World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

AI-driven anomaly detection and root cause analysis: Using machine learning on logs, metrics, and traces to detect subtle performance anomalies, security threats, or failures in complex cloud environments

Raviteja Guntupalli *

Manager, Cloud Engineering, AnnArbor, Michigan, USA.

World Journal of Advanced Research and Reviews, 2025, 26(02), 874-879

Publication history: Received on 18 March 2025; revised on 30 April 2025; accepted on 03 May 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.2.1521

Abstract

Enhanced complexity, together with high service dependencies and dynamic scaling requirements in present-day cloud environments, create both critical and difficult conditions for quick anomaly detection as well as root cause analysis (RCA). The traditional rule-based monitoring framework cannot discover slight and new types of anomalies that occur before system outages or security breaches. The document examines how AI systems alongside Machine Learning (ML) capabilities combined with deep learning processing of logs, metrics, and traces help automatically detect anomalies while performing RCA operations in cloud-native platforms.

The paper examines the utilization of supervised learning with unsupervised and reinforcement methods on diverse telemetry information to perform real-time detection of performance dips and, system errors and anomalous usage patterns. These systems can use AI technology to link distributed system incidents while simultaneously pinpointing foundational problems that human personnel cannot match for speed when recommending solutions. The operational effects of these techniques can be seen through real-life applications at Adobe, Uber, Zalando, and LinkedIn.

Automated RCA systems face ethical and technical challenges, according to the paper, which details problems like model drift, interpretability of complex models, and observability gaps. The ongoing expansion of cloud systems makes AI-driven anomaly detection essential for maintaining resilience and optimizing performance and cyber defense for both multi-cloud and hybrid cloud systems.

Keywords: Cloud monitoring; Anomaly detection; Root cause analysis; Machine learning; Deep learning; Observability; Logs; Metrics; Traces; AI operations; Security threats; Cloud resilience

1. Introduction

Active business infrastructure today has become more complicated than ever before. Cloud-native architectures, which now run global applications with real-time data streams and microservices alongside continuous deployment, have made system health monitoring a major difficulty. The previous methods of threshold-based monitoring and manually written rules fail to deliver sufficient monitoring in cloud environments, which experience non-linear behavior across context-dependent systems that perform continuous change.

The central problem of this issue involves both detecting anomalies and performing root cause analysis (RCA). Performance anomalies, which include CPU leaks, saturation, and I/O bottlenecks, can spread across microservice architectures until they cause service degradation or system outages. System logs, along with trace data, sometimes show minor irregularities that expose security threats by revealing their developmental stages apart from showing later

* Corresponding author: Raviteja Guntupalli

movement along with privilege escalation and data extraction events. Detection of root causes in these environments consumes extensive time. It introduces errors because proper domain understanding and manual processing of different telemetry data, such as traces, logs, and metrics, remain necessary.

The research discusses the adoption of Artificial Intelligence (AI) and its Machine Learning (ML) branch to improve cloud observability functions combined with the automation of root cause analysis. Machine learning systems using data from both historical records and current telemetry measurements maintain abilities to track minimal abnormalities building complex dependency maps, and determine systemic cause origin points through automated learning rules. RNNs and transformers in deep learning practice process high-dimensional time-series data as well as unstructured logs, yet GNNs utilize attention mechanisms to model dependencies across components.

A deep analysis investigates model implementation within five critical areas, which include anomaly discovery from metric and trace data, deep log analysis, automation, and combined safety detection and self-healing infrastructure capabilities. This paper presents real-world production cases that show that AI technology decreases time-to-detect incidents, builds better operational insight, and improves emergency response speed. This work establishes a complete methodology that brings ML-driven observability systems into complex cloud operations while also exploring necessary practical ethical and organizational solutions.

2. Challenges in AI-Driven Anomaly Detection and Root Cause Analysis

AI, together with ML, functions as a revolutionary technology that detects anomalies and determines root causes inside contemporary cloud systems. Modern cloud-native architectures represent environments where these systems achieve success because they process high-dimensional data while managing dynamic conditions and big data sets. The adaptive nature of ML models teaches them to uncover time-based patterns and contextual relationships. At the same time, they evolve with system changes and extract correlations in data to identify anomalies and origin causes with speed surpassing human response times. This passage outlines essential ML-based solutions to tackle the previously described problems by using metrics-based anomaly detection, deep learning on logs, automated RCA, hybrid threat detection, and self-healing infrastructure approaches.

The foundational data signals for cloud monitoring consist of metrics that include CPU utilization together with memory consumption request latency and disk. The analyzed time-series data streams enable the training of ML models that detect abnormal patterns in the recorded data. The anomaly detection techniques of Isolation Forest and Auto encoders, together with DBSCAN, work optimally when operational systems lack abundant labeled anomaly data (Nwachukwu, Durodola-Tunde & Akwiwu-Uzoma, 2024). These methods group correspondences between different behaviors in memory strain nominal patterns for detecting non-normal patterns independently from previous parameter specifications. The training process of a variation autoencoder allows it to reconstruct expected system behavior; reconstruction errors serve as anomaly alerts. These methods deliver strong benefits to systems with changing baseline conditions, including retail applications that experience seasonal traffic fluxes, because they help minimize false alert frequency when compared to static rule systems.

System logs record an account of service activities, including their errors and warnings, with additional messages meant for debugging and user events. Traditional analysis becomes problematic because these structures lack organization. Deep learning methods, particularly transformer-based language models such as BERT and LogBERT, allow the study of log data semantics at a high level. The learning model trains to understand sequence patterns within log messages because it detects anomalies through analysis of message structural and sequential relationships. LogAnomaly, along with DeepLog, proved their effectiveness by understanding normal log sequence patterns to detect minor deviations in actual systems. The detection of anomalies combined with root cause log segment localization through these methods speeds up overall triage procedures.

Root cause analysis represents a fundamental dependency-based problem. The behaviors of a problematic component result in observable consequences all through the system. ML approaches using Graph Neural Networks (GNNs) and attention mechanisms enable the modeling of service-to-service relationships in addition to detecting the root causes between components (Pentyala, 2024). System telemetry enables these analytical models to produce a dynamic service topology that serves as a basis for identifying abnormality occurrences at their probable point of origin. The attention-based models would allow users to track predominant anomaly signals from the nodes with the greatest participation level. These systems deliver better results than linear rule systems and manual impact analysis because they analyze correlations between metrics traces and logs.

3. Solutions in AI-Driven Anomaly Detection and Root Cause Analysis

Cloud environments showcase security threats by presenting themselves as abnormal behavioral patterns that exist throughout various telemetry streams. Security systems that unite rule-based mechanisms and behavioral ML algorithms can detect intricate attack paths, which include credential theft and data stealing activities alongside lateral access phases. Security systems utilize UBA and UEBA for baseline development to discover any abnormal user or entity activity. The combination of clustering analysis reveals anomalies when an EC2 instance downloads excessive data amounts during unusual times, which connects to authentication problems and DNS request abnormalities to confirm security breaches. Deep learning models that analyze multi-modal data consisting of logs and network metrics alongside access control changes improve detection reliability while reducing false positives through their smart learning approach alongside traditional SIEM systems (Hameed & Suleman, 2019).

AI systems use their capabilities to find anomalies and their original causes while simultaneously running automated solutions for correction. Reinforcement Learning delivers a method to teach agents about correction methods through environmental interactions. Cloud systems enable the training of RL agents to execute maintenance operations like pod reboots and service resource scaling depending on the identified anomaly types. The receiving agent receives assessment metrics from performance indicators, which enables it to discover optimal system maintenance policies. RL-inspired mechanisms that operate within Kepler and Amazon DevOps Guru are testing self-healing system approaches for autonomous problem detection alongside diagnosis and autonomous fixes because this approach reduces human involvement and improves system self-resilience.

Multiple data streams demonstrate how anomalies appear separately because they seldom emerge alone during their occurrence. Latency spikes bring about three main side effects: log errors, modified service call patterns, and infrastructure events that cause instance restarts. The manual task of linking anomalies across different data domains proves difficult to execute. Multi-view learning models can analyze multiple telemetry inputs through their joint representation learning process. These models create visualization outputs that show engineers both. Relationship networks and time-based anomaly sequences combine to give engineers a clear understanding of what occurred and its locations and causes. Anodot and AIOps platforms, along with other tools, are now using these methods to assist teams with transitioning from reactive monitoring to proactive event-based operations.

4. Case Studies and Examples

This section illustrates the functional worth of AI-driven detection systems and analysis methods through practical examples from major technology providers. Different organizations from various industries show how they benefit from ML technology to boost observability capabilities together with automated diagnosis systems and cloud-native operation resilience.

Adobe distributes its SaaS product suite Creative Cloud and Adobe Experience Manager, which operate globally in hybrid and multi-cloud environments. The assorted infrastructure environments posed difficulties when trying to identify the underlying causes of issues. Adobe introduced its self-built AI-based root cause analysis system named Cloud Intelligence Platform (CIP) to solve this problem. Real-time service dependency graphs are built by the system when it consumes data from various cloud sources, including log traces and metrics (Olateju et al., 2024). The system localizes faults through multiple ML models, which include gradient boosting alongside time-aware graph analysis. CIP's analysis led to identifying the root cause of a latency issue in a misconfigured CDN node, which cut down RCA time from three hours to fifteen minutes or less. Achieving Adobe's service-level agreements becomes faster through this platform while also reducing consequences for customers.

The cloud infrastructure managed by Meta, which operated under the Facebook name, has emerged as one of the most extensive in the world and handles daily data processing amounts of petabytes. The engineering teams at the company built an anomaly detection framework based on DeepLog, through which they introduced an LSTM-powered model that learned standard log sequences connected to different services. The model identifies failure indications through its detection of deviating log sequence patterns after the completion of its training process. DeepLog served as the monitoring solution for monitoring the backend systems of Messenger and Instagram during the production phases. The system identified unexpected log sequences before service performance issues appeared or user quality diminished even without exceeding metric boundaries. The application of semantic anomaly detection on logs enabled the model to detect future service degradations silently before users encountered any issues.

The Uber platform manages thousands of microservices that span multiple zones and regions through daily operations, which total billions of events. The massive amount of RCA required a system called M3 (Metric Monitoring and Management) to manage it. M3 begins anomaly detection without supervision by identifying metric outliers before utilizing dependency-aware correlation to determine the origin points of anomalies. The system merges metrics alongside traces and logs into an event graph at its core that uses a graph traversal solution boosted through ML-based prioritization. Through their system, M3, Uber pinpointed the source of intermittent trip-booking failures to occur in the fare calculator backend service when it processed edgy input data. The high level of correlation led to RCA accuracy as well as faster responses throughout various company operations (Smith, 2021).

As a prominent digital retailer, Zalando operates its entire containerized platform through Kubernetes across European regions. The team from observability built an anomaly detection system that utilized Prometheus metrics together with auto encoders along with PCA (Principal Component Analysis) for training. Through continuous monitoring of resources pod, activity, and container starts, the system indicated potential deployment instability. Traditional tools failed to notice a memory leak in the payment service because the problem developed slowly until the auto-encoder-based anomaly detector identified the issue before the application crashed. The auto-encoder-based anomaly detection system identified behavioral anomalies through latent space modifications since it detected abnormalities before the application failed while allowing hot-patching without affecting customers. The unsupervised anomaly detection solutions at Zalando cut down downtime occurrences while providing better Kubernetes cluster postmortem analysis.

Alibaba Cloud utilizes multi-modal anomaly detection to monitor all security aspects of its infrastructure-as-a-service platform at scale. The deep learning architecture, which merges CNNs for spatial data evaluation of network flow graphs with RNNs for analyzing temporal behavior of login patterns, enables the system to process access logs API, calls, and network traffic and VM metrics. Insider threats alongside Advanced Persistent Threats (APTs) become detectable using the system since it analyzes service and user behavioral patterns. The system enabled Alibaba Cloud to identify a hidden resource hijacking action through its detection of activity that had typical batch job characteristics. The security team successfully handled the threat incident through deep ML integration, which proved the necessity of applying AI to current security operations.

5. Ethical and Implementation Considerations

Hosting AI implementation in observability programs generates exciting opportunities for organizations. Still, it creates several risks that stem from data management challenges as well as algorithm visibility issues, operational intricacies, and dependency concerns on vendors. Mandatory proactive identification and resolution of these considerations will help establish responsible and sustainable operations for ML-based systems. The following analysis details ethical and technical barriers faced by organizations that implement AI-based anomaly detection with RCA tools.

Anomaly detection and RCA systems need persistent access to logs and traces alongside metrics for their operations, with the potential presence of sensitive information, including PII and user activity logs, authentication records, and request data payloads. The data moves between multiple cloud regions, which follow separate privacy regulations, including GDPR from the EU, CCPA from California, and PIPL from China. The usage of AI systems that depend on this data requires the implementation of privacy protection methods through anonymization, anonymization, and federated learning type approaches (Sambamurthy, 2024). The implementation of complete compliance presents significant difficulties, particularly when using historical material to train the system. The exposure of sensitive information occurs when improper access controls combine with faulty logging misconfigurations to let unauthorized persons and external attackers' access confidential information. Organizations need to design implementation practices that guarantee privacy during their observability pipeline operations because this reduces their exposure to legal and reputational dangers.

The biases present in training data get inherited by any ML model built from those data sets. When anomaly detection models receive training data with statistical imbalances, they tend to produce abnormality alerts that affect specific computational services cl, client populations, or geographically distinct regions. A North American cloud workload training model typically misses normal patterns appearing in European or Asian cloud operations while making them seem anomalous. The performance of models specifically created to decrease as they attempt to work with container-based or serverless frameworks. The issues with biased models prevent their effective use for both efficient alert discrimination and proper execution of remedial actions (Esposito, 2024). Continuous auditing, together with retraining operations, remains vital because it helps address drift issues while maintaining fairness. For acceptable model assessment procedures, operators should validate across different monitoring environments and regional domains. At the same time, explainable alerting systems and Root Cause Analysis outputs must be provided to retain confident operator engagement and operational responsibility.

Deep neural networks and transformers operate as black box models because they constitute some of the most accurate ML systems. Production environments face difficulties when operators need to understand model reasons for detecting anomalies or suggesting remediation because the system lacks interpretability. Engineers tend to avoid restarting containers or throttling traffic when AI systems fail to provide adequate justification for these actions. The development of AI observability systems requires transparent capabilities, model confidence scores, and attention heat maps, to establish trust between users and the platform. The combination of statistical rules with ML outputs enables more transparent detection systems that maintain operational clarity alongside predictive power.

The produced ML models function as dynamic systems because they become less effective when workloads change together with infrastructure and user behaviors transform. A lack of model monitoring combined with improper retraining and versioning methods makes the systems stale while causing the systems to identify wrong patterns or to miss signals. The accumulation of "MLOps debt" occurs when maintaining ML systems becomes more expensive than the actual value they provide to operations. Model tracking needs to use precision-recall metrics made for anomaly detection tasks alongside postmortem evaluations of AI outputs rather than human decisions. The absence of AI operations maturity investment causes organizations to release models that intensify observability challenges instead of resolving them (Adenekan, 2024).

IBM Watson and other AI observability tools need staff from five separate teams to work together: data scientists, SREs, DevOps practitioners, security analysts, and product owners. Numerous organizations face obstacles when trying to merge different roles because they lack essential cross-team infrastructure combined with shared technical vocabulary. The inability of operations teams to understand machine learning anomaly alerts generated from JSON logs matches the data scientist's difficulty in labeling edge cases without understanding specific infrastructure domains. Organizational AI tool utilization requires proper staff documentation alongside unified operations procedures between teams to enable the appropriate usage and interpretation of these tools. A successful trust-building process requires AI systems to function through shadow mode before they can receive ultimate control over automated operations.

6. Conclusion

Directionless cloud infrastructure growth, its dynamic character, and the growing distribution challenge existing system observation and maintenance methods. Nowadays, organizations cannot depend on rule-based alerts together with threshold-driven dashboards or manual log inspection to identify subtle performance problems, security threats, and system failures in real-time. Organizations face a major change in cloud-native environment management because they now integrate Artificial Intelligence (AI) and Machine Learning (ML) into observability pipelines.

The paper analyzed the core obstacles faced during anomaly detection and root cause analysis (RCA) operations, which mainly involve telemetry noise, dependency complexity, false positives, fragmented observability, undetected threats, and slow RCA workflows. These obstacles emerge because modern interconnected systems become too massive to identify problems that appear indirectly or through unexpected means. ML technologies with deep learning capabilities over log data and metrics and trace information solve these difficulties by developing abnormal and normal patterns and establishing domain interconnection, then revealing the root causes much faster than traditional solutions.

Systems based on proposed solutions employ unsupervised learning for metrics anomalies and graph-based RCA and hybrid security detection systems for analysis. AI develops by means of reinforcement learning and self-healing infrastructure to demonstrate its anticipated ability to diagnose problems while repairing them independently. The transition brings several risks to operations. The ethical and implementation concerns related to privacy in data and model biases, explainability issues, MLOps debts, and vendor lock-in need solutions that combine solid governance practices with functional team collaboration supported by transparent and fair AI frameworks. The success of ML-driven observability depends equally on proper implementation within human systems and digital infrastructure that supports current operational practices.

AI-based anomaly detection, along with Reasonable Cause Analysis, will develop into fundamental elements for infrastructure that build self-management capabilities and demonstrate resilience. Organizations that begin implementing these operational capabilities in their models early on will generate substantial advantages regarding operational performance, service reliability, and response speed. Cloud observability will transform into autonomous systems that both understand and predict as well as actively respond to situations.

References

[1] Adenekan, T. K. (2024). Explainable AI Techniques for Root Cause Analysis in Complex Systems.

[2] Esposito, J. (2024). Real-Time Monitoring and Root Cause Analysis in AI-Driven DevOps.

[3] Hameed, A., & Suleman, M. (2019). AI-Powered Anomaly Detection for Cloud Security: Leveraging Machine Learning and DSPM.

[4] Nwachukwu, C., Durodola-Tunde, K., & Akwiwu-Uzoma, C. (2024). AI-driven anomaly detection in cloud computing environments.

[5] Olateju, O., Okon, S. U., Igwenagu, U., Salami, A. A., Oladoyinbo, T. O., & Olaniyi, O. O. (2024). Combating the challenges of false positives in AI-driven anomaly detection systems and enhancing data security in the cloud. *Available at SSRN 4859958*.

[6] Pentyala, D. K. (2024). Artificial Intelligence for Fault Detection in Cloud-Optimized Data Engineering Systems. *International Journal of Social Trends*, *2*(4), 8-44.

[7] Sambamurthy, P. (2024). Advancing Systems Observability through Artificial Intelligence: A Comprehensive Analysis. *ResearchGate, August*.

[8] Smith, B. (2021). Deep Learning for Automated Anomaly Detection in Root Cause Analysis.