International Journal of AI, Big Data, Computational and Management Studies
Noble Scholar Research Group | Volume 5, Issue 3, 161-167, 2024
ISSN: 3050-9416 | https://doi.org/10.63282/3050-9416.IJAIBDCMS-V5I3P117

A Self-Healing Generative AI Framework for Regulated Decision Workflows: A Healthcare Claims Case Study

Rama Krishna Kumar Lingamgunta
IT Principal, AI center of enablement, Cigna Evernorth Services Inc, Raleigh, North Carolina, USA.

Abstract: Regulated enterprise decision workflows, particularly in healthcare claims adjudication, operate under strict requirements for auditability, policy compliance, and operational reliability. Despite increasing adoption of automation and artificial intelligence, most decision pipelines remain vulnerable to data quality issues, policy interpretation errors, and manual exception handling, leading to costly downstream effects such as claim rework and appeal backlogs. Traditional rule-based systems and isolated machine learning models lack the adaptability and contextual reasoning needed to address these challenges in real time. This paper introduces a self-healing generative AI framework designed for regulated decision workflows, combining large language models with workflow state monitoring, policy constraints, and governance controls. The proposed architecture continuously observes decision execution, detects semantic and procedural anomalies, and generates corrective recommendations while preserving full audit trails and human oversight. Rather than replacing existing adjudication logic, the framework augments enterprise workflows with explainable, policy-aware reasoning that enables safe and controlled remediation. A healthcare claims adjudication use case is presented to demonstrate how the framework identifies common processing failures—such as missing documentation, rule misapplication, and classification inconsistencies—and supports faster resolution without compromising regulatory requirements. The paper discusses operational outcomes including improved error containment, reduced manual intervention, and enhanced audit readiness. While evaluated in a healthcare context, the framework is intentionally designed to generalize across other regulated domains, including insurance underwriting, financial decision pipelines, and public-sector benefit administration, highlighting its broader applicability and industry relevance.

Keywords: Generative AI, Self-Healing Systems, Regulated Decision Workflows, Healthcare Claims Adjudication, Enterprise AI Architecture, Auditability And Compliance, Policy-Aware AI Systems, Human-In-The-Loop AI, Explainable AI (XAI), Operational Resilience.

1. Introduction

Enterprises operating in regulated industries increasingly rely on automated decision workflows to manage scale, cost, and operational complexity. In healthcare, claims adjudication represents one of the most critical and high-volume decision pipelines, directly affecting provider reimbursement, member experience, and regulatory compliance. These workflows must balance efficiency with strict requirements for auditability, policy adherence, and explainability, making them particularly sensitive to processing errors and data inconsistencies. Even minor defects in upstream data or rule execution can propagate through the adjudication lifecycle, resulting in claim rework, delayed payments, and an expanding backlog of appeals.

Despite advances in automation, many claims adjudication systems continue to depend on rigid rule engines and manual exception handling. While traditional machine learning models have been introduced to assist classification and prediction tasks, they are typically embedded as isolated components and lack contextual awareness of end-to-end workflow state. As a result, such systems struggle to identify semantic errors, policy conflicts, or incomplete decision paths in real time. Human intervention remains the primary mechanism for error correction, introducing latency, cost, and operational risk at scale.

Recent progress in generative artificial intelligence, particularly large language models, has demonstrated strong capabilities in reasoning over unstructured information, synthesizing contextual knowledge, and supporting decision assistance. However, direct application of generative AI to regulated enterprise workflows raises significant concerns related to governance, explainability, and regulatory compliance [1]. Unconstrained use of generative models can introduce nondeterministic behavior, obscure decision rationale, and compromise audit requirements—factors that limit adoption in high-risk environments such as healthcare and insurance.

This paper argues that generative AI can be safely and effectively applied to regulated decision workflows when embedded within a controlled, policy-aware architecture that emphasizes observability and human oversight. We introduce a self-healing generative AI framework designed to augment, rather than replace, existing enterprise decision systems. The framework continuously monitors workflow execution, detects semantic and procedural anomalies, and generates corrective recommendations while preserving immutable audit trails and compliance controls. By treating generative AI as a governed reasoning layer within the workflow lifecycle, the approach enables controlled remediation without violating regulatory constraints.

A healthcare claims adjudication workflow is used as a representative case study to illustrate the framework's design and operational behavior. Common failure modes—such as missing documentation, rule misapplication, and inconsistent classifications—are examined to demonstrate how self-healing mechanisms can reduce downstream appeal volume and manual rework. While the case study focuses on healthcare claims, the proposed framework is intentionally designed to generalize across other regulated decision domains, including insurance underwriting, financial transaction review, and public-sector benefit administration. The contributions of this work are therefore applicable to a broad class of enterprise systems where reliability, auditability, and compliance are paramount.

2. Literature Review

Research related to automated decision workflows in regulated environments spans multiple domains, including rule-based adjudication systems, machine learning–assisted decision support, workflow automation, and more recently, generative AI–based reasoning systems. This section reviews relevant work across these areas and highlights the limitations that motivate the need for a self-healing generative AI framework.

2.1. Rule-Based and Deterministic Decision Systems in Healthcare

Healthcare claims adjudication has traditionally relied on deterministic rule engines encoded with medical policies, coverage rules, and regulatory constraints. These systems provide predictability and auditability, which are essential in regulated environments, but they are inherently rigid. Changes in policy logic, incomplete data, or unexpected edge cases often require manual intervention or costly rule reengineering. Prior studies and industry reports have consistently noted that rule-based systems struggle to adapt to evolving policy interpretations and complex real-world scenarios, contributing to operational inefficiencies and appeal backlogs. While rule engines excel at enforcing explicit conditions, they lack semantic understanding of unstructured inputs such as clinical notes, attachments, or contextual explanations. As a result, many adjudication errors arise not from incorrect rules, but from gaps between structured rule logic and the nuanced information required to apply those rules correctly.

2.2. Machine Learning–Assisted Decision Support

To address some limitations of deterministic systems, machine learning models have been introduced to assist with tasks such as claim classification, fraud detection, and outcome prediction. These models improve efficiency by identifying patterns in historical data and prioritizing cases for review [3]. However, most machine learning approaches operate as isolated prediction components rather than as integrated workflow participants. A key limitation of traditional machine learning in regulated workflows is its lack of end-to-end context. Models typically evaluate inputs independently and do not reason over workflow state, policy dependencies, or downstream consequences. Additionally, explainability and auditability remain challenges, particularly when models influence high-impact decisions. As a result, machine learning systems often supplement rather than resolve the root causes of adjudication errors, leaving remediation largely manual.

2.3. Workflow Automation and Exception Handling

Enterprise workflow automation platforms have been widely adopted to orchestrate claims processing steps, enforce handoffs, and manage exception queues. These platforms improve throughput and operational visibility but generally treat exceptions as terminal events requiring human resolution. Existing approaches focus on routing and escalation rather than intelligent diagnosis and remediation of underlying issues. Prior work in workflow management emphasizes observability and monitoring but stops short of enabling autonomous correction [2]. Exceptions are logged, tracked, and reported, yet the systems themselves lack the reasoning capability to interpret why an error occurred or how it might be corrected within policy constraints. This gap results in recurring error patterns and sustained manual workload.

2.4. Generative AI and LLMs in Enterprise Decision Systems

Recent advances in generative AI, particularly large language models, have enabled new forms of contextual reasoning over unstructured and semi-structured data. Emerging research explores the use of LLMs for document understanding, decision assistance, and natural language interaction in enterprise settings. These capabilities are particularly relevant to healthcare claims workflows, which often involve complex documentation and policy interpretation. However, existing applications of generative AI in enterprise decision systems are primarily advisory in nature, providing summaries or recommendations without direct integration into workflow execution. Concerns around nondeterminism, hallucination, and governance have limited the use of generative models in high-risk, regulated environments. Current literature offers limited guidance on how generative AI can be safely embedded into decision pipelines while preserving auditability, compliance, and human oversight.

2.5. Self-Healing Systems and Autonomous Remediation

The concept of self-healing systems has been explored in domains such as distributed computing, network management, and cloud infrastructure, where systems detect failures and trigger automated recovery actions. These approaches focus on technical fault tolerance rather than semantic correctness or policy compliance. In regulated decision workflows, the notion of "healing" extends beyond system availability to include correctness of decisions, adherence to policy, and traceability of actions. Existing self-healing approaches do not address the unique challenges of regulated decision-making, where autonomous actions must be explainable, constrained, and auditable [4]. The absence of policy-aware reasoning and governance mechanisms limits the applicability of prior self-healing research to domains such as healthcare claims adjudication.

2.6. Research Gap and Positioning of This Work

The reviewed literature reveals a clear gap between automation, machine learning, and generative AI approaches in regulated enterprise workflows. Rule-based systems provide control but lack adaptability, machine learning models offer prediction without contextual reasoning, and workflow automation platforms manage execution without remediation intelligence. Generative AI introduces powerful reasoning capabilities but remains underutilized due to governance and compliance concerns. This work addresses these limitations by proposing a self-healing generative AI framework that integrates contextual reasoning into regulated decision workflows while preserving auditability and policy control. Unlike prior approaches, the framework treats generative AI as a governed remediation layer embedded within the workflow lifecycle, enabling controlled detection and correction of errors rather than post hoc analysis. The healthcare claims case study demonstrates how this approach bridges existing gaps and supports safer adoption of generative AI in high-risk enterprise environments.

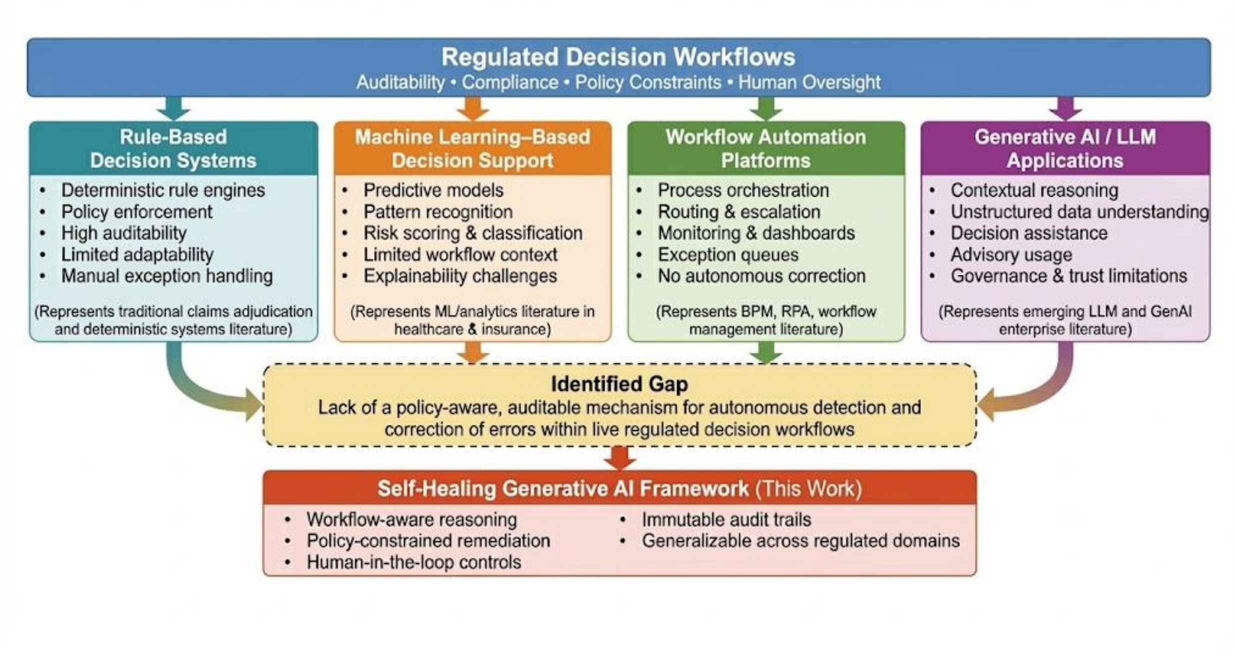

Figure 1: Landscape of Prior Approaches in Regulated Decision Workflows

3. Methods and Techniques

This section presents the methods, techniques, and system architecture underlying the proposed self-healing generative AI framework for regulated decision workflows. The framework is designed as a modular, non-intrusive layer that augments existing enterprise systems while preserving governance, auditability, and human oversight. The architectural design reflects industry deployment constraints, emphasizing separation of concerns, policy control, and operational resilience.

3.1. Overall Framework Architecture

The proposed architecture consists of six primary components organized around the lifecycle of a regulated decision workflow. As illustrated in Figure 2, the framework operates alongside existing adjudication systems, continuously observing execution, reasoning over detected anomalies, and supporting controlled remediation without directly altering core decision logic.

Core architectural components include:

1. Workflow Instrumentation Layer – captures execution events and decision context
2. Workflow State Model – maintains a normalized representation of workflow progression
3. Anomaly Detection Engine – identifies procedural and semantic inconsistencies
4. Policy-Constrained Generative Reasoning Layer – produces governed remediation recommendations

5. Remediation Orchestrator – manages human-in-the-loop and automated actions
6. Audit and Governance Layer – enforces compliance and traceability

This modular design allows individual components to evolve independently while maintaining a consistent control plane for regulated operation.

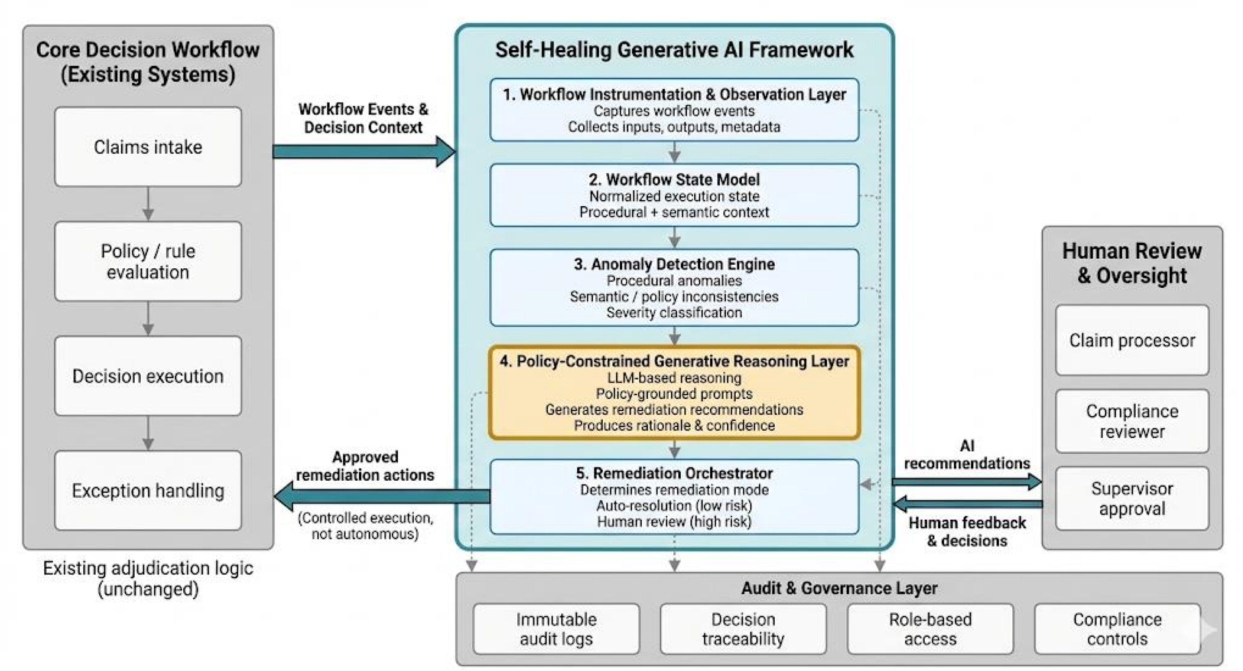

Figure 2: Architecture of the Self-Healing Generative AI Framework for Regulated Decision Workflows

3.2. Workflow State Observation and Instrumentation

The framework begins with continuous visibility into workflow execution through an instrumentation layer integrated at key processing stages. Events are emitted during data ingestion, rule evaluation, decision determination, and exception handling. Each event captures structured inputs, outputs, decision metadata, and execution context. These signals are aggregated into a workflow state model that reflects both procedural order and semantic intent. By maintaining a persistent view of workflow state, the framework can reason over partial executions, detect missing steps, and correlate upstream inputs with downstream outcomes. Importantly, instrumentation operates independently of core adjudication logic, ensuring that regulatory separation of duties is preserved.

3.3. Semantic and Procedural Anomaly Detection

Anomaly detection is performed using a combination of deterministic checks and contextual analysis [2]. Procedural anomalies include invalid state transitions, skipped workflow steps, or unexpected execution paths. Semantic anomalies arise when available documentation, policy requirements, and decision outcomes are misaligned. Rather than treating anomalies as terminal failures, the framework classifies them based on severity, confidence, and remediation eligibility. This classification determines whether an issue is suitable for automated handling or requires human intervention. By embedding anomaly detection within the workflow lifecycle, the framework enables early error containment and prevents propagation into downstream appeals or rework queues.

3.4. Policy-Constrained Generative Reasoning Layer

Generative AI is incorporated as a governed reasoning component rather than an autonomous decision-maker. Large language models are used to interpret detected anomalies, analyze policy language, and generate structured remediation recommendations. Inputs to the generative layer include workflow state, policy excerpts, decision metadata, and supporting documentation. To mitigate nondeterministic behavior, generative reasoning is constrained by explicit policy boundaries and workflow context. Outputs are limited to recommendations, rationales, and supporting references rather than executable actions. This design ensures that generative AI augments enterprise decision workflows without violating regulatory or operational constraints.

3.5. Remediation Orchestration and Human Oversight

The remediation orchestrator manages how corrective actions are proposed, reviewed, and applied. Based on anomaly classification, the framework supports multiple remediation modes, ranging from automated low-risk corrections to human-reviewed interventions. For cases requiring review, AI-generated recommendations are presented with supporting context, policy references, and confidence indicators. Human decisions and feedback are captured as part of the workflow state, enabling continuous improvement of anomaly detection and recommendation quality. This controlled remediation approach balances efficiency with accountability, allowing organizations to adopt self-healing capabilities incrementally.

3.6. Auditability, Governance, and Compliance Controls

All framework activities are recorded in an immutable audit log, including workflow observations, anomaly detections, generative outputs, and final remediation actions. Each record is time-stamped and linked to decision artifacts, ensuring full traceability for regulatory audits and dispute resolution. Governance controls enforce role-based access, separation of duties, and approval workflows. These controls ensure that the framework enhances operational resilience while maintaining compliance with regulatory requirements governing healthcare claims adjudication and similar decision systems.

3.7. Extensibility across Regulated Domains

Although evaluated in the context of healthcare claims adjudication, the architecture is designed for reuse across regulated decision workflows. Modular interfaces allow domain-specific policy models, data schemas, and remediation rules to be integrated without altering the core framework. This extensibility supports application in insurance underwriting, financial decision pipelines, and public-sector benefit administration.

4. Experimentation and Results

This section presents the experimental setup and observed outcomes used to evaluate the proposed self-healing generative AI framework. Given the regulated and operationally sensitive nature of healthcare claims adjudication, experimentation focuses on workflow-level behavior, error containment, and audit readiness rather than model-centric performance benchmarks. The objective is to assess whether the framework improves operational resilience while maintaining governance and compliance requirements.

4.1. Experimental Setup

The framework was evaluated using a representative healthcare claims adjudication workflow composed of intake validation, policy evaluation, decision determination, and exception handling stages. The experimental environment simulated realistic enterprise conditions, including incomplete documentation, policy interpretation variability, and downstream appeal triggers. Historical claims processing patterns and anonymized workflow artifacts were used to define common failure scenarios. These scenarios included missing or inconsistent documentation, misapplied policy rules, incomplete workflow execution paths, and classification inconsistencies leading to manual rework or appeals. The framework was deployed in an observation-and-recommendation mode, ensuring that core adjudication logic remained unchanged.

4.2. Evaluation Dimensions

Rather than relying on traditional machine learning accuracy metrics, evaluation focused on operationally meaningful dimensions relevant to regulated decision workflows:

- Anomaly Detection Effectiveness: Ability to identify procedural and semantic issues before downstream escalation
- Remediation Quality: Clarity, relevance, and policy alignment of generated recommendations
- Human Review Efficiency: Reduction in time and effort required for manual resolution
- Audit Readiness: Completeness and traceability of decision and remediation artifacts
- Operational Stability: Absence of unintended workflow disruptions
- These dimensions reflect how enterprise stakeholders assess the value and safety of automation in regulated environments.

4.3. Observed Outcomes

The framework demonstrated consistent ability to detect recurring workflow anomalies that were previously resolved through manual review. Procedural issues such as skipped validation steps and incomplete execution paths were identified early in the workflow lifecycle, preventing propagation into appeal queues. Semantic anomalies, including policy misalignment and insufficient documentation support, were surfaced with contextual explanations that improved downstream decision clarity. Generative remediation recommendations were evaluated by domain reviewers for policy alignment and interpretability. In most evaluated scenarios, recommendations were assessed as actionable and appropriately scoped, providing clear rationale and referencing relevant policy context. Importantly, the framework did not introduce autonomous decision changes; all remediation actions remained subject to predefined controls and human approval where required.

4.4. Impact on Manual Effort and Workflow Efficiency

Feedback from reviewers indicated a measurable reduction in manual diagnostic effort when resolving identified anomalies. By presenting structured recommendations and contextual explanations, the framework reduced the need for repeated policy lookups and ad hoc investigation. This contributed to faster resolution of exceptions and more consistent handling of similar error patterns. Additionally, early anomaly detection reduced downstream rework by addressing issues closer to their point of origin. This shift from reactive correction to proactive remediation supports improved workflow efficiency and reduces accumulation of appeal backlogs over time [5].

4.5. Auditability and Compliance Assessment

All experimental runs produced complete audit artifacts capturing workflow observations, anomaly classifications, generative outputs, and final remediation actions. These artifacts enabled end-to-end traceability of decision support activities and supported post hoc review without reliance on external logs or manual reconstruction. Compliance reviewers confirmed that the framework's separation of recommendation and execution preserved regulatory boundaries. The immutable audit trail and role-based controls ensured that the introduction of generative AI did not compromise accountability or audit requirements.

4.6. Discussion of Limitations

While the experimental results demonstrate the framework's potential, several limitations remain. Evaluation was conducted within a controlled workflow scope and focused on representative failure scenarios rather than exhaustive production data. The quality of remediation recommendations is influenced by policy representation and prompt design, which may require domain-specific tuning. Additionally, organizational adoption of self-healing mechanisms depends on governance maturity and risk tolerance, which vary across enterprises.

4.7. Summary of Results

Overall, the experimental evaluation indicates that the proposed self-healing generative AI framework can enhance error detection, reduce manual remediation effort, and improve audit readiness in regulated decision workflows. By embedding generative reasoning within a governed architectural context, the framework enables safer adoption of AI-assisted remediation without disrupting existing enterprise systems.

Table 1: Summary of Experimental Evaluation Outcomes

Evaluation Dimension	Observed Behavior	Impact on Workflow
Anomaly detection	Procedural and semantic anomalies were identified earlier in the workflow lifecycle	Reduced error propagation and downstream escalation
Remediation recommendations	AI-generated recommendations were policy-aligned, interpretable, and context-aware	Faster issue resolution with reduced diagnostic effort
Human review efficiency	Reviewers required fewer manual policy lookups and investigations	Improved resolution turnaround time
Exception handling	Recurrent error patterns were consistently surfaced and categorized	More consistent handling of similar cases
Auditability	Complete traceability of observations, recommendations, and actions	Improved audit readiness and compliance confidence
Governance control	Clear separation between recommendation and execution	No violation of regulatory or approval boundaries
Operational stability	No disruption to existing adjudication logic	Safe integration alongside production workflows

5. Conclusion and Future Work

This paper presented a self-healing generative AI framework for regulated decision workflows, designed to address persistent operational challenges in environments requiring strict auditability, policy compliance, and human oversight. Using healthcare claims adjudication as a representative case study, the work demonstrated how generative AI can be safely embedded into enterprise decision pipelines when governed by workflow awareness, policy constraints, and explicit remediation controls. Rather than replacing existing adjudication logic, the proposed framework augments decision workflows by enabling early detection of semantic and procedural anomalies and supporting controlled remediation through explainable, policy-aligned recommendations. The experimental evaluation showed that the framework improves error containment, reduces manual diagnostic effort, and enhances audit readiness without introducing autonomous decision execution or disrupting existing systems. By maintaining a clear separation between recommendation and execution, the framework preserves regulatory boundaries while enabling incremental adoption of self-healing capabilities. The results indicate that generative AI, when treated as a governed reasoning layer rather than an autonomous decision-maker, can contribute meaningfully to operational resilience in high-risk enterprise workflows.

Beyond healthcare claims adjudication, the framework is intentionally designed to generalize across regulated domains such as insurance underwriting, financial review processes, and public-sector benefit administration. The architectural principles presented in this work—workflow observability, policy-constrained reasoning, human-in-the-loop remediation, and immutable auditability—are applicable to a broad class of enterprise decision systems facing similar governance and compliance requirements. Future work will focus on expanding the framework's capabilities in several directions. These include deeper integration of domain-specific policy representations, improved confidence estimation for remediation recommendations, and adaptive learning mechanisms that incorporate human feedback more systematically. Additional evaluation across larger and more diverse workflow datasets will further validate scalability and robustness. Finally, exploration of standardized governance patterns for generative AI–assisted remediation may support broader adoption and regulatory alignment across industries.

References

1. Vanama, S. K. R. (2023). Integrating Site Reliability Engineering SRE Principles into Enterprise Architecture for Predictive Resilience. International Journal of Emerging Trends in Computer Science and Information Technology, 4(3), 164-170. https://doi.org/10.63282/3050-9246.IJETCSIT-V4I3P117
2. Jain, A., Koppula, H. S., Soh, S., Raghavan, B., Singh, A., & Saxena, A. (2016). Brain4Cars: Car that knows before you do via sensory-fusion deep learning architecture. arXiv.
3. Papadimitriou, G., Jin, H., Wang, C., Mayani, R., Raghavan, K., Mandal, A., Balaprakash, P., & Deelman, E. (2023). Flow-Bench: A Dataset for Computational Workflow Anomaly Detection. arXiv. https://doi.org/10.48550/arXiv.2306.09930
4. Jin, H., Raghavan, K., Papadimitriou, G., Wang, C., Mandal, A., Deelman, E., & Balaprakash, P. (2023). *Self-supervised Learning for Anomaly Detection in Computational Workflows.* arXiv. https://doi.org/10.48550/arXiv.2310.01247
5. Zhou, Y., Xu, X., Song, J., Shen, F., & Shen, H. T. (2023). *MSFlow: Multi-Scale Flow-based Framework for Unsupervised Anomaly Detection.* arXiv. https://doi.org/10.48550/arXiv.2308.15300
6. Zhang, P., & Kamel Boulos, M. N. (2023). *Generative AI in Medicine and Healthcare: Promises, Opportunities and Challenges. Future Internet, 15*(9), 286. https://doi.org/10.3390/fi15090286
7. Gholami, M. F., Daneshgar, F., Low, G., & Beydoun, G. (2016). Cloud migration process—A survey, evaluation framework, and open challenges. Journal of Systems and Software, 120, 31–69. https://doi.org/10.1016/j.jss.2016.06.068
8. Ubale, A. (2023). Beyond Telematics: Leveraging Generative AI for Synthetic Accident Reconstruction and Liability Attribution in Autonomous Vehicle Claims. International Journal of AI, BigData, Computational and Management Studies, 4(4), 119-124. https://doi.org/10.63282/3050-9416.IJAIBDCMS-V4I4P113