Zero-Shot Power System Anomaly Detection and Mitigation Using Large Language Models

Saleh Sadeghi Gougheri and Wei Sun
Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL USA

***Abstract*—Reliable detection of cyber–physical anomalies is vital for maintaining the security and resilience of modern power grids. Conventional machine learning detectors often require extensive labeled data and lack interpretability and integrated mitigation capability. This paper introduces a Zero-Shot LLM Anomaly Detection and Mitigation (ZS-LAD) framework that employs Large Language Models (LLMs) as time-series forecasters for unsupervised anomaly detection and mitigation. Multivariate electrical signals are encoded as sequences and processed through the LLM to predict future measurements. Deviations are assessed using a Median-Absolute-Deviation Z-score (MAD-Z), spike detection, and morphological fusion logic. The detected anomalies are interpreted by the LLM to classify event types and recommend mitigation actions aligned with operational standards. Based on experiments on the IEEE 13-bus system, ZS-LAD achieved 0.882 point-wise recall, 100% event recall, and 170 seconds mean detection latency, outperforming other benchmark methods. The results highlight the potential of this model for scalable and explainable grid anomaly management.**

***Index Terms*—Large Language Models, Unsupervised Anomaly Detection, Cybersecurity, and Explainable AI.**

I. Introduction

Modern power systems are becoming large cyber-physical networks relying on sensing and communication infrastructures to exchange data in real time. This digital transformation improves efficiency, automation, and situational awareness, but it also increases the system's exposure to new risks. Both cyber intrusions (such as false data injection or denial-of-service attacks) and physical disturbances (such as equipment faults or sudden load and generation changes) can appear as small deviations in time-series data, including voltage, current, or frequency measurements. Rapid and accurate detection of these anomalies is essential for ensuring the stability, reliability, and resilience of the modern power grid.

Artificial intelligence (AI)–based approaches have gained wide popularity in power system anomaly detection. These methods can be broadly divided into three categories: supervised, unsupervised, and self- or semi-supervised learning approaches. A supervised anomaly detection model was developed in [1], which brings high accuracy but relies heavily on labeled datasets. Gholami *et al.* [2] proposed an outlier-filtered weighted least squares method effective for known conditions but limited in generalization. Since labeled data in power systems are scarce as many faults and attacks are rare or undocumented supervised models often fail to adapt to unseen events. To reduce this dependency, Agottani *et al.* [3] presented a semi-supervised anomaly detection that required few labeled data and threshold set. Although semi- and self-supervised methods improve flexibility, they depend on partial labels and frequent retraining as network conditions evolve. Therefore, unsupervised learning methods are well-suited for dynamic power networks, as they detect anomalies without labeled data. K-Means and Isolation Forest were applied for distribution-level anomaly detection in [4], while Mendia et al. [5] used time-series clustering and outlier analysis for industrial energy monitoring. An adversarial autoencoder was also developed in [6] to enhance robustness in Industrial IoT. However, most of these methods rely solely on numerical data, functioning as black boxes that lack contextual reasoning and interpretability.

Large Language Models (LLMs) such as GPT, Claude, and Gemini demonstrate strong reasoning and generalization. In power systems, Huang *et al.* [7] explored foundation models for tasks like optimal power flow and situational awareness, while Selim *et al.* [8] applied an LLM for cyber-attack detection on smart inverter Volt/VAR commands using textual analysis. LLM reasoning was further enhanced in [9] using a feedback-driven multi-agent framework, operational security risks were examined in [10], and Chen *et al.* [11] investigated LLM–human collaboration for power dispatching. Despite these advances, most studies remain limited to text and overlook high-dimensional numerical data. Power system measurements are multivariate and time-dependent, requiring preprocessing into interpretable text. Moreover, existing works emphasize detection, while real-world operations demand reasoning-based mitigation recommendations. Recent GPT-based defense systems [12] show promise but are not yet suited for real-time grid operation. Accordingly, key gaps include:

- Supervised and semi-supervised models [1]–[3] depend on labeled data and fail to generalize to unseen anomalies.
- Unsupervised methods [4]–[6] detect deviations but lack explainable mitigation strategies.
- Existing models [4], [5] provide anomaly scores without explanation or operational insight.
- Current studies on LLMs [7]–[11] focus on text tasks, not multivariate grid data. GPT-based defense tool [12] is not adapted for real-time grid operation or mitigation.

To address these gaps, we propose the Zero-Shot LLM Anomaly Detection and Mitigation Framework (ZS-LAD), shown in Fig. 1. To the best of our knowledge, this is the first framework that applies a zero-shot LLM to multivariate power system time-series data for both anomaly detection and operator-level mitigation recommendations. The term zero-shot means that the LLM is used in a frozen, off-the-shelf form without any training or fine-tuning on power system data. This avoids the high computational cost of adapting

979-8-3315-8138-1

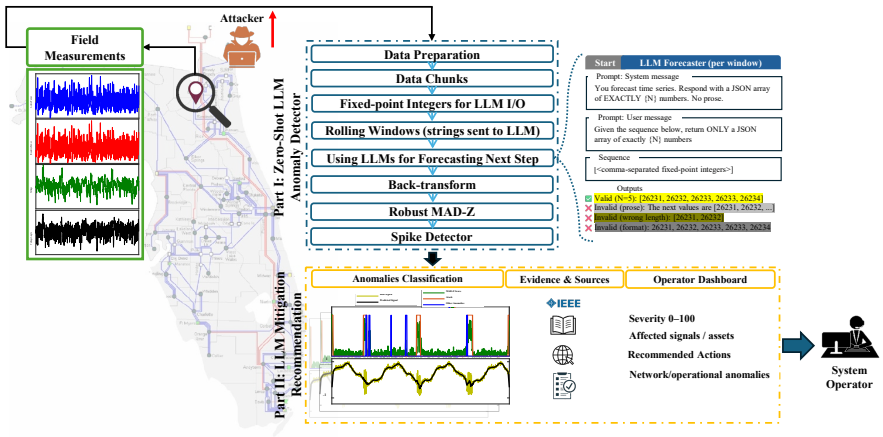

Fig. 1. Overall structure of the proposed ZS-LAD.

large models. Instead, we introduce task specific preprocessing and post-processing steps that convert multivariate time series measurements into LLM compatible inputs and map LLM outputs to reliable anomaly decisions. These steps enable practical zero-shot application of LLMs to power system anomaly detection, while the LLM itself remains unchanged. The proposed method has two parts:

Part I – Zero-Shot LLM Anomaly Detector: This stage identifies abnormal measurement patterns. The raw voltage, current, and power data are divided into overlapping windows and transformed into fixed-point integer sequences suitable for LLM input. Each sequence is sent to the LLM forecaster to predict the next time step. The predicted and actual values are compared using a Median Absolute Deviation Z-score (MAD-Z) to detect gradual deviations. A secondary spike detector captures sharp transients. Both indicators are combined using a morphological fusion algorithm and an adaptive K-of-N logic filter to produce continuous anomaly intervals with severity scores. This hybrid process allows the model to detect both slow and sudden anomalies without retraining.

Part II – LLM Mitigation Recommendation: The second stage leverages the LLM's reasoning capabilities to analyze each event. Each detected anomaly, along with its metadata (time, location, severity, residual, and affected buses), is summarized and sent to the LLM classifier. The model explains the anomaly type, such as cyberattack or load spike and suggests mitigation actions aligned with standard references (e.g., IEEE, NERC, or operational guidelines). This combination of zero-shot detection and interpretive reasoning forms a complete, explainable framework for real-time grid monitoring. In summary, the main contributions are:

- Proposing a novel framework that uses LLM-based approach for power system anomaly detection and mitigation operating entirely in a zero-shot mode.
- A preprocessing pipeline that converts multivariate numerical time-series data into a format suitable for LLM reasoning.
- Integration of MAD-Z residual scoring, spike detection, and morphological logic fusion for robust anomaly interval identification.
- Presenting an LLM-guided recommendation system that maps detected anomalies to standard mitigation actions for operators. Provides interpretable results and adapts to unseen disturbances without retraining or labeled data.

The remainder of the paper is organized as follows. Section II describes the methodology. Section III presents the case study and numerical results. Section IV concludes the paper.

II. Methodology

This section presents the proposed ZS-LAD, which includes two sequential stages: anomaly detection from multivariate feeder measurements without labeled data, followed by LLM-based event classification and operator-level mitigation.

A. Part I — Zero-Shot LLM Anomaly Detector

We consider N nodes, phases $\phi \in \{a, b, c\}$, and parameters $p \in \{P, Q, V, \Theta\}$. Each triplet (p, n, ϕ) yields a time series $x_{p,n,\phi}(t)$ for $t = 1, \ldots, T$, with $n = 1, \ldots, N$. After data preprocessing (normalization and bad data cleaning), synchronized measurements are stacked into a feature vector:

$$\mathbf{X}(t) = \left[x_{p,n,\phi}(t)\right]_{p\in\{P,Q,V,\Theta\},\, n\in\{1,\ldots,N\},\, \phi\in\{a,b,c\}} \in \mathbb{R}^{12N} \tag{1}$$

A fixed-point transformation method reduces the token count and improves stability by quantizing each signal to fixed-point integers. The local baseline and integer mapping are:

$$\mu^{\min}_{p,n,\phi} = \min_t x_{p,n,\phi}(t) \tag{2}$$

$$u_{p,n,\phi}(t) = \text{round}\Big((x_{p,n,\phi}(t) - \mu^{\min}_{p,n,\phi})\, 10^q\Big) \tag{3}$$

where q denotes decimal precision. $u_{p,n,\phi}(t)$ is the quantized fixed-point integer (3). Given a window length W and step $s{=}1$, each normal sequence $(u^{\text{norm}}_{p,n,\phi}(t))$ is divided into overlapping windows by (4). Each window is serialized and passed to the frozen LLM by (5), which predicts the next value, and dequantizes this value using (6). At this stage, a concise and structured prompt design is important for reliable LLM forecasting, as overly long prompts can be inefficient while overly short prompts may not sufficiently guide the model.

$$\mathcal{W}_{i,p,n,\phi} = \left[u^{\text{norm}}_{p,n,\phi}(i), \ldots, u^{\text{norm}}_{p,n,\phi}(i{+}W{-}1)\right] \tag{4}$$

$$\hat{u}^{\text{norm}}_{p,n,\phi}(i{+}W) = \mathsf{LLM}(\mathcal{W}_{i,p,n,\phi}) \tag{5}$$

$$\hat{x}^{\text{norm}}_{p,n,\phi}(i{+}W) = \hat{u}^{\text{norm}}_{p,n,\phi}(i{+}W)\,10^{-q} + \mu^{\min}_{p,n,\phi} \tag{6}$$

Aligning (6) over time yields the reference forecast $\hat{x}^{\text{ref}}_{p,n,\phi}(t)$. By combining the reference forecast data and mixed (normal + anomalous) data $x^{\text{mix}}_{p,n,\phi}(t)$, the residual can be obtained by (7). However, residuals alone are unreliable due to zero-shot forecast bias and drift; hence, they are normalized using Median Absolute Deviation–based Z-score (MAD-Z) and fused with spike and level detectors for robust anomaly detection.

$$r_{p,n,\phi}(t) = \left|x^{\text{mix}}_{p,n,\phi}(t) - \hat{x}^{\text{ref}}_{p,n,\phi}(t)\right| \tag{7}$$

Over rolling residual window $\mathcal{R}_{t,p,n,\phi}$, local median and MAD are [13], and normalized MAD-Z score is derived by (10).

$$m_{p,n,\phi}(t) = \text{median}(\mathcal{R}_{t,p,n,\phi}) \tag{8}$$

$$\text{MAD}_{p,n,\phi}(t) = \text{median}\left|\mathcal{R}_{t,p,n,\phi} - m_{p,n,\phi}(t)\right| \tag{9}$$

$$z_{p,n,\phi}(t) = \begin{cases} 0, \text{ if } \text{MAD}_{p,n,\phi}(t) = 0, \\ 0.6745\dfrac{|r_{p,n,\phi}(t) - m_{p,n,\phi}(t)|}{\text{MAD}_{p,n,\phi}(t)}, \text{ O/W.} \end{cases} \tag{10}$$

As defined in (10), points with $z_{p,n,\phi}(t) > \tau_z$ are labeled residual-based anomalies forming the set $h^{\text{MADZ}}_{p,n,\phi}$. For spike detectio (11)– (15) are used. Transient events are captured via first differences. A spike is accepted if both the score and amplitude exceed thresholds in (15). Where $\tilde{x}_{\text{med}}(t)$ is rolling median of the raw signal, $\widetilde{\text{MAD}}_x(t)$ is its rolling MAD, and γ sets how far a point must deviate to be considered significant.

$$\Delta x_{p,n,\phi}(t) = x^{\text{mix}}_{p,n,\phi}(t) - x^{\text{mix}}_{p,n,\phi}(t{-}1) \tag{11}$$

$$\tilde{m}_{\Delta,p,n,\phi}(t) = \text{median}\big(\Delta x_{p,n,\phi}\big) \tag{12}$$

$$\widetilde{\text{MAD}}_{\Delta,p,n,\phi}(t) = \text{median}\big|\Delta x_{p,n,\phi} - \tilde{m}_{\Delta,p,n,\phi}(t)\big| \tag{13}$$

$$z^{\Delta}_{p,n,\phi}(t) = \begin{cases} 0, \text{ if } \widetilde{\text{MAD}}_{\Delta,p,n,\phi}(t) = 0 \\ 0.674\dfrac{|\Delta x_{p,n,\phi}(t) - \tilde{m}_{\Delta,p,n,\phi}(t)|}{\widetilde{\text{MAD}}_{\Delta,p,n,\phi}(t)}, \text{ O/W} \end{cases} \tag{14}$$

$$\begin{aligned} h^{\text{spike}}_{p,n,\phi}(t) = \big[z^{\Delta}_{p,n,\phi}(t) > \tau_{\Delta}\big] \wedge \\ \Big(|x^{\text{mix}}_{p,n,\phi}(t) - \tilde{x}_{\text{med}}(t)| > \gamma\,\widetilde{\text{MAD}}_x(t)\Big) \end{aligned} \tag{15}$$

To identify persistent, step-like changes in residual behavior, mean residuals in adjacent time windows are compared using (16). Here, $\overline{r}_{\text{left}}(t)$ and $\overline{r}_{\text{right}}(t)$ are mean residuals in adjacent windows, $\text{IQR}(r)$ is the residual interquartile range, and λ controls sensitivity to persistent level changes. A time sample t is anomalous ($h_{p,n,\phi}(t) = 1$) if any of three detectors, MAD-Z, spike, or level-shift, reports an anomaly (Eq. (17)).

$$h^{\text{lvl}}_{p,n,\phi}(t) = \Big[\big|\overline{r}_{\text{right}}(t) - \overline{r}_{\text{left}}(t)\big| > \lambda\,\text{IQR}(r)\Big] \tag{16}$$

$$h_{p,n,\phi}(t) = h^{\text{MADZ}}_{p,n,\phi}(t) \vee h^{\text{spike}}_{p,n,\phi}(t) \vee h^{\text{lvl}}_{p,n,\phi}(t) \tag{17}$$

Finally, Morphological filtering cleans the detection results by removing very short false alarms, connecting nearby detections, and forming clear time intervals for each anomaly as described in (18). Each interval (s_i, e_i) indicates the start and end time of a detected anomaly, and $z^{\max}_i$ gives the maximum MAD-Z score within that interval to represent its severity.

$$\mathcal{I}_{p,n,\phi} = \big\{(s_i, e_i, z^{\max}_i)\big\}, z^{\max}_i = \max_{t\in[s_i,e_i]} z_{p,n,\phi}(t) \tag{18}$$

B. Part II — LLM-Based Event Classification and Mitigation Recommendation

Part II takes the detected anomalies and their scores from Part I, and performs event classification and mitigation guidance. The feeder is represented as an undirected graph:

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}), \qquad \text{Adj}(u) = \{v : (u, v) \in \mathcal{E}\}, \tag{19}$$

where buses are vertices and lines are edges. For each detected event, the system identifies the affected parameter, bus, and phase, and collects its neighboring nodes:

$$\mathcal{N}(\widehat{b}) = \{\mathcal{M}(v) \mid v \in \text{Adj}(\widehat{b})\}, \tag{20}$$

For each anomaly at time t, a temporal window of $K{+}1$ samples is extracted from the affected node and its neighbors across {*P_, Q_, V_, Theta_*} parameters.

$$\mathcal{S}(t; K) = \{t{-}K, \ldots, t\}, \tag{21}$$

This context enables the LLM to infer local and network-wide patterns to classify the anomaly type (e.g., load spike, attack, or line outage). The extracted metadata and context window are sent to a frozen LLM as a structured prompt. Using domain reasoning and standard operational guidelines (e.g., IEEE, NERC), the LLM outputs the event class and recommends mitigation actions following a fixed JSON schema. This links Part I detections with contextual reasoning providing explainable event classification and mitigation strategies.

III. Experimental Setup and Numerical results

The proposed ZS-LAD framework was evaluated on the IEEE 13-node test feeder using the open-source co-simulation platform from [14]. The platform incorporates a hardware-in-the-loop setup with an OPAL-RT real-time simulator, enabling realistic interaction between physical power components and cyber control layers. Time-series measurements were collected for P, Q, V, and θ across all buses and phases. To emulate cyberattacks, a false data injection (FDI) vector was generated via state estimation and applied to ten real-power columns to produce manipulated measurements. The framework employs *gpt-4.1-mini* API in a zero-shot mode with strict JSON output formatting. Each LLM query processes a forecasting window of $W{=}20$ samples, detection thresholds $\tau_z{=}3.5$, $\tau_{\Delta}{=}4.0$, and $\lambda{=}1.5$. For event classification, a temporal context of $K{=}10$ samples from the affected and neighboring buses is used to infer the anomaly type and recommend mitigation actions.

A. Detection Accuracy

Because ZS-LAD flags *all* anomalies (e.g., load spikes, voltage deviations) while the ground truth labels only *attacks*, conventional accuracy/precision/F1 formulas are not meaningful as they would count unlabeled anomalies as false positives. Instead, we report attack-focused metrics: Point-wise Attack Recall (PAR), Event-level Attack Recall (EAR), and Mean Detection Latency (MDL). Let $y_{\text{true}}(t) \in \{0, 1\}$ indicate attack labels and $\hat{y}(t) \in \{0, 1\}$ the detector flag. With

TABLE I
ATTACK-FOCUSED DETECTION METRICS ON $x_{P,692,b}$.

Model (method)	PAR	TP	FN	EAR	MDL (s)
ZS-LAD	**0.882**	45	6	**1**	**170**
MLP Residual	0.667	34	17	1	450
AE Residual	0.451	23	28	0.9	700

$\mathrm{TP} = \sum_t \mathbf{1}[\hat{y}(t) = 1 \wedge y_{\text{true}}(t) = 1]$ and $\mathrm{FN} = \sum_t \mathbf{1}[\hat{y}(t) = 0 \wedge y_{\text{true}}(t) = 1]$, the PAR is:

$$\mathrm{PAR} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \tag{22}$$

For M attack intervals $\{A_M\}_{m=1}^{M}$ and detected intervals $\{P_j\}$ the EAR is:

$$\mathrm{EAR} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\Big[\exists\, j :\ |P_j \cap A_m| > 0\Big] \tag{23}$$

For each attack $A_m = [t_{\text{start}}^{(m)}, t_{\text{end}}^{(m)}]$, let $t_{\text{first}}^{(m)}$ be the first detected timestamp within A_m; then the MDL is:

$$\mathrm{MDL} = \frac{1}{M'} \sum_{m:\, t_{\text{first}}^{(m)} \text{ exists}} \left(t_{\text{first}}^{(m)} - t_{\text{start}}^{(m)}\right) \tag{24}$$

where $M' \leq M$ is the number of detected attack events. For signal $x_{P,692,b}$, which represents the active power at node 692 and phase b, we obtain $\mathrm{PAR} = 0.882$ (TP$=45$, FN$=6$), $\mathrm{EAR} = 1.0$, and $\mathrm{MDL} = 170\,\mathrm{s}$. These results indicate that all attacks were correctly detected while avoiding penalization for correctly identifying unlabeled anomalies. The comparisons in Table I demonstrate the superiority of the proposed ZS-LAD against two learning-based baselines. In addition to its improved detection performance, ZS-LAD leverages an LLM to provide contextual explanations and mitigation suggestions to system operators once an event is detected. Furthermore, the model operates in a zero-shot manner, requiring no prior training or large labeled datasets, which greatly reduces computational efforts while preserving adaptability. Although the MLP and AE baselines can detect anomalies, they are less effective in accurately identifying the temporal boundaries of attacks. Their inability to delineate precise start and end times limits their interpretability compared to the proposed approach.

Fig. 2 illustrates the detection performance of the proposed framework for two representative signals, $x_{P,692,b}$ and $x_{P,675,a}$. The plots compare the mixed measurement data (normal + anomalies), the LLM-based reference forecast, the detected anomaly intervals, and the true attack periods. As shown, the detected regions align closely with the ground-truth attack intervals while suppressing false alarms during normal fluctuations. This confirms that the ZS-LAD approach can effectively capture anomalies across different nodes and phases of the distribution network. Besides, the framework also identifies other non-attack network anomalies, which are classified in Part II of the proposed method to determine their underlying causes and provide mitigation recommendations.

B. Comparison of Anomaly Flagging Techniques

To highlight the importance of accurate anomaly flagging, multiple variants of the proposed method were tested on $x_{P,692,b}$. The comparison quantifies the impact of different design components on accuracy. Table II reports different metrics for various anomaly detection techniques. The plain ChatGPT baseline performs poorly due to its lack of temporal context with a PAR of 0.196 and an average detection delay of 1,200 s. Incorporating residual-based forecasting notably improves performance, as the LLM residual model reaches a PAR of 0.784 and reduces latency to 180 s by leveraging historical patterns. However, after removing the spike and amplitude gating mechanisms, the model's sensitivity to short or subtle disturbances decreases (PAR = 0.51, EAR = 0.6), which confirms the necessity of these components for robust flagging. By integrating spike-aware and amplitude-gated residual analysis, the proposed ZS-LAD attains an optimal balance, with a recall of 0.882 and a latency of just 170 s.

TABLE II
PERFORMANCE COMPARISON OF ANOMALY FLAGGING TECHNIQUES.

Model (method)	PAR	TP	FN	EAR	MDL (s)
Plain ChatGPT (raw chunks)	0.196	10	41	0.3	1200
LLM Forecast + Plain Residuals	0.784	40	11	1	180
ZS-LAD (No Spike + No Amp Gate)	0.51	26	25	0.6	300
ZS-LAD (No Spike)	0.843	43	8	1	180
ZS-LAD (Proposed)	**0.882**	**45**	**6**	**1**	**170**

C. Event Classification and Mitigation Results

An example of the classification and mitigation prompt is provided in the Appendix. Fig. 3 presents second-stage outcomes, where LLM examines network data and results from Part I to classify each detected anomaly, and provides suitable operational recommendations. As depicted in Fig. 3(a), most of the detected events are low-severity cases representing normal or minor load fluctuations. However, the few high-severity events correspond to cyberattacks, which are the most critical anomalies. The time distribution shows that the model can clearly separate routine variations from critical anomalies over multiple days of monitoring. Fig. 3(b) further shows that load-related anomalies such as mild or minor fluctuations are the most frequent, while severe events mainly appear as attack anomalies or sudden load spikes. Finally, Fig. 3(c) highlights that the system not only detects anomalies but also suggests relevant mitigation actions. These recommendations help operators respond appropriately to cybersecurity checks by adjusting power system operations and keeping the system's strength in both detecting and managing anomalies.

IV. CONCLUSION AND FUTURE WORK

This paper presents a zero-shot Large Language Model (LLM) framework for anomaly detection and mitigation in power systems. The proposed ZS-LAD operates without labeled data or retraining, leveraging LLMs as zero-shot time-series predictors to identify gradual and transient anomalies through hybrid residual and spike analysis. By integrating a reasoning-based classification stage, the system translates detections into actionable mitigation strategies. Experimental results show that ZS-LAD achieves high detection recall, fast response time, and interpretability surpassing conventional

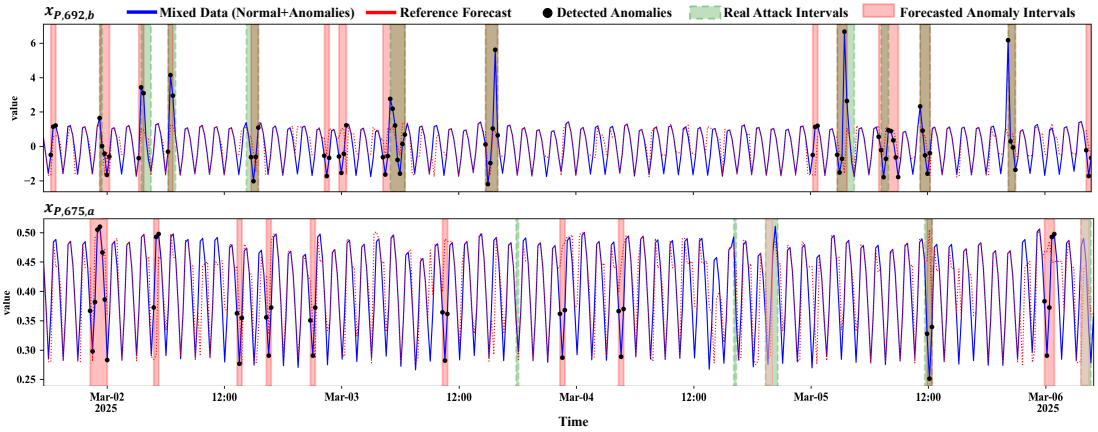

Fig. 2. Anomaly detection results for two representative signals ($x_{P,692,b}$ and $x_{P,675,a}$).

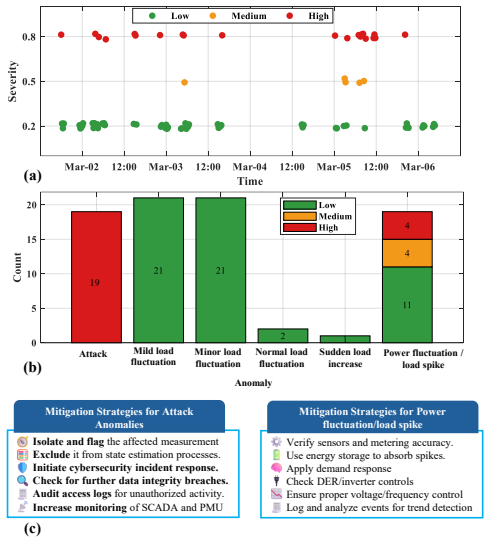

Fig. 3. LLM-based anomaly classification and mitigation: (a) anomaly severity over time; (b) type distribution with urgency; (c) mitigation recommendations.

ML baselines. Future work will explore fine-tuning domain-specific LLMs and integrating multi-agent reasoning to enhance adaptability and real-time performance across larger grid networks. Such fine-tuning is expected to further improve forecasting accuracy, especially for subtle and complex attacks. A practical consideration is latency and cost of frequent LLM queries in large-scale deployments. Future implementations can address this using hierarchical strategies and hybrid architectures that combine LLMs with fast conventional models.

Appendix: LLM Mitigation Strategy Format

For each anomaly from Part I, detector metadata (signal, time, residual, MAD-Z, etc) are combined with topology-adjacent buses. A short temporal window $(K+1)$ of measurements including `P`, `Q`, `V`, and θ for the target bus and its neighbors is extracted and formatted as a compact CSV block. The LLM is instructed to use power-system physics and standard operational references when generating mitigation actions. The structure used for classification and mitigation is:

System: You are an expert power systems engineer and cyber analyst.

User: Reference signal: $x_{p,n,\phi}$; Residual: 2.13; Z-score: 21.6; Neighbors: 671, 675; Data window (timestamps × P,Q,V,θ for target and neighbors)
Return: ONLY JSON with: {`cause, confidence, urgency, physical_explanation, recommended_actions`}.

References

[1] S. Bhattacharya, N. Saqib, and M. Govindarasu, "Ml-based anomaly detection system for iec 61850 communication in substations," in *2024 IEEE Power & Energy Society General Meeting (PESGM)*, 2024.

[2] A. Gholami, A. Tiwari, C. Qin, S. Pannala, A. K. Srivastava, R. Sharma, S. Pandey, and F. Rahmatian, "Detection and classification of anomalies in power distribution system using outlier filtered weighted least square," *IEEE Transactions on Industrial Informatics*, vol. 20, 2024.

[3] L. F. R. Agottani, R. Ferreira, R. S. Teixeira, L. R. de Lima, L. D. S. Coelho, and V. C. Mariani, "A semi-supervised anomaly detection approach applied to solar energy generation," in *ISGT EUROPE*, 2024.

[4] K. SundaraVelrani, K. Palle, B. Mallala, G. Nithya, N. K. K. Rao *et al.*, "Anomaly detection in power distribution networks with unsupervised learning algorithms," in *ICSES*. IEEE, 2024, pp. 1–6.

[5] I. Mendia, S. Gil-Lopez, I. Grau, and J. Del Ser, "A novel approach for the detection of anomalous energy consumption patterns in industrial cyber-physical systems," *Expert Systems*, vol. 41, no. 2, p. e12959, 2024.

[6] G.-Q. Zeng, Y.-W. Yang, K.-D. Lu, G.-G. Geng, and J. Weng, "Evolutionary adversarial autoencoder for unsupervised anomaly detection of industrial internet of things," *IEEE Transactions on Reliability*, 2025.

[7] C. Huang, S. Li, R. Liu, H. Wang, and Y. Chen, "Large foundation models for power systems," in *2024 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2024, pp. 1–5.

[8] A. Selim, J. Zhao, and B. Yang, "Large language model for smart inverter cyber-attack detection via textual analysis of volt/var commands," *IEEE Transactions on Smart Grid*, 2024.

[9] M. Jia, Z. Cui, and G. Hug, "Enhancing llms for power system simulations: A feedback-driven multi-agent framework," *IEEE Transactions on Smart Grid*, 2025.

[10] J. Zhu, S. Cai, F. Deng, B. C. Ooi, and J. Wu, "Do llms understand visual anomalies? uncovering llm's capabilities in zero-shot anomaly detection," in *32nd ACM International Conference*, 2024, pp. 48–57.

[11] X. Chen, X. Lu, Q. Li, D. Li, and F. Zhu, "Integration of llm and human-ai coordination for power dispatching with connected electric vehicles under sagvns," *IEEE Transactions on Vehicular Technology*, 2024.

[12] T. M. Santhi and K. Srinivasan, "Chatgpt-based learning platform for creation of different attack model signatures and development of defense algorithm for cyberattack detection," *IEEE Transactions on Learning Technologies*, vol. 17, pp. 1829–1842, 2024.

[13] J. van Dreven, A. Cheddad, A. N. Ghazi, S. Alawadi, J. Al Koussa, and D. Vanhoudt, "Heat: Hierarchical-constrained encoder-assisted time series clustering for fault detection in district heating substations," *Energy and AI*, p. 100548, 2025.

[14] X. Gao, M. Ali, A. Rahman, M. M. Rahman, and W. Sun, "An open source co-simulation test platform for cyber-physical power system security analysis," in *NAPS*. IEEE, 2024, pp. 1–6.