# Large Language Models for Anomalous Event Detection from Temporal Point Processes

Qinming Zhuang[a], Peng Zhang[a], Hong Yang[a*]
[a] Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou, China
km.zhuang@e.gzhu.edu.cn,{hyang,p.zhang}@gzhu.edu.cn

***Abstract*—Event Sequence Anomaly Detection (ESAD) plays a crucial role in domains such as healthcare, DevOps, and information security, where identifying deviations from normal patterns in event sequences is essential for ensuring stability and mitigating risks. Despite notable progress, existing ESAD methods often struggle when handling continuous-time event streams. Statistical models such as Poisson or Hawkes processes offer efficiency but fail to capture nonlinear temporal dependencies, while deep learning methods demand large-scale labeled data and frequently suffer from poor interpretability. These limitations hinder their deployment in high-stakes applications where reliability and transparency are critical. To address these challenges, we introduce TPP-LLMAD, a novel framework that integrates *Temporal Point Processes* (TPPs) with *Large Language Models* (LLMs) for interpretable anomaly detection. In TPP-LLMAD, neural TPPs model event sequences and estimate intensity functions, which are then transformed into structured triplets of timestamps, event marks, and intensities. These representations are embedded into tailored prompts that guide an LLM to assess deviations, assign anomaly labels, and generate human-readable explanations. By combining the quantitative rigor of TPPs with the interpretive capacity of LLMs, the framework bridges the gap between mathematical modeling and natural-language reasoning. Extensive experiments on real-world datasets demonstrate that TPP-LLMAD achieves performance comparable to or exceeding state-of-the-art baselines while providing explanations that enhance interpretability and usability. This work represents the first systematic integration of TPP intensity modeling with LLM-based reasoning for ESAD, advancing the frontier of interpretable event sequence analysis.**

***Index Terms*—Event Sequence Anomaly Detection, Temporal Point Processes, Large Language Models, Interpretable AI**

## I. Introduction

Event sequence data naturally arise in a wide range of domains. In healthcare, sequences of clinical events must be monitored to detect abnormal patient trajectories that may indicate disease onset or medical errors [1]. In DevOps and system monitoring, streams of server logs or latency traces reveal patterns of normal operation, and detecting anomalies is essential for preventing cascading failures [2]. In cybersecurity, identifying rare but critical deviations in access logs or network flows is central to intrusion detection. Across these domains, *Event Sequence Anomaly Detection* (ESAD) is indispensable for maintaining robustness, reliability, and security.

Despite its importance, ESAD remains a challenging task. Traditional approaches can be broadly divided into three categories. Statistical and feature-based methods, including Poisson and Hawkes processes [3], are lightweight and interpretable but are often unable to capture nonlinear temporal dependencies. Classical machine learning approaches, while flexible, depend heavily on handcrafted features and labeled data [4], which are costly to obtain. Deep learning models such as recurrent neural networks (RNNs) and Transformers [5], [6] excel at capturing complex dynamics but require substantial amounts of annotated sequences, and their black-box nature makes it difficult to trust their predictions in mission-critical settings.

Recently, large language models (LLMs) have emerged as powerful tools for reasoning and interpretation. Their few-shot and zero-shot learning capabilities make them attractive for anomaly detection in situations where labeled data are scarce [7]. However, applying LLMs directly to ESAD is problematic. While LLMs excel at pattern recognition in natural language, they lack inherent temporal reasoning and do not natively encode statistical properties of event streams [8]. This mismatch leads to increased false alarms, reduced accuracy, and diminished trust in the resulting anomaly reports.

To overcome these limitations, we propose **TPP-LLMAD**, a unified ESAD framework that integrates the statistical expressiveness of Temporal Point Processes (TPPs) with the interpretive reasoning power of LLMs. In this framework, neural TPPs model the underlying generative process of event sequences, yielding event-specific intensity functions that capture both temporal dependencies and event-type distributions. These intensity-driven features are then transformed into structured representations, which are embedded into carefully engineered prompts that guide the LLM to identify anomalies and produce human-readable justifications. By explicitly combining mathematical intensity modeling with natural-language reasoning, TPP-LLMAD achieves both accuracy and interpretability.

The contributions of this paper are threefold. First, we introduce TPP-LLMAD, the first ESAD framework that systematically integrates TPPs with LLMs. Second, we design an intensity-based representation and prompting strategy that enables the LLM to reason about statistical deviations in event streams while maintaining interpretability. Third, we validate the proposed approach through experiments on real-world datasets, showing that TPP-LLMAD not only matches or outperforms state-of-the-art baselines in anomaly detection accuracy but also provides concise and actionable explanations. Together, these contributions advance the development

*Corresponding author.

979-8-3315-6689-0/25/$31.00 ©2025 IEEE
DOI 10.1109/ICKG66886.2025.00060

of interpretable and efficient anomaly detection methods for complex event sequences.

## II. Background and Related Work

### A. Event Sequence Anomaly Detection

Event Sequence Anomaly Detection (ESAD) is an active research direction that seeks to identify abnormal patterns in event streams occurring at irregular time intervals [1], [2]. Formally, an event sequence can be defined as $S = \{(t_i, k_i) \mid i = 1, 2, \ldots, N\}$, where $t_i \in [0, T]$ denotes the timestamp of the $i$-th event, $k_i \in \{1, \ldots, K\}$ is its event type (or mark), $T$ is the observation window, and $K$ denotes the number of event types. Given a learned distribution $\mathbb{P}_{\text{data}}$ of typical event streams, ESAD aims to determine whether a new sequence $S$ conforms to this distribution or instead belongs to an anomalous distribution $\mathbb{Q} \neq \mathbb{P}_{\text{data}}$. ESAD plays a critical role in applications such as cybersecurity [9], industrial monitoring [10], healthcare informatics [1], and financial fraud detection [11].

Existing ESAD methods fall into three broad methodological paradigms. The first is statistical modeling, where approaches such as Poisson and Hawkes processes attempt to capture event occurrence patterns through intensity functions [3]. These models are mathematically elegant and computationally efficient, yet they struggle with the nonlinear temporal dependencies that arise in complex real-world event streams [12]. A second direction involves classical machine learning techniques, including supervised classifiers such as SVMs and Random Forests, as well as unsupervised clustering-based approaches [4], [13]. These methods can handle nonlinearities more effectively but rely heavily on handcrafted feature design and labeled datasets, both of which are expensive and prone to bias [14], [15]. More recently, deep learning models have become prominent, with architectures such as recurrent neural networks, LSTMs [5], autoencoders [6], and Transformers [16] demonstrating the ability to capture high-dimensional temporal dependencies. While such models often achieve state-of-the-art detection performance [17], [18], they demand substantial computational resources and large annotated datasets, and their black-box nature limits interpretability [19], [20].

Beyond these three categories, hybrid approaches [21]–[24] and sequence-specific models such as Temporal Point Processes [3], [9] have been introduced to enhance robustness and improve anomaly detection in irregular event streams. However, these methods often increase design complexity and still provide limited transparency for human operators. Balancing predictive power with interpretability remains one of the central challenges for advancing ESAD.

### B. Temporal Point Processes

Temporal Point Processes (TPPs) [3], [9] provide a principled stochastic framework for modeling event sequences over continuous time. At the core of TPPs lies the *conditional intensity function* $\lambda(t|\mathcal{H}_t)$, which specifies the instantaneous event rate given past history $\mathcal{H}_t$. Classical formulations include the Poisson process, the self-exciting Hawkes process, and marked point processes [3]. These approaches naturally accommodate irregular inter-event times and self-exciting dynamics, making them particularly relevant for ESAD in domains such as seismology [9], financial trading, and cybersecurity [3]. Despite these advantages, traditional TPPs often require manual specification of the intensity function and provide limited explanatory power, which restricts their practical adoption by domain experts and system operators [16].

## III. Preliminaries

A *Temporal Point Process* (TPP) [3], [9] is defined over a sequence of event times $\{T_1, T_2, \ldots, T_N\}$, where each $T_i \in [0, T]$ denotes the occurrence time of the $i$-th event within a fixed observation window $[0, T]$. The associated **counting process** $N(t)$ records the cumulative number of events observed up to time $t$. Central to TPPs is the **conditional intensity function** $\lambda(t \mid \mathcal{H}_t)$, which characterizes the instantaneous rate of event occurrence given the event history $\mathcal{H}_t$:

$$\lambda(t \mid \mathcal{H}_t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\left(N(t + \Delta t) - N(t) = 1 \mid \mathcal{H}_t\right)}{\Delta t}, \quad (1)$$

where $\mathcal{H}_t$ denotes the sigma-algebra generated by all events up to time $t$.

Based on this definition, the log-likelihood of observing a sequence of $N$ events under a parametric TPP model is expressed as [3], [9]:

$$L = \left( \prod_{i=1}^{N} \lambda(T_i \mid \mathcal{H}_{T_i}) \right) \exp\left( - \int_0^T \lambda(t \mid \mathcal{H}_t)\, dt \right), \quad (2)$$

where the first term measures the contribution of observed events at their corresponding timestamps, and the compensator term accounts for the absence of events in intervals where none were observed.

**Marked TPPs** extend this formulation by associating each event with a categorical mark $k_i \in \mathcal{K}$, leading to marked events $(T_i, k_i)$ [3], [9]. The log-likelihood generalizes to:

$$L = \left( \prod_{i=1}^{N} \lambda(T_i, k_i \mid \mathcal{H}_{T_i}) \right) \exp\left( - \sum_{k=1}^{K} \int_0^T \lambda(t, k \mid \mathcal{H}_t)\, dt \right), \quad (3)$$

where $\lambda(T_i, k_i \mid \mathcal{H}_{T_i})$ denotes the conditional intensity associated with the specific event type $k_i$, and the compensator integrates intensities over all possible marks.

Within the context of anomaly detection, TPPs provide a natural mechanism for capturing deviations from expected temporal patterns. By modeling both inter-event dependencies and mark-specific occurrence rates, they can represent complex dynamics such as clustering, inhibition, or self-excitation. For instance, Hawkes processes capture self-exciting behavior, making them particularly effective in domains where past events increase the likelihood of subsequent ones, such as social media reposts, seismic aftershocks, or bursts of network traffic. When the observed sequence exhibits significant deviations between empirical intensities and those predicted by the model, the sequence can be regarded as anomalous. This

ability to model fine-grained temporal dependencies makes TPPs especially suitable as a statistical foundation for Event Sequence Anomaly Detection.

## IV. Proposed Method

### A. Preliminaries and Related Concepts

Event Sequence Anomaly Detection (ESAD) can be formally described as the task of learning a decision function $f : S \mapsto \{0, 1\}$, where $S = \{(t_i, k_i)\}_{i=1}^{N}$ denotes an event stream consisting of $N$ events, $t_i \in [0, T]$ is the timestamp, and $k_i$ is the associated event type. The goal is to determine whether a given sequence $S$ deviates significantly from the distribution of normal sequences. A sequence is regarded as anomalous if it exhibits temporal irregularities or abnormal mark distributions that cannot be explained by patterns observed in the training data [1]. This formulation highlights the dual challenge of ESAD: capturing temporal dynamics while simultaneously accounting for categorical event types.

Temporal Point Processes (TPPs) provide a natural probabilistic foundation for modeling such data [9], [25]. Given the event history $\mathcal{H}_t = \{(t_j, k_j) : t_j < t\}$, the conditional intensity function

$$\lambda(t|\mathcal{H}_t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(N(t + \Delta t) - N(t) = 1|\mathcal{H}_t)}{\Delta t}, \quad (4)$$

quantifies the instantaneous arrival rate at time $t$. For an observed sequence $\{t_i\}_{i=1}^{N}$, the log-likelihood under a parametric model is given by

$$L = \sum_{i=1}^{N} \log \lambda(T_i|\mathcal{H}_{T_i}) - \int_0^T \lambda(t|\mathcal{H}_t)\, dt, \quad (5)$$

and parameter estimation proceeds by maximizing $L$. This likelihood-based framework offers a statistically principled way to distinguish normal sequences from anomalous ones through deviations in fitted intensities.

To capture heterogeneity in event types, marked TPPs extend this formulation by introducing a joint intensity function

$$\lambda_k(t|\mathcal{H}_t) = \lambda(t|\mathcal{H}_t)\, f(k|t, \mathcal{H}_t), \quad (6)$$

where $f(k|t, \mathcal{H}_t)$ is the conditional probability mass function over marks. This decoupled representation separates temporal dynamics from categorical distributions, thereby enabling the detection of anomalies along both dimensions: irregular inter-event timings and unexpected mark patterns. Such flexibility makes marked TPPs particularly effective for ESAD tasks in application domains such as cybersecurity and finance, where abnormal event type combinations often carry significant operational implications.

While TPPs provide statistical rigor, they lack interpretability in practical anomaly detection scenarios. Recent advances in prompt engineering for Large Language Models (LLMs) [7] enable bridging this gap by embedding TPP-derived features into structured prompts. These prompts can incorporate task definitions, statistical quantities, and contextual examples, allowing LLMs to generate both binary anomaly decisions and natural-language justifications [8]. This integration of probabilistic modeling with interpretive reasoning constitutes the foundation of our proposed framework, TPP-LLMAD.

### B. LLM-based Anomaly Detection

In TPP-LLMAD, the large language model serves as a reasoning layer that consumes structured features derived from marked TPP intensities. For each observed event $(t_i, k_i)$, the conditional intensity $\lambda_{k_i}(t_i|\mathcal{H}_{t_i})$ is compared with empirical quantiles estimated from normal data. An indicator variable

$$I(t_i, k_i) = \mathbb{1}\{\lambda_{k_i}(t_i|\mathcal{H}_{t_i}) > Q_{90}\} \quad (7)$$

flags extreme deviations, and the proportion of such violations

$$\mathcal{I} = \tfrac{1}{N} \sum_{i=1}^{N} I(t_i, k_i) \quad (8)$$

summarizes the overall level of intensity anomalies within the sequence. Complementary to this, temporal irregularity is quantified by the variance of inter-event intervals

$$\mathcal{T} = \mathrm{Var}(\Delta t_1, \ldots, \Delta t_N), \quad \Delta t_i = t_i - t_{i-1}, \quad (9)$$

while categorical irregularity is characterized by the entropy of the empirical mark distribution

$$\mathcal{M} = -\sum_{m} p(m) \log p(m), \quad p(m) = \tfrac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{k_i = m\}. \quad (10)$$

Together, $\mathcal{I}, \mathcal{T}$, and $\mathcal{M}$ form complementary statistical indicators of anomalous behavior.

The LLM is provided with both event-level triplets $(t_i, k_i, \lambda_{k_i}(t_i))$ and the aggregated statistics $(\mathcal{I}, \mathcal{T}, \mathcal{M})$ in the form of a structured prompt $\mathcal{P}$. Its output is formalized as

$$(y, e, c) = \mathrm{LLM}(\mathcal{P}, \{(t_i, k_i, \lambda_{k_i}(t_i))\}_{i=1}^{N}), \quad (11)$$

where $y \in \{0, 1\}$ denotes the anomaly label, $e$ provides a natural-language explanation, and $c \in [0, 1]$ reflects a confidence score.

This design emphasizes the dual role of the LLM: enforcing decision rules grounded in TPP-based statistics, while augmenting them with context-aware reasoning. In doing so, TPP-LLMAD combines statistical rigor with interpretability, advancing beyond traditional ESAD methods that lack explanations and beyond pure LLM approaches that lack statistical grounding.

### C. TPP-LLMAD Framework

To overcome limitations of existing ESAD methods in data efficiency, interpretability, and adaptability [5], [6], [9], we propose **TPP-LLMAD**, a hybrid framework that integrates neural temporal point processes with large language models.

**Neural TPP Modeling.** Raw event streams are rescaled and embedded into textualized representations. A neural temporal point process (NTPP), parameterized by a recurrent encoder

such as a GRU [5], is trained on normal sequences to estimate conditional intensities $\lambda_k(t|\mathcal{H}_t)$. Training maximizes the regularized log-likelihood

$$\mathcal{L} = -\sum_{i=1}^{N} \log \lambda_{k_i}(t_i|\mathcal{H}_{t_i}) + \alpha\, \mathrm{TV}(\lambda(t))\,, \tag{12}$$

where a total variation penalty encourages smooth intensity trajectories. The resulting fitted intensities form the statistical backbone for anomaly detection.

**Adaptive Feature Extraction.** From the trained TPP, event-level triplets $(t_i, \lambda(t_i), k_i)$ are extracted with an adaptive resolution $\delta(t)$ inversely proportional to local intensity gradients:

$$\delta(t) = \delta_{\text{base}}(1 + \|\nabla\lambda(t)\|)^{-1}. \tag{13}$$

This mechanism ensures fine-grained representation in regions of rapid intensity variation, while preserving computational efficiency elsewhere. Aggregated statistics $(\mathcal{I}, \mathcal{T}, \mathcal{M})$ are then computed to capture complementary evidence of deviations in intensities, timings, and mark distributions.

**LLM-Driven Interpretive Analysis.** Finally, structured prompts are constructed to include (i) task description, (ii) intensity-based rules, (iii) output format, and (iv) representative examples. The LLM processes both event-level and aggregated features, generating a structured output consisting of (1) an anomaly label, (2) a natural-language explanation linking statistical deviations to event semantics, and (3) a confidence score. This integration ensures that anomaly detection is not only accurate but also interpretable and adaptable to diverse application domains.

## V. Experiment

### A. Experimental settings

To evaluate the effectiveness of the proposed TPP-LLMAD framework, we conducted experiments on two real-world datasets: **ServerOverload** and **Latency**. These datasets capture distinct scenarios in system monitoring and cybersecurity, providing a diverse testbed for assessing the generalization of our approach across heterogeneous event sequence patterns [1], [9].

The **ServerOverload** dataset consists of event logs collected from distributed server infrastructures. Events include CPU usage spikes, memory allocation failures, and service request timeouts, with anomalies primarily caused by sudden surges in resource consumption (e.g., CPU utilization exceeding 95%) and cascading failure events [1], [9]. Preprocessing involves transforming raw logs into $(t_i, k_i)$ sequences and estimating their intensities using Hawkes process–based models [3].

The **Latency** dataset captures network latency measurements from cloud services, where events represent request-response delays and timeout occurrences. Anomalies are characterized by abrupt increases in latency (e.g., exceeding 500 ms) or recurring spikes indicative of network congestion [1], [9]. For preprocessing, latency records are aggregated into $(t_i, k_i)$ sequences, and intensities are estimated using Poisson process models [3].

For both datasets, we adopt stratified splits of 70% training, 15% validation, and 15% testing, ensuring that anomalous sequences are proportionally represented across subsets.

**Evaluation Metric.** Model performance was assessed using the *Area Under the Receiver Operating Characteristic Curve (AUC-ROC)*. This metric is particularly suitable for the ESAD setting because it is insensitive to class imbalance [1], captures trade-offs across different decision thresholds [9], and facilitates standardized comparison with prior anomaly detection studies [26].

**Comparison Models.** We compared TPP-LLMAD against three representative supervised learning baselines and one prompt-based baseline. The supervised methods include a feedforward neural network (NN) with three hidden layers (256–128–64 units, ReLU activations) [27], a recurrent LSTM with 128 hidden units designed to capture temporal dependencies [1], and a Support Vector Machine (SVM) classifier with an RBF kernel ($\gamma = 0.1$, $C = 1.0$) [9]. To isolate the contribution of temporal point processes, we also evaluate an *Only LLM* variant in which raw event sequences are textualized and directly processed by GPT-3.5 without TPP-derived features. Finally, we report results for our framework's ablation variant *TPP-LLMAD (no prompt)* and the complete *TPP-LLMAD* model.

### B. Experimental results

We conducted extensive experiments to evaluate the proposed TPP-LLMAD framework on the ServerOverload and Latency datasets. The implementation employed a GRU-128 temporal point process model ($\eta = 0.001$) trained with PyTorch, and integrated GPT-3.5 (temperature = 0.3, max_tokens = 500) through the OpenAI API. Training was carried out for up to 50 epochs with early stopping determined by validation performance, using an NVIDIA A100 GPU. Inference required approximately 1.2 seconds per query, enabling the framework to process about 20 sequences per minute, while sustaining real-time throughput of roughly 1,000 events per second.

Table II shows that TPP-LLMAD consistently outperformed all baselines. The *Only LLM* variant, which excluded intensity-based features, achieved poor performance (0.488–0.502 AUC), highlighting the importance of temporal modeling. Meanwhile, the *TPP-LLMAD (no prompt)* variant surpassed conventional baselines but remained below the full model, confirming that structured prompts substantially enhance both accuracy and interpretability.

## VI. Conclusion

This paper introduced TPP-LLMAD, a novel framework that integrates Temporal Point Processes (TPPs) [3], [9] with Large Language Models (LLMs) [7], [28] for Event Sequence Anomaly Detection (ESAD). By leveraging TPPs to capture temporal dependencies and LLMs to provide human-readable justifications, TPP-LLMAD consistently outperformed state-of-the-art baselines across multiple datasets. Experimental results on ServerOverload and Latency datasets confirmed significant improvements, particularly under complex and irregu-

TABLE I
CHARACTERISTICS OF THE EXPERIMENTAL DATASETS

| Feature | ServerOverload | Latency |
|---|---|---|
| Source | Server logs | Network logs |
| Time Period | 6 months | 3 months |
| Total Sequences | 12,000 | 8,500 |
| Anomalous Sequences | 600 (5%) | 425 (5%) |
| Avg. Sequence Length | 50 | 30 |

TABLE II
MODEL PERFORMANCE COMPARISON (AUC-ROC SCORES)

| Model | ServerOverload | Latency |
|---|---|---|
| NN | 0.491 | 0.496 |
| LSTM | 0.489 | 0.503 |
| SVM | 0.472 | 0.471 |
| Only LLM | 0.502 | 0.488 |
| TPP-LLMAD (no prompt) | 0.564 | 0.605 |
| **TPP-LLMAD (full)** | **0.651** | **0.731** |

lar temporal patterns [9]. Ablation studies further established that both TPP-based feature extraction and prompt-engineered LLM reasoning are essential to the observed performance gains. In summary, TPP-LLMAD provides an effective, interpretable, and adaptable solution for ESAD, with promising extensions to multi-modal and real-time streaming environments [1]. Future work will focus on improving scalability and investigating its generalization across broader application domains.

## REFERENCES


[1] H. Niu, O. A. Omitaomu, M. A. Langston, M. Olama, O. Ozmen, H. B. Klasky, A. Laurio, B. Sauer, M. Ward, and J. Nebeker, "Detecting anomalous sequences in electronic health records using higher-order tensor networks," *Journal of Biomedical Informatics*, vol. 135, p. 104219, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046422002246

[2] M. Du, F. Li, G. Zheng, and V. Srikumar, "Deeplog: Anomaly detection and diagnosis from system logs through deep learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1285–1298. [Online]. Available: https://doi.org/10.1145/3133956.3134015

[3] O. Shchur, A. C. Turkmen, T. Januschowski, J. Gasthaus, and S. Gunnemann, "Detecting anomalous event sequences with temporal point processes," pp. 13 419–13 431, 2021.

[4] R. Chalapathy, N. Khoa, and S. Chawla, "Robust deep learning methods for anomaly detection," *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*, 2020.

[5] C. Wu, S. Shao, C. Tunc, P. Satam, and S. Hariri, "An explainable and efficient deep learning framework for video anomaly detection," *Cluster Computing*, vol. 25, pp. 2715 – 2737, 2021.

[6] J. Ring, C. M. V. Oort, S. Durst, V. White, J. P. Near, and C. Skalka, "Methods for host-based intrusion detection with deep learning," *Digital Threats: Research and Practice*, 2021.

[7] S. Alnegheimish, L. Nguyen, L. Berti-Equille, and K. Veeramachaneni, "Large language models can be zero-shot anomaly detectors for time series?" 2024. [Online]. Available: https://arxiv.org/abs/2405.14755

[8] J. Liu, C. Zhang, J. Qian, M. Ma, S. Qin, C. Bansal, Q. Lin, S. Rajmohan, and D. Zhang, "Large language models can deliver accurate and interpretable time series anomaly detection," 2024. [Online]. Available: https://arxiv.org/abs/2405.15370

[9] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection for discrete sequences: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, pp. 823–839, 2012.

[10] M. Boldt, A. Borg, S. Ickin, and J. Gustafsson, "Anomaly detection of event sequences using multiple temporal resolutions and markov chains," *Knowledge and Information Systems*, vol. 62, pp. 669–686, 2019.

[11] A. B. Nassif, M. A. Talib, Q. Nasir, and F. Dakalbab, "Machine learning for anomaly detection: A systematic review," *IEEE Access*, vol. 9, pp. 78 658–78 700, 2021.

[12] S. Zhang, C. Zhou, P. Zhang, Y. Liu, Z. Li, and H. Chen, "Multiple hypothesis testing for anomaly detection in multi-type event sequences," *2023 IEEE International Conference on Data Mining (ICDM)*, pp. 808–817, 2023.

[13] S. Corli, L. Moro, D. Dragoni, M. Dispenza, and E. Prati, "Quantum machine learning algorithms for anomaly detection: A review," *Future Gener. Comput. Syst.*, vol. 166, p. 107632, 2024.

[14] S. Wang, J. F. Balarezo, S. Kandeepan, A. Al-Hourani, K. G. Chavez, and B. I. P. Rubinstein, "Machine learning in network anomaly detection: A survey," *IEEE Access*, vol. PP, pp. 1–1, 2021.

[15] G. Pang, C. Shen, L. Cao, and A. Hengel, "Deep learning for anomaly detection," *ACM Computing Surveys (CSUR)*, vol. 54, pp. 1 – 38, 2020.

[16] K. Alam, K. Kifayat, G. A. Sampedro, V. Karovic, and T. Naeem, "Sxad: Shapely explainable ai-based anomaly detection using log data," *IEEE Access*, vol. 12, pp. 95 659–95 672, 2024.

[17] T. Mathonsi and T. L. van Zyl, "Statistics and deep learning-based hybrid model for interpretable anomaly detection," *ArXiv*, vol. abs/2202.12720, 2022.

[18] Z. Luo, R. Zuo, and Y. Xiong, "Visual interpretable deep learning algorithm for geochemical anomaly recognition," *Natural Resources Research*, vol. 31, pp. 2211 – 2223, 2022.

[19] D. F. N. Oliveira, L. Vismari, A. M. Nascimento, J. R. de Almeida, P. Cugnasca, J. Camargo, L. Almeida, R. Gripp, and M. M. Neves, "A new interpretable unsupervised anomaly detection method based on residual explanation," *IEEE Access*, vol. 10, pp. 1401–1409, 2021.

[20] G. Kim, Y. Hwang, K. Lee, and Y. Choo, "Generalization performance analysis of anomaly detection-based active sonar classifier using anomaly score landscape," *The Journal of the Acoustical Society of America*, 2024.

[21] Z. Ghrib, R. Jaziri, and R. Romdhane, "Hybrid approach for anomaly detection in time series data," *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, 2020.

[22] W. Lin, S. Wang, W. Wu, D. Li, and A. Zomaya, "Hybridad: A hybrid model-driven anomaly detection approach for multivariate time series," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, pp. 866–878, 2024.

[23] L. Huang, Y. Zhu, Y. Gao, T. Liu, C. Chang, C. Liu, Y. Tang, and C. Wang, "Hybrid-order anomaly detection on attributed networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, pp. 12 249–12 263, 2023.

[24] T. F. Ghanem, W. Elkilani, and H. Abdul-Kader, "A hybrid approach

for efficient anomaly detection using metaheuristic methods," *Journal of Advanced Research*, vol. 6, pp. 609 – 619, 2014.

[25] H. Gong, H. Wang, P. Zhang, S. Zhou, H. Chen, and J. Bu, "Fedmtpp: Federated multivariate temporal point processes for distributed event sequence forecasting," *IEEE Transactions on Mobile Computing*, vol. 24, no. 4, pp. 3302–3315, 2025.

[26] M. Du, F. Li, G. Zheng, and V. Srikumar, "Deeplog: Anomaly detection and diagnosis from system logs through deep learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1285–1298. [Online]. Available: https://doi.org/10.1145/3133956.3134015

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, p. 84–90, May 2017. [Online]. Available: https://doi.org/10.1145/3065386

[28] J. Liu, C. Zhang, J. Qian, M.-J. Ma, S. Qin, C. Bansal, Q. Lin, S. Rajmohan, and D. Zhang, "Large language models can deliver accurate and interpretable time series anomaly detection," *ArXiv*, vol. abs/2405.15370, 2024.