Proceedings of International Conference on "Recent Innovations in Engineering, Technology & Science for Sustainable Development (RIETSSD-2026)"
Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

A Literature Review on Proactive Real-Time Monitoring of Banking IT Infrastructure: A Self-Healing Approach using Agent AI

Mrs.Dakshta Jain[1], Dr.Awanit Kumar[2]

[1]SResearch Scholar, School of Engineering & Technology(CSE), Career Point University, Kota (Raj)

[1]SResearch Supervisor, School of Engineering & Technology(CSE), Career Point University, Kota (Raj)

[1]mittaldakshta@gmail.com; [2]awanit.kumar@cpur.edu.in

Abstract: This paper presents a comprehensive review of the literature on AI-driven proactive monitoring and self- healing mechanisms for banking IT infrastructure [1], [2]. The study systematically examines how AI-enabled models—including supervised and unsupervised machine learning, deep neural networks, and multi-agent systems—are employed to enhance system reliability, reduce operational downtime, and ensure continuous service availability within mission-critical financial environments [3]–[6]. It further analyzes the application of predictive analytics for forecasting infrastructure failures and resource demands, alongside automated remediation systems designed for rapid fault recovery [7], [8]. These technological advancements collectively signify a paradigm shift from reactive IT support toward increasingly proactive and autonomous operational management [1], [2].The review critically assesses emerging AIOps frameworks that seek to unify discrete functions—such as monitoring, predictive analytics, root-cause analysis, and automated remediation—into a cohesive, intelligent pipeline [2]. A key finding is the pressing need for these integrated systems to deliver real-time intelligence, adaptive operational thresholds, and, critically, explainable AI (XAI) capabilities [9], [10]. The requirement for transparency and auditability is paramount in the heavily regulated financial sector, where justifying automated decisions to auditors and regulators is non-negotiable [9], [11]. This underscores a significant gap between advanced algorithmic potential and the practical, compliant deployment of fully autonomous systems.Ultimately, the synthesis of current research reveals that while substantial progress has been made in individual domains—such as anomaly detection or automated scripting—existing solutions often remain fragmented, lacking the scalability, real-time adaptability, and holistic integration required for enterprise-wide banking ecosystems [1], [2]. There- fore, this review identifies a clear and urgent research trajectory toward the development of trustworthy,

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

scalable, and fully autonomous self-healing architectures [5]. Such next-generation systems are essential to address the escalating complexity, security threats, and availability demands inherent in modern digital banking infrastructure [6].

Keywords: Proactive Monitoring, Artificial Intelligence, Banking IT Infrastructure, Anomaly Detection, Self-Healing Systems, Predictive Analytics, Machine Learning, Fault Predic- tion, Automated Remediation, Explainable AI (XAI), AIOps, Operational Resilience, IT Operations Management, Financial Technology (FinTech), Cybersecurity, Root Cause Analysis, Autonomous Systems, Cloud Computing, Service Availability, Regu- latory Compliance, Deep Learning, Real-Time Analytics, System Reliability, Unsupervised Learning, Multi-Agent Systems.

I. INTRODUCTION

In the current digital age, the banking industry heavily depends on sophisticated technological ecosystems to ensure the seamless delivery and security of financial services [6], [11]. Contemporary banking systems—including online and mobile banking platforms, ATM networks, payment gateways, UPI processing systems, fraud detection engines, and core banking solutions—operate on large-scale IT infrastructures that are required to function continuously with near-zero downtime [6]. Even minor disruptions to this infrastructure can lead to significant financial losses, regulatory violations, customer dissatisfaction, and reputational damage [11]. Consequently, the availability, performance, and security of banking IT systems have emerged as critical operational priorities.

Conventional monitoring systems deployed in banking environments largely follow a reactive approach, where issues are detected only after they have already impacted system performance or service availability [1]. These systems rely on static threshold-based alerts, manual supervision, and human- driven troubleshooting, making them inadequate for modern high-volume, distributed, and cloud-enabled banking architec- tures [2]. To overcome these limitations, the IT industry is increasingly adopting Artificial Intelligence for IT Operations (AIOps), with a strong emphasis on proactive monitoring and self-healing mechanisms [1], [2].

A transformative role in this paradigm shift is played by agent-based artificial intelligence. AI agents are intelligent software entities capable of continuously observing

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

system behavior, identifying abnormal patterns, learning from historical data, and predicting potential failures before they occur [3], [5]. Unlike traditional monitoring tools, agent-based AI systems dynamically adapt operational thresholds and autonomously execute corrective actions, thereby minimizing human intervention and enabling faster and more reliable recovery processes [1], [5].

I. Review of Related Work

A. Machine Learning-Based Monitoring

Monitoring approaches based on machine learning employ advanced analytical algorithms to identify deviations from normal operational patterns in financial network infrastructures [4], [6]. Prior research has demonstrated the effectiveness of supervised learning techniques such as Random Forest and Support Vector Machines (SVM) in detecting anomalous network traffic that may indicate fraud, cyber intrusions, or impending system failures [4], [12]. Unlike traditional rule- based monitoring systems that rely on static thresholds, these models learn complex behavioral patterns from historical data, enabling more adaptive and fine-grained anomaly detection and early warning capabilities [1].

In addition to supervised approaches, unsupervised learning techniques offer significant advantages in scenarios where labeled anomaly data is scarce or unavailable. Autoencoder-based models, for example, are widely used to learn the normal distribution of system logs and operational metrics through efficient latent representations [4], [7]. During deployment, reconstruction errors produced by these models serve as strong indicators of abnormal behavior, including previously unseen or zero-day anomalies. Such frameworks enable real-time failure detection without requiring extensive labeled datasets, thereby improving system robustness against evolving and unpredictable operational conditions [7].

The integration of these machine learning paradigms into monitoring frameworks represents a fundamental transition from reactive to proactive system management [1], [2]. Continuous model training and adaptive learning from live network traffic and log streams reduce false positives while improving detection accuracy over time [3]. This capability is critical for maintaining the integrity, availability, and reliability of financial services, where even minor anomalies can propa- gate rapidly and disrupt interconnected banking operations [6], [11]. Consequently, machine learning–driven monitoring forms the foundation of self-

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

evolving and resilient financial IT ecosystems capable of anticipating and mitigating both known and emerging operational threats.

B. Predictive Analytics in IT Infrastructure

Predictive analytics represents a transformative approach to enhancing the resilience and operational efficiency of modern IT infrastructures [1], [2]. It leverages statistical techniques and machine learning models applied to historical and real- time operational data to anticipate future system states, potential failures, and performance bottlenecks [8]. By forecasting anomalies in key metrics such as resource utilization, network latency, and application error rates, predictive models enable early intervention before issues escalate into service disruptions [7]. This paradigm shifts IT operations from a traditional break-fix and reactive model toward a proactive and preventive operational strategy, thereby reducing downtime, optimizing resource allocation, and improving adherence to service-level agreements (SLAs) [6], [11].

Effective predictive analytics relies on robust data pipelines and advanced modeling frameworks that integrate heterogeneous data sources, including system logs, telemetry streams, and configuration management databases, to deliver a holis- tic view of infrastructure health [2]. Commonly employed techniques such as time-series forecasting, regression anal- ysis, and survival modeling are used to estimate disk failure probabilities, memory exhaustion timelines, and network congestion risks [8]. For example, predictive maintenance models trained on Self-Monitoring, Analysis, and Report- ing Technology (SMART) attributes of storage devices can accurately forecast hardware degradation, enabling proactive component replacement and preventing catastrophic data loss or unplanned outages [8].

The adoption of predictive analytics in IT operations facilitates the development of highly anticipatory systems capable of dynamic capacity scaling based on projected demand and automated issue generation through the correlation of predicted anomalies with known failure patterns [1], [3]. This form of proactive intelligence significantly reduces operational costs and manual intervention while enhancing system reliability and overall user experience [6]. Consequently, predictive analytics has emerged as a foundational element of next- generation, agile IT infrastructures designed to support the evolving demands of modern digital enterprises and mitigate risks associated with complex, interdependent systems [2].

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

C. Self-Healing and Automated Recovery

The self-healing and automated recovery paradigm represents a critical advancement toward achieving autonomy and resilience in modern IT infrastructures [1], [13]. This approach integrates automated detection, diagnosis, and reme- diation mechanisms to resolve system faults without human intervention. Such systems are capable of executing corrective actions—including restarting failed services, reallocating re- sources, and failing over to redundant components—based on real-time monitoring data, predefined policies, and orchestra- tion engines [1]. This capability is essential for maintaining continuous service availability in highly distributed environments where manual response times are insufficient to meet stringent recovery time objectives (RTOs) [6].

The architectural foundation of self-healing systems typi- cally follows a closed-loop control model comprising monitoring, analysis, planning, and execution phases [1]. Intelligent agents or centralized orchestration platforms continuously observe key performance indicators (KPIs) and system health metrics. Upon detecting anomalous behavior, root-cause analysis is performed using correlation engines or causal inference models to determine the underlying source of failure [2]. Subsequently, an appropriate remediation strategy is dynamically selected from a repository of automated scripts or infrastructure-as-code (IaC) templates and executed through application programming interfaces (APIs) provided by cloud, virtualization, or container orchestration platforms [13]. This automated workflow significantly reduces mean time to recov- ery (MTTR) and operational overhead [1].

Machine learning techniques are increasingly incorporated into self-healing mechanisms to address uncertain, complex, or previously unseen failure scenarios [3]. Reinforcement learning–based agents, in particular, can learn optimal recov- ery policies through continuous interaction with the operational environment, thereby improving remediation effective- ness over time [5]. The transition toward fully autonomous recovery not only enhances operational efficiency but also enables human engineers to focus on higher-level strategic initiatives rather than routine incident resolution [1]. Con- sequently, automated recovery and self-healing have become foundational components of modern DevOps and Site Reliability Engineering (SRE) practices, enabling the development of resilient systems capable of graceful degradation and rapid.

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

D. AI-Based Security Monitoring

AI-based security monitoring fundamentally transforms traditional security operations by enabling the dynamic and intelligent analysis of large-scale, heterogeneous data streams to detect sophisticated cyber threats [4], [6], [12]. Conventional signature-based intrusion detection systems are increasingly ineffective against modern polymorphic attacks and advanced persistent threats (APTs). In contrast, AI-driven security solutions employ supervised, unsupervised, and deep learning models to establish behavioral baselines across users, devices, applications, and network traffic [12]. These models continuously analyze endpoint events, network flow data, and cloud telemetry to identify subtle deviations indicative of malicious activities such as lateral movement, data exfiltration, or zero- day exploits, while significantly reducing false-positive rates compared to static rule-based approaches [4].

The effectiveness of AI-based security monitoring depends heavily on robust feature engineering and scalable model training pipelines capable of processing diverse telemetry sources, including log files, NetFlow records, and process execution graphs [12]. Unsupervised learning techniques such as clustering and autoencoders enable the detection of anoma- lous behavior without prior labeling, whereas supervised mod- els—such as gradient boosting classifiers—are trained to recognize known attack signatures [4]. Advanced deep learning architectures, including recurrent neural networks (RNNs) and transformer-based models, are particularly well suited for sequential and time-series data, allowing the detection of complex, multi-stage attack campaigns [4], [6]. These AI models are increasingly integrated into Security Information and Event Management (SIEM) and Extended Detection and Response (XDR) platforms, providing contextualized alerts, prioritized incident handling, and a

II. Comparative Analysis

The comparative analysis of the reviewed technological paradigms—namely Machine Learning–Based Monitor- ing, Predictive Analytics, Self-Healing Systems, AI-Based Security Monitoring, and Integrated Intelligent Operations (AIOps)—highlights their distinct yet complementary functional objectives and varying levels of maturity within modern IT operations [1], [2]. Each paradigm addresses a specific layer of the operational lifecycle, ranging from localized anomaly detection to enterprise-wide autonomous

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

decision- making. Machine learning–driven monitoring and predictive analytics primarily operate at diagnostic and predictive levels, focusing on identifying deviations from normal behavior and forecasting future system states based on historical and real- time data trends [3], [7]. These approaches enable early warning mechanisms and informed decision support but often rely on human intervention for remediation.

Self-healing and automated recovery systems represent a prescriptive evolution of these capabilities, wherein detected insights are directly translated into corrective actions [8]. By integrating anomaly detection, root cause analysis, and automated remediation playbooks, these systems reduce mean time to recovery (MTTR) and minimize service disruption in distributed and cloud-native environments. AI-based security monitoring constitutes a specialized yet parallel domain, leveraging similar analytical techniques for threat detection and response. However, it is uniquely focused on adversarial contexts, emphasizing behavioral analysis, intrusion detection, and threat intelligence [4], [12]. In contrast, Integrated Intelligent Operations (AIOps) serves as a unifying architectural framework that integrates monitoring, analytics, security, and remediation into a cohesive platform capable of cross-domain correlation and intelligent automation [1], [2].

The underlying technologies and implementation complexi- ties vary significantly across these paradigms. Monitoring and predictive systems typically rely on supervised learning and time-series forecasting models, which require substantial vol- umes of historical data but offer relatively high interpretability for metric-level predictions such as hardware failure or performance degradation [7]. Self-healing architectures introduce additional complexity through their dependence on orchestra- tion engines, automated workflows, infrastructure APIs, and policy-driven decision logic to ensure safe and controlled re- mediation actions [8]. AI-driven security monitoring demands even greater sophistication, as it must process heterogeneous, high-velocity, and adversarial data sources using advanced feature engineering and deep learning models to uncover concealed or evolving threat patterns [4], [6]. AIOps platforms, as the most comprehensive layer, must integrate all these com- ponents—data aggregation, cross-domain correlation, causal inference, and automated response—within a scalable and governed architecture, making them the most challenging to design and deploy [1], [2].

From a strategic perspective, the return on investment (ROI) in terms of operational efficiency, risk reduction, and business continuity strongly influences the adoption

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

trajectory of these paradigms. Targeted solutions such as predictive maintenance and capacity forecasting often deliver rapid and measurable benefits by reducing unplanned downtime, making them common entry points for organizations beginning their automation journey [7]. In contrast, the transition toward fully integrated AIOps platforms and autonomous recovery systems represents a long-term strategic investment that requires higher upfront costs but offers substantial gains in systemic resilience, scalability, and reduced operational expenditure over time [2], [6]. As contemporary platforms increasingly blur the bound- aries between monitoring, analytics, and action, the future trajectory points toward self-driving IT infrastructures—where the capabilities examined in this study converge into a intelligent operational fabric capable of managing unprece-dented scale, complexity, and service expectations in modern digital enterprises [1], [2].

TABLE I Comparison of Related Research Works

Ref	Focus Area	Method	Main Contribution
[1]	Network anomal detection	ML techniques	Improved anomaly detection in financial networks
[2]	Log anomaly detection	Autoencoder	Real-time log anomaly identification
[3]	Resource utilization prediction	LSTM	Predicts CPU/memory load in banking servers
[4]	IT performance forecasting	ARIMA	Early detection of performance degradation
[5]	Cloud selfhealing	Multi-agent system	Automated recovery in cloud banking
[6]	Fault recovery automation	Rule-based AI	Quick resolution of common software faults
[7]	Intrusion detection	Deep learning	High accuracy threat detection
[8]	Fraud + network security	CNN-LSTM	Hybrid approach for fraud and intrusion detection

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

[9]	AI-driven selfhealing	ML + Predictive AI	Survey of AI techniques for self-healing
[10]	AIOps automation	SLR	Comprehensive AIOps analysis
[11]	Self-adaptive systems	ML methods	ML-enabled system adaptation
[12]	AI system validation	Review	Classification of AI validation methods

III.Research Gap

Gaps observed in the current literature include :

- **3.1 Limited research on fully autonomous self-healing banking systems**-Current studies predominantly focus on isolated, single-task components necessary for an intelligent system, such as anomaly detection, predictive forecasting, or a specific automated recovery mechanism. These works demonstrate strong capabilities in their narrow domains, but they fall short of providing a holistic, end-to-end autonomous self-healing architecture.The deficiency lies in the orchestration layer. A true self-healing system requires a framework capable of: **Detecting** an anomaly or predicting a failure. **Correlating** this event across all connected systems (e.g., from network layer to application layer). **Determining** the root cause (Root- Cause Analysis or RCA) without human input. **Execut- ing** the optimal, risk-assessed remediation action (e.g., resource reallocation, rolling back a change, or restarting a service). **Validating** that the fix was successful. The current literature lacks a unified design for this entire lifecycle, particularly one that is specifically tailored and validated for the high-availability and transactional integrity demands of banking IT systems. The challenge is not merely connecting existing components, but design- ing a reliable and trusted decision-making engine that can manage the entire workflow autonomously.

3.2. Limited Real-Time Operational Capabilities De- spite the demonstrated

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

effectiveness of numerous AI and machine learning techniques in controlled, offline, or batch-processing environments, their performance often deteriorates when deployed within live, high-frequency operational settings characteristic of large-scale banking IT infrastructures [2], [6]. This disparity exposes critical challenges that must be addressed to enable practical, real-world adoption of intelligent monitoring and self- healing systems.

Data Volume and Velocity: Modern banking environments generate enormous volumes of operational data at exceptionally high velocities, including millions of financial transactions, system logs, and performance met- rics per second [1]. Many AI/ML models—particularly deep learning architectures—are computationally intensive and struggle to process such continuous data streams in real time. As a result, these models may fail to deliver timely anomaly detection, accurate prediction, or autonomous decision-making under production-scale workloads, thereby limiting their operational effective- ness [7].

Latency Requirements: Proactive monitoring and self- healing demand ultra-low-latency responses to prevent service degradation or customer impact [8]. In mission-critical banking systems, remediation actions must often be triggered within microseconds to milliseconds. If an AI model requires minutes to infer an impending failure that may materialize within seconds, the system effectively reverts to a reactive mode of operation rather than achieving true proactivity [2]. This latency gap remains a major barrier to the deployment of advanced AI-driven operational intelligence.

Infrastructure Heterogeneity: Banking IT ecosystems are inherently heterogeneous, comprising legacy on- premises mainframes, virtualized data centers, containerized microservices, and public cloud platforms [1]. Ensuring consistent model performance, scalability, and reliability across cloud–edge–on-premise environments presents significant engineering challenges. Real-time AI models must be portable, interoperable, and resilient to variations in data formats, network conditions, and computational resources, which is rarely addressed comprehensively in existing research [6].

- **3.3Lack of real-time AI models optimized for dis- tributed IT environments** As discussed earlier, existing research largely treats the core components of AIOps—namely anomaly detection, forecasting, root cause analysis, and automated healing—as

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

modular and independent functions [1], [2]. This fragmented design results in loosely coupled pipelines where data handoffs, context switching, and inter-component communication introduce latency, increase system complexity, and create additional failure points.

Workflow Integration Gap: A critical limitation is the absence of a fully unified and intelligent operational workflow. In an effective self-healing pipeline, the output of one component should seamlessly and dynamically feed into the next. For instance, a resource utilization forecast should automatically trigger a preventive scaling or reallocation action rather than generating a low-priority alert or service ticket requiring human intervention [8]. **Contextual Information Loss:** When operational functions are isolated, valuable contextual information about incidents is often lost or diluted across system boundaries. This fragmentation hampers accurate root cause analysis and suboptimal decision-making. Consequently, there is a strong need for integrated architectures that preserve a holistic, end-to-end view of system state throughout the entire incident lifecycle, from detection to resolution [2].

itemWorkflow Integration: The critical gap is the absence of a truly unified, intelligent workflow. A successful pipeline must dynamically feed the output of one com- ponent directly into the input of the next. For example, a resource utilization prediction (from the forecasting component) should instantly trigger a preventative reallo- cation action (from the recovery component) rather than merely generating a low-priority ticket.

Contextual Loss: When these functions are separated, contextual information about the event often gets lost or diluted, making accurate RCA and optimal decision- making challenging. The need is for an integrated model that maintains a holistic view of the system state through- out the entire incident lifecycle.

3.4 Minimal integration of predictive maintenance with security analytics Most AI models discussed in the literature are generic; they are designed for broad ap- plication across various IT domains. This generic nature fails to address several highly specific constraints of the banking sector:

Strict Regulatory Compliance: Banking is one of the most heavily regulated industries (e.g., PCI DSS, GDPR, local financial authority mandates). An AI solution must guarantee continuous compliance, which may involve mandatory data segregation, specific data retention policies, or limitations on automated actions. Generic models are

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

not built with these constraints in mind.

High-Availability Mandates: Unlike many other industries, banking systems often require four-nines (99.99)percent or five-nines (99.999)percent availability. Any AI intervention, even an automated fix, must be proven to not risk a widespread outage or transactional loss.

Security Protocol Integration: Banking systems have extremely rigorous security protocols. Solutions must be deeply integrated with existing fraud detection and net- work security systems and must not inadvertently intro- duce new security vulnerabilities. One of the most critical gaps identified in the literature is the near-total absence of explainable AI (XAI) mechanisms in proposed self- healing and AIOps frameworks [6].

Audit and Compliance Requirements: Financial insti- tutions are subject to frequent internal and external au- dits. Any autonomous AI-driven action—such as rolling back a database patch, rerouting transactions, or throttling services—must be accompanied by a clear, human- interpretable explanation detailing the rationale behind the decision [12].

Operational Trust and Accountability: Without explainability, operational teams and senior management cannot fully trust autonomous systems. Black-box mod- els, regardless of predictive accuracy, are unacceptable in mission-critical financial environments where account- ability and risk transparency are mandatory [2].

Regulatory Approval Barriers: Regulatory approval for deploying autonomous AI within core banking systems depends heavily on the ability to provide traceable, ex- plainable, and auditable decision pathways. The lack of XAI integration thus represents a major obstacle to real- world adoption [6].

1.Absence of Fully Autonomous Self-Healing Frame- works:

Current studies focus on isolated tasks such as anomaly detection, prediction, or automated recovery, but none provide an end-to-end autonomous self-healing architec- ture tailored for banking IT systems.

2.Limited Real-Time Operational Capabilities:

Several proposed AI/ML techniques are effective offline; however, real-time monitoring, prediction, and decision- making remain insufficient for large, high- frequency banking environments.

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

1.Lack of Integrated Monitoring–Prediction–Recovery Pipelines: Existing works treat detection, forecasting, root-cause analysis, and healing as separate components rather than a unified, intelligent workflow.

2.Insufficient Banking-Specific Solutions:

Most AI models are generic and do not address sector- specific needs such as regulatory compliance, strict security protocols, and high-availability requirements in financial institutions.

3.Minimal Incorporation of Explainable AI (XAI):

None of the reviewed studies include explainability fea- tures, which are essential for transparency, auditability, and regulatory approval in the banking domain.

3.5 Limited Exploration of Multi-Agent AI Architectures:

Only one study utilizes multi-agent systems, indicating a gap in designing scalable, distributed self-healing solu- tions for complex banking infrastructures.

The complexity and distributed nature of modern banking infrastructure, which often involves thousands of microservices, virtual machines, and network devices, demands a highly decentralized management approach.

- Scalability: Traditional centralized monitoring approaches struggle to scale. Multi-Agent System (MAS) inherently offer a scalable, distributed architecture where autonomous agents monitor specific segments (e.g., a database cluster, a payment gateway, or a core banking module) and collaborate to achieve system-wide self-healing.
- Resilience: Decentralization enhances resilience; if one agent fails, the overall system can continue to operate.
- Gap: The fact that only one study utilizes multi-agent systems suggests a vast research area remains untapped for designing distributed, coordinated, and scalable self- healing solutions tailored for the extreme complexity of modern banking IT. Future work must focus on devel- oping efficient communication protocols and consensus mechanisms for these agents to work together effectively.

- **SUMMARY**

This review indicates the great progress in AI-driven. active control and self-management of banking IT. infras- tructure. Available literature indicates that there is great development in. anomaly detection by use of machine learning, predictive

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

analytics. resource forecasting, rule based and automated fault recovery. AI-assisted security surveillance, multi-agent systems, and AI-assisted secu- rity surveillance. These techniques enhance the availability of systems, decrease downtimes. and contribute to clever decision-making in complicated banking. ecosystems. Nevertheless, without these developments, these are the existing methods. remain disintegrated and do not have real-time adaptability, scalability, and explainability. The need for is evident in the reviewed studies. whole- some, combined, and self-sustaining self-based architec- tures. special-purpose designed to suit mission-critical financial settings.

IV. Conclusion

The use of AI-driven monitoring and self-healing tech- nologies have become a disruptive solution to the banking IT infrastructure of the modern world. The literature at hand has shown great advancement in the field of anomaly detection, predictive analytics, automated recov- ery, and AI-based security. The capacity of the banking systems to identify failures earlier, in order to respond quicker, as well as sustain greater service availability, has been improved by machine learning, deep learning, and agent-based models. Nevertheless, the existing solutions are still mostly incomplete, domain-focused, or confined to the offline setting, which makes them less viable when applied to large-scale banking ecosystems. Further studies are necessary to realize the vision of self-healing banking infrastructure by shifting towards unified and autonomous multi-agent architecture, wherein monitoring and prediction, as well as decision-making and auto- mated remediation, are developed in a single pipeline. Additionally, Explainable AI (XAI) is also mandatory to meet regulatory audit and transparency standards as well as operational confidence in financial institutions. Intelligence in real-time, scalability of models, continuous learning and how to support heterogeneous environments involving clouds and edges and on-premise are very important areas to be developed. The introduction of AI-powered surveillance and repair technologies has be- come the ground-breaking solutions to the contemporary banking IT systems. The available literature shows that there has been a great advance in the area of anomaly detection, predictive analytics, automated recovery and AI-based security. The machine learning, deep learning, and agent-based models have contributed to the possibility of the banking systems to identify failures more quickly, respond to them, and ensure a higher degree of service availability. Nevertheless, the

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

existing methods are mostly piecemeal, local, or not applicable to offline contexts, limiting their application in large, dynamic banking environments. Further investigation is needed to take the vision of a self-healing banking infrastructure to the next level, since unified and autonomous multi-agent models are needed to combine monitoring, prediction, decision-making, and automated remediation into a single and seamless pipeline. These architectures are to be in use throughout the complete technology stack, application logic down to hardware. Additionally, Explainable AI (XAI) is also required to meet the requirements of strict regulatory audits, transparency in its operations and to create the required trust among financial institutions. In addition to transparency there are several critical gaps in research on the attainment of real time intelligence, scalability of models, continuous learning without di- minished performance, and the capability to easily handle heterogeneous cloud-edge-on-premise environments. Comprehensively, although AI has a great positive impact on the reliability and operational resilience of the banking systems, the holistic, real-time, scalable, and explainable self-healing architectures are in dire need. The next generation IT operations platforms should be able to accommodate the increasing complexity, criticality, and regulatory requirements of modern banking infrastructure and have it become a veritable resilience and independent financial utility rather than a collection of frail parts.

REFERENCES

[1] "Self-healing systems: AI for autonomous IT operations and reliability," *Survey of Foundations, Implementations, and Open Issues in AI-driven Self-healing for IT Operations*, 2025.

[2] L. Zhang *et al.*, "A survey of AIOps in the era of large language models," *ACM Computing Surveys*, 2025.

[3] T. R. D. Saputri and S.-W. Lee, "The application of machine learning in self-adaptive systems: A systematic literature review," *IEEE Access*, vol. 8, pp. 205,948–205,967, 2020.

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

[4] M. Emmanuel, "Deep learning architectures for real-time anomaly detection in financial transactions," *ResearchGate Publication*, June 2025.

[5] P. K. Rajput and G. Sikka, "Multi-agent architecture for fault recovery in self-healing systems," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2849–2866, 2021.

[6] A. Ahmed *et al.*, "AI-driven innovations in modern banking: From secure digital transactions to risk management, compliance frameworks, and AI-based ATM forecasting systems," *Journal of Management Science Research Review*, vol. 4, no. 3, pp. 1145–1183, 2025.

[7] V. Anemogiannis *et al.*, "Enhancing Kubernetes resilience through anomaly detection and prediction," *arXiv preprint arXiv:2503.14114*, 2025.

[8] C. A. Prasath, "AI-enabled digital twin framework for predictive maintenance in smart urban infrastructure," *Journal of Smart Infrastructure and Environmental Sustainability*, vol. 2, no. 1, pp. 1–10, 2025.

[9] M. A. Sami, A. Rehman, Z. Ahmad, and N. Bano, "Explainable AIOps: A deep survey on trustworthy and transparent AI in cloud-scale DevOps automation," *Spectrum of Engineering Sciences*, pp. 488–507, 2025.

[10] A. Olivia, "Explainable AI (XAI) in mission-critical systems: Enhancing transparency in financial forecasting and real-time patient monitoring," 2022.

[11] O. N. Ezechi *et al.*, "Service quality improvement in the banking sector: A data analytics perspective," *International Journal of Advanced Multidisciplinary Research and Studies*, vol. 5, no. 1, pp. 958–971, 2025.

[12] A. Mareedu, "Hybrid AI models in network security: Combining ML, DL, and rule-based systems," *International Journal of Emerging Research in Engineering and Technology*, vol. 5, no. 4, pp. 109–121, 2024.

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20685227

[13] B. Magableh and M. Almiani, "A self healing microservices architecture: A case study in Docker Swarm cluster," in *Proc. International Conference on Advanced Information Networking and Applications*, Springer International Publishing, 2019.