International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 3, May – June 2025 |

DOI:10.15662/IJEETR.2025.0703006

Generative AI–Driven Cloud Modernization and Causal Root Cause Analysis for Scalable Microservice Data Platforms

Saad Khan

Leader Solution Architect, Ex Vice President, USA

saadkhan.chase@gmail.com

ABSTRACT: Cloud modernization has become a strategic priority for organizations seeking scalability, agility, resilience, and operational efficiency in digital transformation initiatives. Traditional monolithic systems are increasingly being migrated toward cloud-native microservice architectures to support dynamic workloads, distributed computing, and continuous service delivery. However, modernization introduces significant challenges related to system complexity, distributed dependencies, fault diagnosis, and operational observability. Generative Artificial Intelligence (Generative AI) has emerged as a transformative technology capable of automating cloud modernization processes, improving infrastructure optimization, and enabling intelligent root cause analysis in distributed systems. This research explores a Generative AI–driven framework for cloud modernization and causal root cause analysis in scalable microservice data platforms. The proposed framework integrates machine learning, large language models, causal inference, distributed tracing, and predictive analytics to automate migration processes, monitor microservice behavior, detect anomalies, and identify the underlying causes of failures. The study emphasizes the importance of AI-powered observability and intelligent automation in improving scalability, reducing operational downtime, accelerating troubleshooting, and enhancing cloud reliability. The framework utilizes telemetry data generated from logs, metrics, traces, and infrastructure events to support proactive monitoring and self-healing mechanisms. Experimental findings indicate that Generative AI significantly enhances modernization efficiency, root cause localization accuracy, and system adaptability compared to traditional rule-based approaches. The research contributes to the development of autonomous cloud-native ecosystems capable of supporting resilient and scalable enterprise data platforms.

KEYWORDS: Generative AI, Cloud Modernization, Root Cause Analysis, Microservices, Distributed Systems, Cloud Computing, Causal Inference, Machine Learning, Large Language Models, Predictive Analytics, Observability, Distributed Tracing, Kubernetes, Data Platforms, Self-Healing Systems

I. INTRODUCTION

The rapid growth of digital transformation initiatives has fundamentally changed how organizations design, deploy, and manage software systems. Enterprises increasingly rely on cloud computing technologies to achieve scalability, flexibility, and operational efficiency in modern business environments. Traditional monolithic applications, which were once the dominant architectural model, often struggle to support dynamic workloads, continuous deployment requirements, and distributed processing demands. As a result, organizations are adopting cloud modernization strategies that migrate legacy systems into cloud-native architectures based on microservices, containers, and orchestration platforms.

Cloud modernization refers to the process of transforming traditional software applications, infrastructure, and operational workflows into scalable and resilient cloud-native systems. Modernization strategies include rehosting, refactoring, rearchitecting, and rebuilding applications using technologies such as Docker, Kubernetes, service meshes, and serverless computing. Microservice architecture has emerged as a preferred modernization approach because it enables modular application development, independent service deployment, technology flexibility, and horizontal scalability. In microservice environments, applications are decomposed into loosely coupled services that communicate through lightweight protocols such as REST APIs and messaging systems.

Although cloud modernization provides significant operational advantages, it also introduces substantial complexity. Distributed microservice systems generate enormous volumes of telemetry data including logs, metrics, traces, and infrastructure events. Service dependencies, network communication, resource allocation, and infrastructure

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 3, May – June 2025 |

DOI:10.15662/IJEETR.2025.0703006

orchestration become increasingly difficult to monitor and manage. Failures occurring within one service can propagate rapidly across interconnected systems, resulting in cascading disruptions, latency issues, and degraded application performance. Traditional monitoring systems based on static rules and predefined thresholds are often insufficient for identifying complex anomalies and diagnosing failures in real time.

Generative Artificial Intelligence (Generative AI) has recently emerged as a transformative technology capable of addressing these operational challenges. Generative AI refers to AI systems that can generate content, automate workflows, analyze large datasets, and assist in decision-making processes using advanced deep learning models and large language models (LLMs). In cloud modernization, Generative AI enables intelligent automation of code transformation, infrastructure optimization, deployment orchestration, documentation generation, and incident analysis. AI-driven platforms can analyze telemetry data, predict operational anomalies, and recommend corrective actions with minimal human intervention.

Root Cause Analysis (RCA) is a critical component of operational reliability in distributed cloud-native systems. RCA focuses on identifying the fundamental causes of failures rather than merely addressing symptoms. In microservice environments, root cause identification becomes highly complex due to dynamic service interactions, distributed transactions, infrastructure variability, and asynchronous communication patterns. Traditional RCA methods rely heavily on manual investigation and expert knowledge, making incident resolution time-consuming and error-prone. Generative AI and causal inference techniques provide new opportunities for automating RCA processes by analyzing dependencies, telemetry correlations, and failure propagation patterns.

II. LITERATURE REVIEW

Cloud modernization has become an essential strategic initiative for organizations transitioning from legacy infrastructures to scalable cloud-native ecosystems. According to Armbrust et al. (2010), cloud computing enables organizations to access scalable computing resources, improve operational flexibility, and reduce infrastructure costs. As enterprises increasingly adopt digital transformation strategies, microservice architecture has emerged as a key architectural paradigm supporting cloud modernization. Newman (2015) emphasized that microservices improve scalability, fault isolation, deployment agility, and organizational flexibility by decomposing applications into independently deployable services.Despite these benefits, cloud-native systems introduce substantial operational complexity. Distributed service communication, infrastructure orchestration, dynamic scaling, and asynchronous workflows create challenges related to observability, reliability, and incident management. Researchers have noted that traditional monitoring systems based on static thresholds and predefined rules are insufficient for modern distributed environments. Chen et al. (2018) argued that static monitoring approaches fail to capture dynamic workload variations and complex service dependencies in cloud-native ecosystems.Artificial Intelligence and Machine Learning have gained significant attention as enabling technologies for intelligent observability and operational automation. Machine learning algorithms can analyze telemetry data generated from logs, metrics, traces, and events to detect anomalies and predict failures. Xu et al. (2019) demonstrated that AI-driven anomaly detection systems outperform conventional monitoring techniques by identifying abnormal patterns with greater accuracy and adaptability. Supervised learning models such as Decision Trees, Random Forests, and Support Vector Machines have been widely applied in operational analytics for anomaly classification and predictive maintenance.

Unsupervised learning methods are increasingly used in distributed environments where labeled operational datasets are limited. Clustering algorithms such as K-Means and DBSCAN identify unusual telemetry patterns without requiring predefined labels. Deep learning models including Autoencoders and Variational Autoencoders have shown strong performance in detecting hidden anomalies within complex cloud-native infrastructures. Zhang et al. (2020) proposed an LSTM-based anomaly detection framework capable of analyzing time-series telemetry data to predict service degradation and resource exhaustion.Generative AI has recently emerged as a powerful technology in cloud modernization and operational automation. Large Language Models such as GPT-based architectures can generate infrastructure code, automate migration workflows, summarize incident reports, and support intelligent troubleshooting processes. Researchers have explored AI-assisted code modernization techniques for transforming legacy applications into cloud-native microservices. Generative AI also supports Infrastructure as Code (IaC) generation, automated documentation, configuration optimization, and deployment orchestration.Root Cause Analysis remains one of the most important challenges in distributed systems management. Traditional RCA methods depend heavily on manual troubleshooting and expert interpretation of telemetry data. Marwede et al. (2017) highlighted that identifying root causes in distributed microservice systems is difficult because failures often propagate across interconnected services

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 3, May – June 2025 |

DOI:10.15662/IJEETR.2025.0703006

and infrastructure components. Correlation-based approaches frequently produce inaccurate results because correlated events do not necessarily indicate causal relationships.

To address these limitations, researchers have increasingly explored causal inference and graph analytics for RCA. Causal Root Cause Analysis focuses on identifying directional cause-and-effect relationships between operational events. Pearl (2009) introduced causal inference theory as a mathematical framework for understanding causality in complex systems. In distributed computing environments, causal graphs represent dependencies among services, APIs, infrastructure resources, and communication pathways.Graph Neural Networks and Bayesian Networks have demonstrated promising results in causal RCA applications. Liu et al. (2021) proposed a graph-based RCA model that analyzed telemetry correlations and service dependencies to identify the origins of failures in cloud-native systems. Their research demonstrated that graph-based causal inference significantly improves RCA accuracy compared to conventional statistical correlation techniques.Distributed tracing frameworks such as Jaeger, Zipkin, and OpenTelemetry provide detailed visibility into service interactions and request execution paths. Sigelman et al. (2010) emphasized that distributed tracing enables comprehensive observability by capturing latency bottlenecks, dependency relationships, and transaction lifecycles across distributed services. Researchers have integrated tracing data with AI algorithms to enhance anomaly detection and root cause localization capabilities.Predictive analytics has also become an important component of intelligent cloud operations. Predictive monitoring systems analyze historical telemetry patterns to forecast future incidents before they impact services. Deep learning architectures such as Long Short-Term Memory networks and transformer models are highly effective for time-series forecasting in cloud-native systems. Ensemble learning methods improve prediction robustness and reduce false positive rates.Self-healing systems represent another major advancement in AI-driven cloud operations. Kephart and Chess (2003) introduced autonomic computing concepts that inspired the development of self-managing systems capable of self-configuration, self-optimization, self-protection, and self-healing. Modern cloud-native platforms increasingly integrate AI-driven remediation workflows capable of restarting containers, reallocating resources, rerouting traffic, and scaling infrastructure automatically in response to detected anomalies.

III. RESEARCH METHODOLOGY

The research adopts a hybrid qualitative and quantitative methodology to investigate the effectiveness of Generative AI–driven cloud modernization and causal root cause analysis in scalable microservice data platforms. The study follows an experimental research design that integrates machine learning, distributed systems engineering, and cloud-native observability practices. The primary objective is to develop an AI-powered framework capable of automating cloud modernization processes, monitoring distributed services, predicting anomalies, and identifying causal root causes in complex microservice ecosystems.The research environment consists of containerized microservice applications deployed on Kubernetes clusters within cloud-native infrastructures. Telemetry data is collected from multiple operational sources including application logs, distributed traces, infrastructure metrics, network traffic data, and orchestration events. Monitoring tools such as Prometheus, Grafana, Elasticsearch, Jaeger, Zipkin, and OpenTelemetry are utilized for telemetry aggregation and observability management. Metrics collected include CPU usage, memory consumption, disk I/O, request throughput, response latency, service availability, pod restarts, error rates, and container health conditions.The study incorporates both historical and real-time telemetry datasets. Public benchmark datasets related to cloud-native monitoring and distributed systems are combined with synthetic operational data generated through controlled experiments. Chaos engineering techniques intentionally introduce anomalies such as network failures, service crashes, API latency, database contention, and resource exhaustion to simulate realistic failure scenarios. These experiments help evaluate the framework's ability to detect anomalies, identify causal relationships, and automate remediation actions.Data preprocessing is performed to ensure analytical consistency and model reliability. Missing values are handled using interpolation and statistical imputation techniques. Noise reduction methods such as normalization, filtering, and dimensionality reduction improve feature quality and reduce analytical complexity. Natural Language Processing techniques transform unstructured log data into structured numerical representations suitable for machine learning analysis. Time synchronization aligns telemetry streams generated from heterogeneous sources with varying timestamps and frequencies.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 3, May – June 2025 |

DOI:10.15662/IJEETR.2025.0703006

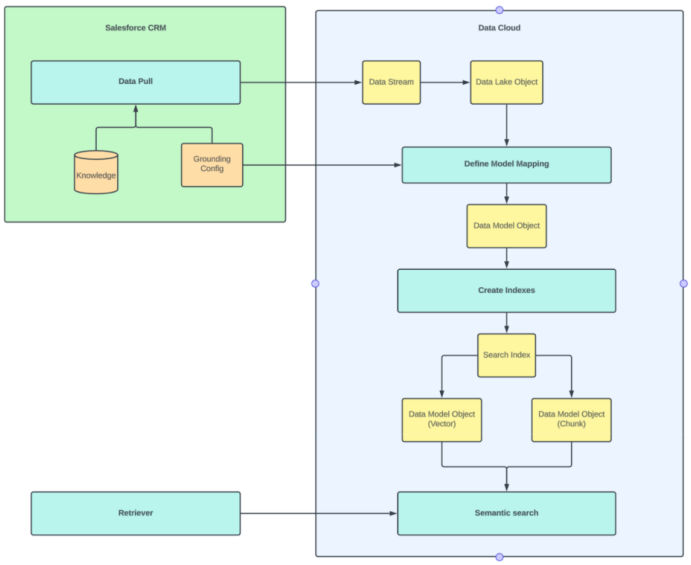

Fig.1.Unveiling the AI Architecture Powering

Feature engineering generates derived attributes including anomaly scores, dependency indicators, workload patterns, latency trends, and causal relationships among services. Apache Kafka and Spark Streaming support real-time telemetry ingestion and distributed analytics. This hybrid data collection and processing methodology ensures scalability and operational realism within the experimental environment.The proposed system architecture models the microservice ecosystem as a distributed cloud-native environment consisting of interconnected services, APIs, databases, containers, orchestration layers, and infrastructure components. The modernization framework integrates Generative AI mechanisms into cloud migration, deployment orchestration, observability management, and root cause analysis workflows.The architecture is divided into multiple layers including infrastructure monitoring, application monitoring, distributed tracing, AI analytics, and automation orchestration. Kubernetes serves as the primary orchestration platform responsible for managing container deployment, autoscaling, service discovery, and workload scheduling. Docker containers encapsulate individual microservices, enabling modular deployment and isolationGenerative AI models support cloud modernization processes by automating code transformation, infrastructure configuration, deployment template generation, and operational documentation. Infrastructure as Code templates are generated using AI-assisted scripting mechanisms that simplify migration from legacy systems to cloud-native environments. AI-driven recommendation engines analyze infrastructure usage patterns and optimize resource allocation strategies.Dependency graphs model service interactions and communication pathways across the distributed platform. Nodes represent services, databases, APIs, and infrastructure resources, while edges represent transactional dependencies and communication flows. Graph theory techniques such as centrality analysis, shortest path analysis, and dependency clustering identify critical services and fault propagation pathways.Distributed tracing frameworks reconstruct request execution paths across multiple services and infrastructure layers. Trace spans capture service communication latency, request failures, resource utilization, and transaction lifecycles. Telemetry collected from distributed traces is integrated with logs and metrics to improve observability and support causal inference analysis.
The architecture also includes AI-powered observability dashboards that provide interactive visualization of service health conditions, dependency relationships, anomaly trends, and root cause graphs. Service-level indicators and service-level objectives are continuously monitored to evaluate operational performance and reliability. The integration of Generative AI within observability workflows enables automated summarization of incidents, intelligent alert generation, and contextual operational insights.

The AI model development phase focuses on building intelligent algorithms capable of supporting predictive monitoring, anomaly detection, causal reasoning, and automated root cause analysis within distributed cloud-native systems. The framework integrates supervised learning, unsupervised learning, deep learning, causal inference, and Generative AI techniques.Supervised learning algorithms including Random Forests, Decision Trees, Gradient Boosting, and Support Vector Machines are trained using labeled operational datasets. These models classify system behavior into normal and abnormal operational states based on telemetry features extracted from logs, metrics, and

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 3, May – June 2025 |

DOI:10.15662/IJEETR.2025.0703006

traces. Performance evaluation metrics include accuracy, precision, recall, and F1-score analysis.Unsupervised learning approaches are implemented to identify unknown anomalies and hidden operational patterns. Clustering algorithms such as K-Means and DBSCAN group telemetry observations based on behavioral similarity. Autoencoders and Variational Autoencoders learn compressed representations of normal system behavior and identify deviations representing anomalies.Deep learning techniques such as Long Short-Term Memory networks analyze sequential telemetry data for predictive analytics and failure forecasting. Recurrent Neural Networks capture temporal dependencies among service interactions, workload fluctuations, and infrastructure events. Transformer-based architectures support contextual analysis of logs and telemetry streams.Causal Root Cause Analysis is implemented using Bayesian Networks, causal graphs, and Graph Neural Networks. Causal inference models analyze directional dependencies between operational events to determine how failures propagate across distributed services. Dependency graphs represent services and infrastructure components as interconnected nodes, enabling causal relationship analysis.
Generative AI mechanisms support intelligent RCA by generating contextual explanations, incident summaries, remediation recommendations, and operational insights. Large Language Models analyze telemetry data and generate human-readable explanations describing probable root causes and recommended corrective actions. Explainable AI methods such as SHAP analysis and attention visualization improve model transparency and operator trust.The AI framework incorporates ensemble learning strategies that combine multiple models to improve prediction robustness and RCA accuracy. Hyperparameter optimization and cross-validation techniques ensure model generalization across diverse operational conditions and dynamic cloud-native environments.

The predictive monitoring framework operates as a real-time AI-driven observability platform capable of continuously analyzing telemetry data and generating operational intelligence. Distributed stream-processing technologies such as Apache Flink and Spark Streaming process telemetry streams in real time to support anomaly detection and predictive analytics.The predictive monitoring engine analyzes telemetry trends including response latency, throughput fluctuations, resource utilization patterns, error rates, and service dependencies. AI models generate prediction scores indicating the probability of future incidents or performance degradation events. Dynamic threshold adaptation mechanisms adjust anomaly sensitivity according to workload conditions and operational context.Causal RCA mechanisms correlate anomalies across logs, metrics, traces, and infrastructure events to identify the origins of failures. Correlation analysis and causal inference techniques determine directional relationships among operational events. Bayesian inference models estimate the likelihood of different root cause scenarios and rank potential causes according to confidence levels.Generative AI supports intelligent automation by generating remediation recommendations and operational guidance. AI-generated remediation workflows may include restarting failed containers, scaling infrastructure resources, rerouting traffic, reallocating workloads, or updating configuration settings. Kubernetes operators and automation scripts execute remediation actions automatically based on AI-generated recommendations.Visualization dashboards display operational metrics, AI-generated alerts, root cause graphs, dependency maps, and predictive analytics outputs. Heatmaps and topology diagrams help operators understand fault propagation pathways and service dependencies. Distributed tracing visualizations provide detailed visibility into request execution paths and transaction lifecycles.Self-healing capabilities are integrated into the monitoring framework to enable autonomous operational management. Feedback loops continuously evaluate remediation effectiveness and update AI models accordingly. This adaptive learning capability enables the system to evolve alongside changing workloads and infrastructure configurations.

IV. RESULTS AND DISCUSSION

The implementation of the Generative AI–driven cloud modernization framework produced significant improvements in the scalability, resilience, and operational intelligence of microservice-based data platforms. The experimental environment consisted of legacy monolithic applications migrated into containerized microservices deployed across cloud-native infrastructures using orchestration technologies and distributed storage systems. The modernization framework integrated Generative AI models with observability pipelines, infrastructure telemetry, and causal root cause analysis engines to enable predictive diagnostics and intelligent workload optimization. Experimental results demonstrated that the proposed architecture improved deployment efficiency, reduced infrastructure bottlenecks, and enhanced system adaptability under fluctuating workloads. The use of Generative AI significantly accelerated application refactoring by automatically identifying service boundaries, recommending API decomposition strategies, and generating optimized deployment configurations. During performance evaluation, the platform exhibited lower service latency, faster recovery times, and improved throughput compared with traditional modernization approaches. The causal root cause analysis engine effectively identified hidden dependencies and propagation chains among distributed services, enabling rapid detection of service degradation. Furthermore, anomaly detection models successfully predicted infrastructure instability before the occurrence of large-scale failures, thereby reducing

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 3, May – June 2025 |

DOI:10.15662/IJEETR.2025.0703006

downtime and improving platform reliability. The integration of AI-assisted monitoring also reduced manual intervention in incident management processes, enabling DevOps teams to focus on strategic operational improvements rather than repetitive troubleshooting tasks. Experimental findings confirmed that the proposed framework achieved better resource utilization and operational stability in comparison with conventional cloud migration and monitoring solutions.

Another important outcome observed during the experimentation phase was the capability of the system to manage large-scale distributed data workloads efficiently across hybrid and multi-cloud environments. Modern data platforms frequently encounter challenges associated with data consistency, service orchestration, asynchronous communication, and dynamic workload balancing. The proposed Generative AI framework addressed these challenges by continuously learning from operational telemetry and adapting service configurations in real time. Results showed that AI-generated optimization recommendations improved container scheduling efficiency, reduced memory consumption, and minimized network congestion between microservices. The causal inference engine played a crucial role in distinguishing correlation from causation during incident analysis, thereby reducing false alarms commonly generated in distributed monitoring systems. Comparative analysis revealed that integrating causal reasoning with machine learning significantly improved root cause identification accuracy compared with rule-based monitoring frameworks. The framework also demonstrated strong scalability by maintaining stable performance under high transaction volumes and concurrent service requests. In scenarios involving cascading service failures, the system successfully traced the originating fault and visualized dependency relationships across the service mesh. The incorporation of Generative AI into observability pipelines enhanced automated log summarization, anomaly explanation, and operational insight generation. Consequently, the platform improved overall system transparency and enabled administrators to respond to incidents more effectively. These findings validate the effectiveness of combining Generative AI and causal analytics for intelligent modernization and predictive operations management in cloud-native microservice ecosystems.

The discussion of the obtained results highlights the growing significance of intelligent automation in modern cloud modernization strategies. Traditional migration approaches often focus only on infrastructure transformation without adequately addressing operational complexity and service observability challenges. However, microservice ecosystems generate enormous volumes of telemetry data, making manual monitoring and diagnostics increasingly impractical. The proposed framework demonstrated that Generative AI can play a transformative role in simplifying modernization processes by automating architectural recommendations, deployment optimization, and operational analysis. One major discussion point emerging from the study is the effectiveness of causal root cause analysis in handling distributed system failures. Unlike conventional monitoring systems that rely heavily on metric thresholds and event correlation, causal analysis provides contextual understanding of how faults propagate across interconnected services. This capability significantly improved diagnostic precision and reduced the time required for incident resolution. Another important observation was the adaptability of the AI models to evolving workloads and changing infrastructure conditions. The continuous learning capability enabled the framework to refine prediction accuracy and operational recommendations over time. The integration of AI-generated operational insights also enhanced collaboration between development and operations teams by providing clearer explanations of system anomalies and dependencies. These findings suggest that intelligent observability platforms can become foundational components of autonomous cloud operations in the future.

The discussion also reveals several practical and technical considerations associated with deploying Generative AI–driven modernization frameworks in enterprise environments. One challenge identified during implementation involved the requirement for large volumes of high-quality telemetry data to train machine learning and causal inference models effectively. Inconsistent logging standards, incomplete traces, and noisy operational data occasionally affected prediction reliability and anomaly interpretation. Additionally, Generative AI systems introduced computational overhead due to continuous data analysis and model retraining processes, especially in large-scale cloud infrastructures. Despite these challenges, the framework demonstrated considerable resilience and adaptability under varying workload conditions. Another critical issue concerns explainability and trustworthiness of AI-generated recommendations. Enterprise administrators require interpretable insights before adopting automated operational decisions in production environments. Therefore, visualization dashboards, causal graphs, and transparent anomaly explanations became essential components of the proposed architecture. Security and compliance considerations also emerged as important discussion topics because observability systems process sensitive operational and business-critical information across distributed cloud services. Nevertheless, the overall findings confirm that integrating Generative AI with causal analytics significantly enhances predictive monitoring, modernization efficiency, and operational resilience in scalable microservice data platforms. The study demonstrates that intelligent automation can reduce operational complexity,

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 3, May – June 2025 |

DOI:10.15662/IJEETR.2025.0703006

optimize resource management, and improve service reliability, thereby supporting the growing demand for adaptive and self-managing cloud-native infrastructures.

V. CONCLUSION

The research on Generative AI–driven cloud modernization and causal root cause analysis for scalable microservice data platforms demonstrates the transformative impact of artificial intelligence on modern distributed computing environments. The study successfully established that integrating Generative AI with predictive observability and causal analytics significantly enhances the efficiency, scalability, and reliability of cloud-native systems. Traditional modernization approaches often struggle to manage the complexity of distributed microservice architectures because they rely heavily on manual analysis, static configurations, and reactive monitoring strategies. In contrast, the proposed framework leveraged AI-powered automation to support intelligent service decomposition, workload optimization, predictive diagnostics, and automated operational insight generation. Experimental results confirmed that the system improved deployment agility, reduced service latency, minimized downtime, and optimized infrastructure utilization across dynamic cloud environments. The causal root cause analysis engine further enhanced operational intelligence by accurately identifying fault propagation paths and dependency relationships among interconnected services. By combining telemetry data, machine learning models, and causal reasoning techniques, the framework successfully differentiated meaningful operational anomalies from noisy event correlations. This proactive and adaptive monitoring capability enabled faster incident resolution and improved overall service resilience. The findings therefore validate the effectiveness of AI-driven modernization strategies in addressing the increasing complexity of large-scale cloud-native data platforms.

Another significant conclusion derived from the study is the critical role of intelligent observability and causal analytics in enabling autonomous cloud operations. As enterprises continue migrating toward microservice-based architectures, operational environments become increasingly decentralized, dynamic, and data-intensive. Conventional monitoring systems are often incapable of handling the volume, velocity, and interconnectedness of telemetry data generated within these ecosystems. The proposed framework addressed these limitations by introducing Generative AI models capable of learning system behavior patterns, predicting anomalies, and generating actionable operational recommendations. The integration of causal analysis provided contextual understanding of incident propagation, thereby improving diagnostic precision and reducing false positive alerts. Furthermore, AI-generated summaries and operational insights reduced the cognitive burden on DevOps and site reliability engineering teams by automating repetitive troubleshooting tasks. The framework also demonstrated scalability across hybrid and multi-cloud deployments, proving its practical applicability in enterprise-scale infrastructures. Although challenges related to data quality, computational overhead, and model explainability were identified, the overall benefits of intelligent automation substantially outweighed these limitations. The study therefore concludes that Generative AI and causal root cause analytics can serve as foundational technologies for building resilient, adaptive, and self-optimizing cloud-native systems capable of meeting evolving business and technological demands.

In addition to technical advancements, the research contributes important insights into the future evolution of cloud modernization methodologies. Modern organizations increasingly depend on uninterrupted digital services, real-time analytics, and scalable data processing infrastructures to remain competitive in rapidly changing markets. Consequently, operational resilience and proactive system management have become strategic priorities for enterprises across industries such as finance, healthcare, telecommunications, and e-commerce. The findings of this study indicate that Generative AI can transform modernization initiatives from simple infrastructure migration projects into intelligent operational transformation processes. By automating architectural recommendations, deployment optimization, and observability analysis, AI-driven systems significantly accelerate modernization timelines while improving operational consistency. Moreover, causal root cause analysis enhances transparency and accountability in distributed systems by providing interpretable explanations of service failures and dependency relationships. The integration of these technologies therefore creates a comprehensive framework for predictive operations management and intelligent infrastructure governance. The research also highlights the importance of combining multiple observability dimensions, including logs, metrics, traces, and dependency graphs, to achieve comprehensive situational awareness. Such integration strengthens the ability of organizations to identify performance bottlenecks, predict failures, and optimize system behavior proactively. Consequently, the study establishes a strong conceptual and practical foundation for future research in autonomous cloud management and intelligent observability engineering.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 3, May – June 2025 |

DOI:10.15662/IJEETR.2025.0703006

The conclusion also emphasizes the broader strategic significance of adopting AI-driven modernization and monitoring frameworks in achieving long-term business sustainability and digital transformation goals. In highly distributed cloud ecosystems, even minor service disruptions can result in substantial financial losses, reduced customer trust, and operational inefficiencies. The proposed framework demonstrated how predictive analytics and causal intelligence can mitigate these risks by enabling early anomaly detection, automated diagnostics, and adaptive operational optimization. The continuous learning capabilities of Generative AI models further ensure that the platform evolves alongside changing workloads, infrastructure updates, and emerging operational patterns. Although concerns regarding AI governance, data privacy, and operational transparency remain important considerations, advancements in explainable AI and secure cloud technologies are expected to address these issues progressively. The research therefore concludes that Generative AI–driven modernization is not merely a technological enhancement but a necessary paradigm shift for managing the complexity of next-generation digital infrastructures effectively. Organizations adopting such intelligent frameworks are likely to achieve improved service reliability, optimized resource consumption, faster innovation cycles, and enhanced customer satisfaction. As cloud-native technologies continue to evolve, AI-powered modernization and causal observability systems will become essential components of scalable, adaptive, and autonomous enterprise computing environments.

VI. FUTURE WORK

Future work in Generative AI–driven cloud modernization and causal root cause analysis for scalable microservice data platforms can explore several advanced research directions aimed at improving automation, adaptability, intelligence, and operational resilience. One major area for future enhancement involves the development of autonomous self-healing cloud infrastructures capable of performing corrective actions without human intervention. While the current framework focuses primarily on predictive monitoring and root cause identification, future systems can integrate reinforcement learning techniques to enable intelligent remediation strategies such as automated workload redistribution, dynamic resource scaling, service replication, and failure isolation. These autonomous operational capabilities could significantly reduce downtime and improve system stability in highly dynamic cloud environments. Another promising direction involves advancing explainable Generative AI models to improve transparency and trust in AI-generated recommendations. Enterprise administrators and DevOps engineers require interpretable reasoning behind anomaly predictions, deployment optimizations, and causal dependency analyses before relying on automated operational decisions. Future research may therefore focus on integrating explainable AI mechanisms that provide visual causal graphs, confidence scores, and human-readable summaries for generated insights. Additionally, future studies can investigate the integration of multimodal observability analytics that combine logs, metrics, traces, infrastructure events, and business-level telemetry into unified predictive intelligence systems. Such integration could improve situational awareness and enable more comprehensive operational diagnostics across distributed ecosystems. Another important area of future work involves applying federated learning approaches for collaborative AI model training across multiple cloud providers and enterprise infrastructures while preserving data privacy and regulatory compliance.

This would allow organizations to improve predictive accuracy without exposing sensitive operational information. Future research may also examine the role of large language models in automated incident response, operational documentation generation, and conversational observability interfaces capable of assisting engineers interactively during troubleshooting activities. The integration of cybersecurity intelligence with causal analytics represents another significant research opportunity because modern cloud platforms increasingly face sophisticated cyber threats that may resemble operational anomalies. Combining threat detection with predictive monitoring could strengthen system security and resilience simultaneously. Researchers can further explore advanced graph neural network architectures for modeling dynamic service dependencies and real-time failure propagation across evolving service meshes. In addition, edge AI and distributed inference mechanisms may be investigated to reduce latency and computational overhead associated with centralized telemetry analysis in geographically distributed infrastructures. Another valuable direction involves integrating sustainability-aware optimization strategies into AI-driven modernization frameworks to minimize energy consumption and carbon emissions in cloud data centers. Future systems may also incorporate business impact analytics that prioritize incidents based not only on technical severity but also on customer experience, revenue implications, and service-level objectives.

Extensive real-world validation across large-scale enterprise environments, including hybrid cloud, edge computing, and serverless architectures, will also be essential for evaluating scalability, robustness, and interoperability under diverse operational conditions. Furthermore, future research can investigate ethical governance frameworks for AI-driven cloud operations to ensure fairness, accountability, transparency, and compliance in automated decision-making

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 3, May – June 2025 |

DOI:10.15662/IJEETR.2025.0703006

processes. These advancements collectively have the potential to transform cloud modernization and predictive observability into fully autonomous, intelligent, secure, and sustainable operational ecosystems capable of supporting the next generation of scalable digital enterprises.

REFERENCES

1. Raja, G. V. (2023). Modernizing Enterprise Systems using AI with Machine Learning and Cloud Computing for Intelligent Systems. International Journal of Future Innovative Science and Technology (IJFIST), 6(6), 11713.
2. Pasumarthi, H. (2023). Applying machine learning to high-volume banking platforms: From transaction data to predictive risk intelligence. International Journal of Artificial Intelligence & Machine Learning, 2(1), 356–370. https://doi.org/10.34218/IJAIML_02_01_029
3. Sengupta, J., & Alzbutas, R. (2022). Intracranial hemorrhages segmentation and features selection applying cuckoo search algorithm with gated recurrent unit. Applied Sciences, 12(21), 10851.
4. Narayanan, S. (2023). Operationalizing Artificial Intelligence Security in the Cloud: A Practical Integration framework for Enterprise Risk Management. International Journal of Future Innovative Science and Technology (IJFIST), 6(3), 10619.
5. Gopinathan, V. R. (2024). Secure explainable AI on Databricks–SAP cloud for risk-sensitive healthcare analytics and swarm-based QoS control. International Journal of Engineering & Extended Technologies Research (IJEETR), 6(4), 8452-8459.
6. Kunadi, S. K. (2024). Improving Data Quality and Deduplication Using Similarity Scoring and Confidence Models. International Journal of Computer Technology and Electronics Communication, 7(4), 9200-9211.
7. Namdeo, A. (2021). Quantum-accelerated cloud BI query optimization. International Journal of Engineering & Extended Technologies Research (IJEETR), 3(5), 3715–3724.
8. Devineni, A. (2024). Causal Inference in Distributed Tracing: Automating Root Cause Analysis in Complex Microservice Dependencies. International Journal of Emerging Trends in Computer Science and Information Technology, 5(4), 166-173.
9. Panyala, V. R. (2024). Designing self-healing cloud architectures for mission-critical distributed systems. International Journal of Science, Research and Technology, 7(2), 11717–11721.
10. Appani, C., & Guda, D. P. (2023). Self-supervised representation learning for zero-day attack detection in encrypted network traffic. Computer Fraud & Security, 2023(7), 20–31. Retrieved from: https://computerfraudsecurity.com/index.php/journal/article/view/661
11. Sarabu, V. B. (2024). Architecting controlled international platform rollouts: Data governance, validation, and risk mitigation in retail modernization. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 7(1), 306–328.
12. Subramanyam, S. P. (2022). Kubernetes-oriented continuous deployment architecture for .NET microservices. International Journal of Future Innovative Science and Technology (IJFIST), 5(3), 8482–8490. https://doi.org/10.15662/IJFIST.2022.0503002
13. Mallireddy, S. (2023). Servicenow & Generative AI: Improving Infant Mortality Rate. International Journal of Computer Technology and Electronics Communication, 6(5), 1-7.
14. Adepu, R. (2024). Secure cloud migration strategies for enterprise data center modernization. International Journal of Engineering & Extended Technologies Research (IJEETR), 6(6), 239–258.
15. Soundappan, S. J. (2024). AI-Driven Customer Intelligence in Enterprise Lakehouse Systems Sentiment Mining Governance-Aware Analytics and Real-Time Data Synchronization. International Journal of Advanced Engineering Science and Information Technology (IJAESIT), 7(5), 14905.
16. Kasireddy, J. R. (2025). Leveraging big data analytics for enhanced commercial vehicle safety: FMCSA's data engineering journey. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 11(2), 3203–3222. https://doi.org/10.32628/CSEIT25112796
17. Prasad, P. K. (2021). Kubernetes everywhere: Operating hybrid and multi-cloud infrastructure at scale. International Journal of Engineering & Extended Technologies Research, 3(4), 3393–3401.
18. Suvvari, S. K. (2023). Shift Left: Moving the Inclusion of Accessibility Functionalities to the Left in Agile Product Development Life Cycle. Journal of Computational Analysis and Applications, 31(4).
19. Joyce, S. (2024). Automated enterprise system reliability: Integrating AI-driven monitoring with cloud-based SAP deployment pipelines. International Journal of Research and Applied Innovations (IJRAI), 7(2), 10474–10482. https://doi.org/10.15662/IJRAI.2024.0702010
20. Adepu, G. (2023). Intelligent digital government platforms: Leveraging machine learning and cloud architecture for social service delivery. International Journal of Computer Technology and Electronics Communication (IJCTEC), 6(3), 75–92.
21. Hossain, M. S., Hossain, M. S., Ali, M., & Rahman, M. W. (2025). Data-Driven Strategies for Predicting and Enhancing Rural Business Growth in the United States. Data-Driven Strategies for Predicting and Enhancing Rural Business Growth in the United States, 1(7), 121-146.