International Journal of Modern Medicine
VOLUME04 ISSUE10
PUBLISHED DATE: - 01-10-2025
Page 42-52
eISSN: 2156-5198 pISSN: 2156-518X

RESEARCH ARTICLE

Toward Resilient Legacy Retail Architectures: A Socio-Technical Integration of Site Reliability Engineering and Machine Learning Observability

Dr. Adrian M. Thorne
Department of Molecular Biology, Cambridge Institute for Biomedical Research, Cambridge, United Kingdom

Abstract: The accelerating digitalization of retail enterprises has placed unprecedented operational strain on legacy infrastructure that was never designed to support real-time analytics, continuous deployment, or machine learning–driven decision systems. As retail organizations increasingly integrate machine learning models into mission-critical workflows such as demand forecasting, pricing optimization, fraud detection, and inventory management, the reliability of both software systems and embedded models has emerged as a strategic concern rather than a purely technical one. Site Reliability Engineering (SRE), originally conceptualized within large-scale cloud-native organizations, offers a principled approach to managing reliability through explicit service-level objectives, automation, and error budgeting. However, the translation of SRE principles into legacy retail environments—characterized by monolithic architectures, heterogeneous data pipelines, and organizational inertia—remains insufficiently theorized. This research addresses that gap by developing a comprehensive socio-technical framework that integrates SRE practices with modern machine learning observability and MLOps methodologies in the specific context of legacy retail infrastructure.

Drawing on an extensive critical synthesis of contemporary scholarship on SRE implementation in retail systems (Dasari, 2025), machine learning drift detection, model monitoring, observability, industrial reliability engineering, and ethical AI governance, this article advances a holistic conceptual model for resilient retail operations. Rather than treating infrastructure reliability and model performance as separate domains, the study conceptualizes them as co-evolving layers within a single operational ecosystem. The methodology is grounded in qualitative comparative analysis of documented industry practices, theoretical extrapolation from reliability engineering literature, and interpretive analysis of production ML monitoring frameworks. The results demonstrate that reliability failures in retail systems are rarely attributable to isolated technical faults; instead, they emerge from systemic misalignments between organizational incentives, data quality regimes, observability maturity, and operational governance structures.

The discussion critically evaluates competing scholarly perspectives on automation, human oversight, and ethical accountability in high-reliability digital systems, arguing that SRE-informed MLOps provides a viable pathway for balancing operational resilience with responsible AI deployment. The article concludes by articulating implications for practitioners, researchers, and policymakers, emphasizing the necessity of reimagining legacy retail infrastructure as adaptive socio-technical systems rather than static technological artifacts. By embedding machine learning observability within an SRE-oriented reliability culture, retail organizations can transition from reactive incident management to proactive, ethically grounded operational excellence.

eISSN: 2156-5198 pISSN: 2156-518X

RESEARCH ARTICLE

Keywords: Site Reliability Engineering; Legacy Retail Infrastructure; Machine Learning Observability; MLOps; Digital Resilience; Socio-Technical Systems

INTRODUCTION

The contemporary retail sector occupies a paradoxical position within the digital economy. On the one hand, retail organizations are among the most data-intensive enterprises in existence, generating vast streams of transactional, behavioral, and logistical information on a continuous basis. On the other hand, much of the technological infrastructure underpinning these data flows is rooted in legacy systems designed for a fundamentally different operational paradigm, one characterized by batch processing, static demand assumptions, and relatively stable business cycles. The tension between these historical architectures and modern expectations of real-time responsiveness, predictive analytics, and continuous availability has become a defining challenge for retail engineering practice (Dasari, 2025). This challenge is further intensified by the growing integration of machine learning models into operational decision-making, a development that introduces new forms of uncertainty, opacity, and failure modes into already complex systems (Lewis et al., 2022).

Site Reliability Engineering has emerged over the past two decades as a dominant framework for managing reliability in large-scale distributed systems. Rooted in the operational philosophies of internet-scale technology firms, SRE reframes reliability as a feature that can be engineered through measurement, automation, and disciplined risk management rather than as an aspirational outcome pursued through ad hoc heroics. Central to this philosophy are constructs such as service-level indicators, service-level objectives, and error budgets, which collectively enable organizations to make explicit trade-offs between innovation velocity and operational stability (Dasari, 2025). While the theoretical elegance of SRE has been widely acknowledged, its practical application outside cloud-native environments remains underexplored, particularly in sectors such as retail where legacy infrastructure, regulatory constraints, and organizational silos complicate implementation.

The incorporation of machine learning into retail operations adds another layer of complexity to this landscape. Unlike traditional software systems, machine learning models are inherently probabilistic and data-dependent, meaning that their performance can degrade over time due to changes in underlying data distributions, business processes, or external conditions. This phenomenon, commonly referred to as model drift, poses significant risks in production environments where decisions must be both timely and accurate (Lewis et al., 2022). In response, the field of MLOps has emerged to address the lifecycle management of machine learning systems, emphasizing automation, reproducibility, and monitoring as core principles (Singla, 2023). However, MLOps frameworks are often implemented in isolation from broader reliability engineering practices, leading to fragmented observability and reactive incident response.

Legacy retail infrastructure magnifies these challenges by introducing constraints that are both technical and organizational in nature. Many retail systems rely on tightly coupled monolithic applications,

eISSN: 2156-5198 pISSN: 2156-518X

proprietary data formats, and aging hardware platforms that resist modularization and automation. At the same time, retail organizations often exhibit deeply entrenched functional divisions between IT operations, data science teams, and business stakeholders, each operating under distinct performance metrics and incentive structures. Dasari (2025) highlights that attempts to implement SRE in such environments frequently falter not because of technical infeasibility but due to misalignment between reliability objectives and organizational culture. This insight underscores the necessity of a socio-technical perspective that situates engineering practices within their broader institutional contexts.

The literature on reliability engineering provides valuable historical context for understanding these dynamics. Traditionally, reliability engineering focused on physical systems such as manufacturing equipment, power grids, and transportation networks, where failure modes could be modeled through statistical analysis of component lifetimes and stress factors (Payette & Payette, 2023). With the advent of software-intensive systems, these principles were adapted to account for non-deterministic behaviors, human intervention, and complex interdependencies. The rise of machine learning further complicates this evolution by introducing adaptive components whose behavior changes in response to data, thereby challenging conventional notions of predictability and control (Huang et al., 2020).

Observability has emerged as a unifying concept that bridges these diverse domains. In the context of software systems, observability refers to the ability to infer the internal state of a system from its external outputs, typically through metrics, logs, and traces (Maverick, 2019). In machine learning systems, observability extends to monitoring data quality, feature distributions, model performance, and decision impacts over time (Evidently AI, 2025). Retail organizations that fail to develop robust observability capabilities risk operating in a state of epistemic blindness, where failures are detected only after they have manifested as customer dissatisfaction, financial loss, or regulatory violations.

Despite the growing body of work on SRE, MLOps, and observability, there remains a notable gap in the literature concerning their integrated application within legacy retail environments. Existing studies tend to focus on isolated aspects of the problem, such as drift detection algorithms, logging best practices, or ethical principles for AI deployment, without addressing how these elements interact within complex organizational systems (UTS, 2020). Dasari (2025) represents a significant contribution in this regard by explicitly examining the challenges of implementing SRE in legacy retail infrastructure, yet even this work stops short of fully integrating machine learning observability into the SRE paradigm. This article seeks to extend that foundation by developing a comprehensive framework that treats infrastructure reliability and model reliability as mutually reinforcing dimensions of operational excellence.

The central research question guiding this study is how Site Reliability Engineering principles can be systematically integrated with machine learning observability and MLOps practices to enhance the resilience of legacy retail systems. Addressing this question requires moving beyond purely technical considerations to examine the organizational, ethical, and epistemological dimensions of reliability. Retail systems do

RESEARCH ARTICLE

not operate in a vacuum; they are embedded within socio-economic contexts that shape data generation, decision-making priorities, and accountability structures. Consequently, any attempt to improve reliability must grapple with issues of governance, transparency, and human oversight alongside questions of automation and scalability (Lewis et al., 2022).

The remainder of this article is structured to progressively build this argument through theoretical elaboration, methodological rigor, and critical discussion. The methodology section outlines the interpretive and comparative approach adopted in synthesizing insights from diverse scholarly and practitioner-oriented sources, acknowledging the limitations inherent in qualitative analysis. The results section presents a descriptive account of key patterns and themes that emerge from the literature, emphasizing the systemic nature of reliability challenges in retail contexts. The discussion section offers an extended theoretical interpretation of these findings, situating them within broader debates on socio-technical systems, ethical AI, and organizational change. The conclusion synthesizes the implications of the study and identifies avenues for future research aimed at operationalizing the proposed framework in empirical settings.

Throughout the article, particular attention is paid to grounding theoretical claims in established scholarship while advancing novel conceptual linkages. By positioning Site Reliability Engineering as a bridge between legacy infrastructure and modern machine learning systems, this research contributes to a more integrated understanding of digital resilience in the retail sector. In doing so, it responds to the pressing need for frameworks that are not only technically sound but also organizationally and ethically sustainable in an era of rapid technological transformation (Dasari, 2025).

METHODOLOGY

The methodological orientation of this research is interpretive, integrative, and theoretically driven, reflecting the complexity of the subject matter and the absence of standardized empirical datasets that adequately capture the socio-technical dimensions of reliability in legacy retail systems. Rather than pursuing a positivist experimental design, the study adopts a qualitative synthesis methodology that draws on comparative analysis of existing scholarly literature, industry case narratives, and conceptual frameworks related to Site Reliability Engineering, machine learning observability, and MLOps. This approach is particularly appropriate given the exploratory nature of the research question and the need to bridge multiple disciplinary domains, including software engineering, reliability engineering, data science, and organizational studies (Payette & Payette, 2023).

At the core of the methodology is a structured literature integration process designed to identify, contextualize, and critically evaluate key themes across the provided references. The priority reference by Dasari (2025) serves as the foundational anchor for the analysis, offering a sector-specific examination of SRE implementation challenges within legacy retail infrastructure. This work is treated not as an isolated case but as a representative articulation of broader structural tensions that characterize the retail industry's digital transformation. Insights from this reference are systematically cross-examined against literature on machine learning production systems, observability practices, and ethical governance to uncover points of convergence and divergence (Lewis et al., 2022).

The analytical process unfolds in three interrelated stages. The first stage involves thematic extraction, whereby key concepts such as reliability objectives, error budgets, drift detection, observability maturity, and organizational alignment are identified across the reference corpus. These concepts are then mapped onto a conceptual matrix that distinguishes between technical, organizational, and ethical dimensions of reliability. This mapping exercise enables the identification of latent assumptions and implicit value judgments embedded within different engineering paradigms, an approach consistent with interpretive research traditions in information systems studies (UTS, 2020).

The second stage focuses on comparative interpretation. Here, the study juxtaposes SRE principles as articulated in the context of large-scale technology firms with the realities of legacy retail environments described by Dasari (2025). This comparison highlights the extent to which canonical SRE practices presuppose infrastructural flexibility, cultural norms of experimentation, and data transparency that may be absent in retail settings. Simultaneously, literature on machine learning observability and MLOps is examined to assess whether these frameworks adequately address the reliability concerns identified in retail contexts or whether they replicate similar assumptions about organizational readiness (Singla, 2023).

The third stage involves integrative theorization, in which insights from the comparative analysis are synthesized into a cohesive socio-technical framework. This framework conceptualizes reliability as an emergent property of interactions between infrastructure, models, humans, and governance structures rather than as a static attribute of individual components. The framework is not presented as a prescriptive blueprint but as a heuristic device intended to guide both scholarly inquiry and practical intervention. Its development is informed by reliability engineering literature that emphasizes systemic thinking and the inevitability of failure in complex systems (Huang et al., 2020).

Methodological rigor is maintained through reflexive consideration of limitations and potential biases. One notable limitation is the reliance on secondary sources, which constrains the ability to validate theoretical claims through direct empirical observation. However, this limitation is mitigated by the depth and diversity of the referenced materials, which encompass academic research, industry white papers, and practitioner blogs documenting real-world experiences with observability and reliability practices (Evidently AI, 2025; Maverick, 2019). Another limitation concerns the generalizability of findings, as retail organizations exhibit significant heterogeneity in size, geography, and technological maturity. Rather than seeking universal applicability, the study emphasizes analytical generalization, aiming to elucidate mechanisms and relationships that may manifest differently across contexts (Dasari, 2025).

Ethical considerations are also integral to the methodological stance of the research. Given the increasing societal impact of machine learning systems in retail, including issues of consumer privacy, algorithmic bias, and accountability, the study explicitly engages with ethical AI frameworks to contextualize reliability beyond narrow operational metrics (UTS, 2020). This ethical lens informs the interpretation of observability practices,

International Journal of Modern Medicine
VOLUME04 ISSUE10
PUBLISHED DATE: - 01-10-2025
Page 42-52
eISSN: 2156-5198 pISSN: 2156-518X

highlighting the importance of transparency and explainability as components of trustworthy systems rather than optional add-ons.

In sum, the methodology reflects a deliberate choice to privilege depth of understanding over empirical breadth, recognizing that the challenges of integrating SRE and machine learning observability in legacy retail systems are as much conceptual and organizational as they are technical. By synthesizing insights across disciplines and grounding them in sector-specific analysis, the study aims to contribute a robust theoretical foundation for future empirical research and practical experimentation (Dasari, 2025).

RESULTS

The results of this study are presented as an interpretive synthesis of patterns, tensions, and systemic dynamics identified across the integrated body of literature, with particular emphasis on how Site Reliability Engineering and machine learning observability intersect within legacy retail infrastructures. Rather than reporting empirical measurements, the results articulate conceptual findings that emerge from comparative analysis, consistent with qualitative research traditions in reliability and socio-technical systems research (Payette & Payette, 2023). Each thematic result reflects recurring insights across multiple sources and is grounded in documented industry practices and scholarly interpretations (Dasari, 2025).

One of the most prominent findings is that reliability failures in legacy retail systems rarely originate from isolated technical defects. Instead, they emerge as cascading phenomena resulting from structural misalignments between infrastructure capabilities, data pipelines, machine learning models, and organizational processes. Dasari (2025) demonstrates that in many retail environments, outages and performance degradations are initially triggered by seemingly minor configuration changes or data anomalies that propagate through tightly coupled systems. When machine learning models are embedded within these systems, the impact of such anomalies is amplified, as models may continue to generate outputs that appear valid while gradually diverging from business reality due to undetected drift (Lewis et al., 2022).

A second key result concerns the uneven maturity of observability practices across system layers. Traditional IT operations within retail organizations often exhibit relatively mature monitoring of hardware utilization, network latency, and application uptime, reflecting decades of operational experience. However, observability at the data and model layers is frequently underdeveloped, with limited visibility into feature distributions, prediction confidence, or downstream decision impacts (Evidently AI, 2025). This asymmetry creates blind spots that undermine the effectiveness of SRE practices, as service-level objectives may be met at the infrastructure level while failing at the decision-quality level, leading to suboptimal or even harmful outcomes for customers and the business (Singla, 2023).

The analysis also reveals that error budgeting, a cornerstone of SRE, is conceptually challenging to apply in the context of machine learning systems within retail. In traditional SRE implementations, error budgets are defined in terms of service availability or latency thresholds that can be objectively measured. By contrast, the "errors" associated with machine learning models—such as prediction inaccuracies, biased recommendations, or delayed adaptation to market shifts—are often probabilistic and context-dependent (Huang et al., 2020). As a result, retail organizations struggle to define meaningful reliability targets for models, leading either to overly permissive tolerances that mask risk or to unrealistic expectations of deterministic performance (Dasari, 2025).

Another significant finding relates to organizational culture and incentive structures. The literature consistently indicates that attempts to introduce SRE and MLOps practices into legacy retail organizations are frequently hampered by entrenched silos between development, operations, and data science teams (Shopify Engineering, 2019). These silos manifest not only in technical interfaces but also in divergent epistemologies regarding what constitutes acceptable risk and success. Operations teams prioritize stability and predictability, data scientists emphasize model accuracy and experimentation, and business stakeholders focus on short-term financial metrics. Without a unifying reliability framework, these perspectives remain misaligned, resulting in reactive incident management rather than proactive resilience building (Dasari, 2025).

The results further highlight the ethical dimensions of observability as an emergent concern in retail machine learning systems. Monitoring practices that focus exclusively on technical performance may overlook issues related to fairness, transparency, and consumer trust, particularly when models influence pricing, credit decisions, or personalized marketing (UTS, 2020). Several sources emphasize that ethical failures often coincide with reliability failures, as both stem from inadequate visibility into system behavior and insufficient accountability mechanisms (Encord, 2024). This convergence suggests that observability should be conceptualized not merely as a technical capability but as a governance practice that supports responsible decision-making.

Finally, the synthesis identifies a recurring pattern of incremental rather than transformative adoption of SRE principles in retail contexts. Organizations tend to selectively implement elements such as on-call rotations or automated alerts without fully embracing the cultural and conceptual shifts required for SRE to function as intended (Dasari, 2025). When machine learning observability is layered onto this partial adoption, the result is often a fragmented toolchain that generates data without delivering actionable insight. This finding underscores the necessity of an integrated approach that aligns technical instrumentation with organizational learning processes and strategic objectives (Lewis et al., 2022).

Collectively, these results point to the conclusion that improving reliability in legacy retail systems requires more than the deployment of advanced monitoring tools or the formal adoption of SRE terminology. It necessitates a reconfiguration of how reliability is defined, measured, and governed across both infrastructural and algorithmic domains. The implications of these findings are explored in depth in the following discussion section, which situates them within broader theoretical debates and examines their significance for future research and practice (Dasari, 2025).

DISCUSSION

The findings of this study invite a deep theoretical reconsideration of how reliability is conceptualized and operationalized in legacy retail systems undergoing machine learning–driven transformation. At a fundamental level, they challenge reductionist views that treat infrastructure reliability, model performance, and organizational

RESEARCH ARTICLE

effectiveness as separable concerns. Instead, the evidence supports a socio-technical interpretation in which reliability emerges from the dynamic interaction of technical artifacts, human actors, and institutional structures (Payette & Payette, 2023). This discussion elaborates on that interpretation by situating the results within existing scholarly debates, addressing counterarguments, and exploring implications for both theory and practice.

A central theoretical contribution of this research lies in its extension of Site Reliability Engineering beyond its traditional infrastructural focus to encompass machine learning systems as first-class reliability objects. Classical SRE literature assumes that system behavior, while complex, is ultimately deterministic within defined operational envelopes. Failures are therefore framed as deviations from expected performance that can be detected, mitigated, and prevented through automation and disciplined process design (Dasari, 2025). Machine learning systems disrupt this assumption by introducing adaptive behavior that evolves over time in response to data, rendering static reliability targets insufficient (Lewis et al., 2022). The challenge, then, is not merely technical but epistemological: organizations must accept uncertainty as an inherent feature of their systems rather than as a temporary aberration.

Some scholars argue that MLOps frameworks already address this challenge by emphasizing continuous monitoring and retraining pipelines (Singla, 2023). However, the results of this study suggest that MLOps alone is insufficient when implemented without an overarching reliability philosophy. MLOps practices often prioritize model lifecycle efficiency—such as faster deployment or automated retraining—without explicitly articulating how these activities relate to broader service-level objectives or organizational risk tolerance (Evidently AI, 2025). By integrating MLOps within an SRE framework, retail organizations can align model monitoring activities with explicit reliability goals, thereby transforming observability data into actionable governance signals rather than passive metrics.

A potential counterargument to this integrationist perspective is that legacy retail systems lack the architectural modularity required to support SRE-style automation and observability. Monolithic applications, proprietary vendor platforms, and regulatory constraints may limit the feasibility of fine-grained instrumentation or rapid iteration (Dasari, 2025). While this concern is valid, it risks reifying legacy constraints as immutable rather than as historically contingent. The literature on incremental modernization demonstrates that even tightly coupled systems can benefit from targeted observability enhancements, such as centralized logging, data quality checks, and shadow monitoring of model outputs (Maverick, 2019). The key insight is that SRE need not be implemented in its canonical form to be effective; its principles can be adapted to local constraints through pragmatic interpretation.

The organizational implications of this adaptation are profound. As the results indicate, misaligned incentives and cultural silos represent some of the most significant barriers to reliability improvement in retail contexts. From a socio-technical perspective, these barriers reflect deeper tensions between competing logics of control, innovation, and accountability (Shopify Engineering, 2019). SRE's emphasis on shared ownership and blameless postmortems offers a mechanism for reconciling these tensions by reframing failures as learning opportunities rather than as grounds for punishment (Dasari, 2025). When extended to machine learning systems, this ethos encourages cross-functional dialogue about model behavior, data assumptions, and ethical trade-offs, fostering a more holistic understanding of system risk.

Ethical considerations further complicate this landscape, as reliability failures in retail machine learning systems can have direct social consequences. Dynamic pricing algorithms, for example, may inadvertently discriminate against vulnerable populations if trained on biased data or deployed without adequate oversight (UTS, 2020). Traditional reliability metrics such as uptime or latency provide little insight into such harms, underscoring the need for expanded observability that incorporates ethical performance indicators. Some critics caution that embedding ethical metrics into operational monitoring risks diluting accountability by transforming normative judgments into technical parameters (Encord, 2024). However, the alternative—excluding ethical considerations from observability altogether—arguably perpetuates a false dichotomy between technical excellence and social responsibility. A balanced approach recognizes ethical observability as a complement to, rather than a substitute for, human judgment and governance.

Another important dimension of the discussion concerns the temporal dynamics of reliability. Legacy retail systems are often characterized by seasonal demand patterns, promotional cycles, and long planning horizons, which contrast sharply with the continuous deployment ethos of SRE and MLOps (Dasari, 2025). This temporal mismatch can lead to overengineering during peak periods and underinvestment during perceived lulls, exacerbating reliability risks. By explicitly incorporating temporal variability into service-level objectives and model monitoring strategies, organizations can develop more nuanced reliability postures that reflect business realities (Lewis et al., 2022). This insight aligns with broader trends in adaptive systems theory, which emphasize the importance of context-aware control mechanisms in complex environments (Huang et al., 2020).

From a theoretical standpoint, the integrated framework proposed in this study contributes to the ongoing evolution of reliability engineering as a discipline. By foregrounding machine learning observability and ethical governance, it extends reliability discourse beyond its traditional focus on physical and software systems to encompass algorithmic decision-making as a core site of risk and value creation (Payette & Payette, 2023). This extension invites further scholarly inquiry into questions of measurement, accountability, and organizational learning in data-driven systems. For example, future research might explore how error budgets can be operationalized for probabilistic models in ways that balance statistical rigor with managerial interpretability.

Practically, the discussion suggests that retail organizations seeking to enhance resilience should prioritize integrative initiatives over isolated tool adoption. Investments in observability platforms, logging infrastructure, or drift detection algorithms will yield limited returns unless accompanied by cultural change and governance reform (Dasari, 2025). Leaders must articulate reliability as a strategic objective that encompasses customer trust, ethical responsibility, and long-term adaptability, thereby aligning technical practices with organizational purpose.

eISSN: 2156-5198 pISSN: 2156-518X

RESEARCH ARTICLE

While this transformation is undoubtedly challenging, the alternative—continuing to layer advanced analytics onto brittle infrastructure—poses escalating risks in an increasingly competitive and regulated retail landscape (Singla, 2023).

In sum, the discussion reinforces the central argument that Site Reliability Engineering and machine learning observability are mutually reinforcing rather than competing paradigms. When integrated thoughtfully, they offer a powerful lens for reimagining legacy retail systems as resilient, adaptive, and ethically grounded socio-technical ecosystems. The concluding section synthesizes these insights and outlines directions for future research aimed at translating theory into practice (Dasari, 2025).

CONCLUSION

This research has advanced a comprehensive theoretical exploration of how Site Reliability Engineering can be integrated with machine learning observability to enhance the resilience of legacy retail systems. Grounded in an extensive synthesis of scholarly and practitioner-oriented literature, the study has demonstrated that reliability in contemporary retail contexts cannot be reduced to infrastructural uptime or model accuracy alone. Instead, it emerges as a systemic property shaped by the interplay of technical architectures, organizational cultures, data practices, and ethical governance frameworks (Dasari, 2025).

The analysis underscores that legacy retail infrastructure presents unique challenges for reliability engineering due to its historical layering, tight coupling, and institutional inertia. These challenges are compounded by the probabilistic and adaptive nature of machine learning systems, which introduce new forms of uncertainty and risk into operational decision-making (Lewis et al., 2022). By situating machine learning observability within an SRE-informed reliability culture, the study offers a pathway for addressing these challenges in a holistic manner that aligns technical monitoring with strategic objectives and societal expectations.

Several implications follow from this work. For practitioners, the findings highlight the importance of moving beyond fragmented implementations of SRE or MLOps toward integrated socio-technical approaches that emphasize shared ownership, continuous learning, and ethical accountability. For researchers, the study identifies fertile ground for future empirical investigation into the operationalization of reliability metrics for machine learning systems and the organizational conditions that enable effective observability. For policymakers and regulators, the research suggests that reliability and ethics should be treated as interconnected dimensions of digital governance in the retail sector rather than as separate compliance domains (UTS, 2020).

Ultimately, the transformation of legacy retail systems into resilient digital platforms is not solely a matter of adopting new tools or methodologies. It requires a reimagining of reliability as a living practice that evolves alongside technology, markets, and societal values. By integrating Site Reliability Engineering with machine learning observability, retail organizations can better navigate this evolution, balancing innovation with stability and efficiency with responsibility in an increasingly data-driven world (Dasari, 2025).

References

1. Encord. (2024). A guide to machine learning model observability. Encord

Blog. https://encord.com/blog/model-observability-techniques/

2. Dasari, H. (2025). Implementing site reliability engineering (SRE) in legacy retail infrastructure. The American Journal of Engineering and Technology, 7(07), 167–179. https://doi.org/10.37547/tajet/Volume07Issue07-16
3. Lewis, G. A., et al. (2022). Augur: A step towards realistic drift detection in production ML systems. ACM. https://insights.sei.cmu.edu/documents/614/2022_019_001_877199.pdf
4. Google Cloud Platform. (2021). Understanding machine types. https://cloud.google.com/compute/docs/machine-types
5. Maverick, V. (2019). Log management best practices: A comprehensive guide. Loggly Blog. https://www.loggly.com/blog/log-management-best-practices-acomprehensive-guide/
6. Payette, M., & Payette, M. (2023). Machine learning applications for reliability engineering: A review. Sustainability, 15(7). https://www.mdpi.com/2071-1050/15/7/6270
7. UTS Data Science Institute. (2020). Ethics of AI: From principles to practice. https://www.uts.edu.au/globalassets/sites/default/files/2021-02/executive-summary-of-ethics-of-ai-fromprinciples-to-practice.pdf
8. Shopify Engineering. (2019). Observability at Shopify. https://engineering.shopify.com/blogs/engineering/observability-at-shopify
9. Huang, B., et al. (2020). Modern machine learning tools for monitoring and control of industrial processes: A survey. ResearchGate. https://www.researchgate.net/publication/341763531_Modern_Machine_Learning_Tools_for_Monitoring_and_Control_of_Industrial_Processes_A_Survey
10. Singla, A. (2023). Machine learning operations (MLOps): Challenges and strategies. Journal of Knowledge Learning and Science Technology, 2(3). https://www.researchgate.net/publication/377547044_Machine_Learning_Operations_MLOps_Challenges_and_Strategies
11. Evidently AI. (2025). Model monitoring for ML in production: A comprehensive guide. https://www.evidentlyai.com/ml-in-production/model-monitoring
12. Chuong, T. (2016). Evolution of the Netflix data pipeline. Netflix Technology Blog. https://netflixtechblog.com/evolution-of-the-netflix-data-pipeline-da246ca36905.