International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

Orchestrating the Future Autonomous Healthcare Data Pipeline Management through Agentic AI Architectures

Venkata Akhilesh Ranga Reddy

Application Architect, USA

venkataakhileshkumar@gmail.com

ORCID ID: 0009-0008-4140-2299

ABSTRACT: Data is undoubtedly the most relevant asset in the digital transformation of businesses. In healthcare, data pipelines enable the automated orchestration and optimization of data flow among diverse systems and stakeholders. Properly designed, they provide a mechanism to abstract and fulfil the resource-related needs of the processes they support, including ingestion, analysis, sharing, and publishing. Nevertheless, healthcare data pipelines are still implemented manually or with ad-hoc automation, which threatens their reliability, flexibility, and efficiency. Integrating Agentic AI (Artificial Intelligence) appears to be the solution for their automation and optimization. Agentic AI systems reproduce the autonomous, self-organizing, decision-making, and multi-role properties of human archaeogenic agentic behaviours. In the context of healthcare data pipelines, these properties enable automatic data ingestion and sharing whenever and wherever needed;participation as data providers, consumers, or orchestrators; dynamic adaptation to changes in resource-related needs, costs, or availability; and multi-agent collaboration with associated role allocation.

Supporters of this approach claim that Agentic AI will favour businesses in their race towards becoming data-driven by allowing managers to focus on their core business and outsource resource-related problems to self-organizing control loops operating over business processes. Nevertheless, Agentic AI systems have yet to be implemented and tested. This gap deserves attention, especially in a domain as heterogeneous and complex as healthcare, where patients' lives depend on the accurate and reliable availability of data. The following presents an in-depth analysis of healthcare data pipelines and identifies how Agentic AI can support their automation and optimization.

KEYWORDS: Agent-based Systems, Autonomous Agents, Agent-oriented Software Engineering, Healthcare Data Orchestration, Healthcare AI.

I. INTRODUCTION

The journey from data collection to analysis and interpretation in modern healthcare involves a series of interdependent steps, including the monitoring of resources, state maintenance, problem detection, and action triggering. Within an organizational context, these steps are orchestrated by humans and assistive technologies that often lack awareness of their environment and interdependencies. While data management and pipeline technologies are widely deployed to assist with these tasks, unforeseen upstream or downstream events often lead to pipeline failures, unexpected delays, or service-level agreement violations.

The introduction of intelligent agent technologies, capable of continuously monitoring ongoing processes, recognizing changes, re-planning resource allocation in real time, and executing corrective actions autonomously, provides an opportunity to increase process resiliency and reduce human supervision effort. These research efforts focus on emerging agent-based approaches that employ autonomous intelligence at the orchestration level for the explicit purpose of autonomous data pipeline operations. The application of intelligent autonomous agents to orchestrate the flow of heterogeneous data among interacting resources creates agentic data pipelines. The term agentic refers to the ability to adaptively sense environmental changes and, when required, take autonomous corrective actions. Agentic data pipelines are characterized by continuous operation and planning and managerial awareness actions that exploit interaction knowledge for real-time performance improvement and task scheduling.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

The emergence of intelligent agent technologies is transforming the way data-intensive systems are designed and managed. By continuously monitoring processes, detecting environmental and operational changes, dynamically reallocating resources, and autonomously executing corrective actions, these agents significantly enhance process resilience while reducing the need for constant human oversight. At the orchestration level, agent-based approaches introduce autonomous intelligence that enables data pipelines to operate in a self-managing and adaptive manner. When applied to the coordination of heterogeneous data sources, services, and computational resources, this paradigm gives rise to *agentic data pipelines*. The term *agentic* emphasizes the system's capacity to perceive contextual shifts and proactively respond through informed decision-making. Such pipelines operate continuously, combining real-time planning, scheduling, and managerial awareness to optimize performance. By leveraging knowledge of interactions across distributed components, agentic data pipelines enable more robust, efficient, and scalable data operations in dynamic environments.

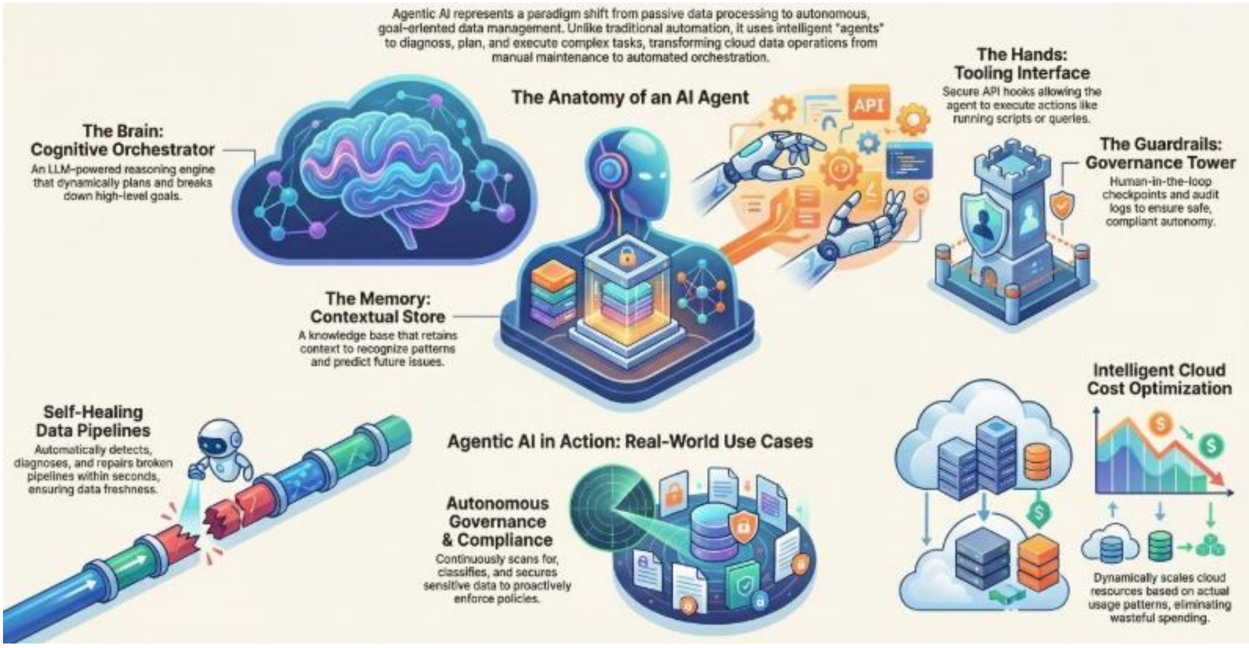


Fig 1: Agentic AI in Cloud Data Management

1.1. Background and Significance

The continuous and equal access to digital devices and health data that characterize the current era has opened new opportunities for individuals and companies operating in the digital health market. Digital health encompasses a variety of tools, products, and services, including mobile, software, hardware, and services related to disease prevention, diagnosis, and pathology treatment, therapy, prognosis, and management. These solutions capitalize on the growing demand for digital Health 2.0 solutions, allowing users more agency, control, and active participation in caring for their health. While the 2.0 phase focused on Web 2.0 technologies and user-community engagement, a more current Health 3.0 phase is maturing, stimulated by the advances of IoT and sensor-based technologies.

At the same time, there is the need for scalable digital transformation across health delivery systems to ensure health equity and resilience. At this level, Smart Health projects support the automated, remote, and continuous monitoring of patients with chronic health conditions and other patients released from hospitals. Patients should be able to take ownership of their own health conditions, receiving timely alerts and education to avert negative health events. Possessing adequate health data increases the chances of having better health outcomes. From a business perspective, enabling better health service delivery results in lower operational costs and increases the potential for new business opportunities. A wide market of third-party developers can create and deploy new health applications, since data are shared through open APIs. However, despite the emergent demand and proposals, there are a lack of fully operational Smart Health services in practice and a lack of commercially-oriented solutions supporting the creation of Smart Health services by third parties.

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

Equation 1: Formalizing an agentic healthcare data pipeline as a scheduling problem
A data pipeline can be modeled as a directed acyclic graph (DAG):

- Tasks: $\mathcal{T} = \{1,2,\ldots,N\}$
- Precedence edges: $(i \rightarrow k) \in \mathcal{E}$meaning task $i$must finish before $k$starts.

Let:

- s_i= start time of task i
- d_i= execution duration of task i

Then "precedence constraints" are:

$$s_k \geq s_i + d_i \forall (i \rightarrow k) \in \mathcal{E}$$

Derivation (step-by-step):

1. If $i \rightarrow k$, task $k$cannot start until $i$completes.
2. Completion time of $i$is $s_i + d_i$.
3. Therefore $s_k \geq s_i + d_i$.

1.2. Research design
Two main lines of inquiry are pursued. First, current pipelines and frameworks are reviewed. A second line of research looks for evidence indicating agentic systems are feasible for delivering autonomy while still respecting prevailing data regulation.

Investigations into the use of agents for orchestrating the components of healthcare data pipelines highlight the design and function of a set of primitives, describe the mechanisms for allocating, controlling, and optimizing groups of agents, and assess the sensitivity of decision support, risk, and scheduling considerations to computational resource availability. The results suggest the orchestration role can be fully automated.

II. FOUNDATIONS OF AGENTIC AI IN HEALTHCARE DATA PIPELINES

The medical data landscape is characterized by diverse sources, types, and structures, complicated further by the full range of well-known interoperability issues pertaining to different clinical, radiology, biomarker, genomic, registry, and patient-generated data. Infrastructures modelled on the data lake, data island, or data hub metaphors promise acquisition from any number of heterogeneous, possibly remote sources, after which the focus shifts to orchestrating processing and storage in an optimized manner for a plethora of currently required, emerging, and future uses.

While such multi-platorm data-integration scenarios exercised the multi-agent systems (MAS) research community a decade ago, the climate crisis and its associated risks with respect to health and well-being now demand a shift from a predominantly supply-side focus on ingestion and accessibility toward a more demand-driven approach. This entails not only a responsive system that acts automagically when an evolving need manifests but also one that allocates different processing and analysis roles to an ensemble of collaborating agents. Such an inversion of the instantiation-deployment paradigm means the question of data availability recedes into the background, with the capacity to pull together whatever data are needed—when and for what—becoming the challenge.

Building on this inversion, a demand-driven paradigm reframes multi-agent systems not as passive orchestrators of pre-integrated data streams but as adaptive sense–reason–act collectives that co-evolve with the problems they are meant to address. Rather than beginning with what data are already available, the system begins with an articulated or inferred decision need, decomposes it into analytical sub-tasks, and dynamically recruits agents with complementary competencies—ranging from data discovery and provenance assessment to modeling, uncertainty quantification, and ethical oversight. In this way, data acquisition becomes a consequence of reasoning rather than its prerequisite. Such architectures privilege timeliness, contextual relevance, and interpretability over sheer volume, enabling MAS to operate as living infrastructures that continuously reconfigure themselves around emerging climate-related risks and societal priorities, ultimately supporting more anticipatory, equitable, and actionable responses.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

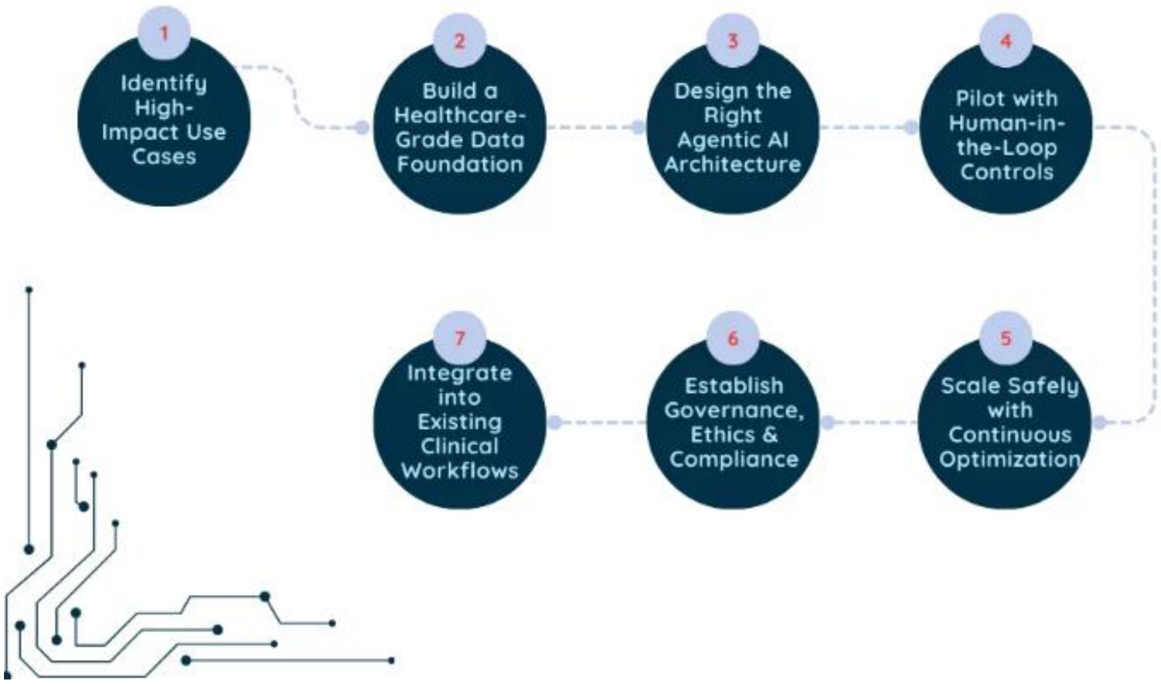


Fig 2: Agentic AI in Healthcare

2.1. Data Ingestion and Interoperability

Agentic systems are designed to orchestrate the healthcare data integration process, enabling automatic monitoring, remediation, and optimization capabilities. However, in real-world scenarios, human intervention is still required due to errors related to data ingestion and source inter-communication, especially when the match between sources is poor. In this sense, the early stage of Data Pipeline formation continues to be manual and complex. New AI technology must enable a more efficient architecture for this process. A remote healthcare patient monitoring system that analyzes sensory data and provides semantic alerts is used as an illustrative case study. The system architecture requires communication between two cloud-edge infrastructures. A data pipeline deployed on a Cloud environment is responsible for ingesting, processing, and storing sensory data generated on the Edge.

Communication is established using Web Services, with sensitive data being stored in a Private Cloud environment. The proposed AI-based technology detects dynamic changes in data sources and pipelines. When a healthcare actor joins the monitoring process, a semantic data pipeline is automatically instantiated. The system performs data supervision and ontology-verification tasks. Agent technology and natural language processing are employed to check actors' qualifications for pipeline participation. When the data exchange between two infrastructures has poor semantics, human intervention is required, and an agent recommends an appropriate algorithm to process the data. The solution has been validated in a real patient monitoring system and has the potential to ease the pipeline establishment of heterogeneous IoT cloud-edge infrastructures.

Equation 2: Agent assignment and "resource-aware scheduling"

Let:

- Agents (workers): $\mathcal{A} = \{1, \dots, M\}$
- Binary assignment variable:
-

$$x_{ij} = \begin{cases} 1 & \text{if task } i \text{ is assigned to agent } j \\ 0 & \text{otherwise} \end{cases}$$

Each task must be assigned to exactly one agent:

$$\sum_{j \in \mathcal{A}} x_{ij} = 1 \forall i \in \mathcal{T}$$

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

Derivation:

1. A task must run somewhere (some agent).
2. “Exactly one agent” means choose one j.
3. Binary selection across all agents sums to 1.

2.2. Semantic Annotation and Provenance

Semantic annotation links content to machine-actionable semantics, crucial for interpretability and retrieval. Resource-aware pipelines timely deploy computing resources across the entire pipeline, ensuring efficiency and efficacy. Knowledge graphs digitize domain expertise to implement proper semantic annotations and enable provenance generation. Data provenance tracks ingested data sources, pipeline processing, and produced output. Data custodians monitor provenance updates to validate semantic annotations.

Agentic systems integrate feedback loops and control mechanisms for constant operational adjustment. Agent action and reaction modules execute pipeline tasks, utilizing the knowledge graph to manage problem-specific requirements encompassed in Quality of Service templates. QoS monitors, modeled as discrete-event systems, guarantee adherence to user-defined pipeline execution constraints, serving as reactive assistant agents

III. ARCHITECTURAL FRAMEWORKS FOR AGENTIC PIPELINES

Architectural frameworks for agentic healthcare data pipelines provide the structures and organization needed to support agentic orchestration. They enable execution across multiple organizations in response to changing circumstances, including the usage of intermittent, non-continuous, or irregularized services. Multi-agent collaboration arises from the roles of the agents involved, covering all aspects of both the data-pipeline operation and external interactions with governance authorities. Institutionally based control loops capture simple agentic weather and scheduling information for use by role-allocation services, while time-driven deployment aligns business and data-generation cycles.

Learning-based detection of external events or conditions enables feedback to influence the orchestration and behavior of waterways and associated gardening resources. The same concepts, when extended across the organization boundary, facilitate the discovery, compliance-checking, and reuse of data across multiple organizations. Role-based operation ensures that the externally visible pipeline remains static, that interactions with partners remain clear and well-defined, and that appropriate decisions on partner selection can be made at schedule- and now-time.

Architectural frameworks for agentic healthcare data pipelines establish a cohesive foundation for orchestrating intelligent, adaptive data flows across complex, multi-organizational ecosystems. By defining clear roles, responsibilities, and interaction patterns, these frameworks enable autonomous and semi-autonomous agents to coordinate data ingestion, transformation, governance, and delivery even when services are intermittent, irregular, or dynamically provisioned. Embedded control loops capture contextual signals—such as operational “weather,” resource availability, and scheduling constraints—to inform role allocation and time-driven deployment strategies that align analytics workloads with clinical, operational, and business cycles. Learning-based mechanisms further enhance resilience by detecting external events or environmental changes and feeding this intelligence back into orchestration logic, allowing pipelines to adjust behavior proactively. Extending these concepts across organizational boundaries supports secure discovery, compliance validation, and reuse of shared data assets, while role-based operation preserves a stable external interface and well-defined partner interactions, ensuring consistent governance, predictable integration, and informed partner selection at both planning and execution time.

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

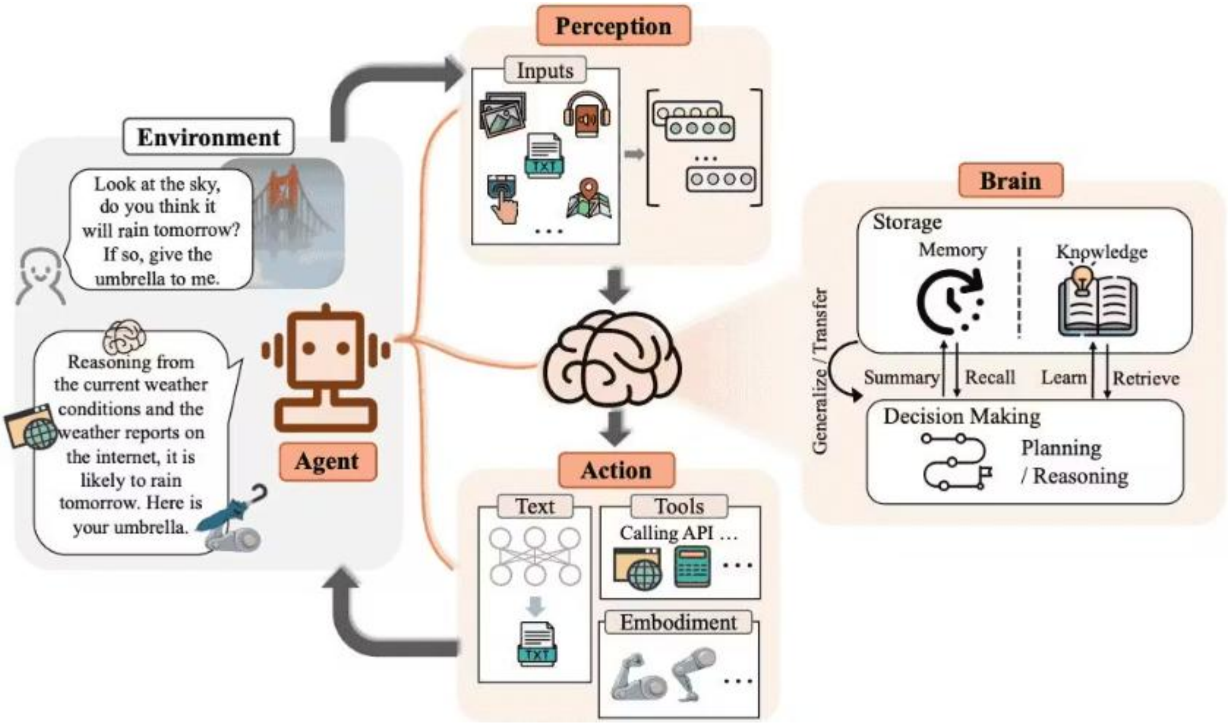


Fig 3: A Guide to Building Agentic AI Workflows

3.1. Multi-Agent Collaboration and Role Allocation

Despite the availability of agent-based programming frameworks, a clear architectural blueprint for agent-based data pipeline control has yet to be established. Feedback kinelligence. Multi-agent frameworks usually support the implementation of different components in distributed settings, allowing for a collaborative approach to data pipeline orchestration. A specific architecture proposed for data pipelines focuses on centralized rollouts for performance or unit commitment optimization but does not explore their control architecture. In the context of traditional control loops—for example, control theory—the distributed optimization problem uses the gradient of a loss function to dynamically incentivize agents toward the central solution without requiring a centralized orchestrator.

The use of multi-agent systems for resource scheduling constitutes a prevailing paradigm in data pipeline control, naturally lending itself to QoS monitoring and control. However, data pipelines differ from traditional grid scheduling in that the data-generating processes and the data haven metadata/annotation providers are independent, and row-by-row construction covers all potential flows. The requirements these data-generating processes impose on datacentric computing environments necessitate a feedback mechanism that teachers monitoring agents how to better serve such requests. A dedicated monitoring agent also enables QoS assurance in terms of read access delay, bandwidth, and response time; the additional role once again activates these features when they are down or underperforming.

Equation 1: Objective functions: QoS-aware, QoI-aware, resource-aware scheduling

- **QoS-aware scheduling** (optimize travel cost + response time)

Agentic AI Systems for Autonomo…

- **QoI-aware scheduling** (minimize integration cost remember minimum QoI)

Agentic AI Systems for Autonomo…

- **Resource-aware scheduling** (respect budgets/constraints allocated by higher-level manager)

Agentic AI Systems for Autonomo…

A unified multi-objective (common in practice) can be written as a weighted sum:

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

$$\min_{x,s} \ \alpha \underbrace{\sum_i \sum_j x_{ij} \ \ell_{ij}}_{\text{data travel cost}} + \beta \underbrace{\sum_i \sum_j x_{ij} \frac{w_i}{v_j}}_{\text{compute/response time}} + \gamma \underbrace{\sum_i C_i^{\text{grant}}}_{\text{privacy/access granting cost}} - \delta \underbrace{\sum_i Q_i}_{\text{data quality (QoI)}}$$

Where:

- $Q_i$is the achieved QoI contribution of task i
- C_i^{grant}reflects the paper's point that privacy-preserving scheduling can include **cost of data-access granting**.

3.2. Control Loops and Feedback Mechanisms

Multi-agent systems enable autonomous execution of healthcare data pipelines in a decentralized manner. Such a system requires a defining architecture and set of control loops that support its various control objectives. Multi-agent systems require roles, which can either be defined statically or determined dynamically at runtime.

Control loops realize feedback mechanisms for the underlying functions of the agentic systems and link the agents to the layers of a healthcare data pipeline. Feedback loops regulate shorter-term decisions such as scheduling, alter resource and node configuration, aggregate multiple scheduling constraints in a resource-aware manner, and delegate these resource allocation and scheduling tasks to the feeder agents whose function is primarily resource management. Longer-term or periodic optimization of larger parts of the pipeline for QoS, reliability, compliance, and candidate provider selection for invocation can also be integrated. Conceptual Cloud technologies enable Cloud environments to be utilized for the hosting, management, and execution of workloads of various nature, and application domains such as SaaS, PaaS, and DaaS have emerged.

IV. DECISION-MAKING AND OPTIMIZATION

Successful operation of autonomous pipelines necessitates a suite of online decision-making and optimization techniques. The scheduling of tasks and control loops adversely affecting resources use and data quality must be managed. Facilities for respecting quality-of-service and compliance constraints must also be provided. If resource-aware scheduling implemented is implemented correctly, it can also indirectly optimize the quality of data produced, even if data quality is not explicitly considered in the pipeline workflows. These methods complement other aspects of agentic pipelines, yet these aspects do not require agentic AI support. However, scheduling solutions depend not only on the agentic nature of pipelines but also on the designated application domain.

The scheduling of multi-task pipelines is an essential consideration for autonomous multi-agent data pipelines operating without human intervention. Such systems typically incorporate a two-level architecture, where an upper-level manager allocates execution budgets to lower-level response agents. Resource-aware scheduling attempts to optimize the monitoring phase of a multi-task monitoring pipeline while satisfying the planned monitoring costs assigned by the higher-level resource manager. The aim of resource-aware scheduling is to maximize the quality of the data being collected while satisfying user-defined quality-of-service constraints and completing the task within the assigned monitoring budgets. Scheduling solutions in agentic pipelines are resource-aware because their decisions respect the resource allocations made at a higher decision level.

Resource-aware scheduling assumes that both task execution and data delivery processes can be managed and planned independently. The master agent monitors the statuses of the data-delivery agents and allocates them to response tasks while considering users' quality-of-service requirements. The allocation of data-delivery agents to monitoring tasks can be seen as a resource-assignment problem. Alternatively, the quality of service of data-delivery agents can be viewed as a type of cost that monitoring tasks need to incur to fulfill their data-delivery requests. Properly managing this quality-of-service assignment aspect can also facilitate the optimization of data quality without requiring an explicit optimization criterion in the pipeline task.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

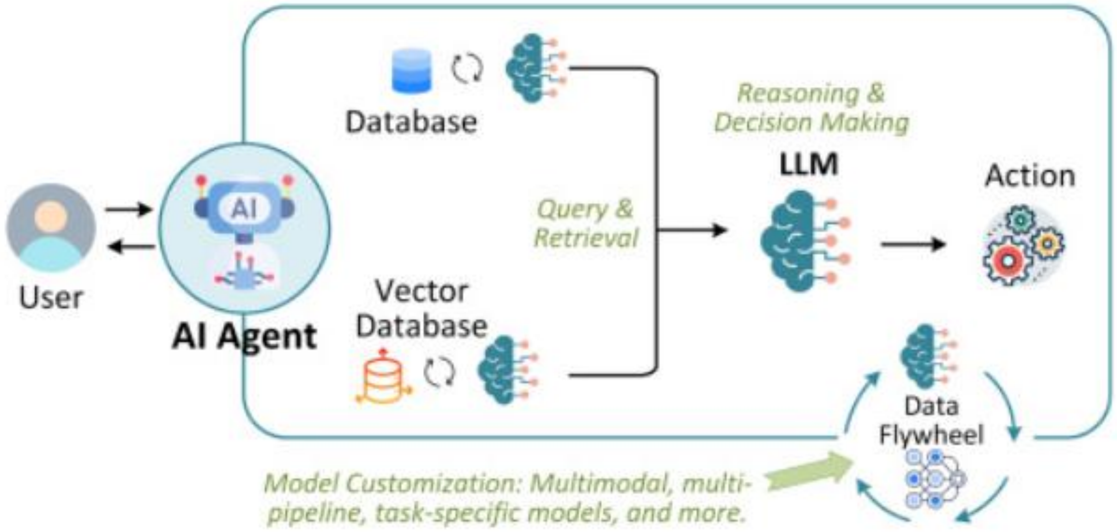


Fig 4: Decision-Making and Optimization of Agentic AI Systems

4.1. Resource-Aware Scheduling

Rescheduling of awoken Data AgEntry Roles must be based on the Quality of Service criteria guaranteed to the respective Data Pipelines. Timeliness requirements can be expressed in the form of a scheduling discipline. The orography of the communication network can be exploited by Lane Agent Interactivation, by means of Path Calculation and Optimization, to minimize delay in time-critical Data Pipelines.

The scheduling policy decides the Timing of next activations of the Data AgEntry Roles with respect to their Quality of Service requirements. Such Timing can be expressed in the form of Periodic Activation (P), Sporadic Activation (S), Aperiodic Activation (A), or Other Forms of Activation (O). The periodic activation discipline enforces certain properties on the interactions within the Data Pipeline, and is therefore the most suitable for Pipelines whose Delay Tolerance Level is Critical.

Equation 4: Control loops and feedback (agentic adaptation)

- State (what the pipeline "looks like now"): s_t
- Action (what agents decide): a_t
- Exogenous events (workload changes, failures): w_t

$$s_{t+1} = f(s_t, a_t, w_t)$$

A typical optimization policy chooses $a_t$to minimize expected future cost:

$$a_t = \pi(s_t) = \arg\min_{a} \ \mathbb{E}[J(s_t, a, w_t)]$$

Derivation:

1. Control loop observes state s_t.
2. Takes corrective action a_t(paper: "autonomously executing corrective actions").
3. Environment evolves into s_{t+1}.
4. Choose actions that minimize a cost capturing QoS/QoI/compliance.

4.2. Quality of Service and Compliance Constraints

Approaches to agent-based orchestration and optimization of healthcare data pipelines often consider Quality-of-Service (QoS) requirements, which address the performance and availability of the operations conducted by the various components, or the Quality-of-Information (QoI) requirements, which specify the properties of the data at a higher level of abstraction than the properties of the services. Agentic AI also enables the modeling of other forms of

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

constraints and criteria, for example security compliance constraints, or high-level policies needed to manage information-sharing among participants with different roles involved in the pipeline execution.

QoS-aware scheduling handles the allocation of computing resources such as Central Processing Unit (CPU) and memory by optimizing the total travel cost for input data and the system response time. QoI-aware scheduling selects a subset of input datasets to minimize the total integration cost while preserving the minimum user-specified QoI-value for the output QoI-type. Resource-aware scheduling focuses on optimizing the execution schedule with respect to userspecified and agent-managed resource constraints and QoS requirements. For a healthcare scenario, the privacy of patients is also critical, due to data-sharing regulations, and a privacy-preserving scheduling technique considers the cost of data-access granting when allocating operations to the agents that execute the operations and manage the exchanged data.

V. DATA GOVERNANCE, PRIVACY, AND SECURITY

The analysis, sharing, and publication of sensitive data demands distinct governance processes, which must be supported by AI-based methods and tools. These processes should cover both the orchestration and optimization of pipelines as well as the optimal trade-off between data accessibility and privacy protection. Privacy-preserving computations—e.g., Secure Multi-Party Computation (SMPC), Homomorphic Encryption, or Differential Privacy—enable the analysis of sensitive data without direct disclosure. Furthermore, agents in charge of decision making should minimize or avoid thus sensitive data travel.

Privacy-preserving techniques do not render distributed data sharing entirely safety-free, though. In fact, they can make sensitive data even more vulnerable, because users may share their sensitive data hinting at, e.g., Encrypted Database as a Service. For this reason, guarantees on the security of the agents and data providers involved in the orchestration and execution of the operation are of the utmost importance. Auditability and explainability methods and techniques help auditors and users understand the how and why of an AI-based decision (e.g., in the sharing or publication of sensitive data). Security considerations aim to convey the “securely” property of the proposed AI agents and solutions, ensuring that no unauthorized party is able to change or to sample the data of sensitive data holders.

Equation 5: Privacy-preserving computation (equation-level foundations)

- True query: $f(D)$
- Sensitivity: $\Delta f = \max_{D,D'} \parallel f(D) - f(D') \parallel$for neighboring datasets

Laplace mechanism:

$$\mathcal{M}(D) = f(D) + \eta, \eta \sim \mathrm{Lap}\left(\frac{\Delta f}{\varepsilon}\right)$$

Step-by-step derivation idea:

1. Privacy goal: hide any single patient’s influence.
2. Worst-case change from adding/removing one record = Δf.
3. Add noise scaled to $\Delta f / \varepsilon$: smaller ε→ more noise → stronger privacy.

5.1. Privacy-Preserving Computation

Privacy-Preserving Computation represents a subdomain of Data Privacy that applies to specific operations in the Healthcare and Life Sciences sectors. Its primary purpose is to protect the confidentiality of a patient’s data while allowing it to be shared with the appropriate participants within the networking environment for analysis in accordance with the role-based access policies assigned to those specific participants. Such processes involve the usage of cryptographic techniques, secure hardware, and randomization, often referred to as third-party computation, or Trusted Third Party (TTP) privacy-preserving solutions. Solutions that do not involve the presence of a TTP rely on cryptographic primitives that allow calculations to occur without direct knowledge of the input data, while maintaining particular assurances in the final output. Common cryptographic primitives supporting these solutions include: Secure Multi-Party Computation (MPC) and Homomorphic Cryptography.

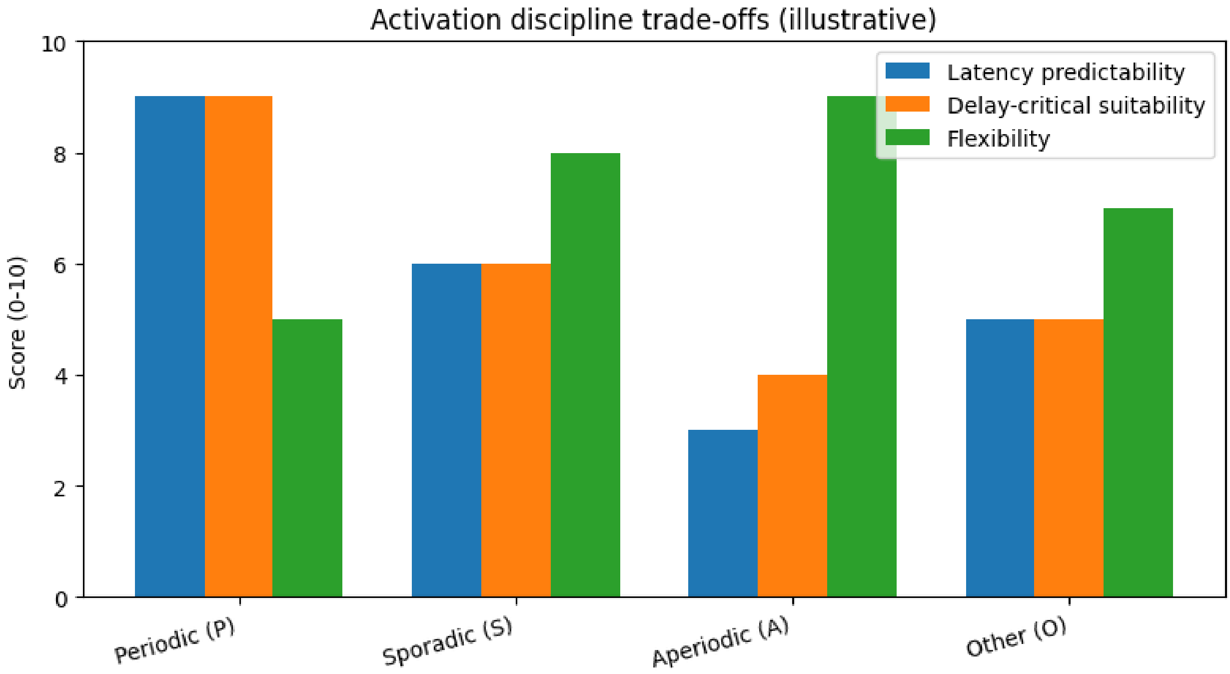
Secure Multiparty Computation, or Secure Multi-Party Computation, is a family of cryptographic techniques that expose a specific publicly verifiable function of all the inputs without revealing the individual inputs to any other entity. Entities share their data with a collection of other parties, where each party has access to a subset of the data and

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

are not trusted with the data they collected from other participants. A trusted third party or an institution is chosen to derive the output from the inputs of the participants and the participants can only verify the result of the calculation. Secure Multi-Party Computation has an inherent solution to the problem of a TTP since the TTP has enough information to reconstruct the original input data. Homomorphic Cryptography permits the execution of a variety of calculations on cryptographic censorship without actually decrypting the ciphertext. A fundamental property of homomorphic encryption is that the level of a ciphertext does not affect the operation to be executed but only the size of the resulting ciphertext.



5.2. Auditability and Explainability

The inner workings of Agentic systems must be transparent to facilitate control by humans or AI-based regulators. For this purpose, an infrastructure of social networks of trust or social ALGs can be established, whose purpose is to provide trustworthy information to users, regulators, or third parties that inquire about the operations of an organization and about the AAIs that have learned from them and are orchestrating their operations, and that need to prove compliance with GDPR rules. Such auditability is warranted through the use of process mining techniques. With respect to explainability, it has been shown that the execution of multi-agent systems can be semantically annotated such that explanations of opaque actions can be produced by revisiting the semantic annotation.

Nonetheless, enhancing transparency through audit reports, exposing the role allocation process, and providing guidance on how to compose more explainable Agentic systems and generate evaluations of their probative power, especially in critical settings, are still open research questions. Techniques belonging to XAI are expected to have a significant role in this respect, as will approaches based on current social standards of intelligence and ethical principles that consider not only the action outcomes, but also their motivations and intentions.

VI. CONCLUSION

The development and deployment of agentic systems supporting the autonomous orchestration and optimization of healthcare data pipelines represent an important step towards overcoming current limitations of the field. Agents have been shown to address key challenges throughout the data lifecycle: facilitating interoperability between disparate data sources, ensuring data quality through suitable validation and cleaning procedures, making automated decisions about storage and processing resource provisioning, and adapting the pipeline's behaviour in response to changing operating conditions through proactive feedback loops. These capabilities empower healthcare data pipelines to operate independently, reducing the burden on human operators or system administrators and enabling continuous analysis of

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

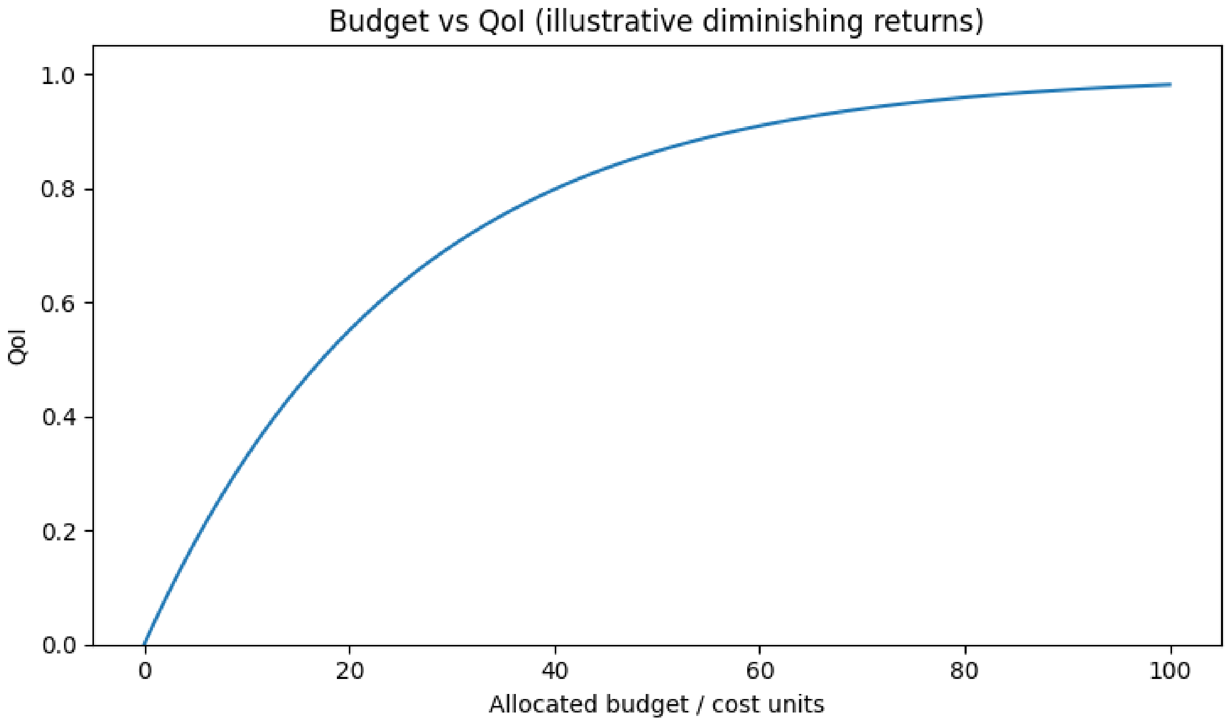
| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

incoming data streams. Nevertheless, further research is needed to explore additional dimensions required by an agentic approach, such as the integration of optimization techniques for resource-aware scheduling, the ability to guarantee Quality of Service and compliance with security and privacy regulations, and mechanisms to ensure system auditability and explainability.

In addition to such conditions for full autonomy, the use of multi-agent systems that collaborate towards a shared goal must also be extended to more advanced roles and patterns of interaction. Emerging trends in data engineering, such as the growing use of Data Mesh and Data Fabric concepts, are opening new avenues for agentic healthcare data pipeline architectures capable of seamlessly responding to the challenges posed by the massive increase in data volume, variety, and velocity and the proliferation of heterogeneous use cases across multiple stakeholders in the healthcare ecosystem.



6.1. Emerging Trends

On the far horizon, quantum computing will shift the physical limits of problem-solving computational systems from the speedup of exponentially large problems by a specific solver to the simultaneous solution of widely different types of problem for data analysis and pipeline orchestration. Massive quantum circuits will evolve, supporting completely new paradigms of quantum-enhanced cost function optimization, helping multi-agent pipelines jointly converge and trial-and-error convergence in quantum-enhanced training of task and domain-reinforced learning agent ai-processes. Beyond operation, a complex pipeline covering locations used by its agents will be feasible with sensor support. Unmanned agent-combot teams, capable of carrying their own networks and microbs, would provide the ultimate practical Maschine-AI candidate: a general purpose persistent agent-combot that could culturally reflect the population served—like war robots on a grand scale. In this bridge region, deception, self-dependence, and planning in agents than in their own staff will be remedies for current human-organizational and institutional toxicity. Quantum-and-sensor enhanced information middle agencies, white and black hat, will quickly emerge in the service of every agent-capitalist. In the service of crime detection, prevention, and prosecution preventive-and-blue hat-mids will counteract, probe, and by quire-the mandates, infrastructures, and behaviours of preventive-and-cybercrime machines.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

Table: Activation disciplines (illustrative)

Discipline	Latency predictability (0-10)	Best for delay-critical pipelines (0-10)	Operational flexibility (0-10)
Periodic (P)	9	9	5
Sporadic (S)	6	6	8
Aperiodic (A)	3	4	9
Other (O)	5	5	7

REFERENCES

[1] Adler-Milstein, J., & Jha, A. K. (2017). HITECH Act drove large gains in hospital electronic health record adoption. Health Affairs, 36(8), 1416–1422.
[2] Gottimukkala, V. R. R. (2020). Energy-Efficient Design Patterns for Large-Scale Banking Applications Deployed on AWS Cloud. power, 9(12).
[3] Ahmad, M. A., Eckert, C., & Teredesai, A. (2018). Interpretable machine learning in healthcare. Proceedings of the ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, 559–560.
[4] Ancker, J. S., Edwards, A., Nosal, S., Hauser, D., Mauer, E., & Kaushal, R. (2017). Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Medical Informatics and Decision Making, 17(1), 36.
[5] Kolla, S. K. (2021). Designing Scalable Healthcare Data Pipelines for Multi-Hospital Networks. World Journal of Clinical Medicine Research, 1(1), 1-14.
[6] Birkhead, G. S., Klompas, M., & Shah, N. R. (2015). Uses of electronic health records for public health surveillance to advance public health. Annual Review of Public Health, 36, 345–359.
[7] Blumenthal, D. (2010). Launching HITECH. New England Journal of Medicine, 362(5), 382–385.
[8] Inala, R. Advancing Group Insurance Solutions Through Ai-Enhanced Technology Architectures And Big Data Insights.
[9] Burwell, S. M. (2015). Setting value-based payment goals—HHS efforts to improve U.S. health care. New England Journal of Medicine, 372(10), 897–899.
[10] Carayon, P., Wooldridge, A., Hose, B. Z., Salwei, M., & Benneyan, J. (2020). Challenges and opportunities for improving patient safety through human factors and systems engineering. Health Affairs, 39(11), 1862–1869.
[11] Mangalampalli, B. M. (2021). Scalable Data Warehouse Architecture for Population Health Management and Predictive Analytics. World Journal of Clinical Medicine Research, 1(1), 1-18.
[12] Classen, D. C., & Bates, D. W. (2011). Finding the meaning in meaningful use. New England Journal of Medicine, 365(9), 855–858.
[13] Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: Management, analysis and future prospects. Journal of Big Data, 6(1), 54.
[14] Peddi, R. K. (2021). Optimizing Case Management Workflows in Global Data Center Colocation Services. Universal Journal of Computer Sciences and Communications, 1(1), 1-21.
[15] Dixon, B. E., Grannis, S. J., & McAndrews, C. (2018). Leveraging data standards to improve interoperability in healthcare. Healthcare Informatics Research, 24(2), 91–101.
[16] Evans, R. S. (2016). Electronic health records: Then, now, and in the future. Yearbook of Medical Informatics, 25(S1), S48–S61.
[17] Reddy, V. A. R. (2021). Challenges in Standardizing Member Eligibility Data Across Multi-Payer Healthcare Ecosystems. International Journal of Medical Toxicology and Legal Medicine, 24(3), 1-19.
[18] Friedman, D. J., Parrish, R. G., & Ross, D. A. (2013). Electronic health records and population health. American Journal of Preventive Medicine, 44(6), 535–536.
[19] Segireddy, A. R. (2020). Cloud Migration Strategies for High-Volume Financial Messaging Systems.
[20] Gliklich, R. E., Leavy, M. B., & Dreyer, N. A. (2020). Registries for evaluating patient outcomes: A user's guide (4th ed.). Agency for Healthcare Research and Quality.
[21] Halamka, J. D., & Tripathi, M. (2017). The HITECH era in retrospect. New England Journal of Medicine, 377(10), 907–909.
[22] Haux, R. (2010). Medical informatics: Past, present, future. International Journal of Medical Informatics, 79(9), 599–610.
[23] Himmelstein, D. U., Wright, A., & Woolhandler, S. (2010). Hospital computing and the costs and quality of care. New England Journal of Medicine, 362(18), 1628–1638.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

[24] Hripcsak, G., & Albers, D. J. (2015). Next-generation phenotyping of electronic health records. Journal of the American Medical Informatics Association, 22(1), 117–121.
[25] Kolla, S. H. (2022). Knowledge Retrieval Systems for Enterprise Service Environments. International Journal of Intelligent Systems and Applications in Engineering, 10, 495-506.
[26] Jha, A. K., DesRoches, C. M., Campbell, E. G., Donelan, K., Rao, S. R., Ferris, T. G., Shields, A., Rosenbaum, S., & Blumenthal, D. (2009). Use of electronic health records in U.S. hospitals. New England Journal of Medicine, 360(16), 1628–1638.
[27] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology, 2(4), 230–243.
[28] Mangala, N. (2021). CI/CD Pipeline Automation for Enterprise Data Artifacts Using Azure DevOps. Universal Journal of Business and Management, 1(1), 1-18.
[29] Keesara, S., Jonas, A., & Schulman, K. (2020). Covid-19 and health care's digital revolution. New England Journal of Medicine, 382(23), e82.
[30] Kern, L. M., Kaushal, R., Barrón, Y., & Abramson, E. L. (2015). Electronic health records and ambulatory quality of care. Journal of General Internal Medicine, 30(4), 496–503.
[31] Inala, R. Designing Scalable Technology Architectures for Customer Data in Group Insurance and Investment Platforms.
[32] Kho, A. N., Cashy, J. P., Jackson, K. L., Pah, A. R., Goel, S., Boehnke, J., Humphries, J. E., Kominers, S. D., Hota, B., Sims, S. A., Malin, B. A., French, D. D., & Meltzer, D. O. (2015). Design and implementation of a privacy-preserving electronic health record linkage tool in Chicago. Journal of the American Medical Informatics Association, 22(5), 1072–1080.
[33] Yandamuri, U. S. (2021). A Comparative Study of Traditional Reporting Systems versus Real-Time Analytics Dashboards in Enterprise Operations. Universal Journal of Business and Management.
[34] Kuperman, G. J. (2011). Health-information exchange: Why are we doing it, and what are we doing? Journal of the American Medical Informatics Association, 18(5), 678–682.
[35] Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S., & Ramoni, R. B. (2016). SMART on FHIR: A standards-based interoperable apps platform for electronic health records. Journal of the American Medical Informatics Association, 23(5), 899–908.
[36] 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.
[37] Mehta, N., Pandit, A., & Shukla, S. (2019). Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. Journal of Biomedical Informatics, 100, 103311.
[38] Menachemi, N., & Collum, T. H. (2011). Benefits and drawbacks of electronic health record systems. Risk Management and Healthcare Policy, 4, 47–55.
[39] Murphy, S. N., Weber, G., Mendis, M., Gainer, V., Chueh, H. C., Churchill, S., & Kohane, I. (2010). Serving the enterprise and beyond with i2b2. Journal of the American Medical Informatics Association, 17(2), 124–130.
[40] Kolla, S. H. INTELLIGENT SYSTEMS! ND! PPLIC! TIONS IN ENGINEERING.
[41] National Academy of Medicine. (2015). Improving diagnosis in health care. National Academies Press.
[42] O'Malley, A. S., Grossman, J. M., Cohen, G. R., Kemper, N. M., & Pham, H. H. (2010). Are electronic medical records helpful for care coordination? Journal of General Internal Medicine, 25(3), 177–185.
[43] Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216–1219.
[44] Rajesh Mattaparthi (2021). Unified Data Lineage and Quality Governance Framework for Multi-Source Sensor Streams in Heavy-Duty Powertrain Manufacturing. Online Journal of Mechanical Engineering, 1(1), 1-15. https://doi.org/10.31586/ojme.2021.1365
[45] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. Health Information Science and Systems, 2(1), 3.
[46] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.
[47] Loganathan, R. (2021). Integrated Risk and Compliance Frameworks for Global Data Center Operations: A Governance-Centric Approach. Universal Journal of Computer Sciences and Communications, 1(1), 1-26.
[48] Rumsfeld, J. S., Joynt, K. E., & Maddox, T. M. (2016). Big data analytics to improve cardiovascular care: Promise and challenges. Nature Reviews Cardiology, 13(6), 350–359.
[49] Shah, S. J., Katz, D. H., Selvaraj, S., Burke, M. A., Yancy, C. W., Gheorghiade, M., Bonow, R. O., Huang, C. C., & Deo, R. C. (2015). Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation, 131(3), 269–279.

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 6, Issue 2, March-April 2023||

DOI:10.15662/IJARCST.2023.0602008

[50] Reddy, V. A. R. (2022). Designing Fault-Tolerant Data Ingestion Pipelines for High-Volume Healthcare Transactions. Frontiers in Health Informatics, 11, 861-889.

[51] Steyerberg, E. W. (2019). Clinical prediction models (2nd ed.). Springer.

[52] Kolla, S. H. (2023). Large Language Model-Driven Enterprise Service Intelligence for Digital Workflow Transformation. *International Journal of Research and Applied Innovations*, *6*(1), 8380-8391.

[53] Vest, J. R., Campion, T. R., Kaushal, R., & HITEC Investigators. (2013). Challenges, alternatives, and paths to sustainability for health information exchange efforts. Journal of Medical Systems, 37(6), 9987.

[54] Mangalampalli, B. M. (2022). Automated Invoice Validation Systems Using Advanced SQL Analytics in Healthcare Insurance. Front Health Inform, 11.

[55] Walker, J., Pan, E., Johnston, D., Adler-Milstein, J., Bates, D. W., & Middleton, B. (2005). The value of health care information exchange and interoperability. Health Affairs, 24(Suppl. 1), W5-10–W5-18.

[56] Wang, Y., Kung, L., Wang, W. Y. C., & Cegielski, C. G. (2018). An integrated big data analytics-enabled transformation model: Application to healthcare. Information & Management, 55(1), 64–79.

[57] Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment. Journal of the American Medical Informatics Association, 20(1), 144–151.

[58] Zhou, L., Soran, C. S., Jenter, C. A., Volk, L. A., Orav, E. J., Bates, D. W., & Simon, S. R. (2007). The relationship between electronic health record use and quality of care. Journal of General Internal Medicine, 22(5), 693–697.

[59] Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. A. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data. Circulation: Cardiovascular Quality and Outcomes, 10(11), e003477.

[60] Davuluri, P. N. (2022). Cloud-Native Data Platform Modernization for Regulatory Compliance in Global Banking.

[61] Friedman, C. P., Wong, A. K., & Blumenthal, D. (2010). Achieving a nationwide learning health system. Science Translational Medicine, 2(57), 57cm29.

[62] Kolla, S. H., & Loganathan, R. (2023). Cloud-Native Deep Learning Architectures For Secure Generative AI Deployment In Enterprise Workflow Platforms. *Journal of International Crisis and Risk Communication Research*, 603-618.

[63] Adler-Milstein, J., Everson, J., & Lee, S. Y. D. (2015). Sequencing of EHR adoption among U.S. hospitals. American Journal of Managed Care, 21(11), e607–e614.

[64] Friedman, C. P., Allee, N. J., Delaney, B. C., Flynn, A. J., Silverstein, J. C., Sullivan, K., & Brantley, K. L. (2020). The science of learning health systems: Foundations for a new journal. Learning Health Systems, 4(1), e10203.

[65] Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Affairs, 33(7), 1123–1131.

[66] Hersh, W. R. (2020). Health informatics: Practical guide for healthcare and information technology professionals (7th ed.). Informatics Education.

[67] Cohen, I. G., Amarasingham, R., Shah, A., Xie, B., & Lo, B. (2014). The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Affairs, 33(7), 1139–1147.

[68] Ben-Assuli, O. (2015). Electronic health records, adoption, quality of care, legal and privacy issues. Health Policy, 119(3), 287–294.

[69] Adler-Milstein, J., Holmgren, A. J., Kralovec, P., Worzala, C., Searcy, T., & Patel, V. (2017). Electronic health record adoption in U.S. hospitals. Journal of the American Medical Informatics Association, 24(6), 1142–1148.

[70] Hoerbst, A., & Ammenwerth, E. (2010). Electronic health records: A systematic review on quality requirements. Methods of Information in Medicine, 49(4), 320–336.