AGENTIC AND MULTIMODAL LLM-BASED CYBER THREAT DETECTION IN ELECTRIC VEHICLE CHARGING SYSTEMS

by

Ritesh Revansiddappa Honnalli

Master of Science in Software Engineering
Department of Computer and Information Science
University of Michigan–Dearborn

Thesis Committee:

Dr. Junaid Farooq, Ph.D., Assistant Professor (Thesis Advisor)
Dr. Ang Li, Ph.D., Assistant Professor
Dr. Xiao Zhang, Ph.D., Assistant Professor

A thesis presented in partial fulfillment of the requirements for the degree of Master of Science in Software Engineering, administered through the College of Engineering and Computer Science, University of Michigan–Dearborn, Winter 2026.

Ritesh Revansiddappa Honnalli

riteshrh@umich.edu

ORCID ID: 0009-0002-8037-2623

© Ritesh Revansiddappa Honnalli, 2026

Dedication

This thesis is dedicated to my family, whose constant support, patience, and encouragement made this journey possible. Their belief in my abilities provided strength during challenges and motivation throughout this academic endeavor.

I am deeply grateful to my advisor, Junaid Farooq, for his guidance, insight, and continued support, which played a vital role in shaping my research and intellectual development.

List of Figures

1.1 High-level architecture of the electric vehicle charging ecosystem showing power flow, communication interfaces, backend coordination, and representative cyber attack exposure points. . . . 2
1.2 High-level overview of the two-part LLM-based cybersecurity framework proposed in this thesis. . . . 5

3.1 Overview of the electric vehicle charging infrastructure, highlighting communication interfaces, backend services, and potential cybersecurity attack surfaces. . . . 12
3.2 Key components and interaction pathways within the EV charging ecosystem, including vehicles, charging stations, cloud services, and grid interfaces. . . . 13
3.3 Architecture of the LLM-powered agent for anomaly detection, illustrating the integration of EV charging input data, retrieval-augmented contextual knowledge, and reasoning-driven threat assessment. . . . 14
3.4 Workflow for generating anomalous EV charging sessions using real-world data and LLM-guided attack scenario synthesis. . . . 17
3.5 Architecture of the LSTM autoencoder used as a baseline model for anomaly detection. . . . 18
3.6 Effect of temperature settings on anomaly detection accuracy across different LLM models. . . . 20
3.7 Impact of retrieval-augmented generation on anomaly detection accuracy across different LLM models, demonstrating the benefit of contextual grounding. . . . 20
3.8 Detection accuracy of different LLM models across multiple EV charging attack categories. . . . 21

4.1 Cumulative detection accuracy as a function of prediction window size expressed as a percentage of session length. . . . 29

4.2 Mean detection speed as a function of prediction window size. Higher values indicate earlier anomaly detection relative to total session duration. 30
4.3 Speed–accuracy utility as a function of prediction window size. The metric represents the product of cumulative detection accuracy and detection speed. 31
4.4 Illustration of the trade-off between cumulative detection accuracy and detection speed as prediction window size increases. 32
4.5 Average cumulative energy delivered across anomalous sessions with and without early detection. The shaded region represents average energy loss prevented due to timely intervention. 32

List of Tables

3.1 Example session-level records from the Level 3 EV charging dataset. 16
3.2 Example minute-level measurement records for a charging session. 16
3.3 Performance metrics for baseline and LLM-based anomaly detection models. . 19

Contents

List of Figures **I**

List of Tables **II**

Abstract **VI**

1 Introduction **1**

1.1 Introduction 1
1.2 Motivation and Challenges in EV Charging Cybersecurity 3
1.3 Limitations of Existing Detection Approaches 3
1.4 Overview of the Proposed LLM-Based Framework 4
1.5 Thesis Contributions 5
1.6 Thesis Organization 6

2 Background and Related Work **7**

2.1 Electric Vehicle Charging Systems Overview 7
2.2 Cyber Threats in EV Charging Infrastructure 8
2.3 Traditional Intrusion Detection Techniques 8
2.4 Machine Learning-Based Cybersecurity Approaches 9
2.5 Large Language Models and Agentic AI for Security 9
2.6 Multimodal and Sequential Threat Detection 10

3 Part 1: LLM-Powered Agentic Cybersecurity Framework for EV Charging Systems **11**

3.1 Introduction to Agentic Security Framework 11
3.2 Threat Model and System Assumptions 12
3.3 Attack Detection Agent for EV Charging System 13
3.3.1 Detection Workflow and Data Processing 14

3.3.2 Threat Classification and Agent Output . . . 14
3.3.3 Algorithmic Description . . . 14
3.4 Performance Analysis . . . 15
3.4.1 Data Source and Synthetic Attack Generation . . . 15
3.4.2 Experimental Setup . . . 17
3.5 Experimental Results . . . 18
3.5.1 Comparative Performance Across Models . . . 19
3.5.2 Impact of Temperature Settings on LLM Performance . . . 19
3.5.3 Effect of RAG on Detection Performance . . . 19
3.5.4 Accuracy Across Different Attack Types . . . 21
3.6 Discussion . . . 21
3.7 Conclusion . . . 22

4 Part 2: Multimodal LLM-Guided Sequential Detection of Cyber Threats in EV Charging Systems 23
4.1 Introduction to Multimodal Sequential Detection . . . 23
4.2 Motivation for Temporal and Multimodal Modeling . . . 24
4.3 System Model and Threat Landscape . . . 25
4.4 Proposed Multimodal Sequential Detection Framework . . . 26
4.4.1 Framework Overview . . . 26
4.4.2 Sequential SOC Forecasting Using LSTM . . . 26
4.4.3 Progressive Percentage-Based Windowing . . . 26
4.4.4 Multimodal LLM-Based Visual Classification . . . 27
4.4.5 Detection Speed Metric . . . 27
4.5 Algorithmic Description . . . 27
4.6 Experimental Setup . . . 28
4.7 Results and Discussion . . . 29
4.7.1 Cumulative Detection Accuracy vs. Window Size . . . 29
4.7.2 Detection Speed vs. Window Size . . . 30
4.7.3 Speed–Accuracy Utility Analysis . . . 30
4.7.4 Joint Visualization of Speed and Accuracy . . . 31
4.7.5 Operational Impact: Energy Loss Prevention . . . 31
4.8 Conclusion . . . 33

5 Conclusions and Future Work **34**
5.1 Summary of Contributions 34
5.2 Key Findings and Insights 35
5.3 Limitations of the Proposed Framework 35
5.4 Future Research Directions 36
5.5 Concluding Remarks 36

Abstract

Electric vehicle (EV) charging infrastructure is evolving into a highly networked cyber-physical ecosystem that integrates vehicles, electric vehicle supply equipment (EVSE), charge station management systems, and cloud-based services. This connectivity increases the attack surface and enables threats such as billing fraud, energy theft, denial-of-service, and protocol manipulation, while the scarcity of labeled attack data limits the effectiveness of conventional machine learning-based intrusion detection. This thesis develops a unified, LLM-driven cybersecurity framework for EV charging systems that progresses from an agentic detection foundation to a multimodal sequential detection extension. First, an LLM-powered agent augmented with retrieval-augmented generation (RAG) integrates domain knowledge with live charging-session telemetry to classify anomalous and malicious behavior with improved generalization and reduced false positives. Second, a multimodal detection pipeline combines sequential state-of-charge (SOC) forecasting with LLM-based visual reasoning by converting predicted versus observed SOC trajectories into plots that a vision-language model can interpret for zero-shot anomaly classification. A progressive windowing strategy enables early detection while quantifying the trade-off between timeliness and reliability. Experimental results on real-world and augmented datasets demonstrate that the proposed approach improves detection capability and operational impact, including earlier intervention to mitigate energy loss during malicious sessions.

Chapter 1

Introduction

1.1 Introduction

The global transition toward electrified transportation has positioned electric vehicles (EVs) as a central component of modern smart energy and mobility ecosystems [1]. Rapid growth in EV adoption has been accompanied by the large-scale deployment of electric vehicle charging infrastructure across residential, commercial, and public environments [2]. These charging systems are no longer isolated power delivery devices but are increasingly integrated with communication networks, cloud-based management platforms, and grid-level control systems. As a result, EV charging infrastructure has evolved into a complex cyber-physical system that combines power electronics, embedded control, communication protocols, and software-defined services.

While this connectivity enables advanced functionalities such as smart charging, load balancing, demand response, and remote monitoring, it also expands the system's attack surface [3]. EV charging stations exchange operational, billing, and authentication data with vehicles, backend management systems, cloud services, and grid operators over heterogeneous communication channels. These channels often rely on standardized protocols such as ISO 15118 and OCPP [4, 5] and may traverse public or semi-trusted networks [6, 7].

Figure 1.1 presents the high-level architecture of the electric vehicle charging ecosystem, illustrating physical power flow, communication pathways, backend coordination, and integration with utility operators. The figure also highlights how exposed interfaces and distributed control logic introduce multiple potential entry points for cyber adversaries.

Cyber attacks targeting these interconnected components can disrupt charging operations, degrade grid stability, compromise user privacy, and in extreme cases cause physical

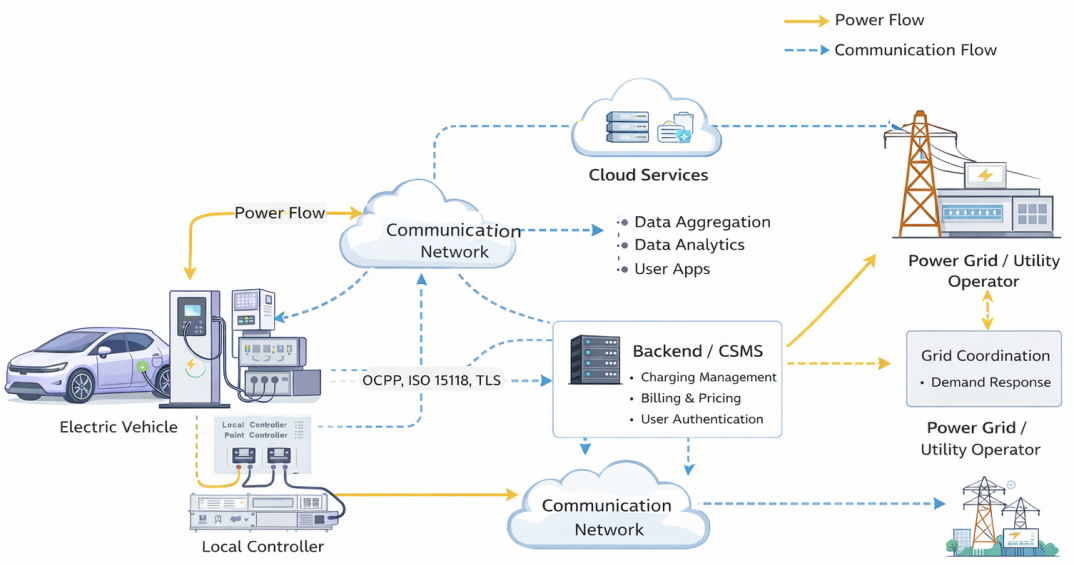


Figure 1.1: High-level architecture of the electric vehicle charging ecosystem showing power flow, communication interfaces, backend coordination, and representative cyber attack exposure points.

damage to charging equipment or connected vehicles [8, 9]. Because charging infrastructure operates at the intersection of cyber and physical domains, security failures may propagate beyond information systems into real-world energy systems.

Traditional cybersecurity mechanisms deployed in EV charging networks largely rely on rule-based intrusion detection systems, signature matching, or conventional machine learning classifiers trained on static feature representations [10]. Although these approaches have shown promise in controlled environments, they struggle to cope with the dynamic, evolving, and multimodal characteristics of real-world cyber threats. Attack patterns may evolve over time, span multiple communication layers, and exploit temporal dependencies that are difficult to capture using static or single-stage detection pipelines. These limitations motivate the development of adaptive, context-aware, and explainable security frameworks, which form the foundation of this thesis.

Recent advances in large language models (LLMs) have demonstrated remarkable capabilities in reasoning, contextual understanding, and decision-making across complex domains [11]. When combined with agentic architectures and multimodal data processing, LLMs offer a new paradigm for cybersecurity systems that can reason over heterogeneous data sources, adapt to evolving threat landscapes, and provide interpretable security decisions. This thesis investigates the use of LLM-powered agentic and multimodal frameworks to enhance the detection and mitigation of cyber threats in EV charging systems.

1.2 Motivation and Challenges in EV Charging Cybersecurity

EV charging systems operate at the intersection of transportation networks, power grids, and information technology infrastructures. This convergence introduces several unique cybersecurity challenges. First, EV charging stations communicate using multiple protocols and interfaces, including vehicle-to-charger communication, backend management protocols, and grid interaction signals. These interfaces generate diverse data streams such as network traffic, operational logs, sensor measurements, and visual indicators, each of which may carry evidence of malicious activity.

Second, cyber attacks targeting EV charging infrastructure often unfold sequentially rather than instantaneously. Adversaries may probe systems, exploit vulnerabilities gradually, and escalate privileges over time [12]. Detecting such attacks requires models that can capture temporal dependencies and evolving behavioral patterns rather than relying solely on instantaneous feature snapshots.

Third, the operational constraints of EV charging systems impose strict requirements on detection latency, reliability, and interpretability. False positives may interrupt legitimate charging sessions and degrade user trust, while false negatives may allow attacks to persist undetected. Moreover, security decisions must often be explainable to operators and regulators to support corrective actions and compliance requirements.

Finally, the scarcity of labeled attack data and the continuous emergence of novel threat vectors limit the effectiveness of supervised learning approaches that rely on fixed training distributions. These challenges motivate the need for intelligent security systems that can generalize beyond seen attacks, reason under uncertainty, and integrate information across time and modalities.

1.3 Limitations of Existing Detection Approaches

Conventional intrusion detection systems deployed in EV charging networks primarily employ rule-based engines or classical machine learning models trained on handcrafted features. While these methods are computationally efficient and relatively easy to deploy, they exhibit limited adaptability to unseen attacks and evolving system behaviors. Rule-based systems require frequent manual updates and are prone to high false positive rates in dynamic environments.

More recent machine learning-based approaches leverage supervised classifiers such as decision trees, support vector machines, and deep neural networks to improve detection accuracy [13, 14]. However, these models typically operate on static feature vectors extracted from individual time windows, thereby neglecting sequential dependencies and long-term contextual information. Furthermore, most existing solutions focus on a single data modality, such as network traffic, without exploiting complementary information available from physical sensors or system logs.

Another critical limitation lies in the lack of uncertainty awareness and interpretability. Deterministic classifiers provide point predictions without quantifying confidence, making it difficult to assess the reliability of detection outcomes in safety-critical applications [**?**]. In addition, black-box models offer limited insight into the reasoning process behind security decisions, hindering trust and operational adoption.

These limitations underscore the need for advanced detection frameworks that combine temporal modeling, multimodal data fusion, uncertainty quantification, and explainable reasoning.

1.4 Overview of the Proposed LLM-Based Framework

This thesis proposes a unified cybersecurity framework for EV charging systems based on LLM-powered agentic intelligence and multimodal sequential detection. The framework is developed in two tightly connected parts. A high-level overview of the proposed two-part framework and the information flow between the agentic reasoning layer and the multimodal sequential detection layer is shown in Figure 1.2.

In Part 1, an agentic LLM-based security architecture is introduced to serve as the foundational detection and decision-making layer. In this part, the LLM functions as an intelligent agent that orchestrates multiple subtasks, including data interpretation, threat reasoning, contextual retrieval, and response generation. The agent integrates structured cybersecurity knowledge with real-time system observations to identify anomalous behaviors and potential attacks. By leveraging retrieval-augmented generation and probabilistic reasoning, the agentic framework enhances adaptability, robustness, and explainability in threat detection.

Building upon this foundation, Part 2 extends the framework to a multimodal, sequential detection setting. This part incorporates heterogeneous data sources, including time-series operational data and visual representations of system states, to capture complex attack signatures that evolve over time. A multimodal LLM-guided pipeline is developed to analyze sequential windows of data, progressively refine threat assessments, and generate actionable

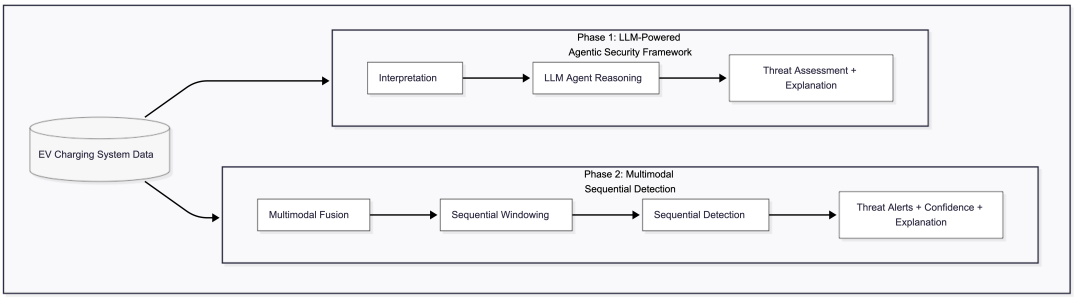


Figure 1.2: High-level overview of the two-part LLM-based cybersecurity framework proposed in this thesis.

security insights. This extension enables early detection of stealthy attacks and improves resilience against sophisticated adversarial strategies.

Together, these parts form a cohesive security solution that addresses both high-level reasoning and low-level sequential analysis within EV charging environments.

1.5 Thesis Contributions

The primary contributions of this thesis are summarized as follows:

- Development of an LLM-powered agentic cybersecurity framework for EV charging systems that integrates real-time charging telemetry with retrieval-augmented generation (RAG) to enable context-aware anomaly detection.
- Design and implementation of a multimodal sequential detection framework that combines LSTM-based state-of-charge (SOC) forecasting with vision-language model reasoning for visual anomaly classification.
- Formulation of a progressive percentage-based windowing mechanism to enable early-stage cyber threat detection while quantifying detection speed and reliability across charging sessions.
- Introduction of a unified evaluation methodology incorporating detection accuracy, detection speed, and a joint utility metric to analyze the trade-off between timely intervention and classification confidence.
- Comprehensive experimental validation using real-world EV charging data augmented with synthetic cyber attack scenarios, demonstrating improved detection performance and significant reduction in energy loss through early intervention.

1.6 Thesis Organization

This thesis is organized as follows: Chapter 2 reviews the background and related work on EV charging systems, cybersecurity threats, intrusion detection techniques, and recent advances in machine learning and LLM-based security approaches. Chapter 3 presents Part 1 of the proposed framework, detailing the LLM-powered agentic architecture, threat modeling assumptions, and experimental evaluation. Chapter 4 introduces Part 2, which extends the framework to multimodal and sequential threat detection using LLM-guided analysis. Finally, Chapter 5 concludes the thesis by summarizing key findings, discussing limitations, and outlining directions for future research.

Chapter 2

Background and Related Work

2.1 Electric Vehicle Charging Systems Overview

Electric vehicle charging systems form the backbone of modern electric mobility infrastructure and serve as the primary interface between electric vehicles and the power grid [1]. A typical EV charging ecosystem consists of electric vehicles, charging stations, backend management systems, and grid-side control entities [3]. Charging stations vary in power levels and functionalities, ranging from low-power residential chargers to high-power public fast-charging units. These systems incorporate power electronics, embedded controllers, communication modules, and software platforms to enable efficient and flexible charging operations.

Modern EV charging stations support advanced features such as smart charging, dynamic pricing, load balancing, and remote diagnostics. These capabilities are enabled through continuous data exchange between chargers, vehicles, and centralized management platforms. Communication protocols such as Open Charge Point Protocol (OCPP) and ISO 15118 facilitate information exchange related to charging status, authentication, billing, and control commands [4, 5]. While these protocols improve interoperability and operational efficiency, they also expand the attack surface of EV charging systems [3].

From a systems perspective, EV charging infrastructure represents a cyber-physical system where cyber components directly influence physical processes [15]. Malicious manipulation of control signals or data streams can lead to unsafe operating conditions, reduced charging performance, or physical damage to equipment. Consequently, understanding the architecture and operational characteristics of EV charging systems is essential for designing effective cybersecurity solutions.

2.2 Cyber Threats in EV Charging Infrastructure

The integration of communication and control technologies in EV charging systems exposes them to a wide range of cyber threats [16, 17]. Attack vectors may target charging stations, backend servers, communication channels, or even the vehicles themselves [18]. Common threats include denial-of-service attacks that disrupt charging availability, man-in-the-middle attacks that intercept or modify communication messages, and spoofing attacks that impersonate legitimate devices or users [4, 5].

More sophisticated adversaries may exploit protocol vulnerabilities, misconfigurations, or software flaws to gain persistent access to charging infrastructure [3]. Such attacks can enable unauthorized control of charging sessions, manipulation of billing information, or coordinated disruptions that impact grid stability [8]. In large-scale deployments, compromised charging stations can be leveraged as entry points for broader attacks on energy management systems or utility networks [12].

Cyber attacks in EV charging environments often exhibit temporal and contextual dependencies. Attackers may perform reconnaissance over extended periods, gradually escalate privileges, and adapt their strategies in response to system defenses [12]. These characteristics complicate detection and mitigation efforts, particularly when relying on static or rule-based security mechanisms.

2.3 Traditional Intrusion Detection Techniques

Traditional intrusion detection systems have been widely adopted in conventional information technology networks and industrial control systems [10]. These systems typically employ rule-based engines or signature matching techniques to identify known attack patterns. In EV charging environments, such approaches are often deployed at network gateways or backend servers to monitor traffic for suspicious activity [16].

While rule-based systems offer low computational overhead and straightforward implementation, they require continuous manual updates to remain effective against evolving threats [10]. They are also limited in their ability to detect previously unseen attacks or subtle deviations from normal behavior. In dynamic environments such as EV charging networks, these limitations can result in high false positive rates or missed detections.

Anomaly-based detection methods attempt to address some of these shortcomings by modeling normal system behavior and flagging deviations as potential threats. However, traditional anomaly detection techniques often rely on simplistic statistical assumptions and

struggle to distinguish benign operational variability from malicious activity. This challenge is particularly pronounced in EV charging systems, where load patterns and user behaviors can vary significantly over time [19].

2.4 Machine Learning-Based Cybersecurity Approaches

Machine learning has emerged as a powerful tool for enhancing intrusion detection by enabling data-driven modeling of complex system behaviors [10]. Supervised learning approaches leverage labeled datasets to train classifiers that distinguish between normal and malicious activity. Techniques such as decision trees, support vector machines, random forests, and deep neural networks have been applied to cybersecurity tasks across various domains.

In the context of EV charging systems, machine learning-based detectors have been used to analyze network traffic, operational logs, and sensor measurements [13, 14]. These models often outperform rule-based systems in terms of detection accuracy and adaptability. However, many existing approaches operate on fixed-length feature vectors extracted from individual time windows, thereby neglecting temporal dependencies and long-term contextual information [20].

Unsupervised and semi-supervised learning methods aim to reduce reliance on labeled attack data by learning representations of normal behavior [21]. While these methods improve generalization, they frequently suffer from limited interpretability and sensitivity to noise. Moreover, most machine learning-based solutions focus on a single data modality and do not exploit complementary information available across different system components.

Another important limitation of conventional machine learning models is their deterministic nature. Point predictions without associated uncertainty estimates can be problematic in safety-critical applications such as EV charging, where confidence in detection outcomes is essential for informed decision-making [**?**].

2.5 Large Language Models and Agentic AI for Security

Large language models have recently demonstrated strong capabilities in reasoning, contextual understanding, and knowledge integration across diverse domains [11, 22]. Unlike traditional machine learning models that operate on fixed feature spaces, LLMs process unstructured and semi-structured data and can incorporate prior knowledge through pretraining on large collections of text.

In cybersecurity applications, LLMs have been explored for tasks such as log analysis, incident response automation, vulnerability assessment, and threat intelligence synthesis [23, 24]. Their ability to reason over complex narratives and generate interpretable explanations makes them valuable for operational security settings. When combined with agentic architectures, LLMs can orchestrate multiple subtasks, interact with external tools, and adapt their behavior based on feedback [25].

Agentic AI frameworks enable LLMs to function as autonomous or semi-autonomous decision-making entities. In a security context, such agents can retrieve relevant knowledge, analyze observations, assess uncertainty, and recommend mitigation actions. This paradigm offers a promising alternative to static detection pipelines by supporting adaptability, explainability, and continuous learning.

2.6 Multimodal and Sequential Threat Detection

Many cyber attacks in EV charging systems manifest across multiple data modalities and evolve over time [17]. Network traffic anomalies, changes in power consumption patterns, abnormal control signals, and visual indicators from system interfaces may collectively reveal malicious activity. Multimodal detection approaches aim to fuse these heterogeneous data sources to improve detection robustness and reduce ambiguity.

Sequential modeling techniques capture temporal dependencies by analyzing data streams over successive time windows. Methods such as recurrent neural networks and temporal convolutional networks have been applied to intrusion detection tasks to identify evolving attack behaviors [20]. However, these models often require extensive tuning and large labeled datasets, and they provide limited interpretability.

Recent research has begun exploring the use of LLM-guided multimodal reasoning for sequential threat detection [26]. By leveraging the contextual reasoning capabilities of LLMs, these approaches can integrate information across modalities and time, generate coherent threat narratives, and refine detection decisions as new evidence becomes available.

Collectively, the limitations of existing approaches and the emerging capabilities of LLM-based agentic and multimodal systems motivate the unified framework proposed in the subsequent chapters.

Chapter 3

Part 1: LLM-Powered Agentic Cybersecurity Framework for EV Charging Systems

3.1 Introduction to Agentic Security Framework

The increasing complexity and connectivity of electric vehicle (EV) charging systems necessitate cybersecurity solutions that extend beyond static detection mechanisms. Traditional pipelines that rely on fixed rules or single-stage machine learning models often lack the adaptability and contextual awareness required to detect sophisticated and evolving threats [10, 3]. This chapter presents an LLM-powered agentic cybersecurity framework designed for EV charging environments, with emphasis on context-aware anomaly detection under limited labeled attack data.

The proposed framework leverages large language models (LLMs) as intelligent agents capable of reasoning over heterogeneous security observations, integrating prior cybersecurity knowledge, and making detection decisions with improved generalization. Rather than functioning as a passive classifier, the LLM operates as an active decision-making entity that orchestrates multiple subtasks, including data interpretation, contextual retrieval, threat assessment, and explanation generation. This phase establishes the foundational reasoning layer of the overall thesis framework, which is extended in Chapter 4 to multimodal and sequential detection.

3.2 Threat Model and System Assumptions

The threat model reflects realistic adversarial capabilities within EV charging infrastructures. Attackers are assumed to be able to interact with exposed communication interfaces, exploit software vulnerabilities, and manipulate traffic exchanged among charging stations and backend services [5]. These adversaries may conduct both disruptive attacks and stealthy campaigns that blend malicious actions with normal traffic to evade detection.

Figure 3.1 provides an overview of the EV charging infrastructure considered in this work, illustrating key communication pathways and representative attack surfaces.

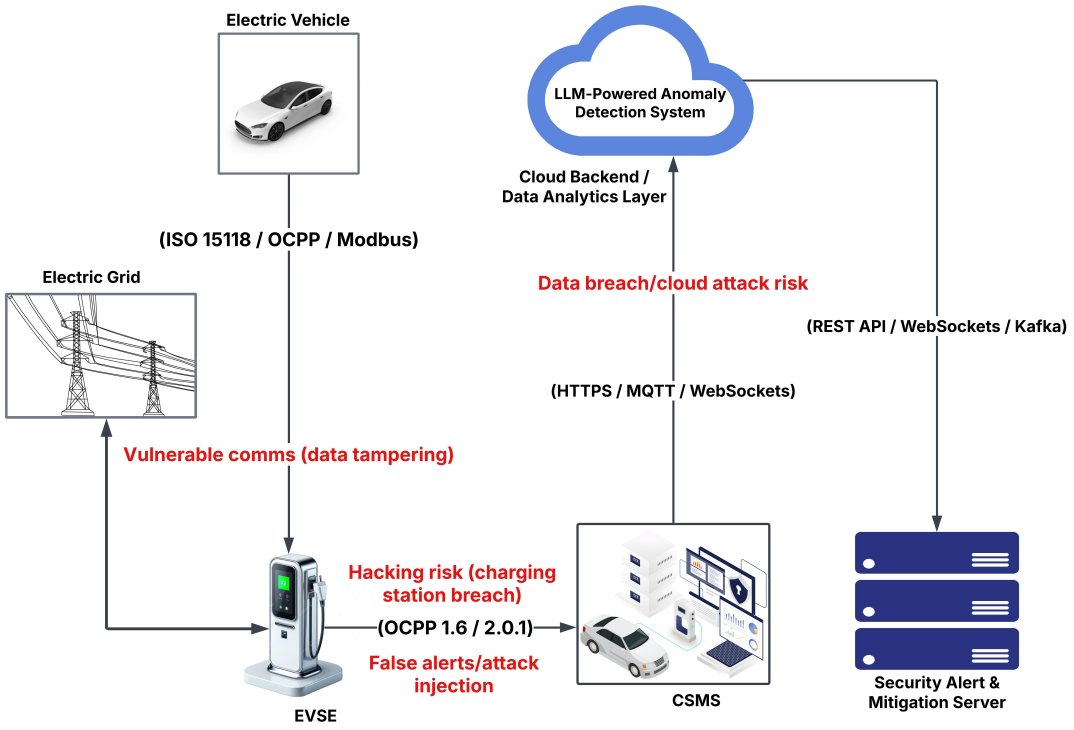


Figure 3.1: Overview of the electric vehicle charging infrastructure, highlighting communication interfaces, backend services, and potential cybersecurity attack surfaces.

The framework assumes that raw operational data from EV charging systems is available for analysis. This includes charging session telemetry and protocol-level metadata, such as power delivered, requested power, state-of-charge (SOC), session duration, and message sequencing behavior. Standardized protocols such as OCPP (versions 1.6 and 2.0.1), ISO 15118, and cloud-based service interfaces are typical within the ecosystem [4, 5]. Although encryption and authentication mechanisms may be present, operational metadata and behavioral patterns remain observable and can provide indicators of malicious activity [3].

The internal structure and interaction mechanisms within the EV charging ecosystem are further detailed in Figure 3.2, motivating the observability assumptions adopted in this chapter.

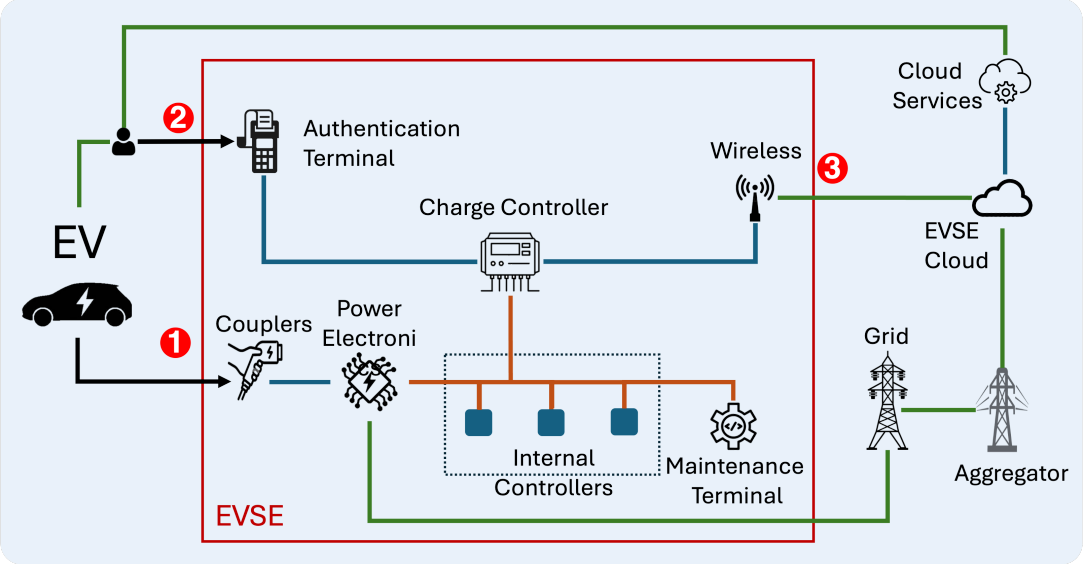


Figure 3.2: Key components and interaction pathways within the EV charging ecosystem, including vehicles, charging stations, cloud services, and grid interfaces.

The proposed approach does not rely on prior knowledge of specific attack signatures. Instead, it is designed to generalize across known and previously unseen threats through contextual reasoning grounded by retrieval-augmented generation (RAG) [27]. Detection outputs are intended to support timely mitigation under operational constraints, where both false positives and false negatives can be costly.

3.3 Attack Detection Agent for EV Charging System

This section presents the proposed agentic detection framework, which integrates real-time session monitoring with contextual retrieval and LLM-based reasoning. The agent continuously processes charging session data and classifies each session as normal or malicious. When malicious behavior is detected, the agent assigns the session to threat categories such as billing fraud, energy theft, or overcharging.

Figure 3.3 illustrates the high-level architecture of the LLM-powered agentic framework, highlighting the integration of data processing, contextual knowledge retrieval, and reasoning-driven classification.

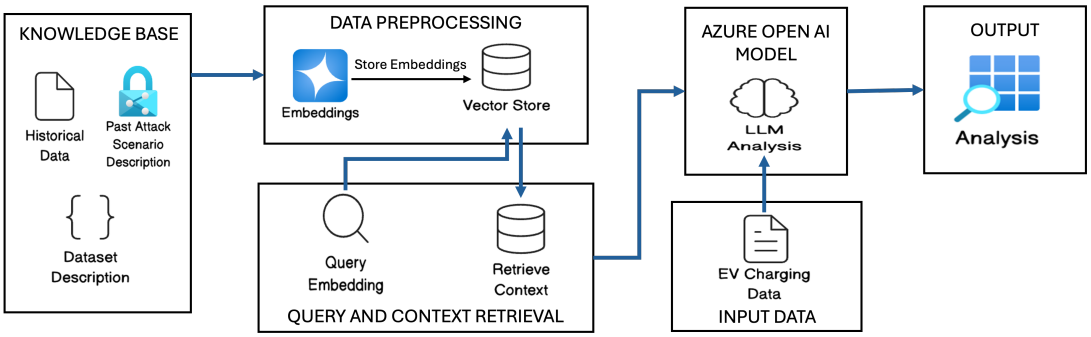


Figure 3.3: Architecture of the LLM-powered agent for anomaly detection, illustrating the integration of EV charging input data, retrieval-augmented contextual knowledge, and reasoning-driven threat assessment.

3.3.1 Detection Workflow and Data Processing

The detection process begins with ingestion of charging session data from the CSMS. Inputs include session identifiers, delivered power P, requested power P_{req}, SOC, and energy consumption. Preprocessing is applied to ensure consistency, and derived metrics such as ΔP and ΔSOC are computed to capture deviations indicative of malicious behavior.

Next, contextual information is retrieved from a domain-specific knowledge base using a RAG pipeline. Retrieved context includes descriptions of normal operating behavior, protocol expectations, and representative attack patterns. The enriched session data and retrieved context are then encoded into a structured format (for example JSON) and provided to the LLM for reasoning-based classification.

3.3.2 Threat Classification and Agent Output

The LLM-based engine classifies sessions using prompt-driven analysis grounded in retrieved context. Sessions classified as malicious are assigned threat categories such as billing fraud, energy theft, or overcharging. In addition to a categorical decision, the agent produces an explanation summarizing the evidence supporting the classification. This explanation is intended for operator interpretability and for audit and forensic workflows.

3.3.3 Algorithmic Description

Algorithm 1 summarizes the high-level workflow of the proposed attack detection agent.

Algorithm 1: Attack Detection Agent

Input: Real-time EV charging session data from CSMS
Output: Session classification and threat category
Initialization: Load session data and preprocess (clean, compute Power Deviation, SOC Change Rate);
while *there are sessions to process* **do**
    Retrieve contextual knowledge using RAG;
    Convert session data and context to JSON format;
    Analyze the session data using LLM;
    if *Session is classified as malicious* **then**
        Classify the session as Billing Fraud, Energy Theft, or Overcharging;
        Trigger security response: block session, notify administrators, and log forensics;
    else
        Mark session as Normal;
    end
    Save classification results and update dataset;
end
return Classified session and threat category;

3.4 Performance Analysis

This section describes the dataset, the synthetic anomaly generation process, the experimental setup, and the comparative evaluation against a baseline sequence model. The objective is to evaluate the accuracy, reliability, and adaptability of the proposed LLM-based approach in detecting anomalies in EV charging sessions.

3.4.1 Data Source and Synthetic Attack Generation

The dataset used in this study originates from a Level 3 EV charging station in southwestern Switzerland and contains minute-resolution charging session records spanning April 12, 2022 to July 4, 2023 [28]. The dataset includes 64,277 charging records from 1,878 sessions across two CCS plugs.

The dataset is organized into two hierarchical components:

- **Session-level metadata**, where each row corresponds to a complete charging session and includes attributes such as session identifier, plug identifier, start and end timestamps, and total delivered energy.

- **Minute-level measurement data**, where each row corresponds to a time-stamped measurement within a session and includes operational telemetry such as state-of-charge (SOC), delivered power, requested power, and cumulative energy.

Each session is uniquely identified by a session ID that links session-level metadata with minute-resolution measurements. This hierarchical structure enables both session-level anomaly classification (Part 1) and sequential time-series forecasting (Part 2).

Table 3.1: Example session-level records from the Level 3 EV charging dataset.

Session ID	Plug ID	Start Time	End Time	Total Energy (Wh)
1001	CCS1	2022-04-12 08:15	2022-04-12 09:02	18450
1002	CCS2	2022-04-12 10:21	2022-04-12 11:05	16230
1003	CCS1	2022-04-12 13:48	2022-04-12 14:33	17890

Table 3.2: Example minute-level measurement records for a charging session.

Session ID	Timestamp	SOC (%)	Delivered Power (kW)	Requested Power (kW)
1001	08:15	32	42.1	45.0
1001	08:16	33	43.0	45.0
1001	08:17	34	42.8	45.0

Since real-world cybersecurity anomalies are rare, the dataset was augmented with synthetically generated anomalous sessions to simulate realistic attack scenarios. The anomaly generation process followed three steps: (i) defining attack prompts for energy theft, overcharging, and billing fraud; (ii) generating anomalous charging sessions using an LLM-driven synthesis process; and (iii) integrating anomalies into the real dataset while preserving statistical integrity. Figure 3.4 illustrates the anomaly generation and integration workflow.

A total of 251 synthetic attack sessions comprising 7,986 anomalous entries were generated and integrated into the dataset. This augmentation improves coverage of diverse attack patterns and enables evaluation under threat scenarios that are otherwise underrepresented in real-world data.

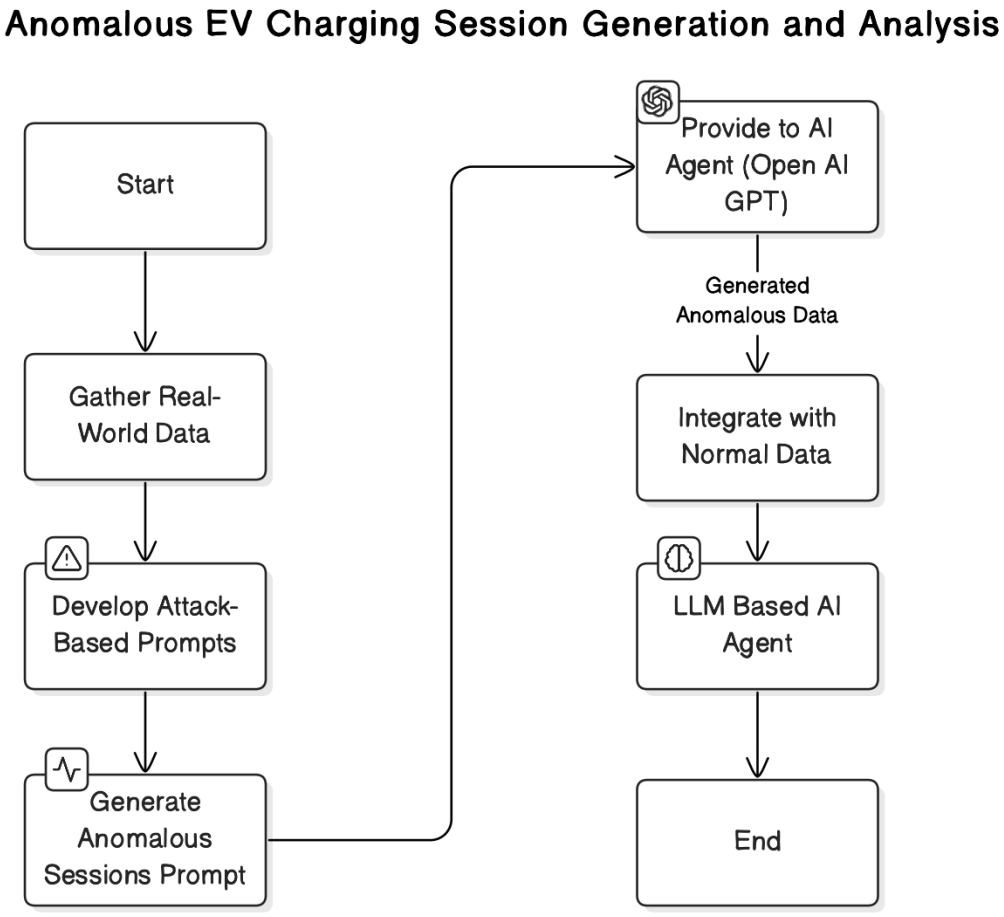


Figure 3.4: Workflow for generating anomalous EV charging sessions using real-world data and LLM-guided attack scenario synthesis.

3.4.2 Experimental Setup

The experimental evaluation compares the proposed LLM-based agent against a baseline LSTM autoencoder anomaly detector. Evaluation metrics include accuracy, precision, recall, and F1-score. In addition, sensitivity analyses are conducted for LLM temperature settings and for RAG integration.

RAG Retrieval Mechanism

Context retrieval uses embedding similarity to select the most relevant knowledge items [29]. Given a query embedding q representing the session prompt and a stored embedding v_i from

the knowledge base, cosine similarity is computed as

$$C(q, v_i) = \frac{q \cdot v_i}{|q||v_i|}. \tag{3.1}$$

Retrieved context is injected into the LLM prompt to improve grounding and reduce false alarms.

Baseline LSTM Autoencoder

The baseline LSTM autoencoder models sequential dependencies in charging session data. Feature selection includes power delivered, requested power, SOC, and energy. MinMax scaling normalizes features to [0, 1], and a sliding window of 50 time steps forms input sequences. The encoder compresses sequences into a latent representation and the decoder reconstructs them. Anomalies are detected by reconstruction error, where sequences exceeding a threshold (set as the 95th percentile of errors on normal sessions) are flagged as anomalous.

Figure 3.5 shows the baseline LSTM autoencoder architecture.

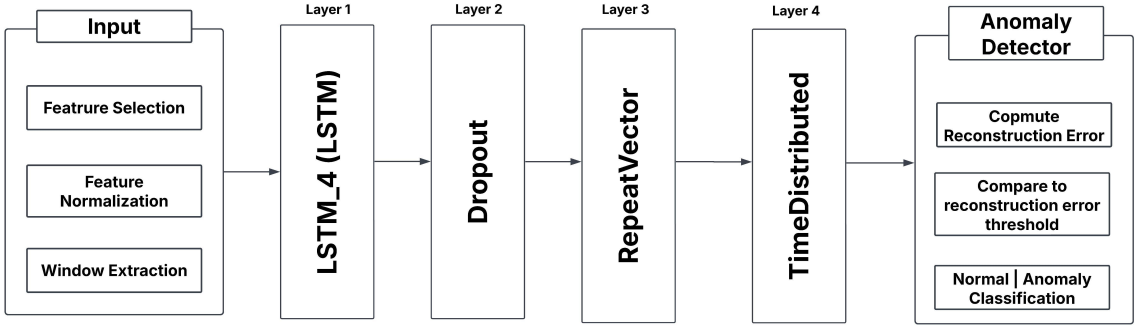


Figure 3.5: Architecture of the LSTM autoencoder used as a baseline model for anomaly detection.

The reconstruction error is quantified using RMSE:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{x}_t)^2}, \tag{3.2}$$

where x_t and $\hat{x}_t$ denote original and reconstructed sequence elements at time step t.

3.5 Experimental Results

This section reports comparative performance results, sensitivity to temperature, impact of RAG, and accuracy across different attack categories. The discussion emphasizes interpretabil-

ity and operational relevance, in addition to aggregate metrics.

3.5.1 Comparative Performance Across Models

Table 3.3 summarizes performance for the baseline LSTM autoencoder and three LLM variants. The LSTM autoencoder exhibits relatively high precision (0.80) but low recall (0.50), indicating a conservative detector that misses many anomalous sessions. Among the LLM-based models, GPT-4o-Mini achieves the highest overall performance with accuracy 0.77 and recall 0.77, suggesting stronger sensitivity to anomalous behavior while maintaining competitive precision.

Model	Accuracy	Precision	Recall	F1-Score
LSTM (Baseline)	0.50	0.80	0.50	0.62
GPT-4	0.57	0.68	0.57	0.61
GPT-4o	0.61	0.68	0.61	0.64
GPT-4o-Mini	0.77	0.66	0.77	0.70

Table 3.3: Performance metrics for baseline and LLM-based anomaly detection models.

These results support the central hypothesis of this phase: LLM-based reasoning, particularly when grounded with contextual retrieval, can improve detection of subtle and evolving threats compared to sequence-only baselines trained on limited labeled anomaly data.

3.5.2 Impact of Temperature Settings on LLM Performance

LLM temperature influences the diversity and confidence of model outputs. Figure 3.6 reports detection accuracy under four temperature settings (0.1, 0.5, 0.7, and 1.0). Lower temperatures tend to produce conservative outputs that can reduce recall, whereas higher temperatures can increase false positives. The default temperature setting (0.7) provides a better balance across evaluated LLM variants.

3.5.3 Effect of RAG on Detection Performance

Figure 3.7 evaluates detection accuracy with and without retrieval-augmented generation. Incorporating RAG improves performance across all models, indicating that contextual grounding enables the LLM to better discriminate between legitimate operational variability and malicious activity. For example, GPT-4o-Mini accuracy increases from 0.65 to 0.85 when RAG is enabled, representing an improvement of approximately 20%.

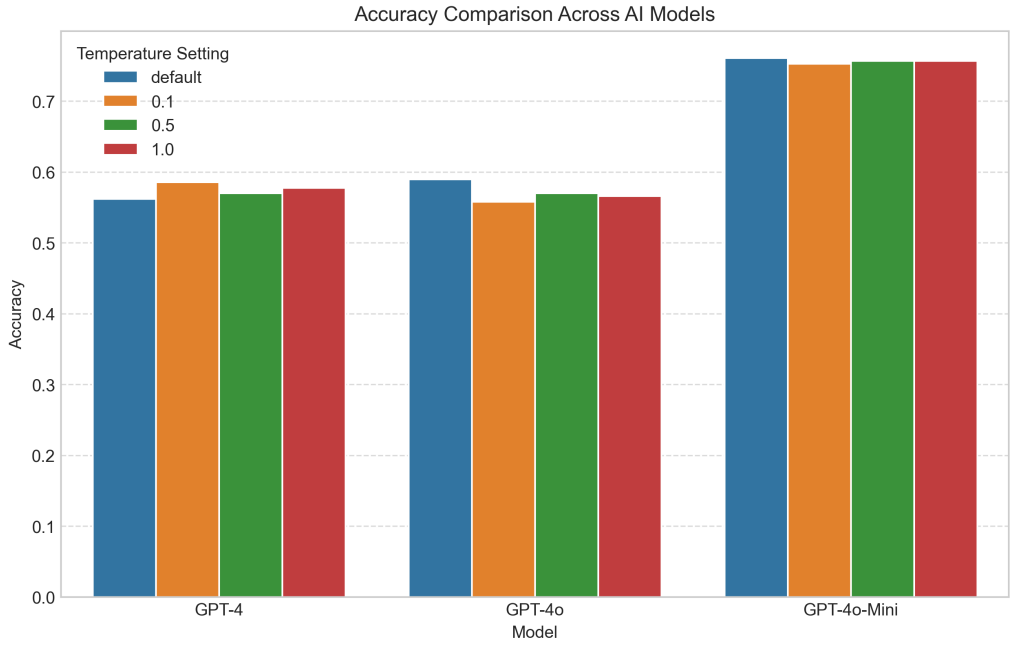


Figure 3.6: Effect of temperature settings on anomaly detection accuracy across different LLM models.

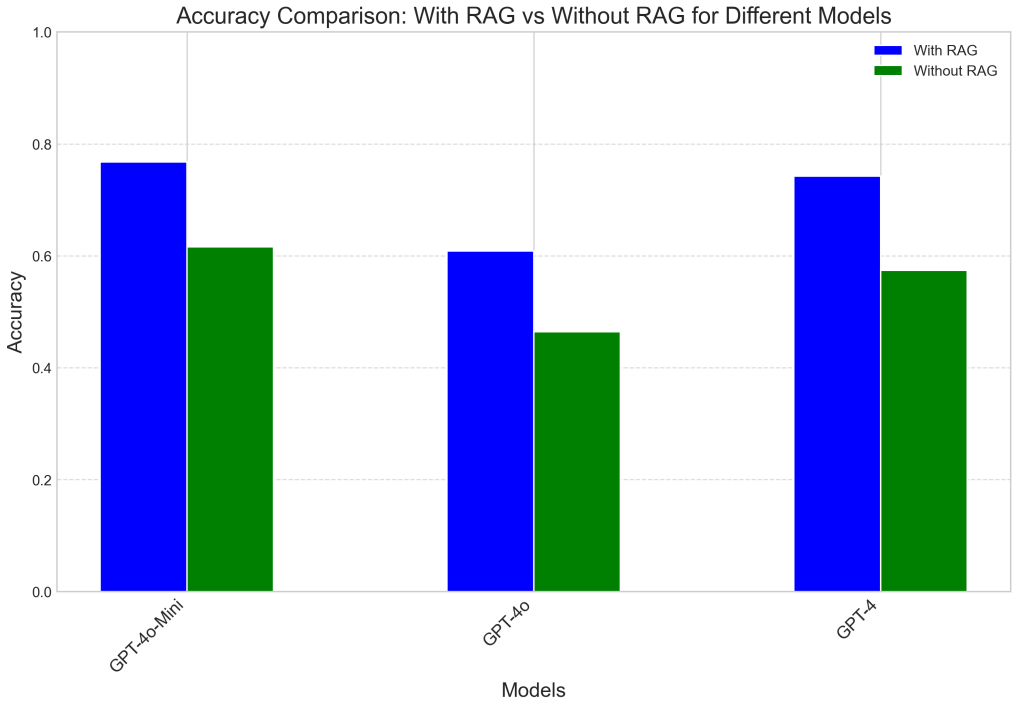


Figure 3.7: Impact of retrieval-augmented generation on anomaly detection accuracy across different LLM models, demonstrating the benefit of contextual grounding.

3.5.4 Accuracy Across Different Attack Types

To analyze robustness across threat categories, Figure 3.8 reports accuracy for normal sessions, billing fraud, and energy theft. GPT-4o provides more balanced performance across categories, while GPT-4o-Mini achieves higher accuracy for energy theft detection, indicating stronger sensitivity to power manipulation patterns. However, this strength can coincide with reduced accuracy on other categories, suggesting that model selection should consider the threat priorities of a deployment environment.

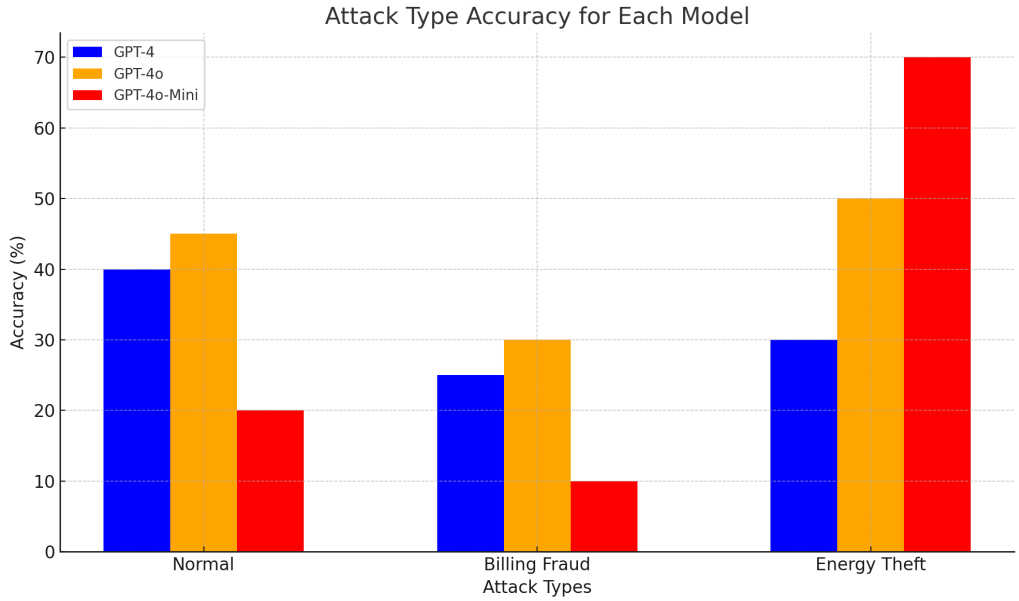


Figure 3.8: Detection accuracy of different LLM models across multiple EV charging attack categories.

3.6 Discussion

This phase demonstrates that an LLM-powered agentic framework, particularly when grounded with retrieval-augmented contextual knowledge, can improve anomaly detection capability in EV charging systems under limited labeled attack data. Compared to an LSTM autoencoder baseline, the LLM-based approach improves sensitivity to stealthy anomalies and enables explainable outputs that can support human-in-the-loop validation.

The temperature sensitivity results highlight the importance of inference configuration in operational deployment, and the RAG ablation demonstrates that contextual grounding is a key factor in improving reliability and reducing false alarms. Attack-type analysis further

suggests that different LLM variants may exhibit different strengths depending on the threat category, motivating adaptive or hybrid strategies in practical deployments.

Limitations of this phase include computational overhead and reliance on the quality and coverage of the knowledge base used for retrieval. These considerations motivate the development of complementary mechanisms for temporal aggregation and multimodal analysis, addressed in Chapter 4.

3.7 Conclusion

This chapter presented the Part 1 foundation of the thesis: an LLM-powered agentic cybersecurity framework for EV charging systems that integrates real-time session data with retrieval-augmented contextual knowledge. A comprehensive evaluation demonstrated improved detection performance relative to a sequence-only LSTM baseline, including measurable gains from RAG integration and informative sensitivity trends with respect to LLM temperature settings. This establishes the reasoning-centered detection foundation that is extended in the next phase to multimodal and sequential detection settings.

Chapter 4

Part 2: Multimodal LLM-Guided Sequential Detection of Cyber Threats in EV Charging Systems

4.1 Introduction to Multimodal Sequential Detection

Part 1 established an agentic LLM-based anomaly detection framework that reasons over structured charging session telemetry and contextual cybersecurity knowledge to classify sessions as normal or malicious. While effective for session-level classification, many cyber threats in EV charging systems evolve progressively during a charging session and may not be reliably detectable from a single static snapshot of data [3, 8].

Attacks such as denial-of-service behavior, overcharging manipulation, energy theft, and billing fraud often manifest as temporal deviations in the charging trajectory and protocol-level behavior [5, 4]. Early indicators may appear in the form of subtle irregularities in state-of-charge (SOC) evolution, power deviations, or unexpected shifts in charging rate before the full impact of the attack becomes evident.

To address this limitation, this chapter introduces a multimodal, sequential detection framework that integrates time-series forecasting with multimodal LLM-based visual reasoning. The key idea is to use a sequence prediction model to forecast the evolution of SOC and convert prediction deviations into a visual representation that can be interpreted by a multimodal LLM. This design combines statistical temporal modeling with high-level reasoning, enabling early-stage detection while preserving interpretability and explainability. SOC trajectory modeling has been widely studied for battery monitoring and energy management

systems [30].

This part builds directly upon the agentic reasoning architecture developed in Part 1. Whereas Part 1 focused on structured telemetry and contextual reasoning at the session level, Part 2 explicitly models temporal dynamics and progressively accumulates evidence during a charging session to improve detection timeliness and reliability.

4.2 Motivation for Temporal and Multimodal Modeling

Cyber attacks in EV charging infrastructure rarely occur as isolated events. Instead, adversaries frequently execute manipulations gradually, including progressive energy diversion, incremental parameter tampering, or communication-level disruptions. Detecting such behavior requires models that exploit temporal dependencies rather than relying solely on instantaneous features.

EV charging ecosystems generate heterogeneous information streams, including:

- Time-series telemetry such as SOC, power delivered, and requested power
- Derived metrics such as power deviation and SOC rate
- Protocol-level communication traces
- Operational and billing context

Each modality provides partial evidence. Sequential models can detect deviations in time-series behavior but often provide limited interpretability and may require careful threshold tuning [14, 20]. In contrast, multimodal large language models are capable of reasoning over structured prompts and visual artifacts, providing explanation-oriented classification [26].

The proposed framework therefore adopts a hybrid strategy:

1. A sequential forecaster models expected SOC evolution.
2. Prediction errors are rendered as visual plots comparing predicted and actual SOC trajectories.
3. A multimodal LLM performs visual reasoning over these plots to classify anomalies.

This converts a purely numerical anomaly detection task into a structured visual reasoning problem, leveraging the strengths of multimodal LLMs without requiring them to directly model raw time-series data.

4.3 System Model and Threat Landscape

The EV charging ecosystem considered in this chapter follows the system model introduced in Chapter 3. The infrastructure includes EVSE devices, backend charge station management systems, and cloud-based monitoring services. During a charging session, operational parameters are exchanged between the EV and EVSE, including SOC, requested power, and delivered power.

The following derived metrics are used as features or interpretation context:

$$\Delta P = |P_{\text{set}} - P|, \tag{4.1}$$

$$\Delta SOC = \frac{SOC_i - SOC_{i-1}}{t_i - t_{i-1}}, \tag{4.2}$$

$$E_{\text{session}} = \sum_{i=1}^{n} E_i. \tag{4.3}$$

These metrics provide insight into abnormal charging dynamics. For example:

- Sustained power deviation may indicate manipulation or interference.
- Irregular SOC rate changes may reflect energy theft or tampering.
- Unexpected energy accumulation may signal billing anomalies.

The anomaly classes considered in this part include:

- normal
- denial-of-service (dos)
- billing_fraud
- overcharging
- energy_theft

The objective is to detect such anomalies as early as possible during a charging session while maintaining high classification accuracy.

4.4 Proposed Multimodal Sequential Detection Framework

4.4.1 Framework Overview

The proposed detection pipeline consists of three primary stages:

1. Sequential SOC forecasting using an LSTM model.
2. Visual generation of predicted versus actual SOC trajectories.
3. Multimodal LLM-based classification of deviation patterns.

At each stage of a charging session, the system evaluates whether the observed SOC trajectory deviates from expected behavior. If an anomaly is identified, the session is flagged and further analysis is halted to minimize latency.

4.4.2 Sequential SOC Forecasting Using LSTM

Let a charging session be represented as a time-indexed sequence of telemetry measurements. At time step t, the LSTM model receives historical features up to t and predicts SOC values for a future horizon.

Input features may include:

- SOC history
- Delivered power
- Requested power
- Derived metrics such as ΔP

The LSTM is trained using supervised learning on normal and synthetically augmented sessions. Under normal conditions, predicted SOC closely tracks actual SOC. Under anomalous behavior, divergence patterns become visible [31].

4.4.3 Progressive Percentage-Based Windowing

To enable early detection, the framework employs a progressive windowing strategy. Rather than evaluating a fixed-length window, the session is analyzed using windows defined as percentages of the total session length, for example:

$$10\%, 20\%, 30\%, 40\%, 50\%, 60\%.$$

Smaller windows allow earlier decisions but may reduce reliability due to limited context. Larger windows improve stability but delay detection. This strategy enables systematic evaluation of the trade-off between detection speed and accuracy.

4.4.4 Multimodal LLM-Based Visual Classification

For each evaluated window:

1. The predicted SOC trajectory is plotted against the true SOC.
2. The plot is provided to a multimodal LLM along with a structured classification prompt.
3. The LLM assigns one of the predefined anomaly labels.

The LLM reasons over visual features such as:

- Persistent divergence
- Sudden discontinuities
- Abnormal slope variations

Once an anomaly label other than normal is produced, the session is marked as anomalous.

4.4.5 Detection Speed Metric

To quantify timeliness, detection speed is defined as:

$$\text{Speed} = 1 - \left(\frac{t}{T}\right), \tag{4.4}$$

where T is the total session length and t is the time index at which the first anomaly is detected. A higher value indicates earlier detection relative to session duration.

4.5 Algorithmic Description

Algorithm 2 formalizes the progressive detection procedure.

Algorithm 2: Progressive Multimodal Sequential Anomaly Detection

Input: Preprocessed session dataset $\mathscr{D}$, trained LSTM model $\mathscr{M}$, percentage intervals P

Output: Anomaly classification and detection speed

foreach *percentage* $p \in P$ **do**
  foreach *session* $s \in \mathscr{D}$ **do**
    $T \leftarrow$ total records in s;
    $w \leftarrow \max(1, \lfloor p \cdot T \rfloor)$;
    if $T < 2w$ **then**
      Mark session as insufficient_data;
      continue;
    end
    for $t \leftarrow w$ **to** $T - w$ w **do**
      $X \leftarrow$ features from index 0 to t;
      $Y_{\text{pred}} \leftarrow \mathscr{M}(X)$;
      $Y_{\text{true}} \leftarrow$ actual SOC from t to $t + w$;
      Generate SOC comparison plot;
      Send plot to multimodal LLM and receive label c;
      if $c \neq$ *normal* **then**
        Record anomaly at index t;
        Compute $1 - \frac{t}{T}$;
        break;
      end
    end
  end
end

4.6 Experimental Setup

The framework is evaluated using the same real-world EV charging dataset described in Section 3.4.1, augmented with synthetically generated anomalous sessions to ensure representation of diverse attack types [28]. This maintains experimental consistency between the session-level agentic detection framework (Part-1) and the sequential multimodal extension (Part 2).

The LSTM model is trained using supervised learning on normal and synthetically augmented sessions. Under normal conditions, predicted SOC closely tracks actual SOC. Under anomalous behavior, divergence patterns become visible. Progressive windows ranging from 10% to 60% of session length are evaluated.

Performance is assessed using:

- Cumulative detection accuracy
- Detection speed
- Speed–accuracy utility, defined as the product of speed and accuracy

4.7 Results and Discussion

This section presents a detailed evaluation of the progressive multimodal sequential detection framework. The analysis focuses on cumulative detection accuracy, detection speed, and their joint utility across progressive window sizes ranging from 10% to 60% of session length.

4.7.1 Cumulative Detection Accuracy vs. Window Size

Figure 4.1 illustrates cumulative detection accuracy across all anomalous sessions as a function of prediction window size.

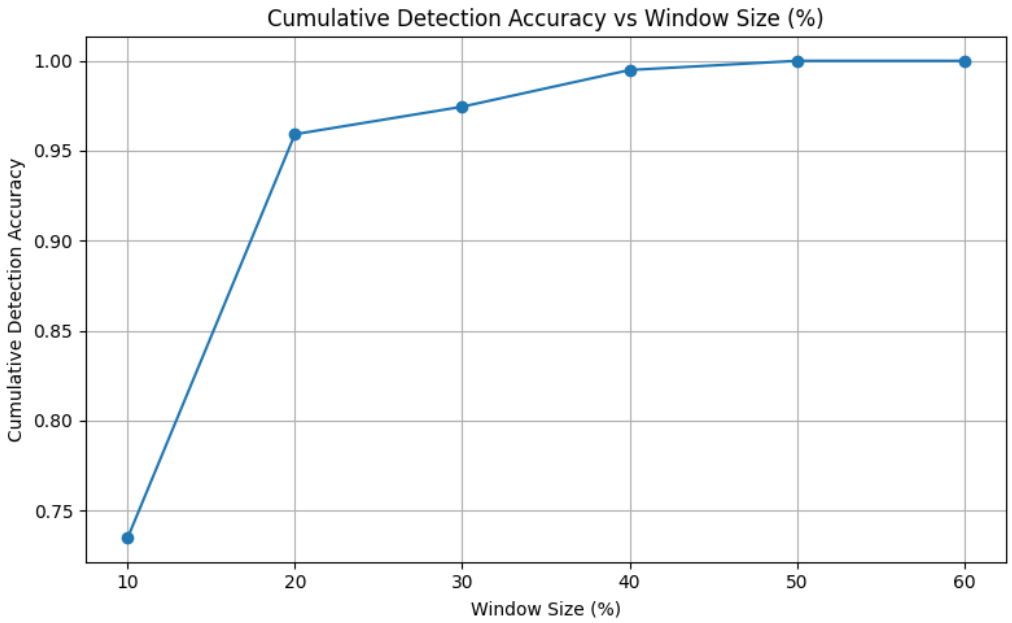


Figure 4.1: Cumulative detection accuracy as a function of prediction window size expressed as a percentage of session length.

At a 10% window size, detection accuracy is approximately 73%. Increasing the window to 20% results in a substantial rise to approximately 96%. Accuracy continues to improve gradually between 30% and 40%, reaching approximately 99%, and saturates near 100% at 50% to 60%.

These results indicate that most anomalous SOC trajectory deviations become distinguishable early in the charging session. However, extremely small windows may lack sufficient temporal context for stable classification.

4.7.2 Detection Speed vs. Window Size

Figure 4.2 shows the mean detection speed for sessions where anomalies were successfully identified.

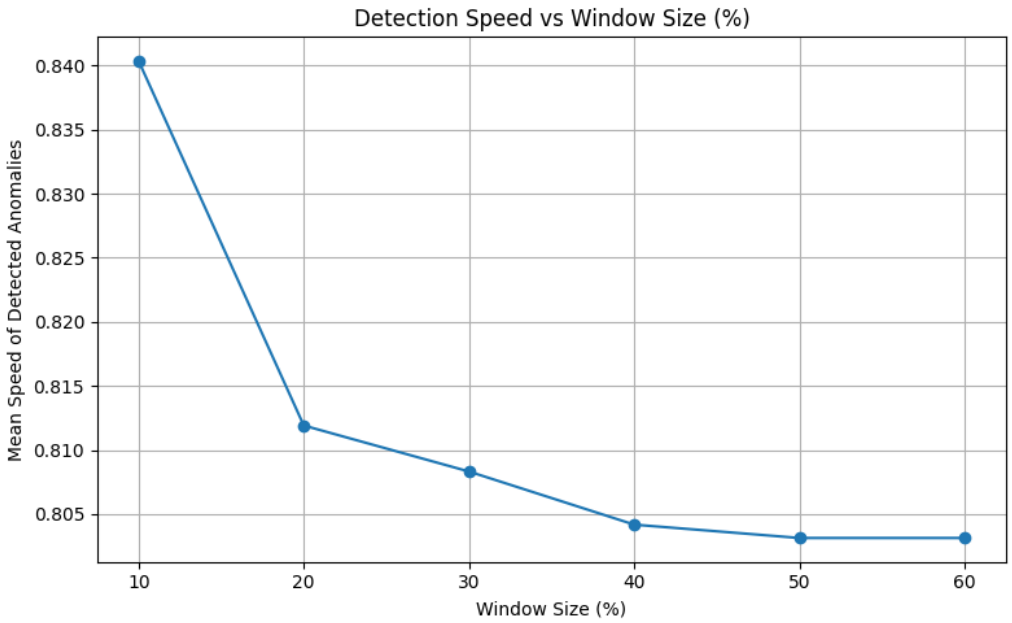


Figure 4.2: Mean detection speed as a function of prediction window size. Higher values indicate earlier anomaly detection relative to total session duration.

At 10%, detection speed is approximately 0.84, reflecting very early anomaly identification. As the window size increases to 20%, speed decreases to approximately 0.81. Further increases to 40% and beyond result in speed values near 0.80.

This decreasing trend is expected because larger windows require more observed session data before classification is attempted. Thus, improved contextual completeness comes at the cost of delayed detection.

4.7.3 Speed–Accuracy Utility Analysis

To jointly evaluate correctness and timeliness, a composite utility metric is defined as the product of cumulative detection accuracy and detection speed. Figure 4.3 presents this metric.

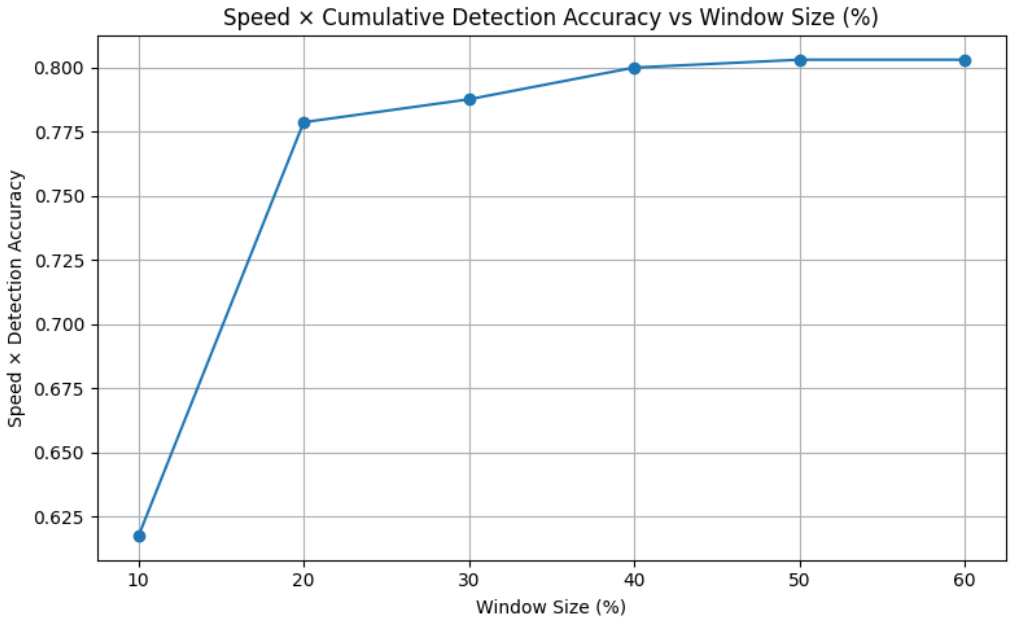


Figure 4.3: Speed–accuracy utility as a function of prediction window size. The metric represents the product of cumulative detection accuracy and detection speed.

At 10%, the utility is approximately 0.62 due to limited classification accuracy. At 20%, it increases sharply to approximately 0.78. The metric continues to increase gradually between 30% and 50%, stabilizing near 0.80.

These results indicate that intermediate window sizes between 20% and 40% provide the most favorable balance between early detection and reliable classification.

4.7.4 Joint Visualization of Speed and Accuracy

Figure 4.4 provides a combined visualization of cumulative detection accuracy and detection speed.

As window size increases, accuracy improves monotonically while detection speed decreases gradually. The intersection of these trends highlights the operational design region where both timeliness and reliability are acceptable.

4.7.5 Operational Impact: Energy Loss Prevention

Beyond classification metrics, the operational benefit of early detection is evaluated in terms of energy misuse mitigation.

Figure 4.5 compares cumulative delivered energy under two scenarios:

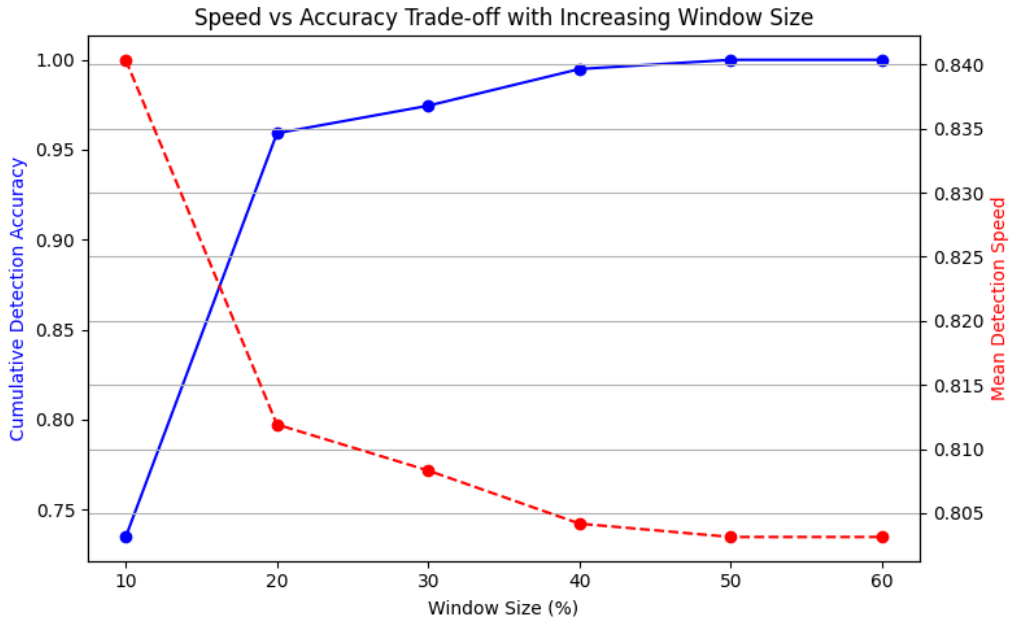


Figure 4.4: Illustration of the trade-off between cumulative detection accuracy and detection speed as prediction window size increases.

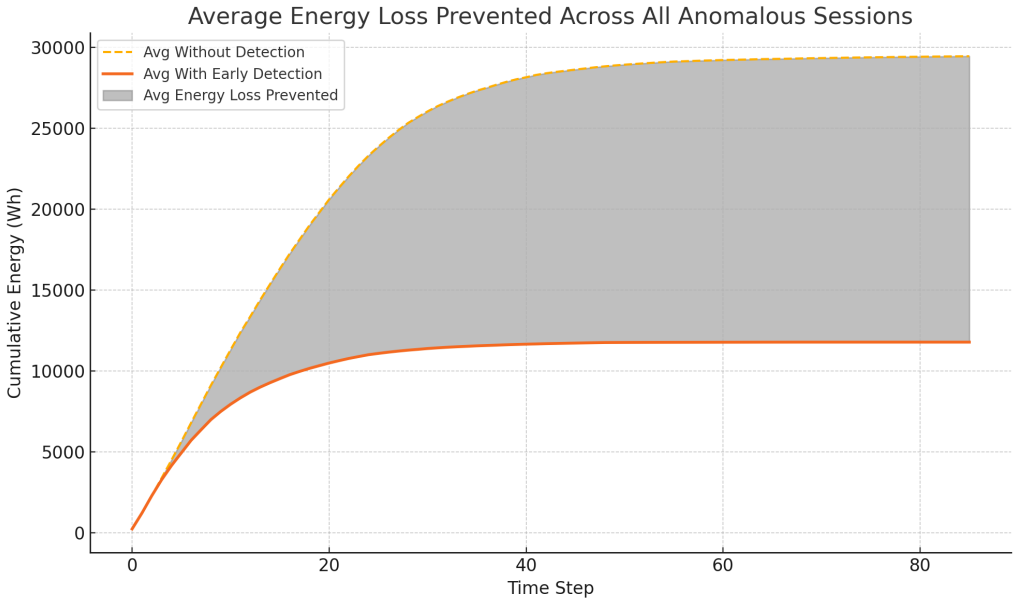


Figure 4.5: Average cumulative energy delivered across anomalous sessions with and without early detection. The shaded region represents average energy loss prevented due to timely intervention.

1. No detection mechanism

2. Early detection with intervention enabled

Without detection, anomalous sessions result in cumulative misuse approaching approximately 30,000 Wh. With early detection enabled, misuse is limited to approximately 12,000 Wh.

This corresponds to an average reduction in energy misuse of approximately 60%, demonstrating that early detection yields measurable operational and financial benefits.

4.8 Conclusion

This chapter introduced a multimodal sequential detection framework that integrates LSTM-based SOC forecasting with multimodal LLM-based visual reasoning. By employing progressive percentage-based windowing, the system enables early anomaly detection while maintaining high reliability. Vision-language models have demonstrated strong capabilities in interpreting visual patterns using natural language reasoning [32].

This part extends the agentic reasoning architecture developed in part 1 by incorporating explicit temporal modeling and progressive evidence accumulation. The combined framework forms a unified security architecture for EV charging systems, capable of both contextual reasoning and temporally grounded anomaly detection.

Chapter 5

Conclusions and Future Work

5.1 Summary of Contributions

This thesis investigated the use of large language models (LLMs) to enhance cybersecurity in electric vehicle (EV) charging systems. Motivated by the increasing connectivity and cyber-physical complexity of modern charging infrastructures, this work proposed a unified detection framework that combines agentic LLM reasoning with multimodal sequential analysis. The objective was to address key limitations of conventional intrusion detection and anomaly detection approaches, including reliance on labeled attack data, limited generalization to novel threats, and reduced interpretability under operational uncertainty.

The first contribution is an LLM-powered agentic anomaly detection framework for EV charging sessions. This framework integrates real-time session telemetry with retrieval-augmented generation (RAG) to ground inference in domain knowledge, enabling context-aware session classification and explainable threat reasoning. Unlike static signature-based systems, the Part 1 framework emphasizes generalization to unseen attack patterns using prompt-driven reasoning supported by external contextual retrieval.

The second contribution extends this agentic foundation to a multimodal sequential detection framework. Part 2 integrates time-series forecasting and multimodal LLM-based visual reasoning to enable early-stage detection of evolving threats. Specifically, an LSTM forecaster predicts state-of-charge (SOC) trajectories, predicted-versus-actual SOC plots are generated, and a multimodal LLM performs visual classification of anomalous behaviors. A progressive, percentage-based windowing strategy is introduced to quantify the trade-off between early detection and classification reliability.

Together, these contributions form a cohesive and extensible security architecture tailored

to EV charging systems, linking context-aware reasoning (Part 1) with temporally grounded and multimodally interpretable detection (Part 2).

5.2 Key Findings and Insights

The results from Part 1 demonstrate that LLM-powered reasoning can improve anomaly detection performance compared to traditional sequence-only baselines in settings where labeled cyberattack data are scarce. Incorporating retrieval-augmented contextual grounding demonstrates improved detection reliability in the evaluated dataset by providing domain-specific information that helps distinguish malicious behavior from benign operational variability. The evaluation further highlights that inference configuration affects performance, and that careful selection of LLM parameters can reduce false alarms while maintaining sensitivity to anomalies.

Part 2 shows that incorporating temporal structure can improve detection timeliness for attacks that unfold progressively during a charging session. The progressive windowing strategy reveals a clear trade-off between early detection and reliable classification: smaller windows enable earlier detection, while larger windows provide more context and improve accuracy. The results indicate that intermediate window sizes provide favorable speed–accuracy utility, supporting deployment configurations that balance early intervention with classification robustness. In addition, the energy impact analysis illustrates that early detection can yield meaningful operational benefit, including substantial reduction of energy loss across anomalous sessions through proactive intervention.

An important overall insight is that combining forecasting-based temporal modeling with multimodal LLM interpretation provides an interpretable pathway for detecting time-evolving threats. By converting time-series deviations into visual evidence, the framework enables multimodal LLMs to classify anomalies using reasoning over trajectory patterns, while preserving a clear audit trail for operators.

5.3 Limitations of the Proposed Framework

Despite promising results, the proposed framework has several limitations. First, the use of large models introduces computational overhead that may constrain deployment in latency-sensitive or resource-limited environments. Although the design can operate in centralized or edge-assisted settings, real-time deployment at scale requires careful optimization of inference frequency, windowing policy, and model selection.

Second, the reliability of LLM-based reasoning depends on prompt design and the coverage and quality of the contextual knowledge base. While RAG improves grounding, incomplete or outdated knowledge sources can affect robustness. In addition, the synthetic anomaly generation process strengthens evaluation coverage but may not fully capture the diversity and complexity of real adversarial behavior. Broader validation on large-scale field datasets remains limited by data availability and operational access constraints.

Third, this thesis focuses primarily on detection and analysis rather than automated response. Integrating closed-loop mitigation actions requires careful safety constraints and operational governance, since incorrect automated actions could disrupt charging availability or interfere with grid operations.

5.4 Future Research Directions

Several directions can extend this work. A key avenue is improving deployment feasibility through computational optimization, including lightweight model variants, caching strategies for retrieval, and adaptive inference policies that invoke the LLM only when lower-cost detectors indicate elevated risk. This would support deployment on resource-constrained edge devices or within EVSE controllers.

Another direction is to expand the multimodal pipeline to incorporate additional modalities such as grid-level measurements, charger firmware logs, vehicle telemetry, pricing signals, and user behavior patterns. Multi-station cooperative detection is also a promising extension, enabling coordinated threat detection across geographically distributed charging stations and improving resilience against synchronized attacks such as load-altering attacks.

Future work can also integrate detection with response by designing safe decision support mechanisms that recommend mitigation actions under human supervision. Finally, continual learning and adaptive knowledge updating can be incorporated to address concept drift and evolving attack patterns, including automated expansion of the retrieval knowledge base using newly observed incidents and forensic reports.

5.5 Concluding Remarks

This thesis demonstrated that LLMs, when embedded within agentic and multimodal sequential detection frameworks, provide a practical and interpretable approach to cybersecurity for EV charging systems. Part 1 established a context-aware agentic detection framework

grounded by retrieval-augmented knowledge, while Part 2 extended the system to sequential multimodal detection using forecasting-driven visual evidence and progressive early detection.

The combined results indicate that LLM-guided security frameworks can improve both detection robustness and interpretability, while also enabling earlier detection of evolving threats with measurable operational benefit. As EV charging infrastructure continues to expand and integrate with smart grid ecosystems, the methodologies developed in this thesis provide a foundation for future research and deployment of intelligent, resilient cybersecurity solutions for cyber-physical charging networks.

Bibliography

[1] International Energy Agency (IEA), "Global ev outlook 2024," 2024. Licence: CC BY 4.0.

[2] A. Bui, P. Slowik, and N. Lutsey, "Power play: Evaluating the u.s. position in the global electric vehicle transition," tech. rep., International Council on Clean Transportation, 2021. Accessed: Mar. 3, 2025.

[3] J. Antoun, M. E. Kabir, B. Moussa, R. Atallah, and C. Assi, "A detailed security assessment of the ev charging ecosystem," *IEEE Network*, vol. 34, no. 3, pp. 200–207, 2020.

[4] A. Ahalawat, S. Adepu, and J. Gardiner, "Security threats in electric vehicle charging," in *IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2022.

[5] D. Elmo, G. Fragkos, J. Johnson, K. Rohde, S. Salinas, and J. Zhang, "Disrupting ev charging sessions and gaining remote code execution with dos, mitm, and code injection exploits using ocpp 1.6," in *Resilience Week (RWS)*, 2023.

[6] O. C. Alliance, "Open charge point protocol (ocpp) 1.6 specification," 2016.

[7] International Organization for Standardization, "Iso 15118: Road vehicles — vehicle to grid communication interface," 2019.

[8] S. Acharya, H. A. U. Khan, R. Karri, and Y. Dvorkin, "Madeviot: Cyberattacks on ev charging can disrupt power grid operation," in *IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2024.

[9] R. Walton, "Simultaneous hack of ev chargers could cause manhattan blackout, nyu researchers find." Utility Dive, Aug. 2019. Accessed: Mar. 3, 2025.

[10] E. Benkhelifa, T. Welsh, and W. Hamouda, "A critical review of practices and challenges in intrusion detection systems for iot: Toward universal and resilient systems," *IEEE Communications Surveys & Tutorials*, 2018.

[11] H. Xu, S. Wang, N. Li, K. Wang, Y. Zhao, K. Chen, T. Yu, Y. Liu, and H. Wang, "Large language models for cyber security: A systematic literature review," *arXiv preprint*, 2024. arXiv:2405.04760.

[12] I. Skarga-Bandurova, I. Kotsiuba, and T. Biloborodova, "Cyber security of electric vehicle charging infrastructure: Open issues and recommendations," in *IEEE International Conference on Big Data (Big Data)*, 2022.

[13] S. Dey and M. Khanra, "Cybersecurity of plug-in electric vehicles: Cyberattack detection during charging," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 1, pp. 478–487, 2021.

[14] A. Hussain, A. Yadav, and G. Ravikumar, "Anomaly detection using bi-directional long short-term memory networks for cyber-physical electric vehicle charging stations," *IEEE Transactions on Industrial Cyber-Physical Systems*, vol. 2, pp. 508–518, 2024.

[15] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on cyber security for smart grid communications," *IEEE Communications Surveys & Tutorials*, vol. 14, pp. 998–1010, Jan. 2012.

[16] J. Johnson, B. Anderson, B. Wright, J. Quiroz, T. Berg, R. Graves, J. Daley, K. Phan, M. Kunz, R. Pratt, *et al.*, "Cybersecurity for electric vehicle charging infrastructure," tech. rep., Sandia National Laboratories, July 2022.

[17] D. Ronanki and H. Karneddi, "Electric vehicle charging infrastructure: Review, cyber security considerations, potential impacts, countermeasures, and future trends," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2024.

[18] P. Gawron *et al.*, "Security analysis of electric vehicle charging infrastructure," *Energy Informatics*, 2018.

[19] D. Kern, C. Krauß, and M. Hollick, "Detection of anomalies in electric vehicle charging sessions," in *Proceedings of the 39th Annual Computer Security Applications Conference (ACSAC '23)*, (New York, NY, USA), Association for Computing Machinery, 2023.

[20] I. Benfarhat, V. T. Goh, C. Lim Siow, M. Sheraz, and T. Chee Chuah, "Temporal convolutional network approach to secure open charge point protocol (ocpp) in electric vehicle charging," *IEEE Access*, vol. 13, pp. 15272–15289, 2025.

[21] H. S. Mavikumbure, V. Cobilean, C. S. Wickramasinghe, T. Phillips, B. J. Varghese, B. Carlson, C. Rieger, T. Pennington, and M. Manic, "Physical anomaly detection in ev charging stations: Physics-based vs resnet ae," in *IEEE 32nd International Symposium on Industrial Electronics (ISIE)*, 2023.

[22] T. B. Brown *et al.*, "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, 2020.

[23] A. Diaf, A. A. Korba, N. E. Karabadji, and Y. Ghamri-Doudane, "Beyond detection: Leveraging large language models for cyber attack prediction in iot networks," in *20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*, 2024.

[24] Y. Zhang, Y. Cao, X. Xu, and W. Shen, "Logicode: An llm-driven framework for logical anomaly detection," *IEEE Transactions on Automation Science and Engineering*, 2024.

[25] A. Russell-Gilbert, A. Sommers, A. Thompson, L. Cummins, S. Mittal, S. Rahimi, M. Seale, J. Jaboure, T. Arnold, and J. Church, "Aad-llm: Adaptive anomaly detection using large language models," in *IEEE International Conference on Big Data (BigData)*, 2024.

[26] X. Xu, H. Wang, Y. Liang, P. S. Yu, Y. Zhao, and K. Shu, "Can multimodal llms perform time series anomaly detection?," 2025.

[27] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive nlp tasks," *NeurIPS*, 2020.

[28] DESL-EPFL, "Level 3 ev charging dataset," 2024. Accessed: 2025-02-18.

[29] J. Chen, H. Lin, X. Han, and L. Sun, "Benchmarking large language models in retrieval-augmented generation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, Mar. 2024.

[30] H. He, R. Xiong, and J. Fan, "Evaluation of lithium-ion battery equivalent circuit models for state of charge estimation," *IEEE Transactions on Industrial Electronics*, 2019.

[31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[32] A. Radford *et al.*, "Learning transferable visual models from natural language supervision," *ICML*, 2021.