International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

Beyond Medallion: Next-Generation Lakehouse Architectures for Real-Time AI-Driven Enterprise Decision Systems

Shashikala Valiki

Independent Researcher, India

shashikala.valiki.researcher@gmail.com

Publication History: Received: 18.03.2026; Revised: 10.04.2026; Accepted: 13.04. 2026; Published: 18.04.2026.

ABSTRACT: A comprehensive analysis explores the evolution of lakehouse architectures and their applicability for real-time AI-driven decision systems. State-of-the-art architectures for data ingestion and streaming processing, as well as a hybrid extension of the medallion framework, form the foundation for semantics-aware online feature engineering, high-availability model serving, and comprehensive drift monitoring and detection. Provenance tracking and preservation, consistency model selection and conflict handling, access and sharing control, privacy and compliance requirements, and benchmark construction for end-to-end performance evaluation are discussed. Case studies demonstrate how next-generation enterprise real-time decision systems satisfy at least all of data quality, freshness, and access control.

Real-time AI capabilities are increasingly being adopted in enterprise systems for a variety of purposes, including customer experience enhancement, risk mitigation, fraud detection, and service optimization. AI solutions in these domains dynamically adapt using data generated in real time, relying on various online and near-real-time algorithms that systematically consume and produce data and decisions. Such solutions find their origins not only in classic fraud detection, recommender engines, and online bidding systems, but also in AI domains such as learning-to-rank, multi-armed bandits, reinforcement learning, transfer learning, deep reinforcement learning, online learning, and continual learning. Real-time feature engineering pipelines in these sectors address a wide range of challenges, such as content-based spam detection, sentiment analysis, stock market prediction, and stock price analysis.

KEYWORDS : lakehouse; medallion architecture; data-driven decision making; data pipelines; data management; streaming data; continuous AI; streaming analytics; real-time feature engineering; model serving; model monitoring

I. INTRODUCTION

The Medallion Architecture has emerged as a groundbreaking technology for advanced analytics and Business Intelligence (BI) pipelines by overlaying Data Warehousing paradigms on top of Data Lakes and their inherent Data Engineering capabilities. However, it does not yet cover the full Analytics & AI life-cycle and is not well-adapted for Real-Time AI applications. A more realistic representation is to overlay Data Warehousing and Streaming concepts onto Data Lake and Data Mesh paradigms. This finding leads to the identification of a Two-Stream Process Framework, consisting of Online and Nearline paths, that maps natively onto different Data Representations. This framework provides a broader perspective on the practical requirements, solutions, and components needed to deploy Real-Time AI decision systems at Enterprise scale.

All analytical data remains subject to the Data Medallion but is broadened with additional components for exhaustive Feature Engineering, Model Serving support, and auto-Supervision. Doing so facilitates the integration of online and nearline Real-Time AI streams with those of common analytics and A/B Testing; thereby ensuring full bi-directional support for validated Model Drift Detection, monitoring, and corresponding retraining triggers. Moreover, the specific needs for Provenance and Lineage tracking, Consistency Models, Access Control, and Privacy by Design/WORLD Logic must also be addressed in this context. Their articulation and resolution yield a taxonomy of candidate technologies that is well-populated for Enterprise workloads yet less so for smaller users.

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

1.1. Research design

The paper presents a conceptual design of an enterprise-grade real-time-ready Decision System Lakehouse to fulfil the vision of real-time AI-driven decision making. The Decision System Lakehouse extends the well-known Medallion Architecture with the Hybrid Medallion extension for context-dependent performance, it combines Online and Nearline data representation for sub-second query response, and it incorporates real-time feature engineering pipelines. Aligning the design with the principles outlined in Designing a Real-time AI-driven Enterprise Decision System, the Decision System Lakehouse represents an intermediate artefact: a conceptual design blueprint whose adequacy is assessed through cross-domain evaluation with real-world case studies. The presented design capabilities address key components and considerations for delivering operationalised real-time AI-driven Decision Systems: model serving, monitoring and drift detection; sub-second query response time; access control and data sovereignty; provenance, lineage and reproducibility; consistency models and conflict resolution; privacy-preserving computation and compliance; and an evaluation framework for real-time AI pipeline benchmarks.

The study identifies a sequence of Transform-Democratise-Operationalise steps to achieve the real-time AI-driven Decision System. In the first step, organisations continue to invest in Data Democratization and build their Data Lakehouses as Medallion Architectures. Once these foundations are in place, they can take an incremental migration approach powered by Modern Data Stack tools to build their Decision System Lakehouse and deliver an Operationalised Real-time AI-driven Decision System that supports a business-ready enterprise.

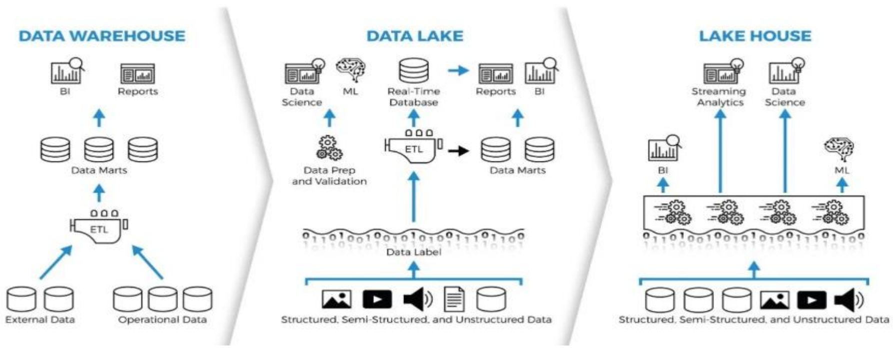


Fig 1: Medallion architecture in a Data Lakehouse

1.2. Background and Significance

Enterprise decision systems increasingly depend on real-time AI pipelines that are engineered to produce prediction features by leveraging internal, external, and social data at the right time with respect to a decision or action. Ideally, such pipelines should undergo persistent online learning (i.e., self-adaptation) and be prepared for rapid retraining. Consequently, the systems begin to resemble closed-loop control systems engineered for stability, safety, and security in the presence of uncertainties. These real-time AI pipelines are different from traditional ML pipelines used in retrospective analysis and are subject to different requirements and constraints.

Real-time AI-driven enterprise decision systems differ from traditional ML-driven systems in important respects. First, instead of predicting long-term behaviour directly, they periodically produce prediction features that are consumed by lightweight predictor models triggered by business processes, such as enhancing the overall customer journey (e.g., through better tailored customer offers) or achieving faster time-to-service (e.g., via intelligent resource allocation). In supervised mode, these prediction features are used to bridge the gap between relevant internal and external data, such as consuming sub-second social media feeds to position offerings more effectively within milliseconds

.

Equation 1. Streaming feature computation

Let the raw event stream be

$$x_1, x_2, x_3, \dots$$

where each event $x_i$arrives at time t_i.

Suppose we want a rolling-window feature over the last Wseconds.

Define the active window at time $t$as

$$\mathcal{W}(t) = \{ i \mid t - W < t_i \le t \}$$

Now let a feature be the average of some event attribute $g(x_i)$over that window.

Step-by-step derivation
Start from the ordinary average:

$$\text{Average} = \frac{\text{sum of values}}{\text{number of values}}$$

Inside the active window, the sum is

$$\sum_{i \in \mathcal{W}(t)} g(x_i)$$

and the number of active events is

$$|\mathcal{W}(t)|$$

Therefore the online rolling feature is

$$\boxed{f_t = \frac{1}{|\mathcal{W}(t)|} \sum_{i \in \mathcal{W}(t)} g(x_i)}$$

II. BACKGROUND AND MOTIVATIONS

The next generation of real-time AI-driven enterprise decision systems demands a new generation of data management implementations beyond the Medallion architecture. Despite its success, the Medallion architecture is still targeted on historical analysis and mirrored across data-mart ecosystems modeled after the previous generation of batch-oriented enterprise data warehouses and data lakes. The Lakehouse paradigm codifies core capabilities of Medallion architectures across these ecosystems while supporting different processing contexts and enabling data-layering approaches that reflect the diverse access patterns of enterprise data consumers. But as the Lakehouse architecture realizes Medallion on a broader scale, a one-size-fits-all approach to layer semantics is less compelling and a more modular approach opens newer avenues. In particular, Upper adjacent layers can introduce new modular services. For enterprise decision systems where decisions need to be made in real-time, a different Data ingestion layer–configuration emerges, enabling access to new types of Data Representations in both online and nearline sections.

Real-time AI pipelines embody the features required of such an architecture and the growing number of Use Cases in this area clearly indicates that Real-time AI is on the Enterprise Horizon. But for real-time decision systems, the classic ML pipeline needs extending beyond model-training and model-serving to incorporate Model-Monitoring, Model-Drift Detection and Real-Time Feature-Engineering. So far, it has only been possible to execute these extended pipelines in disparate ecosystems, necessitating time-consuming and error-prone data-copying and data-transformation tasks. The natural next step is to realize these extended pipelines in the same ecosystem and automate the Data-Availability, Data-Conformance and Data-Transformation requirements of all four tasks at once.

2.1. Evolution of Lakehouse Paradigms

The evolution of the lakehouse paradigm is traced from the pioneering Medallion architecture through the many provided by commercial platforms, such as Databricks' Delta Lake, Confluent's and Snowflake's products and services. Medallion, though conceptually simple, has been shown useful by numerous publications, prototypes and production systems. Its operation in the nearreal-time space and the capability of long-running batch queries are standard practice in most advanced data organizations. Still, there is a clear and pressing demand for the true real-time pipelines and AI/ML-driven decision systems described below. Such systems also require the flexibility of Mixed Signal architectures, the ability to draw data from many sources in diverse representations, and support for processes beyond ingestion and serving.

Combining these characteristics pushes hybrid Medallion into being a family of patterns. The core class defines access-to-query pipelines suitable for AI/ML-driven Decision Systems, connected into a graph by models and their features. Transformations from these and the primary sources occur in real time or near real time, enabling prompt feature monitoring and drift detection, and interactive model evaluation. Beyond models and features, the class also allows custom transformations; supporting, for example, replication, sessionization and data masking, and accommodating anything that online architectures struggle to do. These, in turn, can oftentimes remain the main form of data access,

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

due to their appeal of easy scaling and cost-effectiveness. Support for any sources and representations, the mixing of streaming and batch elements, and the implicit serving of models by their features facilitates natural completion of multiple steps in one pass.

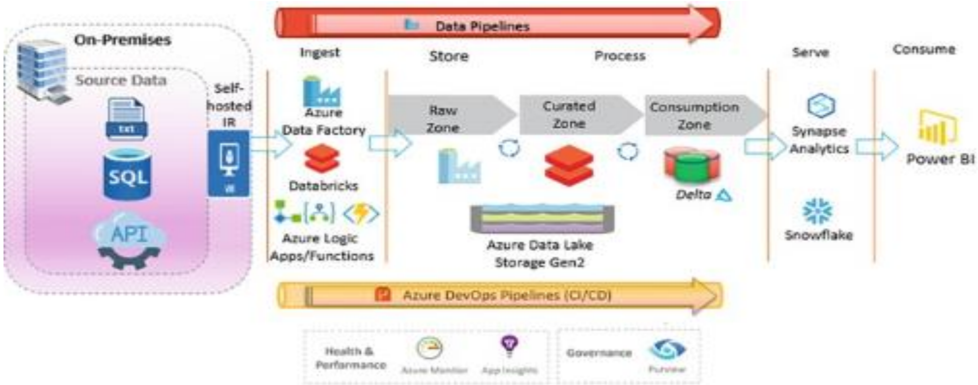


Fig 2: The Data Lakehouse Paradigm

2.2. Real-Time AI-Driven Decision Making: Requirements and Challenges

The convergence of Artificial Intelligence and Big Data might be heralding a new dawn for technology-centered business. Yet many industries are still lagging behind, as emerging digitally native companies embrace these technologies to rapidly evolve and automate their supply chains, enabling real-time decision-making while traditional corporations struggle with traditional technology stacks. Think of the accelerations seen in the Financial Services industry by the likes of Revolut and Monzo, yet a huge focus has also been brought to Food and Retail Services by cloud born companies like Deliveroo, Amazon Fresh etc. An obvious need acknowledged by many C-level executives is the investment in capabilities that enable the enterprise to adopt a real-time decision-making model powered by an AI/ML engine.

The idea of enabling companies to make AI/ML driven decisions based on a streaming data flow has been around for some time. But a few practical implementations around the true ability to transform enterprise processes where a constant input through a stream concept is mandatory have yet to emerge. The adoption of a real-time, AI enabled decision-making model brings fresh requirements and hurdles that need to be addressed.

Equation 2. Hybrid online–nearline feature fusion

Let

- $f_t^{(on)}$= online feature from the live stream
- $f_t^{(nl)}$= nearline feature from micro-batch / scheduled computation

A simple hybrid fusion is a weighted combination.

Step-by-step derivation

Assume the final feature should blend both sources:

$$f_t^{(hyb)} = a\,f_t^{(on)} + b\,f_t^{(nl)}$$

To keep the feature on the same scale, require the weights to sum to 1:

$$a + b = 1$$

Let $a = \alpha$, then $b = 1 - \alpha$, where $0 \le \alpha \le 1$.

Substitute back:

$$\boxed{f_t^{(hyb)} = \alpha f_t^{(on)} + (1-\alpha) f_t^{(nl)}}$$

Interpretation

- If $\alpha \approx 1$, the system favors freshness.
- If $\alpha \approx 0$, it favors stability from nearline aggregates.

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

III. ARCHITECTURAL FOUNDATIONS

Next-generation AI-driven real-time decision-making systems require a rethinking of the requirements and architectural foundations of Lakehouse architectures. Two key attributes are addressed. First, the evolution of data ingestion patterns towards heterogeneous streaming, real-time and micro-batching workloads securing freshness in both input data and the features delivered to ML training exhibits the need for a progressive enhancement of the Medallion layer model; a requirement reinforced by the emergence of **AI feature stores**. Second, the modelling of real-time ML inference pipelines must also be reflected in the Lakehouse architectural foundation. Traditional serving approach—for batch inference and thus continuous scoring in the ML context—are extended with principles more align to real-time streaming pattern.

Traditional Medallion lakehouse architecture focuses on daily or weekly ELT pipelines followed by batch model training. Evolution whithin cloud environments has seen companies like **Databricks** or **Google** augment the Medallion architecture with real-time streaming ingestion patterns or specialized streaming log-systems (Pubs-Sub, Kafka...). Emerging AI applications are however incresingly relying on continuous feature freshiness: both for feeding the model with low-latency feeds and preventing concept drifts during serving. Not only the AI Feature Store emerged but also real-time Feature Engineering Pipelines controlling with finel granuality the temporary representation of the real-time Feature Store. **Continuous Feature Engineering** pipelines aim at implementig those requirements; briding offline and online Environments, Controling Computation, Results Freshness and AI Feature Store synchronization. The study described associated architecture and required changes in the Lakehouse Medallion Model.

3.1. Data Layering and Ingestion

Architecture decisions strongly affect integration of an AI-driven decision system with a lakehouse. The original Medallion Architecture applies a three-layer data representation pattern—Bronze, Silver, and Gold—each supporting different jobs—FastLoad, FastTransform, and FastQuery. Such separation serves well when considering batch workloads and different execution speeds for these jobs. Recent efforts propose to extend the Medallion Architecture with support for real-time features and decision systems. Incorporation of real-time aspects expresses the goal of feeding the decision engines as soon as possible; that is, when low-latency and low-to-medium-throughput ingestion pipelines are in place.

Three different configurations arise: a hybrid Medallion extension, contributing to the decision engines but being different from FastLoad pipelines; an on-disk representation built as a by-product of the decision-engine features; and a nearline representation built to improve the performance of the Silver layer while being fed by FastLoad pipelines. The hybrid extension serves as a non-***medallion*** layer when ingestion of low-to-medium-throughput data becomes necessary to feed the decision engines in a near-real-time fashion.

Equation 3. Real-time model serving function

Let the request at time $t$be represented by feature vector

$$\mathbf{z}_t = [z_{t1}, z_{t2}, \ldots, z_{td}]^\top$$

For a linear decision model, prediction begins with a score

$$s_t = \mathbf{w}^\top \mathbf{z}_t + b$$

where $\mathbf{w}$is the weight vector and $b$is the bias.

Step-by-step derivation

Expand the dot product:

$$\mathbf{w}^\top \mathbf{z}_t = w_1 z_{t1} + w_2 z_{t2} + \cdots + w_d z_{td}$$

So the raw decision score is

$$s_t = \sum_{j=1}^{d} w_j z_{tj} + b$$

For binary real-time decisions, convert score to probability using sigmoid:

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

$$\hat{y}_t = \sigma(s_t) = \frac{1}{1 + e^{-s_t}}$$

Substitute s_t:

$$\boxed{\hat{y}_t = \frac{1}{1 + \exp\left(-\left(\sum_{j=1}^{d} w_j z_{tj} + b\right)\right)}}$$

3.2. Real-Time Processing and Streaming Architectures

Real-time AI-driven enterprise decision systems require continuous data collection from monitored processes, which is complemented by data from social media and other online information sources, processed and cleaned for natural language processing tasks. These data streams support processes such as user profiling and product recommendation systems, which have generally served the retail sector. Organizations are increasingly interested in establishing continuous operational data pipelines and service applications, with data models emerging as a key part of a real-time AI-driven decision system. Nevertheless, current data architectures are not designed to support rapid and evolving changes in data models, require data engineers for adjustment, and often involve undetected anomalies.

Decomposing medallion architectures into online and nearline data representations, a hybrid solution identifies continuous feature engineering pipelines as a third ordering layer. Temporal and modeling aspects provide real-time focus for serving, monitoring, and drift detection. The integration of online and nearline processing allows for increased demand from non-deterministic and user-driven decision systems. However, hybrid solutions remain elusive. In addition, provenance and lineage for real-time AI decision systems has received little attention in the research literature. A comprehensive approximation in the form of a provenance and lineage representation checklist is offered, anchoring provenance and lineage discussions in requirements of the operational phase of enterprise decision systems involving AI modeling.

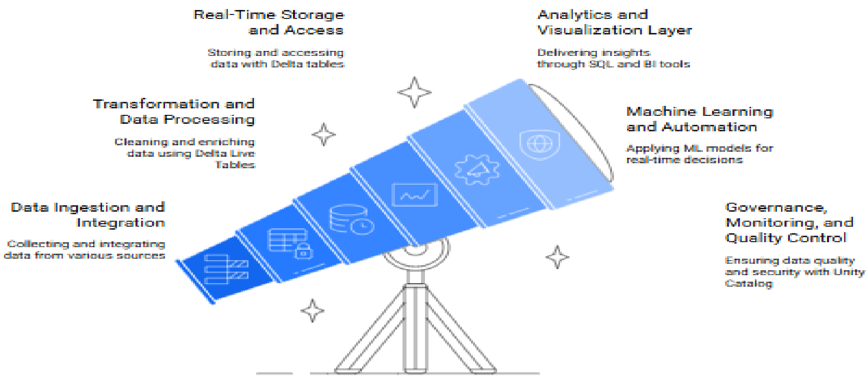


Fig 3: Real-Time Processing and Streaming Architectures

IV. METHODOLOGY

The hybrid medallion extension approach considers the real-time AI pipeline context, defines four transformations for online feature engineering and two for nearline feature readiness. Online feature engineering supports low-latency and low-throughput transformations, using arrivals detected by the streaming ingestion service; nearline computations deal with higher-latency, higher-throughput transformations, scheduling continuous batch mode. Enabling nearline processing within the streaming framework can benefit from getting data prepared at sufficient temporal frequency.

The per-request context behind request-response Model Serving is a simple need to apply the model at request and use the application-level features present in the serving system. Routine model serving pipeline is extracting feature sub-space of interest from the arriving requests, sending it to the model for prediction and making the prediction available at the application layer. Monitoring always and drift-detection whenever required are added to the normal prediction service and are important engine serving components.

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

4.1. Hybrid Medallion Extensions for Real-Time Context

Real-time AI pipelines for enterprise decision-making require context-aware data representations. However, prior explorations into hybrid Medallion implementations focus solely on offline training data. Extension to the online context supports real-time feature engineering pipelines within the Medallion framework, enabling consistent, controlled, and monitored data representation prior to model serving. Concept drift is continuously detected through monitoring of incremental feature distributions. Automated detection triggered by synthetic features serves to cutoff remote locations in a semi-supervised manner. Proposed Medallion extension molds online representation prior to serving as part of an AI pipeline for the financial sector.

The wide adoption of AI models across multiple sectors highlights an increasing reliance on information-driven decision-making. To ensure that decisions are based on accurate, timely, and relevant information, real-time AI pipelines that govern enterprise real-time decision systems have emerged. Such pipelines include modules for real-time feature engineering, model serving, monitoring, and drift detection.

Fig 4: Hybrid Medallion Extensions for Real-Time

4.2. Online and Nearline Data Representations

Model-based solutions enable complex scaling requirements to be met by generating context- and usage-specific derivations of the base tables. Streaming data and near-real-time pipelines complete the structural support for the online operational feature store. A nearline operational feature store is maintained through a three-layer pipeline combining online, batch-mode, and demand basic ETL pipelines. Both models offer feature derivations ahead of business detection latency requirements and support business-driven explorations of data outside the data mart ontology.

The ongoing transition from a lake to a lakehouse architecture for the real-time AI-driven decision system foreseen in a medium- to long-term horizon includes transformations of the data subsystems. Whereas the basic schema and a foundational area have been defined for the lakehouse transition, the components for near-real-time and near-operational decision systems supporting feature engineering and operational feature stores have only begun to emerge. The near-operational pipeline for real-time feature engineering is now explicitly defined, enabling the model-based generation of nearline and online representations of the data and of the basic BI ETL pipelines. As such, it is a basis for addressing the lakehouse architecture of the real-time decisions.

Equation 4. Concept-drift detection

Let

- $P(x)$= baseline distribution
- $Q_t(x)$= live distribution at time t

A natural divergence measure is KL divergence.

Step-by-step derivation

For discrete bins $k = 1, \ldots, K$,

$$P_k = \Pr(X \in \text{bin } k \text{ during baseline})$$

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

$$Q_{t,k} = \Pr(X \in \text{bin } k \text{ during live window at time } t)$$

KL divergence from $Q_t$to $P$is defined as

$$D_{KL}(Q_t \parallel P) = \sum_{k=1}^{K} Q_{t,k} \log \frac{Q_{t,k}}{P_k}$$

So the drift score is

$$\boxed{D_t = \sum_{k=1}^{K} Q_{t,k} \log \frac{Q_{t,k}}{P_k}}$$

Trigger drift alarm when

$$D_t > \tau$$

for threshold τ.

V. OBJECTIVE OF THE STUDY

The discussion focuses on run-time requirements for AI-ready decision systems in real time. A Hybrid Medallion paradigm provides an early view for new data as it arrives, complemented by an online context that supports feature engineering stream processing; finally, a nearline context supports monitoring of deployed models and detection of drift. The work does not address model training, an activity driven by a legacy or nearline context, offering a yet-new view over provenance of training data but solving neither technical nor organisational problems.

These pipelines allow the design of (near) real-time decision systems on data from a Hybrid Medallion lake. Training data can be revalidated or rejected, predictions from data that failed the validation checks can be marked as “caution” status, and data aged beyond the drift-monitoring window can be removed from the models’ monitoring context. The case studies do not address conflicts on training data possible for current Medallion systems nor replaying of predictions due to new truth information; these gaps remain open for (near) real-time AI-driven decision systems and need assessment of the associated impact on the business processes and drivers covered.

5.1. Real-Time Feature Engineering Pipelines

Feature engineering can be classified as continuous, defined as already encapsulated logic, and `on-demand,' defined as yet-to-be-created capabilities. Continuous pipelines often rely on the same or similar code in feature stores and training pipelines. Therefore, the main concern is to ensure that they consume data from consistent snapshots of data sources. To define partitions over those changes, either managed tables updated through SCD at the ingestion layer or metadata tables for nearline views can be used. Also, some processing and analysis of real-time data may be done outside of the real-time AI pipeline context: e.g., Data Quality and Business Value Assessment analyses. The results of these analyses can create use cases for quality databases; e.g., outlier detection for product recommendations.

Equation 5. Provenance / lineage mapping

Let

- r_i= output record
- S_i= set of source records used to produce r_i
- T_i= transformation logic applied
- t_i= timestamp

Then a provenance entry can be written as

$$\Pi(r_i) = (S_i, T_i, t_i)$$

Step-by-step derivation

Suppose an output feature record is produced as

$$r_i = T_i(s_{i1}, s_{i2}, \dots, s_{im})$$

The source set is

$$S_i = \{s_{i1}, s_{i2}, \dots, s_{im}\}$$

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

To make the result reproducible, store:

- the input set S_i,
- the transformation T_i,
- the execution time t_i.

Hence provenance is the tuple

$$\boxed{\Pi(r_i) = (\{s_{i1}, \dots, s_{im}\}, T_i, t_i)}$$

5.2. Model Serving, Monitoring, and Drift Detection

Companies need to regularly deploy, serve, and monitor machine learning models in production environments to support AI-driven decision-making. In most enterprises, however, the corresponding data pipelines still rely on mostly batch processes. Even if the features used by the model are computed in real time, there is still a risk that other features that change dynamically within their validity windows (e.g., one week for user behaviour, one month for economic indicators) are not reflected in ML model predictions since they are not retrained regularly. These models potentially suffer from concept drift: the function that relates X to Y changes over time such that the current map, learned on previous feature values and their corresponding labels, is no longer appropriate. Out-of-Date models adversely impact the most complex and sophisticated recommendations. ML models may therefore require retraining and redeployment daily for each online query or in a periodic batch approach (e.g., once nightly). Specialised serving layers and platforms allow predictions to be coupled with automatic slices, visualisation dashboards, and monitoring of concept drift.

Actual consumption patterns should therefore be exploited for online retraining with basic features and popular models. New data may also be introduced to improve model performance. Building new features in near real time in the same format as the online query predictions allows users to provide business-oriented feedback on model predictions based only on these basic features. Model performance on these key features may permit modelling of consumption patterns by users and may enhance the training data accordingly. In turn, dimensional monitoring for regularly changing business-entity-specific features should be used to indicate the degree of concept drift for the most popular models with complex features that are less frequently updated.

VI. RESEARCH SUMMARY

The objective of the research proposes a set of practical, next-generation principles for building modern real-time, AI-driven enterprise decision systems. The principles derive from key requirements levied by real-time AI decision systems—detecting and responding to changes as they happen, serving ML models in a controlled but open-ended manner, and tracing and lineage tracking. The principles extend the popular Medallion architecture into less mature areas of real-time decision systems by applying concepts and techniques from other research domains. The resulting hybrid Medallion architecture offers useful directions for both real-time decision-making pipelines and the data layer that underpins them.

As organizations increasingly shift their decision systems toward real-time operation, they likewise want their data platforms to operate in real time. In real-time AI systems, data undergoes real-time feature engineering before decision-makers use models to determine actions. How these operations are catalyzed and the underlying architecture is constructed are less explored topics. To help IT architects and data engineers, a collection of principles for engineering real-time AI pipelines is articulated. The principles also address how organizations can tailor online and nearline data representations to meet the needs of low-latency machine-learning inference—without sacrificing the convenience, cost-effectiveness, and maintainability of Medallion-like architectures.

6.1. Provenance, Lineage, and reproducibility

The medallion architecture strongly links every data record to the original data source to allow for reproducibility of results. For every new data record inserted into Silver/Nearline data storage, an entry in a provenance table can be created, with an associated timestamp. Alongside feature engineering pipelines producing data structures required for AI model training, a secondary pipeline could be created to handle candidate lineage data structures for AI model drift detection. Similar to the training metadata, the insertion of records into the drift candidate structure can trigger an entry into a provenance table.

Real-time AI-driven enterprise decision systems operate within a rapidly changing environment, where priorities shift minute by minute. The models underpinning those decisions need to adapt to new situations, just like the humans making decisions based on them. Traditionally, a trigger or condition would be defined to initiate a new model training,

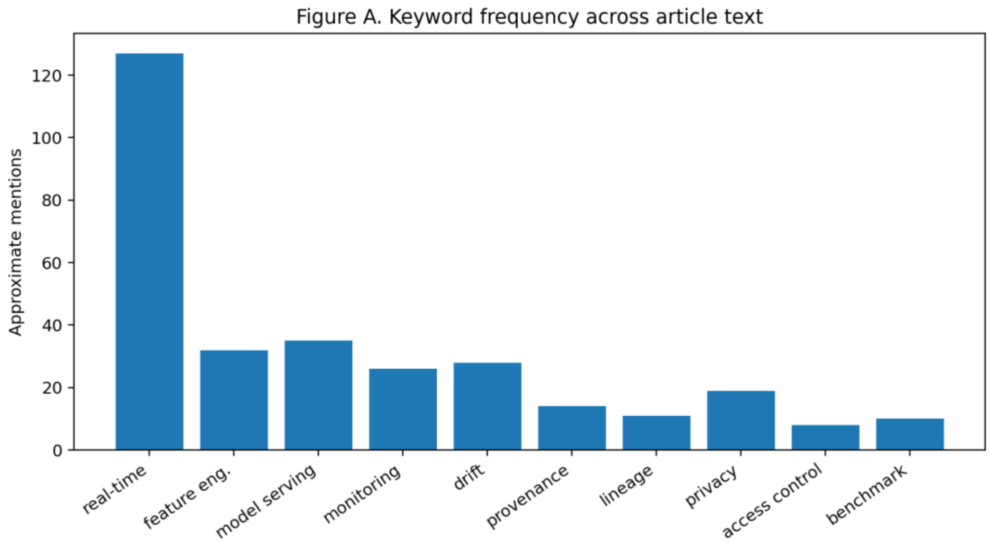
International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

but such triggers are not suitable for a near-real-time environment. Instead, every time new training data becomes available, an entry in the drift candidate structure would suggest that the model's performancebe monitored, and that the model be retrained if it drifts beyond an acceptable level.



6.2. Consistency Models and Conflict Resolution

A major challenge of real-time, multi-user data systems is the trade-off between data freshness and consistency. For systems solely supporting read-write access, two main guarantees are usually supported: linearizability, which ensures that all operations appear to take effect instantaneously at some point between their start and end times; and serializability, which ensures that all operations appear to execute in some sequential order that is consistent with the order of operations by each individual client. Natural choices for a consistency model in a multi-user context include linearizability and session guarantees, for example. Both choices provide an unequivocal interfacing contract, since they guarantee that all clients observe the same sequential order.

Forentina et al. consider theorems to facilitate the implementation of the linearizable data structure for replicated implementations and distinguish between logical and physical operations. Linearizability could lead to performance degradation due to the enforced order of interaction. Non-linearizable models such as eventual consistency, weak consistency, and causal consistency relax the order of interaction between clients. These models allow clients, not through the system, to define the order of their operations.

VII. RESULT

To support enterprise decision processes that function on a real-time basis, the requirements for a next-generation lakehouse architecture go beyond those identified by the Medallion data layer. A fundamental requirement for any data system is to safeguard sensitive information that could lead to violation of regulations such as GDPR and HIPAA, or give competitive advantage to adversaries. The inclusion of provenance or lineage tracking ensures that data and metadata are stored with sufficient detail to allow for reliable reproducibility of subsequent data usages, that is, originality concerns associated with the creative aspect of such usages are adequately addressed. Formal consistency models address scenarios in which multiple users need raw data at the same point in time but in different locations. The {CAP theorem} states that a distributed system cannot guarantee consistency, availability and partition tolerance concurrently; the need for production quality feature engineering pipelines that compute the parameters for ML model input. Data pipelines supporting these use cases also need to be able to perform periodic batch processing in an online or nearline manner.

Feature Engineering for ML models is an online process, it takes reference data from historical data streams and processes it to form the Model Feature Engineering “Feature” data stream at the same rate as the Feature data stream serving the ML Model with input features. The feature pipeline continuously monitors the feature data for drift, stores the statistical parameters of the Feature data stream from the previous execution and compares them with current

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

statistics. Privacy concerns constrain the usage of sensitive raw data and several forms of privacy preservation techniques have been developed to address this. Nevertheless, the underlying algorithm on the aggregated or anonymized data can result in outputs that are still privacy sensitive.

Area	What the paper stresses	Practical control	Expected benefit
Provenance / lineage	Every important transformation should remain traceable to source and timestamp	Provenance tables, lineage metadata, audit logs	Reproducibility and investigation support
Consistency models	Freshness must be balanced against stronger guarantees	Choose linearizability, serializability, session guarantees, or weaker models by workload	Clear client expectations and safer conflict handling
Access control / sovereignty	Restricted data access must be enforced and logged across the AI chain	Federated policy controls and access auditing	Regulatory conformance and tenant isolation
Privacy-preserving computation	Sensitive computations may require masking, perturbation, or secure MPC	Differential privacy, masking, cryptographic protocols, private UDF execution	Protected analytics over sensitive data

Table: Governance, consistency, and compliance matrix

7.1. Access Control and Data Sovereignty

Validating data access rights in an enterprise decision system is a paramount task. Data privacy regulations such as GDPR, HIPAA, and CCPA mandate that organizations safeguard user privacy and enforce compliance at all system levels, from data processing to distribution. Data governance plays a critical role in establishing a federated network of access control and data privacy policies that can ensure even multi-tenant AI-systems do not breach data sovereignty. Provenance and data lineage help track data usage throughout the complete life cycle of the enterprise AI-processing chain. Thus, all access to limited or restricted data must be logged for auditing purposes.

Managing policies for data access and privacy must be integrated along with model deployment for both batch and real-time situations. Without secure multi-stage data processing serving limited datasets, the risks of AI misuse escalates. In real-time scenarios, ensuring privacy in generated models entails more than determining whether the training dataset is limited. For every deployment involved in model drift, maintaining the privacy preserving guarantees in the final model at all times must be verified.

7.2. Privacy-Preserving Computation and Compliance

The fine-grained access control model is extended with a privacy-preserving computing framework based on cryptographic primitives to address operationalization concerns regarding the General Data Protection Regulation (GDPR). Protecting sensitive datasets usually requires access protocols that ensure no sensitive information is leaked and can actively prevent data leakage during query execution. Existing solutions not only require a trusted data controller and third-party cryptographic service providers, but also prevent any privacy-sensitive analysis over encrypted data. To address these gaps, an approach for privacy-preserving secure multi-party computation (MPC) over local computations is proposed, that integrates secure MPC with a formalized protocol for privately executing queries over arbitrary user-defined and privacy-sensitive UDFs. Such UDFs can include data access and publishing operations that apply local perturbation, differentially private mechanisms, cryptography-based masking, or K-anonymization techniques. Secure MPC guarantees that the sensitive assets disclosed by users during query execution remain private and prevents unauthorized exposures of non-sensitive data assets from the locally-private data controller and third-party services.

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

Furthermore, different active-preservation approaches are adopted for processing under GDPR. A line of such active-preservation approaches integrate general data perturbation or differentially-private mechanisms for sensitivity protection, whereas two propose full-functional sensitive data masking. Legal compliance verification is supported by extending procedural horizons. The solution is general, applicable across different settings of sensitive datasets, and can be seamlessly integrated in queries whose execution requires sensitivity protection.

VIII. EVALUATION FRAMEWORKS AND CASE STUDIES

The evolving requirements of real-time AI-driven decision making, summarized by the three layers of the Hybrid Medallion Architecture, necessitate advanced evaluation frameworks for enterprise systems. Such frameworks should provide: (1) benchmarks for assessing the performance of a full real-time AI pipeline in a lakehouse and (2) case studies that validate the proposed architectural dimensions through the design and implementation of real-time decision systems that meet the underlying requirements and challenges.

The need to deliver real-time AI-based recommendations to human decision makers leads to the definition of an end-to-end benchmark for real-time AI pipelines. The benchmark comprises the three infrastructure layers of the Hybrid Medallion Architecture (HMA) and decomposes the AI pipeline into the three major tasks of real-time feature engineering, model serving, and monitoring. The benchmark supports the identification of performance bottlenecks through deployment and load-testing in a lakehouse for multiple concurrent users. A use case driven by bank customer data demonstrates an enterprise-level, real-time decision-making system that enables personalized product offers for new customers. In order to prepare future data scientists for the challenges of production environments, a second use case implements an HMA-compliant real-time AI pipeline as part of the AI4Good course in the Master of Production Engineering.

8.1. Benchmarks for Real-Time AI Pipelines

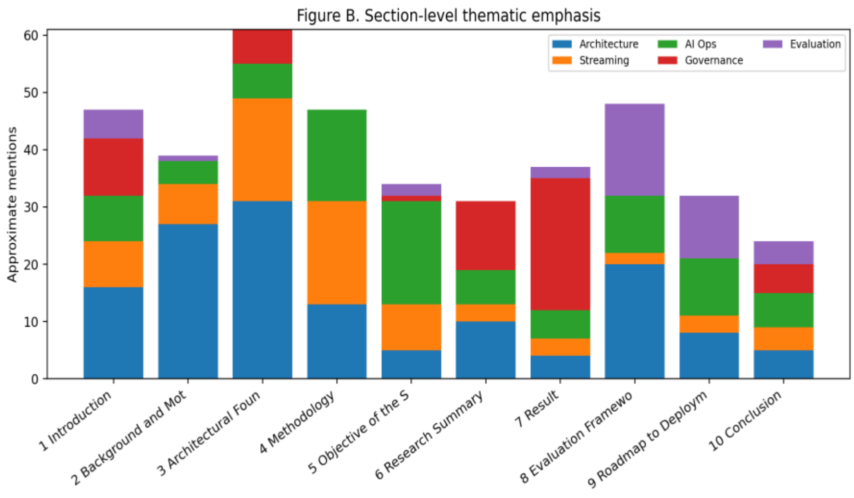
Data industry analysts are not known for benevolence: the Data Management Solutions Lab of the University of Austin has characterized Medallion approaches as "a wall of shame" to avoid and dissuade adoption. Yet a large community not-yet-publicly-adverted adoption within some of the world's largest enterprises suggests that it is not called "Medallion" for nothing. Claims of an end to data warehousing appear premature, and the Medallion discontent is rooted less in Microsoft than in batch-oriented decision pipelines. The underlying concept of "open data lake" and a proper-attribution-led discussion of Azure functionality still have merit. Medallion or similar methodologies are, however, rarely pursued for complex Real-Time AI Decision contexts. Such heterogeneous and divergent requirements for Data Lagoons — and their freshly appearing architectures — suggest that specialized taxonomies, processes and systems are overdue.

Conceiving Reference Benchmarks not to be overseen by the SREs who covet them, this Composite Benchmark seeks to consolidate a middle-road approach. It highlights quality aspects to be measurably tackled in a Real-Time Decision context; promotes a horizontally aligned consolidation in the Lakehouse vat of compound quality criteria, and lift the right Quality characteristics into Benchmarking, Validation, Monitoring and other Quality Preservation layers, hence relieving the Data Pipeline layer. Mindful that the Pipeline itself is less a service than the composition of services for which SLOs apply, it forswears product-page charms in favouring Teeth. The Composite becomes a Taxonomy for A-B-C-D+ Test sequences of the whole Real-Time-AI Decision System — including data generation as many initiatives wish to avoid — one Test to validate increasingly autonomous Test-Driven Monitoring & Drift Detection, and introduces into Discussion the recently over-heating topic of Model Serving. A By-Domed.

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017



8.2. Case Studies: Enterprise Real-Time Decision Systems

Next-generation real-time architectures are evaluated in various settings to demonstrate how their individual characteristics contribute to the elicitation of different dimensions of AI-driven enterprise decision systems. Real-time AI pipelines orchestrate feature engineering and model monitoring with the goal of extracting signals from streaming data and communicating them to downstream applications with minimal latency. The first case study implements such a pipeline with a focus on delivering low-latency feature engineering on a hybrid medallion design. It deploys an analysis-by-synthesis transfer-services architecture on an Apache Spark–AWS infrastructure to extract structured, synoptic information from multimodal streaming data sources.

A supplementary case study on AI-driven retail replenishment applies a non-trivial deployment in a near-real-time context to assess closer-to-production operational and business characteristics. The case study highlights the subtle complexities of the pipeline components and the importance of orchestrating them to expose the necessary business signals to AI-consuming applications—TPE values checked against thresholds and house values checked against deviations from historical norms—in a perceptually plausible and semantically meaningful way. It also elucidates AI-drivers currently in test mode and additional capabilities planned for the solution. Real-time AI decision systems also are applied in a distinct business layer: sensitivity analysis estimating the effects of lead time and holding cost on the order volume of a shipment.

IX. ROADMAP TO DEPLOYMENT

Grounded in academic research with real-world validation, a roadmap detailing migration strategies to next-generation, real-time-supported lake-house architectures is proposed. These architectures are designed to support enterprise AI and ML decision-making and enable businesses to become "real-time AI-driven." The roadmap provides insights into the incremental transformation, indicated lessons and safeguards, and addresses key requirements, implementation considerations, and deployment issues.

The proposed migration strategies are particularly valuable for industrial setups that have become heavily dependent on conventional batch-oriented data pipelines and are now looking for a radically different paradigm that fulfils the requirements needed for real-time AI-driven enterprise decision systems while addressing the shortcomings of Medallion architecture. Moreover, a principled migration strategy that provides a structured framework for gradual transition lowers the risk of switching to the new development. Many complex systems face unique challenges when attempting to apply any next-generation architecture in a "big bang" manner; thus, a small incremental approach for the deployment of real-time systems is a proven way to reduce risk.

9.1. Migration Strategies from Legacy Architectures

Migrating workloads from legacy architectures into real-time decision systems is a complex task, mainly due to the differences in the underlying characteristics of these paradigms. Therefore, it is reasonable to take a phased approach to transformation, allowing for incremental change and risk mitigation. Real-time processing and batch modes often

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|
||Volume 9, Issue 2, March - April 2026||
DOI:10.15662/IJRPETM.2026.0902017

coexist, making it easier for earlier and later stages of enterprise decision logic to stay in separate execution plans while automation is added to detection, retraining, and validation.

Traditional batch-mode processing continues to supply older exhibits and support model retraining, while automated triggers manage and control the entire process, including healthy feature drift detection. Following model retraining, a drift-test phase serves to compare the predictive performance of the new and previous incarnations. Depending on that comparison's outcome, the retrained model is either put into service or held in a test mode until the next retraining cycle. The normal validation of A/B testing also applies when the supported decision logic can incorporate the test, and the potential change is deemed sufficiently large to justify the added uncertainty.

9.2. Incremental Transformation and Risk Mitigation

To mitigate operational risk, even minor and incremental changes to online AI pipelines should be deployed to production via controlled canary releases, with traffic gradually shifting from the old to the new version. In a production system however, the same level of traffic cannot typically be directed at both the old and new implementation for canary releases. Hence, detection of potential regressions of feature engineering pipelines becomes crucial. True drift detection tracks the data characteristics of the online features as they flow through the pipeline to the AI model, independently of the model predictions. When drift is detected by this monitoring aspect of the pipeline, a fast-feedback process should be used to replicate and substantiate the detected changes. If regression testing indeed proves to be necessary, it must happen as soon as possible, without delaying the release and breaking the fast feedback of AI-driven enterprise decision systems. Potential regressions in models serving deployment can be confronted with sound model performance monitoring wherever possible.

X. CONCLUSION

Next-generation AI-Driven Decision Systems can push the lakehouse paradigm to the next level, bringing the time dimension beyond mere job scheduling to online stratification of the policies that support enterprise operations. Current research work has highlighted the requirements and open problems related to a hybrid extension with tighter integration of logic and physical deployment closely tailored to real-time decision making, and a supporting library of AI-ready tools. These encompass focused support for nearline pipelines for feature engineering, online-dedicated data representations, model serving monitoring and drift detection, provenance and lineage tracking, model-based access control, privacy-preserving decomposition, and evaluation frameworks. The proposal also provides a roadmap for migration, allowing incrementally changing the underlying infrastructure toward the new architecture.

The proposed architecture extends the Medallion structure and the r3-framework in the lakehouse system to support real-time reasoning through a combination of multi-representation repository, real-time feature engineering, model deployment and monitoring, enhanced support of explainable AI, and access control based on online data dynamics, thereby widening the spectrum of real-time automated AI-driven enterprise decision systems deployable over real-time AI pipelines.

REFERENCES

[1] Kalisetty, S., & Singireddy, J. (2023). Optimizing Tax Preparation and Filing Services: A Comparative Study of Traditional Methods and AI Augmented Tax Compliance Frameworks. Available at SSRN 5206185
[2] Kumar, B. H., Nuka, S. T., Recharla, M., Chakilam, C., Suura, S. R., & Pandugula, C. (2025, July). Addressing Ethical Challenges in AI-Driven Health Predictions. In 2025 2nd International Conference on Computing and Data Science (ICCDS) (pp. 1-6). IEEE.
[3] Recharla, M., & Nuka, S. T. (2025). Translational Approaches To Commercializing Neurodegenerative Therapies: Bridging Laboratory Research With Clinical Practice. South Eastern European Journal of Public Health, 121–144.
[4] Singireddy, S. (2025, May). AI-Driven Comprehensive Insurance and AAA Membership Benefits Overview. In 2025 2nd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE) (pp. 1-13). IEEE.
[5] Sivanand, R., Kumar, D. P., Nagabhyru, K. C., Natarajan, E. P., Pamisetty, V., & Kapila, D. (2025). IoT and AI for Real-Time Monitoring in Substation Automation. In 2025 International Conference on Computing and Communications (COMPUTINGCON) (pp. 1–5). IEEE. 2025 International Conference on Computing and Communications (COMPUTINGCON). https://doi.org/10.1109/computingcon64838.2025.11377780.

[6] Kummari, D. N., Challa, S. R., Pamisetty, V., Motamary, S., & Meda, R. (2025). Unifying Temporal Reasoning and Agentic Machine Learning: A Framework for Proactive Fault Detection in Dynamic, Data-Intensive Environments. Metallurgical and Materials Engineering, 31(4), 552-568.
[7] Sudha Rani, P. R., Amistapuram, K., Pamisetty, V., Singireddy, S., Kummari, D. N., & Sheelam, G. K. (2025). Hybrid Knowledge Graph–Deep Learning Framework for Automated Exception Handling and Investigation in Complex Insurance Claims. In 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN) (pp. 1–6). IEEE. 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN). https://doi.org/10.1109/gcwcn66157.2025.11448301.
[8] Mangalampalli, B. M., & Kolla, T. (2026). FHIR-Based Interoperability Frameworks For Real-Time Healthcare Data Exchange: Architecture Patterns And Performance Optimization. International Journal Of Advances in Signal and Image Sciences, 1514-1536.
[9] Kummari, D. N., & Burugulla, J. K. R. (2023). Decision Support Systems for Government Auditing: The Role of AI in Ensuring Transparency and Compliance. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 493-532.
[10] Amistapuram, K. (2025). GENERATIVE AI FOR CLAIMS EXCEPTIONS AND INVESTIGATIONS: ENHANCING RESOLUTION EFFICIENCY IN COMPLEX INSURANCE PROCESSES. Available at SSRN 5785482.
[11] Mangalampalli, B. M. Intelligent Data Profiling for Healthcare Data Lakes Using AI-Enhanced Analytics.
[12] Kolla, S. K. (2023). Big Data–Driven Machine Learning Frameworks for Clinical Risk Prediction. International Journal of Medical Toxicology and Legal Medicine, 26(3), 44-59.
[13] Sheelam, G. K., & Koppolu, H. K. R. (2024). From Transistors to Intelligence: Semiconductor Architectures Empowering Agentic AI in 5G and Beyond. Journal of Computational Analy- sis and Applications(JoCAAA), 33(08), 4518-4537.
[14] Aitha, A. R., & Jyothi Babu, D. A. (2025). Agentic AI-Powered Claims Intelligence: A Deep Learning Framework for Automating Workers Compensation Claim Processing Using Generative AI. Available at SSRN 5505223.
[15] Meda, R. (2024). Agentic AI in Multi-Tiered Paint Supply Chains: A Case Study on Efficiency and Responsiveness. Journal of Compu-tational Analysis and Applications (JoCAAA), 33(08), 3994-4015.
[16] Inala, R. (2025). A Unified Framework for Agentic AI and Data Products: Enhancing Cloud, Big Data, and Machine Learning in Supply Chain, Insurance, Retail, and Manufacturing. EKSPLORIUM-BULETIN PUSAT TEKNOLOGI BAHAN GALIAN NUKLIR, 46(1), 1614-1628.
[17] Paleti, S., Kummari, D. N., Garapati, R. S., Sheelam, G. K., Adusupalli, B., & Singireddy, J. (2025, December). Building a Cyber-Resilient Payment Infrastructure: Transforming Payment Security with Zero Trust Architecture. In 2025 3rd International Conference on IoT, Communication and Automation Technology (ICICAT) (pp. 1-7). IEEE.
[18] Pamisetty, A., Paleti, S., Adusupalli, B., Singireddy, J., Inala, R., & Nagabhyru, K. C. (2025, September). Explainable AI Systems for Credit Scoring and Loan Risk Assessment in Digital Banking Platforms. In 2025 IEEE 13th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS) (pp. 1478-1483). IEEE.
[19] Kolla, S. H. (2022). Knowledge Retrieval Systems for Enterprise Service Environments. International Journal of Intelligent Systems and Applications in Engineering, 10, 495-506.
[20] Pote[1], X. R., Pamisetty, A., Karthikeyan, G., & Gupta[1], D. (2025, May). Artificial Intelligence Enabled Smart Energy Conservation Systems for Intelligent Resource Management and Sustainable Future Power Grids. In Proceedings of the International Conference on Sustainability Innovation in Computing and Engineering (ICSICE 24) (p. 196). Springer Nature.
[21] Mangala, N. (2021). Optimizing Large-Scale ETL Pipelines Using Medallion Architecture on Azure Data Lake. Journal of Artificial Intelligence and Big Data, 1(1), 1-20. https://doi.org/10.31586/jaibd.2021.1361.
[22] Bandi, V. D. V. K. (2024). Intelligent Data Platforms For Personalized Retail Analytics At Scale. Metallurgical and Materials Engineering, 30 (4), 1011–1027.
[23] Chakraborty, S., Pamisetty, A., Chandana, N., & CS, B. (2025, October). Depth-Wise Temporal Convolutional Networks with Layer Normalization for Waste Food Prediction. In 2025 2nd International Conference on Software, Systems and Information Technology (SSITCON) (pp. 1-6). IEEE.
[24] Deepika, G., Recharla, M., Deepika, S., P, Ilanchezhian., & G, Nirupashri. (2025). Adaptive Lightweight Autoencoder with Noise Estimation Module for Noise Reduction in ECG Signals. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 1–6). IEEE. 2025 International Conference on Communication, Computer, and Information Technology (IC3IT). https://doi.org/10.1109/ic3it66137.2025.11340876.
[25] Tripathi, A., Kalhapure, B., Bhattacharya, T., Singireddy, J., & Ghosh, R. K. (2026). Behavioural biases in retail investing and legal drivers of tax compliance: A post-pandemic comparative study across OECD Nations. The

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

International Tax Journal, 53(1), 425–435. Retrieved from https://internationaltaxjournal.online/index.php/itj/article/view/533.

[26] Annapareddy, V. N., Singireddy, J., Preethish Nandan, B., Lakarasu, P., & Burugulla, J. K. R. (2025). Emotional intelligence in artificial agents: Leveraging deep multimodal big data for contextual social interaction and adaptive behavioral modelling. Available at SSRN 5241039.

[27] Agrawal, S., Kumar, S. N., Singh, D. K., Sai Niharika, D., Nandan, B. P., & Asati, D. (2025). Dynamic Access Management and Authentication Mechanisms for Enhancing 5G Security Against Heterogeneous Adversaries. In 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG) (pp. 1–6). IEEE. 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG). https://doi.org/10.1109/ictbig68706.2025.11323683.

[28] Yandamuri, U. S. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706.

[29] Krishnan, M., Aitha, A. R., Amistapuram, K., Nandan, B. P., Kaulwar, P. K., & Singireddy, J. (2025). Human-in-the-Loop Hybrid Neuro-Symbolic AI Model for Reliable Data Engineering in High-Stakes Industrial Systems. In 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN) (pp. 1–7). IEEE. 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN). https://doi.org/10.1109/gcwcn66157.2025.11448516.

[30] Bandi, V. D. V. K. (2024). Automated Feature Engineering Systems in Large-Scale Healthcare Data Environments. Journal of Neonatal Surgery, 1.

[31] Challa, K., Singireddy, J., Pamisetty, A., Garapati, R. S., Kannan, S., & Sriram, H. K. (2025, December). Harnessing Agentic AI and IT Infrastructure in Banking to Drive Consumer Insights, Operational Excellence, and Intelligent Financial Innovation. In 2025 3rd International Conference on IoT, Communication and Automation Technology (ICICAT) (pp. 1-7). IEEE.

[32] Naik, A. V., Sheelam, G. K., Panchakatla, N., Muthukumaran, K., & Saranya, K. (2025). Comprehensive Analysis on Depression Detection From Social Media Using Deep Learning and Transformer Architectures. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 1–8). IEEE. 2025 International Conference on Communication, Computer, and Information Technology (IC3IT). https://doi.org/10.1109/ic3it66137.2025.11341160

[33] Kolla, S. H. (2021). Rule-Based Automation for IT Service Management Workflows. Online Journal of Engineering Sciences, 1(1), 1-14.

[34] Mangalampalli, B. M. Generative AI Applications In Healthcare Data Mart Design And Optimization.

[35] Segireddy, A. R. (2022). Terraform and Ansible in Building Resilient Cloud-Native Payment Architectures. International Journal of Intelligent Systems and Applications in Engineering, 10, 444-455.

[36] Mangala, N. (2022). Real-Time Data Quality Monitoring and Gating Frameworks in Cloud-Based Data Pipelines. International Journal of Research and Applied Innovations, 5(6), 8197-8219.

[37] Yandamuri, U. S. (2022). Big Data Pipelines for Cross-Domain Decision Support: A Cloud-Centric Approach. International Journal of Scientific Research and Modern Technology (IJSRMT).

[38] Kolla, T. (2025). The Future of Healthcare Analytics: Leveraging AI and Data Engineering for Personalized Medicine. Journal of Computer Science and Technology Studies, 7(4), 634-640.

[39] Gottimukkala, V. R. R. (2025). Generative AI for Exceptions and Investigations: Streamlining Resolution Across Global Payment Systems. Journal of International Commercial Law and Technology, 6(1), 969-972.

[40] Kumar, K. M., Parasar, A., Walia, A., Inala, R., & Thulasimani, T. (2025, August). Enhancing Risk Management Strategies in Financial Institutions Using CNN and Support Vector Regression. In 2025 5th Asian Conference on Innovation in Technology (ASIANCON) (pp. 1-6). IEEE.

[41] Vajpayee, A., Khan, S., Gottimukkala, V. R. R., Sharma, D., & Seshasai, S. J. (2025). Digital Financial Literacy 4.0: Consumer Readiness for AI-Driven Fintech and Blockchain Ecosystems. International Insurance Law Review, 33(S5), 963-973.

[42] Madhavi, K. R., Rongali, S. K., Polineni, T. N. S., Kummari, D. N., Challa, K., & Challa, S. R. (2026). Explainable AI (XAI)-Driven Predictive Analytics Framework for Ethical and Scalable Automation in Cloud-Native Architectures with Enterprise and Healthcare Interoperability. In 2026 International Conference on Electronics and Renewable Systems (ICEARS) (pp. 31–36). IEEE. 2026 International Conference on Electronics and Renewable Systems (ICEARS). https://doi.org/10.1109/icears67481.2026.11416589

[43] Selvaraj, F. J., Rani, S., Nagubandi, A. R., Chawla, C., & Sekar, G. (2026). Beyond Traditional Ledgers: A Blockchain-Integrated Accounting Model for Seamless Digital Transformation in Retail Economies. Advances in Consumer Research, 3(1), 934-941.

[44] Kummari, D. N. (2022). AI-Driven Audit Frameworks For Enhancing Compliance In Modern Manufacturing Systems. Migration Letters, 19, 2150-2177.

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

[45] Kolla, S. K. (2023). Explainable AI and ML Models for Transparent Clinical Decision Support. Journal for ReAttach Therapy and Developmental Diversities, 6, 2444-2460.
[46] Singreddy, S., Challa, K., Sriram, H. K., & Gadi, A. L. (2025). Leveraging AI, ML, and Gen AI in Automotive and Financial Services: Data-Driven Approaches to Insurance, Payments, Identity Protection, and Sustainable Innovation. Kishore and Sriram, Harish Kumar and Gadi, Anil Lokesh, Leveraging AI, ML, and Gen AI in Automotive and Financial Services: Data-driven Approaches to Insurance, Payments, Identity Protection, and Sustainable Innovation (April 01, 2025).
[47] Valiki, D., & Segireddy, A. R. (2023). Deep Learning Architectures Deployed on Cloud Platforms for Dynamic Financial Risk Evaluation and Market Prediction. American International Journal of Computer Science and Technology, 5(5), 12-24.
[48] Alshar, M. M., Shahdadpuri, N., Rajeshwari, M., Gupta, M., Joshi, N. R., & Singireddy, J. (2025). Enhanced Management & Performance of Remote Workforce with Cloud and AI-Driven HR Analytics. In 2025 3rd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT) (pp. 631–636). IEEE. 2025 3rd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT). https://doi.org/10.1109/icaiccit68829.2025.11434104
[49] Paleti, S., Gadi, A. L., Singreddy, S., & Preethish Nandan, B. (2025). Optimizing Edge Computing for Big Data Processing in Smart Cities.
[50] Seenu, A., Aitha, A. R., Gottimukkala, V. R. R., Singireddy, J., Meda, R., & Garapati, R. S. (2025). Hybrid Multi-Agent Reinforcement Learning and Blockchain Framework for Real-Time Transaction Integrity in Cloud-Driven Financial Systems. In 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN) (pp. 1–6). IEEE. 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN). https://doi.org/10.1109/gcwcn66157.2025.11448456
[51] Sanku, R., Singireddy, S., Nandini, M. R., Dhanamalar, M., & Soni, M. (2025, October). Comprehensive Insights Into Multimodal Emotion Recognition Using Machine Learning and Deep Learning. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 01-08). IEEE.
[52] Davuluri, P. N. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems.
[53] Amistapuram, K. (2024). Federated Learning for Cross-Carrier Insurance Fraud Detection: Secure Multi-Institutional Collaboration. Journal of Computational Analysis and Applications (JoCAAA), 33(08), 6727-6738.
[54] Dwaraka Nath Kummari,. (2022). Machine Learning Approaches to Real-Time Quality Control in Automotive Assembly Lines. Mathematical Statistician and Engineering Applications, 71(4), 16801–16820. Retrieved from https://philstat.org/index.php/MSEA/article/view/2972
[55] Segireddy, A. R. (2025). GENERATIVE AI FOR SECURE RELEASE ENGINEERING IN GLOBAL PAYMENT NETWORK. Lex Localis: Journal of Local Self-Government, 23.
[56] Aitha, A. R. (2024). Generative AI-Powered Fraud Detection in Workers' Compensation: A DevOps-Based Multi-Cloud Architecture Leveraging, Deep Learning, and Explainable AI. Deep Learning, and Explainable AI (July 26, 2024).
[57] Amistapuram, K. (2025). Agentic AI for Next-Generation Insurance Platforms: Autonomous Decision-Making in Claims and Policy Servicing. Journal of Marketing & Social Research, 2, 88-103.
[58] Ashokkumar, S., Amistapuram, K., C, Bharathi., M, Dhanamalar., & J, Gokulraj. (2025). Attention-Guided Spatial Temporal Framework for Deepfake Detection on Social Video Platforms. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 1–6). IEEE. 2025 International Conference on Communication, Computer, and Information Technology (IC3IT). https://doi.org/10.1109/ic3it66137.2025.11341690
[59] Kumar, I., Nagabhyru, K. C., G, Naveen. I., V, Prabhakaran. M., & V, Sruthy. K. (2025). Adaptive Meta-Knowledge Transfer Network with Feature Hallucination and Attention for Low-Shot Object Detection in Aerial Images. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 1–6). IEEE. 2025 International Conference on Communication, Computer, and Information Technology (IC3IT). https://doi.org/10.1109/ic3it66137.2025.11341447
[60] Pallapu, S. R., Aitha, A. R., K, Sudhakar., Vandhana, K., & Chelladurai, S. (2025). GAN-Augmented Transformer Framework for Cross-Domain Video Style Transfer. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 1–6). IEEE. 2025 International Conference on Communication, Computer, and Information Technology (IC3IT). https://doi.org/10.1109/ic3it66137.2025.11341104
[61] Seenu, A., Sheelam, G. K., Motamary, S., Meda, R., Koppolu, H. K. R., & Inala, R. (2025). AI-Driven Innovations in Infrastructure Management with 6G Technology. In 2025 2nd International Conference on Computing and Data Science (ICCDS) (pp. 1–6). IEEE. 2025 2nd International Conference on Computing and Data Science (ICCDS). https://doi.org/10.1109/iccds64403.2025.11209649

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902017

[62] Kummari, D. N. (2025). Advanced Practices in Auditing, Regulatory Compliance, and Smart Manufacturing Systems. Deep Science Publishing.

[63] Davuluri, P. S. L. N. . (2024). AI-Driven Data Governance Frameworks for Automated Regulatory Reporting and Audit Readiness. Metallurgical and Materials Engineering, 30(4), 996–1010. Retrieved from https://metall-mater-eng.com/index.php/home/article/view/1936

[64] Yandamuri, U. S. AI-Driven Decision Support Systems for Operational Optimization in Hospitality Technology.

[65] FinOps Strategies for AI-Enabled Real-Time Compliance Platforms in Cloud Native Environments. (2025). MSW Management Journal, 35(2), 2080-2088.

[66] Pandiri, L. (2025, May). Exploring Cross-Sector Innovation in Intelligent Transport Systems, Digitally Enabled Housing Finance, and Tech-Driven Risk Solutions A Multidisciplinary Approach to Sustainable Infrastructure, Urban Equity, and Financial Resilience. In 2025 2nd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE) (pp. 1-12). IEEE.

[67] Thutari, R. T., Garapati, R. S., B M, Manjula., R K, Supriya., & M, Senbagan. (2025). Adaptive Access Control and Authentication Management for IoT Using Attention-GRU and Reinforcement Learning. In 2025 2nd International Conference on Software, Systems and Information Technology (SSITCON) (pp. 1–6). IEEE. 2025 2nd International Conference on Software, Systems and Information Technology (SSITCON). https://doi.org/10.1109/ssitcon66133.2025.11342003