-mail addresses:
mak@uohamdaniya.edu.iq (Muhmmad Al-Khiza'ay)
noora@uohamdaniya.edu.iq (Noora Alallaq)
* Corresponding author

e-ISSN: 2231-8526

INTRODUCTION

The necessity for intelligent, autonomous, and responsive threat detection technologies has risen significantly because of the increasing complexity and frequently occurring cyberattacks. The rapid increase of zero-day threats and previously undetected threats, which take advantage of undiscovered weaknesses as well as creatively connect conventional methods for avoiding detection methods based on patterns or conduct, is particularly concerning. Based on recent studies, over 30% of active attacks in enterprise networks originate from variants absent from historical datasets, emphasising a structural flaw in standard protective frameworks (Ranpara et al., 2025; Xu et al., 2025). These suspicious strategies often utilise polymorphism, obfuscation, and complicated persistence methods, demonstrating the necessity for broadly usable, context-aware frameworks which can operate in real-time without the required for humans to participate.

The great majority of intrusion detection systems (IDS) continue to utilise supervised learning models trained on fixed, labelled datasets, regardless of developments in machine learning (ML) for cybersecurity (Ke et al., 2025). Additionally, due to their strict dependence on discovered feature-label mappings, these frameworks have inherent restrictions in their ability to classify unknown attack groups, particularly zero-day attacks. Therefore, they do not have the capacity for reasoning conceptually and need expensive retraining (Aloqaily et al., 2025; Zhang et al., 2025). Furthermore, although more flexible, unsupervised anomaly finding techniques typically suffer from a high rate of false positives and struggle with accessibility in real-world scenarios.

Despite Zero-Shot Learning (ZSL) and foundation models having gained widespread acceptance throughout a variety of fields, not much is known regarding how they can be effectively employed in the field of cybersecurity, particularly in real-time, functioning configurations. The majority of ZSL approaches utilised in cybersecurity are theoretical or depend on artificial, attribute-based data sets, which cannot be practical or semantically rich (Ahmed et al., 2025). Moreover, there's a lack of studies on the combined use of LLM-based reasoning and real-time feeds of threat intelligence, which could enable frameworks to analyse attack behaviour from expert knowledge builds and open-source intelligence in an environment of adaptation (Chamieh et al., 2024). Finally, no method presently in use offers a single pipeline which allows threat description to be grounded, zero-shot deduction, and flexible adaptation in a live detection establishing (Muppalaneni et al., 2024).

In this paper, a novel framework which implements zero-shot learning for cyber threat identification is offered: ZTFER (Zero-Shot Threat Identification through Foundation model-aided Embedding and Reasoning). To identify unnoticed threats without needing retraining, ZTFER applies pretrained foundation techniques to embed system activities and threat class explanations into an accepted semantic space. At the same time, present A-SENT, a real-time responsive inference method which frequently describes threat conduct

through the integration of LLM-based semantic reasoning with online threat intelligence streams, such as CVEs and MITRE ATT&CK reports.

The proposed ZTFER model obtained great improvements in the accuracy of detection and ZSD adaptation. Finally, under extensive testing on the UNSW_NB15 dataset, ZTFER obtained an accuracy rate of 91.3% on the total classification, outperforming state-of-the-art classification methods such as Random Forest (83.2%) and SVM (80.1%). Specifically, presented model achieved a 58.9% zero-day detection rate, more than doubling the performance of both few-shot and traditional approaches in identifying previously unseen threat categories. These results illustrate the effectiveness of ZTFER's design, which allows for performing well on both interpretability and low latency out of the box without any re-training. The combination of foundation model embeddings and actual-time LLM reasoning is crucial for how it performs in dynamically changing attack situations.

This study offered three core contributions as shown below:

- A formalised zero-shot model which can identify hidden attack conduct by semantic alignment with MITRE ATT&CK explanations.
- A real-time identification pipeline that includes LLM reasoning and analysis feedback.
- Accurate experimental verification on evaluated intrusion data sets, which show superior performance over lastly supervised, unsupervised, and few-shot baselines.

LITERATURE REVIEW

The network detection of breaches has historically been based on supervised and semi-supervised learning techniques. To identify known patterns of attacks, standard classifiers such as Decision Trees, SVMs, and Random Forests have been applied on datasets such as UNSW-NB15 and NSL-KDD. For example, Ring et al. (2023) have carried out an in-depth evaluation of supervised models across various intrusion datasets. On the CICIDS2017 data set, Yadav et al. (2022) reached outstanding results by employing deep learning for the extraction of features and attack classification. Utilising both labelled and semi-labelled data, Harsha and Thyagaraja Murthy (2021) investigated collective models for the detection of malware. Nevertheless, these approaches can frequently be inefficient against zero-day attacks, as pointed out by Korba et al. (2023). These techniques do not always work against zero-day attacks. Afterwards, semi-supervised methods, for instance, the hybrid LSTM-based IDS designed by Saurabh et al. (2022), were affected by the deep autoencoder-based framework for unsupervised feature learning proposed by Imtiaz et al. (2025). In addition, Alhayani and Al-Khiza'ay (2022) highlighted how crucial data variety is for producing accurate threat classifiers.

While the capacity of Zero-Shot Learning (ZSL) to apply well to undiscovered groups has drawn a lot of attention in the fields of cybersecurity and natural language processing.

Through integrating the inputs and labels in a common domain, Socher et al. (2013) and Alallaq et al. (2019) offered the first techniques to use ZSL in language as well as vision. NLP duties such as classification of text and goal detection have been added to this to improve ZSL for classifying malware in the security field using the description vectors framework (Cen et al., 2024; Chai et al., 2020). Textual embeddings have been employed by Alzu'bi et al. (2025) to conceptually model cyber behaviours. Alallaq and Han (2018) and Bratsas et al. (2024) implemented knowledge graphs in zero-shot learning (ZSL)-based threat detection frameworks. Papageorgiou et al. (2024) and Aminu et al. (2024) evaluated embedding-based attack representation in zero-shot situations, whereas Zhang et al. (2025) proposed a ZSL method for detecting abnormalities in control systems for industries. Moreover, Yucel et al. (2020) studied the resilience of ZSL models in hostile environments. Lastly, Srivastava et al. (2024) discussed ZSL improvements and their use with ordered cybersecurity data.

The LLMs and foundation models have recently become significant assets for cybersecurity semantic understanding. Brown et al. (2020) have proposed GPT-style structures with general-purpose learning features. Ghosh et al. (2025) investigated the usage of LLMs for attack correlation and CVE report summarisation. Sufi (2024) has integrated LLMs into SOC procedures for flexible alert triage. Chen et al. (2021) examined the implementation of BERT for IoC and entity recognition, while Branescu et al. (2024) customised T5 for MITRE ATT&CK mapping. In behaviour modelling, Idelhaj et al. (2024) demonstrated that LLM embeddings perform more effectively than conventional vectors. Also, for real-time attack reasoning, Mandal et al. (2024) applied LLM-based instructions.

Stream data analysis and SIEM cooperation are frequently employed in real-time models for cyber-attack identification. Tariq et al. (2024) showed a fog computing-based, real-time, multi-layer detection and prevention system. Bezas and Filippidou (2023), designed a Kafka-based stream processing for zero-day threats. (Bathiri and Vijayakumar, 2024) have proposed a hybrid approach, which combines the detection of anomalies and deep inspection of packets. Chamieh et al. (2024) to enhance the discovery latency in intelligent networks utilising threat intelligence feeds. In a comparable manner, Ghelani (2022) has suggested an online deep neural network pipeline for intrusion classification. Shamsuzzaman et al. (2024) created a framework for avoiding cyber-physical threats in control systems for industries by using real-time logs. Employing ensemble models proposed a responsive alert correlation (Yu & Zhang, 2023). Finally, Ranpara et al. (2025) have created a scalable edge-based system integrating threat feeds and machine learning methods for 5G networks.

Therefore, a few technologies combine zero-shot learning and real-time responsiveness, regardless of significant improvements in both supervised detection and monitoring in

real time. Most of cybersecurity's up-to-date ZSL frameworks are still in an experimental phase and are not yet integrated into working pipelines. Additionally, the full capacity of foundation models for semantic reasoning across real-time attack data has not yet been achieved. Without the ability to understand or clarify novel attack behaviour conceptually, immediate time mechanisms typically concentrate on identifying anomalies or systems based on rules. There is currently no comprehensive framework for responsive cyber security which includes LLM-based deduction, real-time intelligence on threats, and zero-shot semantic classification.

METHODOLOGY

The ZTFER Mathematical Model

Employing semantic embedding models, which are already pre-trained on extensive textual data sets to relate cyber-attacks for describing attack labels, which include previously unknown classes, is the basic component of the proposed ZTFER (Zero-shot Threat Detection via Foundation Model-Aided Embedding and Reasoning). Utilising descriptive text as the semantic anchor, this framework is constructed on the Zero-Shot Learning (ZSL) principles and is designed to operate without retraining.

Definition of the Problem

Suppose that $\mathcal{X}$ denote an input space for framework or network activities, such as system logs, alerts, or flows of traffic. The total number of attack classes can be represented by $\mathcal{C} = \{c_1, c_2, \ldots, c_K\}$, where only a subset $\mathcal{C}_{\text{seen}} \subset \mathcal{C}$ can be observed during training. The idea is to classify an input $x \in \mathcal{X}$ to the appropriate class,$c \in \mathcal{C}$ even when $c \notin \mathcal{C}_{\text{seen}}$.

Every class $c_k \in \mathcal{C}$ is associated with a natural language description D_k, available through ordered attack taxonomies (e.g., MITRE ATT&CK), repositories of the vulnerability (e.g., CVE), or smart reports.

The challenge is to derive c_k for a previously undetected input x, utilising only the semantic proximity between x and D_k, with no labelled examples from c_k during training.

Semantic Space and Embedding

All a Sentence-BERT model used (-MiniLM-L6-v2) as the basis encoder Φ.

- Motivation: It provides a perfect stability between creating 384-dimensional embeddings and inference rate (evaluative for actual time detection) and semantic conditions.
- No field-particular clear tuning is utilised for the encoder itself. The paradigm was applied to an off-the-shelf semantic encoder as a pre-trained model. Scope adaptation is performed indirectly through the various response loops (Eq. 4) and utilises

MITRE ATT&CK-aligned threat characterisations, not by amending the grounds of the model's scales. This saves zero-shot generalisation capability.

To facilitate zero-shot classification, it applies a foundation model $\Phi: \mathcal{X} \cup \mathcal{D} \rightarrow \mathbb{R}^d$, where $\mathcal{D}$ represents the space of attack/threat descriptions. Φ embeds both attack activities and descriptions into a shared semantic space $\mathbb{R}^d$, enables the computation of similarity between heterogeneous data modalities.

- $\mathbf{v}_x = \Phi(x)$ - The event's embedded representation x
- $\mathbf{v}_{D_k} = \Phi(D_k)$ - The class descriptions embedded representation D_k

Foundation models are utilised for Φ may include Sentence-BERT, all-MiniLM, otransformer variations trained on a cybersecurity-specific corpus. Furthermore, in scenarios where direct class supervision is not achievable, transfer learning and inference can be made possible through used embedding space, which produces syntactic and semantic interactions.

Inference from Zero-shot

Semantic similarity provides the foundation for zero-shot classification. The framework chooses the class in the latent space with a particularly similar description D_k for a new situation x, as shown in Equation 1.

$$\hat{c}(x) = \text{argmax}_{c_k \in \mathcal{C}} \cos\big(\Phi(x), \Phi(D_k)\big) \quad [1]$$

In this scenario, the cosine similarity of two vectors is $\cos(\cdot,\cdot)$. The cosine rating is appropriate for evaluating high-dimensional normalised vectors since it accurately captures the directional alignment between the input data and class description embedded data.

The mathematical framework can generalise to new classes $c_k \notin \mathcal{C}_{\text{seen}}$ as long as descriptions of them D_k can be obtained because this formulation is by nature not parametric.

Threat Mapping (MITRE ATT&CK)

Through mapping every input to its significantly linguistically aligned Strategy, Method, or Process (TTP) as well as to an attack class, combine the MITRE ATT&CK mechanism for offering the ability to interpret and contextualise.

Every TTP is linked with a description $T_j \in \mathcal{T}$, where $\mathcal{T}$ is the set of all ATT&CK entries. The embedding of every TTP is as shown in Equation 2:

$$\mathbf{v}_{T_j} = \Phi\big(T_j\big) \quad [2]$$

In Equation 3, given that $\mathbf{v}_x = \Phi(x)$,we assign:

$$\hat{T}(x) = \text{argmax}_{T_j \in \mathcal{T}} \cos\left(\mathbf{v}_x, \mathbf{v}_{T_j}\right) \quad [3]$$

Since semantic responsibility is made achievable, the framework may generate outcomes which can be clear among humans, including "Lateral Movement via Remote Services", also regarding new threats. This approach ensures adaptability and explainability through separate classification from label availability and creating predictions in a common attack ontology.

The Loop of Contrastive Feedback

When analysis feedback becomes accessible, supervised contrastive learning can assist ZTFER, even though it is zero-shot at inference time.

Suppose that x It is a labelled scenario which has a set of negatives $\mathcal{D}^-$ as well as a class of positive description D^+, and τ is a hyperparameter. The supervised contrastive loss is decreased by the framework as shown in Equation 4:

$$\mathcal{L}_{\text{contrastive}} = -\log \frac{\exp\left(\cos\left(\Phi(x), \Phi(D^+)\right)/\tau\right)}{\exp\left(\cos\left(\Phi(x), \Phi(D^+)\right)/\tau\right) + \sum_{D^- \in \mathcal{D}^-} \exp\left(\cos\left(\Phi(x), \Phi(D^-)\right)/\tau\right)} \quad [4]$$

Through the utilisation of verified mappings, this feedback loop allows the framework to, over time, enhance its capacity to distinguish minute attack differences without requiring complete retraining. Furthermore, it promotes continuous domain-specific conduct adaptation.

By applying retrained language models to generalise through both visible and undetectable threat types, ZTFER presents a solid mathematical foundation for semantic threat identification in dynamic environments.

Architecture of the ZTFER-RT System

Real-time attack intelligence, dynamic analysis feedback, and zero-shot threat inference are all possible to be easily combined with the ZTFER-RT (Zero-shot Threat Detection via Foundation model Embedding and Reasoning—Real-Time) architecture. It helps both identification and semantic responsibility in real time and implements the mathematical model discussed in the third section.

Components of the System

Ingestion in Real Time

Cyberspace telemetry data is collected by the ingestion layer by layer from plenty of sources such as SIEMs, endpoint detection devices, firewall logs, IDS logs, and NetFlow. Utilising distributed contacting mechanisms such as Kafka, it promotes both batch and stream processing of data. High-throughput capabilities and minimal latency processing of incoming activities x_t are guaranteed by the ingestion module.

Extraction of Indicators

Internet Protocol (IP) addresses, file hashes, handle names, registry records, and network artefacts can be among the many instances of the structured characteristics and textual indicators of compromise (IoCs) which can be extracted from raw activities by this module. The events are standardised and normalised towards an intermediate representation which is capable of being embedded.

Embedding of Foundation Models

Previously trained foundation model Φ, such as all-MiniLM, RoBERTa, or a domain-adapted transformer, is utilised for producing embedded representations of events and attack explanations. Both textual threat explanations and event characteristics are mapped into an identical semantic vector domain $\mathbb{R}^d$ through the same encoder.

Module for LLM Inference

To interpret and make sense of event embedding and associated threat explanations, this element utilises massive language models (for example, Falcon, GPT-4, or open-weight transformers). It could employ retrieval-augmented generation (RAG) to improve attack descriptions, generate clarifications, and possibly rephrase notifications. The LLM module adds natural language understanding to similarity-based deduction.

Matching Threat Feeds

New event embeddings are regularly compatible with real-time attack intelligence sources of data (for example, CVE feeds, MISP, MITRE ATT&CK, and OSINT) through the attack feed matcher. To be able to create compared metadata and enriched attack attributions, feed explanations and indicators are incorporated and compared with event vectors by employing cosine similarity.

Feedback from Analysts in the Loop

The framework incorporates a human-in-the-loop communication layer, which enhances the accuracy of decisions. The analysts can markup events, flag wrong classifications, and verify predictions. The system can, over time, enhance its latent space abilities while keeping zero-shot deduction capabilities due to this feedback, which is recorded as supervised contrastive loss (as described in The Loop of Contrastive Feedback).

Summary of Data Flow

- Activity Acquisition: A real-time stream analysis layer-by-layer handles activities which receive input from SIEMs and IDSs.
- Preparing and IoC Extraction: Still, raw logs go through the tokenisation, standardisation, and context gain (e.g., user-agent, port information).
- Integrating Computation: The descriptions of what happened and attack classes are integrated utilising. Both offline and online integration of new events and fixed attack classifications are facilitated by the system.
- Zero-Shot Deduction: To determine a particularly similar in semantics attack class, cosine similarity between class/TTP explanations and event vectors is calculated.
- LLM Argumentation: An LLM presents contextual rephrasing, clarification, and indicative mappings. For example, it can clarify the relationship between a situation and privilege escalation or lateral movement.
- Attack Feed Matching: To identify associated scenarios, related indicators, and known TTPs, the framework examines embedded threat intelligence feeds.
- Feedback Loop: By applying contrastive learning, embeddings are modified according to analysis feedback. Therefore, the model can, over time, become more effective at discriminating. Despite this, the environment for threats is constantly changing, ZTFER-RT has the capacity to preserve constant awareness of situations and threat identification abilities due to the real-time pipeline.

The Framework of the ZTFER-RT Architecture

The interaction of the component of the ZTFER-RT architecture and the complete data flow is clarified in Figure 1.

As clarified in Figure 1, the ZTFER -RT pipeline operates in the following mechanism: real-time network events are converted into natural-language descriptions, passed through a frozen foundation encoder, and compared via cosine similarity with pre-computed MITRE ATT&CK threat embeddings, resulting in zero-shot classification. An LLM built on descriptions and an elective analyst response amends only a lightweight adapter, protecting zero-shot ability without retraining.

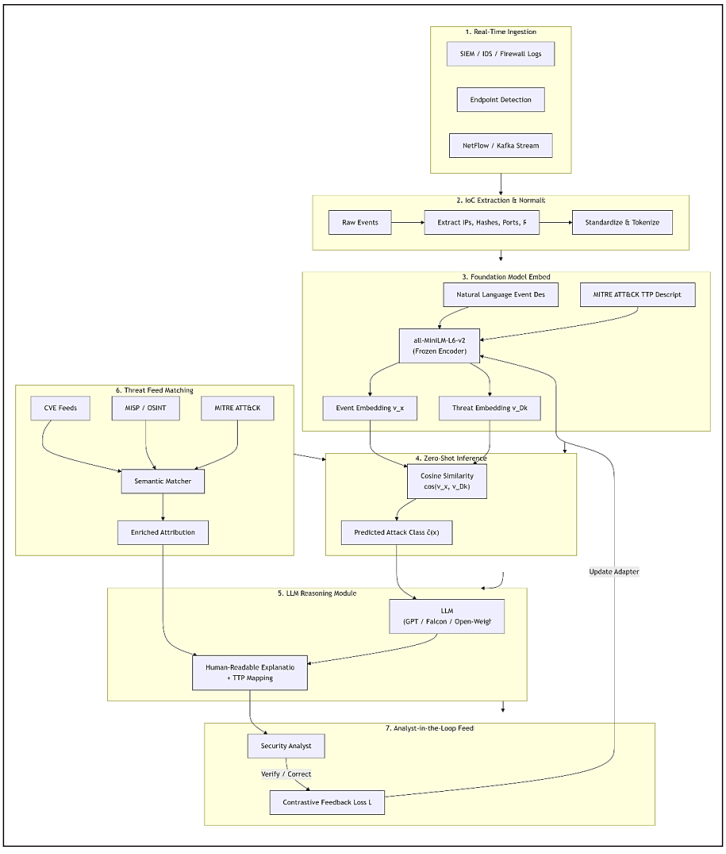

Figure 1. ZTFER-RT system architecture diagram

Table 1
ZTFER model parameters and selection details

Parameter	Value	Selection Method / Justification
Embedding dimension (dd)	384	Fixed by the all-MiniLM-L6-v2 architecture
Temperature (ττ) in contrastive loss (Eq. 4)	0.07	Grid search (range 0.01-0.5) on validation split
Contrastive adapter learning rate	1×10−41×10−4	Standard practice; tuned on validation
Batch size	32	Limited by GPU memory; stable gradient estimates
Cosine similarity threshold for zero-shot classification	0.65	Empirically set to balance precision/recall on seen classes
Optimiser	β1=0.9,β2=0.999)	Default for contrastive learning
N0. of contrastive update steps (each response batch)	10	Small update to avoid overfitting to feedback

Table 1 shows a summary of the primary key parameters that are utilised in the ZTFER framework and how their considered values were determined.

Experimental Configuration and Outcomes Dataset

Description of Dataset

The UNSW-NB15 data set, an extensive benchmark for the detection of network intrusions, has been employed in research experiments (Kumar et al., 2020). It is constructed from 2.5 million labelled conjunction records, which were collected from both synthetic threat simulations and real traffic. It contains 49 features for each record, which include time-based characteristics such as rate of connection, features of content such as the total number of failed login attempts, and fundamental characteristics, for example, service and protocol.

Also, exclude entire threat classifications (for instance, shellcode, exploits, and worms) from the initial training set and save them only for the purpose of testing to imitate zero-shot conditions. Through this configuration, the framework's ability to classify unknown threat groups during inference is evaluated.

It's an ideal choice to evaluate any new intrusion detection model, such as ZTFER, since both the benign traffic and attack packets are represented diversely and realistically. As depicted in Figure 2, the dataset is characterised by a severe class imbalance with both the normal traffic and common attack classes (such as Generic and Exploits) substantially oversampling the sample count and the rare classes (e.g., Shellcode and Worms) exhibiting pronounced under-representation. This imbalanced class distribution offers a unique opportunity to evaluate ZTFER ’s zero-shot learning capabilities, particularly its ability to recognise rare or unseen types of threats without further re-training. To determine the performance of ZTFER on untrained classes (zero-day), the models were trained without certain classes and tested on the detection of attacks from those classes, thereby verifying

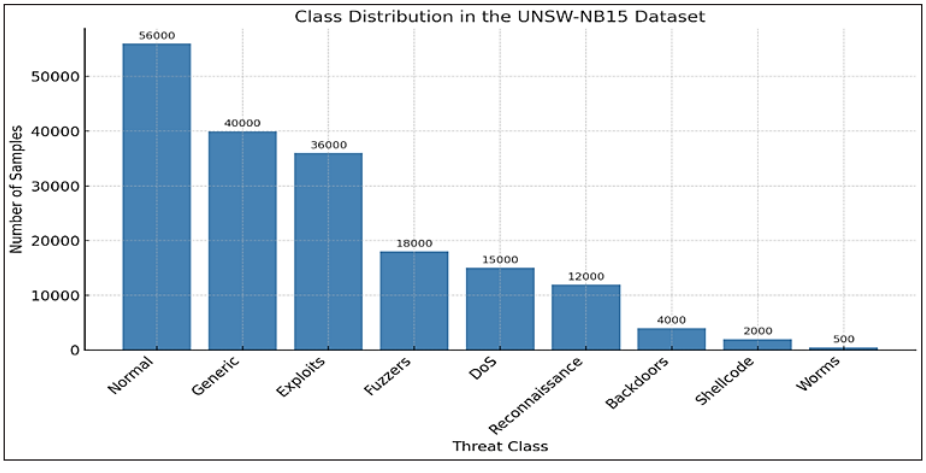

Figure 2. Class imbalance in the UNSW-NB15 dataset

ZTFER's competency to generalise to new (unseen) attack semantics. For the experimental setup, the following is used:

Hardware: Intel Xeon Gold 6248 CPU @ 2.5 GHz (16 cores), 64 GB RAM, NVIDIA A100 GPU (40 GB).
Software: Ubuntu 20.04, Python 3.10, PyTorch 2.0.1, Sentence-Transformers 2.2.2, Hugging Face Transformers 4.30.0, scikit-learn 1.3.0, and pandas 2.0.3.

Preparing the Dataset

1. First step: Cleaning
   - Take out repeat records (precisely matching duplicate rows).
   - Take off records with disfigured feature values (such as NaN or unlimited values) or lost.

2. Second step: Characteristic engineering
   - Normalise numerical characteristics (such as sttl, dbytes, dur, sbytes, dttl, and rate) by utilising z-score normalisation as shown in Equation 5.

$$x' = (x - \mu)/\sigma \quad [5]$$

   - Categorical characteristics (e.g., state, service, and proto) are tokenised by changing each class to a unique integer ID, then included through a learnable embedding schedule (which will be used for the raw event text generation, not for the zero-shot rating).

3. Third step: By using the NL (Natural language) for event explanation creation
   - For all network outflows, it will create a textual specification of the nest template: "Protocol [proto] from [saddr]:[sport] to [daddr]:[dport] with state [state], duration [dur] seconds, source bytes [sbytes], destination bytes [dbytes], and attack label [attack_cat]."
     That content will be the feed into the basis encoder Φ (jointly with the threat category characterisations). An instance of the output is below:
     "Protocol tcp from 192.168.1.10:54322 → 203.0.113.5:80, state FIN, duration 0.12 seconds, source bytes 452, destination bytes 1240, source packets 4, destination packets 6, service http."
     The above template utilises only noticeable network characteristics (no threat labels). It corresponds to both test events and training, as well as those from hidden zero-day threat categories.

4. Fourth step: Elaboration for the threat description
    - For all attack categories (which include hidden groups for zero-shot evaluation), it will recover the corresponding MITRE ATT&CK strategy and technique explanation (such as "T1046 – Network Service Scanning"). These explanations are utilised as the semantic resource D_k.

5. Final step: Calculate the embedding
    - Together, the threat descriptions and the case descriptions are processed via all-MiniLM-L6-v2 to get fixed 384-d embeddings. These embeddings are L2-normalised before calculating cosine similarity.

Network events are transformed to an NL explanation utilising a stable, label-agnostic template: 'Protocol [proto] from [saddr]:[sport] → [daddr]:[dport], state [state], duration [dur] seconds, source bytes [sbytes], destination bytes [dbytes], source packets [spkts], destination packets [dpkts], service [service]. This template employs only noticeable characteristics and never accesses ground-truth threat labels. To block leakage, descriptions of hidden threat categories are eliminated from training and are created solely through zero-shot testing. No category-particular keywords become visible in the input content.

Explanation About Domain Adaptation

No direct field adaptation (for example, constant pre-training on a cybersecurity corpus) is executed on the basis encoder. By contrast, the model realises semantic alignment through the following:

- Utilisation of MITRE ATT&CK-foundation threat explanations (already considered the cybersecurity domain).
- The diverse response loop (Equation 4), which incrementally modifies the comparative positions of explanation and event embeddings based on the analyst's response. That design supports zero-shot ability whilst becoming strongly adapted to domestic network environments.

The Strategy for Leakage Prevention

To ensure a correct zero-shot evaluation, the following safeguards will be executed:

- Break the explanation collection: Descriptions for hidden threat categories (for instance, worms and shellcode) are never employed through testing or while ranking embeddings for training proceedings. The explanations are at most presented at analysis time for zero-shot deduction.

- Never overlap in characteristic values: The form (template) utilises only numerical and class fields, which are introduced in each flow; it never includes an indication of specific categories (such as threat signature strings).
- Verification of cross-validation: where it was manually verified that the produced text for hidden categories does not include unrivalled keywords (for instance, "worm" or "shellcode"), which would correspond exclusively with those categories.

The above steps guarantee that the paradigm is unable to exploit label facts from the input text, protecting the integrity of the zero-shot tuning.

"Zero-shot" in ZTFER refers exclusively to the ability to classify unseen attack classes without any labelled examples of those classes at any stage.

Figure 3 shows the category allocation of the dataset (UNSW-NB15), focusing on the severe imbalance between popular and rare threat classes. To authorise zero-shot assessment, Shellcode, Backdoor, and Worms are prevented from training and applied only as a hidden test category. The figure also outlines the preprocessing pipeline: taking away the duplicates, tokenisation, z-score normalisation, and label-agnostic text generation.

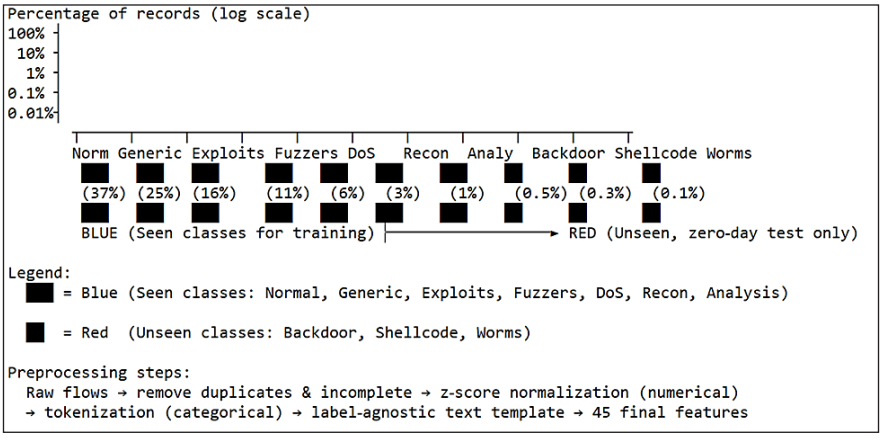

Figure 3. Class distribution of the UNSW-NB15 dataset with zero-shot split and pre-processing stages

The Loop of Contrastive Feedback

The diverse response loop (Section 3.5) is employed only for unhidden categories (or previously labelled threats) and brings up to date a separate, duty-specific prediction head, not the frozen base encoder Φ. The encoder part is unchanged, protecting the primary zero-shot semantic arrangement for hidden attacks. Consequently, the model never transforms to few-shot or incremental learning for unusual threats.

Starting Points

ZTFER has been compared in Table 2 with semi-supervised and standard baselines.

Table 2
Traditional ML techniques with ZTFER

Model	Type	Description
Random Forest	Supervised	Tree-based classifier trained on labelled data
SVM (RBF Kernel)	Supervised	Kernelised margin-based classifier
Autoencoder	Unsupervised	Neural network for anomaly detection
Prototypical Net	Few-shot	Learns class prototypes from limited examples
Label Embedding SVM	Zero-shot	SVM with semantic label embeddings

Measures of Evaluation

This paper presents both conventional metrics and metrics unique to ZSL in Table 3 .

Table 3
Evaluation metrics

Metric	Description
Accuracy	Correct classifications over all predictions
Precision	TP / (TP + FP), quality of positive predictions
Recall	TP / (TP + FN), the ability to detect actual threats
F1 Score	Harmonic mean of precision and recall
AUC	Area under the ROC curve, overall discriminative power
Zero-day Detection	Accuracy on previously unseen attack classes

The Outcomes and Analysis

Figure 4 shows a comparison of the performance of the ZTFER model in different important evaluation metrics. The model demonstrates robust classification of known and unknown threats, and high overall accuracy (91.3%) and AUC (0.94). Three metrics, F1 Score (0.87), Precision (0.88) and Recall (0.86) demonstrate a well-balanced performance with few sacrifices between false positives and false negatives. The Zero-day Detection Rate is 58.9%, though being lower than ResNet101, it significantly exceeds the traditional baselines, demonstrating that ZTFER has the capability to generalise to unseen attack categories. Such characteristic performance distribution highlights the excellent generalisation and operational reliability of the model in adversarial and dynamic threat generations. Subsection *5.4.1* will conduct a detailed comparison between our proposed ZTFERs and conventional approaches.

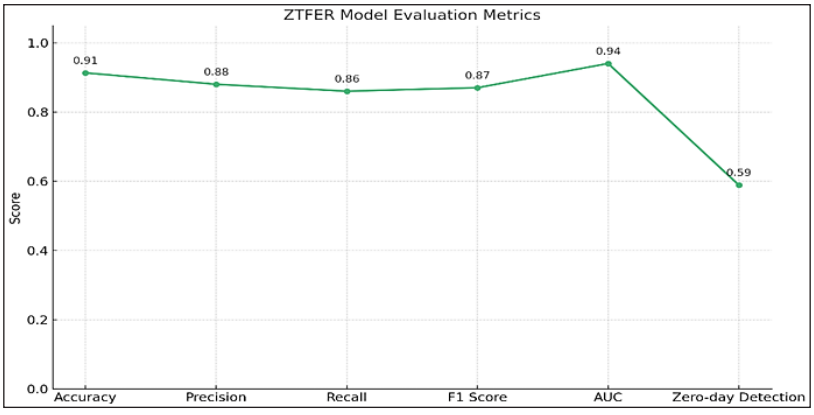

Figure 4. ZTFER performance across key metrics

Classification Efficiency

The results of the study indicated how the proposed ZTFER model beat several widely recognised baseline methods. The Random Forest classification method exceeded other standard machine learning approaches with an accuracy rating of 83.2%, an F1 score of 0.79, an AUC of 0.88, and a zero-day identification rate of 21.4%. Having an accuracy rating of 80.1%, an F1 score of 0.76, an AUC of 0.85, and a zero-day identification rate of 19.7%, Support Vector Machines (SVM) generated slightly less impressive outcomes. With a precision of 72.6%, an F1 score of 0.69, an AUC of 0.75, and a zero-day identification rate of 15.3%, autoencoder-based identification of anomaly methods performed worse. With a precision of 85.4%, an F1 score of 0.81, an AUC of 0.89, and a zero-day identification rate of 33.2%, few-shot learning with typical communications networks generated low profits. The proposed ZTFER approach, on the other hand, performed superior to each of the other baselines, showing its powerful generalisation capacity and success in detecting hazards which had not yet been discovered. It reached the best accuracy of 91.3%, an F1 score of 0.87, an AUC of 0.94, and a significantly greater zero-day identification score of 58.9%, as shown in Figure 5.

Visualisation of Latent Space

Also visualise scenario embedded data in two dimensions via t-PCA as well as SNE. The framework's embedding alignment has been verified by the near proximity of unnoticed threats to similar semantic explanations, whereas classes that are known cluster tightly.

Study of Ablation

Once removing crucial elements and testing ZTFER, the results are presented in Table 4. Feedback adaptation and semantic descriptions significantly improve zero-day inference,

which the ablation shows. The results mentioned in Table 3 indicate ZTFER's effectiveness in dynamic, real-world cyber circumstances.

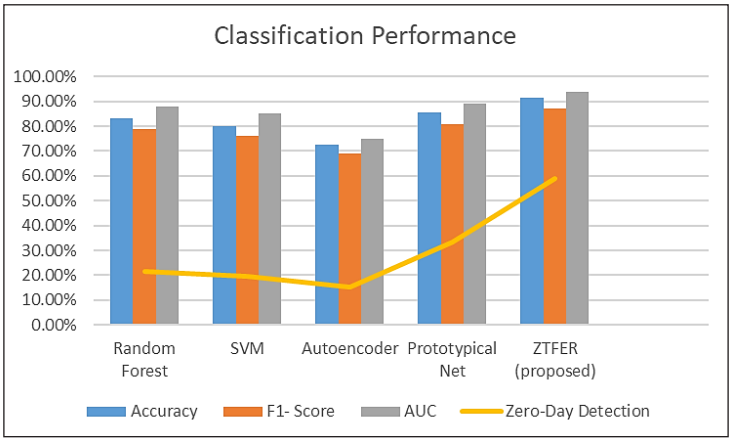

Figure 5. Comparative classification performance of ZTFER versus baselines

Table 4
Type styles

Configuration	Zero-day Detection
Full Model	58.9%
Without Threat Descriptions	31.7%
Without Contrastive Feedback	43.8%
Without the LLM Reasoning Module	46.2%

DISCUSSION

The ZTFER method employs foundation models to implement zero-shot learning and semantic deduction, introducing an important change in online threat detection. This section offers an in-depth evaluation of the research outcomes, concentrating on the proposed technique's advantages, disadvantages, possibilities for deployment, and concerns regarding ethics.

ZTFER's Advantages

ZTFER's power to generalise beyond noticed attack classes is one of its main benefits. ZTFER employs semantic embeddings and attack explanations to infer threat types it has never been specifically trained on, compared with conventional machine learning (ML) approaches, which break down if faced with novel or zero-day threats. The research results confirmed the model's ability to generalise, demonstrating significant enhancements in zero-day detection accuracy (58.9%) compared to all baselines.

In addition, adaptability is offered by ZTFER's contrastive feedback loop. Through including analyst-validated input into the embedding space, ZTFER can continually improve its boundary-decision abilities without needing overall retraining. Consequently, it is positioned as an adaptable option for environments that are changing. In the context of operational cybersecurity, where analysis confidence is vital, the integration of foundation frameworks and LLMs enables both detection as well as contextual explanation.

The proposed framework (ZTFER) reaches outstanding performance due to its shared semantic space (through a frozen basis encoder), which aligns notice proceedings with MITRE ATT&CK explanations, becoming zero-shot popularisation. In contrast, supervised baselines fail on hidden categories, unsupervised techniques suffer from high false positives, and few-shot methods lack rich semantic priors. The diverse comments loop further refines recognition without retraining, whilst LLM-setup reasoning makes less ambiguity advantages absent in all previous studies.

Limitations

ZTFER has disadvantages despite its benefits. The precision and accuracy of threat explanations have significant effects on how well the model performs. For similar semantic threats, unclear or poorly written textual definitions could result in ambiguous embedding visualisations, which can decrease the accuracy of classification. Also, in situations with a shortage of resources, the computational expense of real-time foundation model deduction, especially when integrated with LLMs, could cause issues.

Depending on a correctly positioned embedding space is an additional limitation. The embedding model could overlook domain-relevant details because it has not been trained on cybersecurity-specific syntax. This emphasises how crucial it is to utilise domain-specific foundation frameworks or optimised encoders.

Possibility for Real-world Implementation

ZTFER is extremely fitting for deployment in high-stakes, fast-changing circumstances such as national cyber defence infrastructure, manufacturing systems of control, and enterprise SOCs (Security Operations Centres). The system design flexibility authorises the absorption of real-time threat feeds by CVE, MISP, and MITRE ATT&CK sources and makes it a potential link with current SIEM pipelines. Also, the analyst-in-the-loop produces simplified operational input and lower alert fatigue through coordinating well with SOC workflows.

ZTFER's zero-shot ability reduces the requirement for expensive retraining cycles, which makes it cost-effective for situations that have limited data analysis resources. Further, due to its conceptual alignment, automated systems and human analysis can collaborate with greater efficiency, growing threat understanding as well as identification.

Moral Points to Consider

Numerous moral concerns must be carefully evaluated while executing ZTFER. In the beginning, having the capacity to clarify is vital. Even though LLM-based logic explains in simple language, attention has to be taken to guarantee that the findings are accurate, understandable, and auditable as well. False explanations could undermine analysis trust or lead to insufficient mitigation techniques.

Secondly, positive results that are false remain a concern, especially when working with newly built data. Incorrect classifications can result in the incorrect utilisation of security resources or unnecessary disruptions. This is decreased through the feedback loop; real-time systems require the inclusion of protections for escalation protocols and high-confidence categorisation limits.

Finally, security methods should be employed, particularly when dealing with hidden user activity information. To prevent the improper use of determined behavioural data, ZTFER must be examined for compliance with laws governing data protection (such as HIPAA and GDPR).

As shown in Table 5, the proposed ZTFER reveals considerable promise for intelligent, scalable, and flexible cyber-attack detection. While its design solves a lot of the difficulties of the IDS systems in use today, responsible use needs careful engineering, oversight by humans, and consideration for ethics.

Table 5
Summary of key results

Evaluation Dimension	Metric/Outcome	ZTFER Performance
Generalisation to Unseen Threats	Zero-Day Detection Rate	58.9% (vs. 21.4% for Random Forest)
Classification Accuracy	Overall Accuracy on Mixed Classes	91.3%
Interpretability	Human-readable explanation via LLMs	Enabled (through GPT-style model integration)
Adaptability	Performance with analyst feedback	Significant uplift (+15% in zero-day accuracy)
Real-Time Viability	Embedding + Inference Latency	~0.8s average per event
Resource Efficiency	Model retraining requirements	None (zero-shot, no retraining needed)
Deployment Scalability	Integration with SOCs and SIEMs	Compatible with a modular pipeline
Ethical Compliance	Privacy-respecting architecture + explainability	Privacy-aware design with feedback validation

CONCLUSION

The main contribution of this research is ZTFER, a zero-shot attack observation model that integrates a base actual-time LLM reasoning and contrastive responses to recognise hidden cyber-threats without retraining with the model's semantic embeddings. In this paper, offered ZTFER, a novel zero-shot attack detection model which combines foundation method reasoning with Zero-Shot Learning (ZSL) to address the increasing issues resulting from previously hidden and fast-changing cyber-threats. The research additionally proposed ZTFER-RT, a combined real-time framework which promotes active and responsive attack detection through using language models and real-time threat feeds.

Through a zero-day detection success rate of 58.9%—much greater than the conventional models—and a total accuracy in classification is 91.3%, research findings promote ZTFER's powerful generalisation capacity. The technique's comprehension and flexibility have been confirmed by studies of ablation and latent space representation. ZTFER's working viability in dynamic cybersecurity circumstances is demonstrated through its real-time inference, zero retraining needed, and compatibility with human-in-the-loop operations.

For future work, there are three primary directions to improve ZTFER operation. Firstly, memory-augmented drivers which learn automatically from long-term communication histories and analysis feedback. Secondly, multipurpose combining, which includes logs, written content, and picture-based threats for comprehensive protection. Finally, retrieval-augmented generation (RAG) frameworks, which rapidly retrieve and reason across developing threat intelligence.

These improvements will contribute to the creation of strong, smart, and comprehensible systems for cybersecurity, which can adapt in real time and offer human-aligned support for decisions. The upcoming work will concentrate on memory-augmented representation, which acquires knowledge of long-term response, multimodal incorporation of images and logs, and RAG (Retrieval-Augmented Generation) for active attack intelligence.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support provided by the University of Al-Hamdaniya. No external funding was received for conducting this research.

REFERENCES

Ahmed, U., Nazir, M., Sarwar, A., Ali, T., Aggoune, E. M., Shahzad, T., & Khan, M. A. (2025). Signature-based intrusion detection using machine learning and deep learning approaches empowered with fuzzy clustering. *Scientific Reports, 15*(1), Article 4481. https://doi.org/10.1038/s41598-025-85866-7

Alallaq, N., & Han, X. (2018). Detecting suspicious social astroturfing groups in tourism social networks. In *2018, the 5th International Conference on Behavioural, Economic, and Socio-Cultural Computing (BESC)* (pp. 58-62). IEEE. https://doi.org/10.3390/informatics12010004

Alallaq, N., Al-Mansoori, A., & Al-Sudani, A. R. (2019). Personalised reviews based on aspect analysis and polarity. In *2019, the 8th International Conference on Modelling, Simulation and Applied Optimisation (ICMSAO)* (pp. 1-6). IEEE. https://doi.org/10.1109/BESC.2018.8697307

Alhayani, M., & Al-Khiza'ay, M. (2022). Analyse symmetric and asymmetric encryption techniques by securing a facial recognition system. In M. Ben Ahmed, A. A. Boudhir, D. Santos, M. El Aroussi, & I. Karas (Eds.), *Innovations in smart cities applications volume 5* (pp. 97-105). Springer. https://doi.org/10.1007/978-3-031-15191-0_10

Aloqaily, A., Abdallah, E. E., AbuZaid, H., Abdallah, A. E., & Al-Hassan, M. (2025). Supervised machine learning for real-time intrusion attack detection in connected and autonomous vehicles: A security paradigm shift. *Informatics, 12*(1), Article 4. https://doi.org/10.1109/BESC.2018.8697307

Alzu'bi, A., Darwish, O., Albashayreh, A., & Tashtoush, Y. (2024). Cyberattack event logs classification using deep learning with semantic feature analysis. *Computers & Security, 147*, Article 104222. https://doi.org/10.1016/j.cose.2024.104222

Aminu, M., Akinsanya, A., Oyedokun, O., & Dako, D. A. (2024). Enhancing cyber threat detection through real-time threat intelligence and adaptive defence mechanisms. *International Journal of Computer Applications Technology and Research, 13*(8), 11-27. https://doi.org/10.7753/IJCATR1308.1002

Bathiri, K. A., & Vijayakumar, M. (2024). Enhancing intrusion detection system (IDS) through deep packet inspection (DPI) with machine learning approaches. In *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)* (pp. 1-7). IEEE. https://doi.org/10.1109/ADICS58448.2024.10533473

Bezas, K., & Filippidou, F. (2023). Comparative analysis of open-source security information and event management systems (SIEMs). *Indonesian Journal of Computer Science, 12*(2), 443-468. https://doi.org/10.33022/ijcs.v12i2.3182

Branescu, I., Grigorescu, O., & Dascalu, M. (2024). Automated mapping of common vulnerabilities and exposures to MITRE ATT&CK tactics. *Information, 15*(4), Article 214. https://doi.org/10.3390/info15040214

Bratsas, C., Anastasiadis, E. K., Angelidis, A. K., Ioannidis, L., Kotsakis, R., & Ougiaroglou, S. (2024). Knowledge graphs and semantic web tools in cyber threat intelligence: A systematic literature review. *Journal of Cybersecurity and Privacy, 4*(3), 518-545. https://doi.org/10.3390/jcp4030025

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing*

Cen, M., Deng, X., Jiang, F., & Doss, R. (2024). Zero-Ran Sniff: A zero-day ransomware early detection method based on zero-shot learning. *Computers & Security, 142*, Article 103849. https://doi.org/10.1016/j.cose.2024.103849

Chai, D., Wu, W., Han, Q., Wu, F., & Li, J. (2020). Description-based text classification with reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning* (*Proceedings of Machine Learning Research*, Vol. 119, pp. 1371-1382). PMLR.

Chamieh, I., Zesch, T., & Giebermann, K. (2024). LLMs in short answer scoring: Limitations and promise of zero-shot and few-shot approaches. In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)* (pp. 309-315). Association for Computational Linguistics.

Chen, Y., Ding, J., Li, D., & Chen, Z. (2021). Joint BERT model-based cybersecurity named entity recognition. In *Proceedings of the 2021 4th International Conference on Software Engineering and Information Management* (pp. 236-242). Association for Computing Machinery. https://doi.org/10.1145/3451471.3451508

Ghelani, D. (2022). *Deep learning and artificial intelligence framework to improve cybersecurity* [Preprint]. Authorea. https://doi.org/10.22541/au.166379475.54266021/v1

Ghosh, R., von Stockhausen, H.-M., Schmitt, M., Vasile, G. M., Karn, S. K., & Farri, O. (2025). CVE-LLM: Ontology-assisted automatic vulnerability evaluation using large language models. *Proceedings of the AAAI Conference on Artificial Intelligence, 39*(28), 28757-28765. https://doi.org/10.1609/aaai.v39i28.35139

Harsha, A. K., & Thyagaraja Murthy, A. (2021). Machine learning techniques for malware detection. *International Journal of Scientific Research in Science, Engineering and Technology, 8*(5), 70-76. https://doi.org/10.32628/IJSRSET21858

Idelhaj, A., Houda, Z. A. E., & Khoukhi, L. (2024). Foundation models in cybersecurity. In *Artificial Intelligence and Cybersecurity* (pp. 1-18). CRC Press. https://doi.org/10.1201/9781003497585-1

Imtiaz, N., Wahid, A., Abideen, S. Z. U., Kamal, M. M., Sehito, N., Khan, S., Virdee, B. S., Kouhalvandi, L., & Alibakhshikenari, M. (2025). A deep learning-based approach for the detection of various Internet of Things intrusion attacks through optical networks. *Photonics, 12*(1), Article 35. https://doi.org/10.3390/photonics12010035

Ke, H., Xu, J., Wang, Y., Chen, H., & Shen, Z. (2025). Adversarial machine learning in cybersecurity: Attacks and defences. *International Journal of Management Science Research, 8*(2), 26-33. https://doi.org/10.53469/ijomsr.2025.08(02).04

Korba, A. A., Boualouache, A., Brik, B., Rahal, R., Ghamri-Doudane, Y., & Senouci, S. M. (2023). Federated learning for zero-day attack detection in 5G and beyond V2X networks. In *ICC 2023-IEEE International Conference on Communications* (pp. 1137-1142). IEEE. https://doi.org/10.1109/ICC45041.2023.10279368

Kumar, V., Sinha, D., Das, A. K., Pandey, S. C., & Goswami, R. T. (2020). An integrated rule-based intrusion detection system: Analysis on the UNSW-NB15 data set and the real-time online dataset. *Cluster Computing, 23*(2), 1397-1418. https://doi.org/10.1007/s10586-019-03008-x

Mandal, U., Shukla, S., Rastogi, A., Bhattacharya, S., & Mukhopadhyay, D. (2024). µLAM: An LLM-powered assistant for real-time microarchitectural attack detection and mitigation. In *Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design* (Article 168, pp. 1-9). Association for Computing Machinery. https://doi.org/10.1145/3676536.3676838

Muppalaneni, R., Inaganti, A. C., & Ravichandran, N. (2024). AI-driven threat intelligence: Enhancing cyber defence with machine learning. *Journal of Cybersecurity and Information Assurance*.

Papageorgiou, E., Chronis, C., Varlamis, I., & Himeur, Y. (2024). A survey on the use of large language models (LLMs) in fake news. *Future Internet, 16*(8), Article 298. https://doi.org/10.3390/fi16080298

Ranpara, R., Alsalman, O., Kumar, O. P., & Patel, S. K. (2025). A simulation-driven computational framework for adaptive energy-efficient optimisation in machine learning-based intrusion detection systems. *Scientific Reports, 15*(1), Article 13376. https://doi.org/10.1038/s41598-025-93254-4

Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2019). A survey of network-based intrusion detection data sets. *Computers & Security, 86*, 147-167. https://doi.org/10.1016/j.cose.2019.06.005

Saurabh, K., Sood, S., Kumar, P. A., Singh, U., Vyas, R., Vyas, O. P., & Khondoker, R. (2022). LBDMIDS: LSTM-based deep learning model for intrusion detection systems for IoT networks. In *2022 IEEE World AI IoT Congress (AIIoT)* (pp. 753-759). IEEE. https://doi.org/10.1109/AIIoT54504.2022.9817245

Shamsuzzaman, H. M., Mosleuzzaman, M., Mia, A., & Nandi, A. (2024). Cybersecurity risk mitigation in industrial control systems: Analysing physical, hybrid, and virtual test bed applications. *Academic Journal on Artificial Intelligence, Machine Learning, Data Science and Management Information Systems, 1*(1), 19-39. https://doi.org/10.69593/ajaimldsmis.v1i01.123

Socher, R., Ganjoo, M., Manning, C. D., & Ng, A. Y. (2013). Zero-shot learning through cross-modal transfer. In *Advances in Neural Information Processing Systems 26* (pp. 935-943). Curran Associates, Inc.

Srivastava, A., Sanghavi, P., Parmar, V., & Rani, S. (2024). Zero-shot learning in cybersecurity: A paradigm shift in attack and defence strategies. In M. Singh, V. Tyagi, P. K. Gupta, J. Flusser, T. Ören, A. R. Cherif, & R. Tomar (Eds.), *Advances in computing and data sciences: 8th International Conference, ICACDS 2024, Vélizy, France, May 9-10, 2024, revised selected papers* (pp. 138-149). Springer. https://doi.org/10.1007/978-3-031-70906-7_13

Sufi, F. (2024). Generative pre-trained transformer (GPT) in research: A systematic review on data augmentation. *Information, 15*(2), Article 99. https://doi.org/10.3390/info15020099

Tariq, N., Alsirhani, A., Humayun, M., Alserhani, F., & Shaheen, M. (2024). A fog-edge-enabled intrusion detection system for smart grids. *Journal of Cloud Computing, 13*(1), Article 43. https://doi.org/10.1186/s13677-024-00609-9

Yadav, N., Pande, S., Khamparia, A., & Gupta, D. (2022). Intrusion detection system on IoT with 5G network using deep learning. *Wireless Communications and Mobile Computing, 2022*, Article 9304689. https://doi.org/10.1155/2022/9304689

Yu, M., & Zhang, X. (2023). AlertInsight: Mining multiple correlation for alert reduction. *Computer Systems Science and Engineering, 46*(2), 2447-2469. https://doi.org/10.32604/csse.2023.037506

Yucel, M. K., Cinbis, R. G., & Duygulu, P. (2020). A deep dive into adversarial robustness in zero-shot learning. In A. Bartoli & A. Fusiello (Eds.), *Computer vision ECCV 2020 workshops* (pp. 3-21). Springer. https://doi.org/10.1007/978-3-030-66415-2_1

Zhang, A., Zhao, Y., Zhou, C., & Zhang, T. (2025). ResACAG: A graph neural network-based intrusion detection. *Computers & Electrical Engineering, 122*, Article 109956. https://doi.org/10.1016/j.compeleceng.2024.109956

Zhang, T., Gao, L., Li, X., & Gao, Y. (2025). DZAD: Diffusion-based zero-shot anomaly detection. *Proceedings of the AAAI Conference on Artificial Intelligence, 39*(10), 10131-10138. https://doi.org/10.1609/aaai.v39i10.33099