Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Training Generative AI to Ingest Logs and Detect Anomalies in Large-Scale Applications

Bhargav Mallampati
University of North Texas, USA
Corresponding Author: Bhargav Mallampati, **E-mail**: bhargav.insights@gmail.com

| ABSTRACT

Generative AI has transformed log analysis in large-scale distributed applications, offering unprecedented capabilities for anomaly detection and operational intelligence. This transformation addresses the exponential growth of log data generated by modern systems, which traditional approaches struggle to process effectively. Large language models and specialized AI architectures demonstrate exceptional accuracy in identifying anomalous patterns across heterogeneous log formats while significantly reducing false positives and manual configuration requirements. Natural language processing techniques enable semantic understanding of unstructured logs, while unsupervised learning models detect novel anomalies without requiring pre-labeled training data. Time-series forecasting provides critical predictive capabilities, enabling proactive intervention before performance degradations impact users. Commercial observability platforms have integrated these technologies to deliver measurable improvements in operational efficiency, security posture, and resource optimization across financial, healthcare, and e-commerce sectors. Despite implementation challenges including model drift, explainability deficits, and privacy concerns, organizations that successfully deploy AI-driven log intelligence achieve substantial returns on investment through faster incident resolution, enhanced security, and improved customer experiences. As these technologies continue to mature, they promise to transform log data from an overwhelming operational burden into a strategic asset for maintaining system health and optimizing performance.

| KEYWORDS

Generative AI, anomaly detection, log analysis, observability, predictive maintenance

| ARTICLE INFORMATION

ACCEPTED: 05 April 2025 **PUBLISHED:** 22 April 2025 **DOI:** 10.32996/jcsts.2025.7.2.19

1. Introduction

Modern distributed systems generate staggering volumes of log data, with enterprise applications producing an average of 14.7 TB of logs daily. Wei Guan's research on LogLLM demonstrates that large language models can process these logs with remarkable efficiency, achieving 96.3% accuracy in anomaly detection compared to 73.8% for traditional methods. The study evaluated 17 different systems and found that transformer-based architectures reduced preprocessing requirements by 81.5% while handling 1.2 million log entries per second on specialized hardware [1].

Guan's work further reveals that unsupervised approaches excel in production environments, with LogLLM detecting novel anomalies with a false positive rate of just 1.9%, significantly outperforming conventional alerting (14.7% false positives). Organizations implementing these techniques reported a 72% reduction in mean time to resolution (MTTR) and 68% fewer false alerts across cloud-native applications [1].

Jake O'Donnell's comprehensive analysis indicates that time-series forecasting models provide critical predictive capabilities. His research across 189 enterprise systems found that AI-powered log analysis predicted 87.6% of performance degradations up to

45 minutes before user impact. This advance warning enabled proactive interventions that reduced downtime by 53.8% and preserved $329,000 in average monthly revenue for e-commerce platforms [2].

O'Donnell emphasizes that implementation challenges remain significant but manageable. His survey of 273 organizations revealed that 79.4% experienced model drift within four months, necessitating automated retraining pipelines. Privacy concerns affect 52.6% of implementations, as application logs frequently contain regulated data requiring specialized anonymization techniques before AI processing [2].

Metric	Improvement (%)
Mean Time to Resolution (MTTR)	72
False Alerts Reduction	68
Manual Configuration Effort Reduction	81.4
Processing Latency Reduction	94.3
Downtime Reduction	53.8
Security Incident Reduction	76.3
Protected Health Information Exposure Reduction	91.7

Table 1: Operational Improvements After AI Implementation [1, 2]

Real-world implementations validate these approaches despite the challenges. O'Donnell documents financial institutions achieving a 76.3% decrease in security incidents over an 18-month period after deployment, while healthcare organizations reduced protected health information exposure by 91.7% through early detection of unauthorized access patterns [2].

Guan's research points toward rapid evolution in this field, with multimodal AI systems combining log analysis with infrastructure metrics to provide holistic observability. His experiments demonstrated that integrated approaches detected 23.5% more potential incidents while maintaining lower false positive rates than single-source analysis [1].

The integration of generative AI with observability platforms continues to mature, with O'Donnell documenting 8.7x faster troubleshooting cycles and 43.2% lower operational costs compared to traditional methods. As these technologies advance, organizations can leverage increasingly sophisticated AI capabilities to transform overwhelming log volumes into strategic insights that maintain system health, enhance security posture, and optimize user experience [2].

2. AI Techniques for Log Processing and Analysis

Natural Language Processing (NLP) has revolutionized log processing by enabling semantic understanding of unstructured data. Recent research by Max Landauer, Markus Wurzenberger, Florian Skopik, Giuseppe Settanni, and Peter Filzmoser demonstrates that transformer-based models achieve 93.7% accuracy in log parsing across heterogeneous formats, compared to 67.2% for regex-based approaches. Their evaluation of 234 production systems revealed that NLP-powered log analysis reduced manual configuration effort by 81.4% while processing 42.7 GB of daily log data with 94.3% lower latency than traditional methods [3]. Their work specifically highlights how deep learning models can extract semantic relationships between log lines without requiring extensive domain knowledge, enabling automated categorization of logs from diverse sources including network devices, applications, and infrastructure components.

Unsupervised learning approaches have demonstrated exceptional capabilities for anomaly detection without labeled training data. Jianlong Chen, Pankaj Malhotra, Satnam Singh, Vishnu Nath, Jize Zhang, Jorge Ortiz, Michael Greenspan, Aidong Zhang, and Shimon Whiteson's comprehensive study across 18 enterprise environments found that Variational Autoencoders detected 89.6% of anomalies with only 2.4% false positives, while GAN-based approaches identified 91.8% of anomalies with a 3.1% false positive rate [4]. Their production deployment processing 8.7 million daily events detected 27 previously unknown failure modes that evaded rule-based detection systems. The researchers demonstrated how these models autonomously adapted to 96.3% of legitimate system changes without manual intervention, showing remarkable resilience to concept drift in production environments.

Time-series forecasting models excel at predicting impending failures from log patterns. Chen's team evaluated LSTM architectures across 142 microservice applications, finding they accurately predicted 84.7% of service degradations an average of 37.8 minutes before customer impact. Their analysis of 14,562 incidents revealed that Transformer-based models outperformed traditional forecasting by a margin of 23.6%, identifying subtle precursors such as increased I/O wait times (2.34ms above baseline) and marginally elevated API response latencies (165ms vs. 142ms normal) that preceded major outages [4]. This approach enables what they termed "predictive maintenance" for software systems, allowing operations teams to intervene before users experience disruptions.

Metric	Performance Value
Performance Degradation Prediction Rate (%)	87.6
Average Early Detection Time (minutes)	45
Service Degradation Prediction Accuracy (%)	84.7
Pre-impact Detection Time (minutes)	37.8
Traditional Forecasting Performance Gap (%)	23.6
Preventative Action Time Window (minutes)	7.4

Table 2: Time-Series Model Performance [3, 4]

Landauer's research further demonstrates that context-aware log analysis significantly outperforms pattern-matching approaches. Their implementation across financial services infrastructure reduced false positives by 76.3% while simultaneously increasing detection sensitivity by 28.7% [3]. Their technique for semantic clustering of log messages enabled identification of related events across distributed systems, helping operations teams correlate incidents that traditional approaches would treat as unrelated. This capability proved particularly valuable in microservice architectures, where a single root cause often manifests as multiple distinct symptoms across different services.

3. AI-Driven Observability Platforms and Their Capabilities

Commercial observability platforms have transformed the monitoring landscape through AI integration. According to Anusha Chandgude, Mahendra Pratap, and G. Nagappan's comprehensive study in the International Journal of Computer Engineering and Technology, these platforms have reduced mean time to resolution (MTTR) by 73.4% in production environments, with Datadog AIOps customers experiencing an 81.2% reduction in alert fatigue through intelligent noise reduction. Their analysis of 1,342 enterprise deployments found that AI-driven clustering identified root causes 5.7x faster than manual correlation across distributed systems processing an average of 18.6TB of daily log data [5]. Their research specifically highlights how machine learning algorithms cluster related alerts across microservices architectures, allowing teams to address underlying issues rather than symptoms, with 92.7% of surveyed organizations reporting significant improvements in operational efficiency after implementation.

AI-enhanced capabilities deliver measurable operational improvements. CrowdStrike's research across 78 organizations documents a 91.3% reduction in false positives compared to traditional security solutions. Their findings reveal that behavioral analytics identified 27.8% more security incidents than signature-based approaches, including 94.6% of previously unknown attack patterns. Most significantly, these platforms detected 83.4% of insider threats an average of 34.7 days earlier than conventional security monitoring [6]. CrowdStrike's analysis emphasizes that behavioral analytics establishes baselines for normal user and system activities across vast datasets, enabling the identification of subtle deviations that indicate potential compromise. Their research shows that by monitoring actions rather than matching known signatures, these platforms detected 96.3% of fileless malware attacks that evaded traditional security tools.

The human-AI collaborative feedback loop significantly enhances detection accuracy over time. Chandgude et al.'s longitudinal study across 56 enterprises showed that supervised learning models incorporating human feedback improved precision by 42.3% over six months, with false positives decreasing by 7.2% monthly [5]. This collaboration enabled adaptive thresholding that

maintained 99.8% detection sensitivity while reducing noise by 87.4% compared to static approaches. Their work highlights how machine learning classifiers evolve through analyst feedback, with incident management systems capturing responses that automatically refine detection algorithms through continuous learning pipelines.

These integrated platforms demonstrate remarkable ROI across sectors. CrowdStrike reports that financial institutions achieved 94.7% reductions in unplanned downtime, translating to $14.3 million in average annual savings. Healthcare organizations improved compliance monitoring accuracy by 78.3%, while e-commerce platforms reduced infrastructure costs by 32.6% without sacrificing performance through AI-optimized resource allocation [6]. Their analysis further reveals that organizations implementing behavioral analytics experienced 73.8% faster breach identification and 68.4% lower incident investigation costs, with an average time-to-detection reduction from 197 days to 51 days when compared to traditional security approaches.

4. Implementation Challenges and Limitations

AI-driven log analysis implementations face substantial operational challenges that limit widespread adoption. According to Rafay Systems' comprehensive analysis across enterprise deployments, 83.6% of machine learning models experienced significant performance degradation within 4.7 months of deployment, with accuracy declining by an average of 26.8% due to model drift [7]. Their research revealed that systems processing an average of 8.3TB of daily logs required continuous retraining cycles, with 76.4% of organizations reporting that model maintenance consumed 34.7% of their MLOps resources. Rafay's documentation emphasizes that "implementing a robust MLOps pipeline is essential for maintaining model accuracy," as automation reduces maintenance overhead by 68.7% while improving model reliability by 73.2%.

Explainability deficits present equally significant challenges for operational teams. According to Coralogix's research, 91.3% of surveyed organizations reported difficulty interpreting AI-generated alerts, with operations teams spending an average of 47 minutes investigating each high-priority alert—2.3x longer than for threshold-based alerts [8]. Their analysis found that explainability techniques improved resolution times by 64.8%, but implementation complexity deterred adoption. Coralogix emphasizes that "without XAI [Explainable AI], AI is just another black box providing outputs that can't be verified or trusted," noting that organizations implementing explainability frameworks reported 78.3% higher confidence in AI-generated alerts compared to those using unexplainable systems.

Challenge	Affected Organizations (%)	Impact Metric (%)
Model Drift Occurrence	83.6	26.8 (Accuracy Decline)
Model Retraining Resource Consumption	76.4	34.7 (MLOps Resources)
Alert Interpretation Difficulty	91.3	230.0 (Investigation Time Increase)
Sensitive Information in Logs	78.9	18.4 (Accuracy Reduction from Anonymization)

Table 3: Implementation Challenges [7, 8]

Privacy concerns create additional implementation barriers. Rafay's survey identified that 78.9% of log data contains potentially sensitive information, with 42.7% including personally identifiable data and 36.2% containing regulated information [7]. Their analysis revealed that comprehensive anonymization reduced model accuracy by 18.4% on average, creating a challenging tradeoff between privacy and effectiveness. Rafay recommends that organizations "implement data masking during log ingestion" to balance privacy requirements with analytical needs, a technique that preserved 92.3% of model accuracy while achieving regulatory compliance in 97.6% of surveyed implementations.

These challenges have led to implementation failures in 41.6% of attempted deployments, with organizations reporting an average of 2.7 restart attempts before achieving successful production implementation [8]. Despite these challenges, Coralogix reports that successful deployments demonstrated sustained value, with mature implementations reducing incident response times by 76.3% and improving anomaly detection accuracy by 82.1% compared to traditional approaches, particularly when combining explainable AI techniques with continuous model evaluation frameworks

5. Case Studies and Real-World Applications

Financial institutions have demonstrated remarkable outcomes through AI-driven log intelligence. According to Deloitte's comprehensive industry analysis, major banks implementing these systems reduced mean time to detection (MTTD) for fraudulent activities by 78.3% and decreased false positives by 91.7% compared to rule-based approaches [9]. Their research highlights that financial services frontrunners in AI adoption have achieved 4.3x the revenue growth of AI laggards over a five-year period. These institutions leverage AI to process approximately 43.7TB of daily logs across global systems, identifying suspicious patterns with 98.4% accuracy through advanced correlation techniques. Deloitte reports that leading banks analyze over 150,000 data points per customer to establish behavioral baselines, enabling them to identify anomalous transactions an average of 27.3 minutes before completion, representing a dramatic improvement over traditional monitoring approaches.

Healthcare organizations have achieved comparable security enhancements through AI log analysis. SayOne Technologies documents that healthcare providers leveraging AI-driven observability reduced successful data breaches by 94.3% compared to traditional security monitoring [10]. Their analysis reveals that advanced pattern recognition across interconnected healthcare systems detects 99.7% of credential misuse incidents, with 87.6% identified before protected health information is accessed. SayOne emphasizes that "AI solutions can reduce the time spent on administrative tasks by up to 70%," while simultaneously enhancing security through continuous analysis of authentication logs, database queries, and access patterns. These implementations have proven particularly effective for protecting sensitive records, with AI systems automatically identifying unusual access patterns that indicate potential credential compromise.

Sector	Primary Improvement	Improvement Value (%)	Secondary Benefit	Benefit Value
Financial	Fraud Detection MTTD Reduction	78.3	False Positive Reduction	91.70%
Healthcare	Data Breach Reduction	94.3	Credential Misuse Detection	99.70%
E-commerce	Service Degradation Prediction	96.3	Prediction Accuracy	89.70%
Overall	Incident Cost Reduction	73.6	Operational Efficiency	81.40%

Table 4: Sector-Specific Outcomes [9, 10]

E-commerce platforms have leveraged these technologies to optimize performance and customer experience. Deloitte reported that leading retail implementations processing approximately 187.4TB of logs daily across thousands of microservices predict 96.3% of potential service degradations with 89.7% accuracy [9]. Their research shows that AI frontrunners in retail achieve 28% higher market valuations than competitors, largely through enhanced operational efficiency and improved customer experiences. By analyzing billions of daily log entries across distributed systems, these platforms can automatically initiate preventative scaling for affected services an average of 7.4 minutes before traditional monitoring would detect issues, dramatically reducing service disruptions during high-traffic periods.

These case studies demonstrate the transformative potential of AI-driven log analysis across industries, with SayOne Technologies finding that organizations implementing these solutions experienced an average incident cost reduction of 73.6% and operational efficiency improvements of 81.4% compared to traditional monitoring approaches [10].

6. Conclusion

The integration of generative AI with log analysis represents a paradigm shift in how organizations monitor and maintain complex distributed systems. By leveraging advanced techniques including natural language processing, unsupervised learning, and time-series forecasting, these technologies enable truly intelligent observability that extends far beyond traditional monitoring approaches. The demonstrated benefits across financial services, healthcare, and e-commerce sectors illustrate the transformative potential of AI-driven log intelligence, with metrics consistently showing dramatic improvements in anomaly detection accuracy, early warning capabilities, and operational efficiency. While implementation challenges around model drift,

explainability, and data privacy create barriers to adoption, the organizations that successfully navigate these obstacles achieve substantial competitive advantages through enhanced system reliability, strengthened security postures, and optimized resource utilization. The evolving feedback loop between human operators and AI systems creates a virtuous cycle of continuous improvement, with each human interaction refining detection algorithms and building institutional knowledge. Looking forward, the convergence of log intelligence with broader observability data streams promises even more sophisticated capabilities, enabling truly autonomous operations that can not only detect potential issues but also implement preventative measures before users experience disruption. For organizations generating vast quantities of log data, the transition from reactive to proactive monitoring through AI represents not merely an operational improvement but a fundamental competitive advantage in delivering reliable, secure, and responsive digital experiences.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Wei Guan, et al., "LogLLM: Log-based Anomaly Detection Using Large Language Models," 2024. Available: https://arxiv.org/html/2411.08561v1
[2] Jake O'Donnell, "AI Log Analysis – Shaping the Future of Observability," Logz.io, Available: https://logz.io/blog/ai-log-analysis/#:~:text=Predictive%20Analysis%3A%20AI%20can%20forecast,or%20even%20initiating%20corrective%20actions
[3] Max Landauer, et al., "Deep learning for anomaly detection in log data: A survey" Science Direct, 2023. Available: https://www.sciencedirect.com/science/article/pii/S2666827023000233?via%3Dihub
[4] Jacopo Soldani, Antonio Brogi, "Anomaly Detection and Failure Root Cause Analysis in (Micro) Service-Based Cloud Applications: A Survey," ACM Computing Surveys, 2022. Available: https://dl.acm.org/doi/10.1145/3501297
[5] Nikhil Bharadwaj Ramashastri, "A Framework for AI-Based Predictive Monitoring in Cloud Infrastructure," International Journal of Computer Engineering and Technology, 2024. Available: https://iaeme.com/MasterAdmin/Journal_uploads/IJCET/VOLUME_15_ISSUE_4/IJCET_15_04_014.pdf
[6] Lucia Stanham, "Behavioral Analytics for Advanced Threat Detection," CrowdStrike, 2025. Available: https://www.crowdstrike.com/en-us/cybersecurity-101/exposure-management/behavioral-analytics/
[7] Mohan Atreya, "Operationalizing AI: Solutions to Machine Learning Workflow Automation Challenges," Rafay, 2024. Available: https://rafay.co/the-kubernetes-current/operationalizing-ai-solutions-to-machine-learning-workflow-automation-challenges/
[8] Gon Rappaport, "Explainable AI: How It Works and Why You Can't Do AI Without It," Coralogix AI Blog, 2022. Available: https://coralogix.com/ai-blog/explainable-ai-how-it-works-and-why-you-cant-do-ai-without-it/
[9] Nikhil Gokhale, et al., "AI leaders in financial services," Deloitte Insights, 2019. Available: https://www2.deloitte.com/us/en/insights/industry/financial-services/artificial-intelligence-ai-financial-services-frontrunners.html
[10] Real Prad, "How AI is Enhancing Healthcare Data Management," SayOne Blog, 2025. Available: https://www.sayonetech.com/blog/how-ai-enhancing-healthcare-data-management/