

ARCHITECTURE FOR HEALTHCARE ANALYTICS
EXTENDING THE FULL-STACK PRODUCTION-PLATFORM REFERENCE
ARCHITECTURE TO CLINICAL AND OPERATIONAL DECISION-TO-ACTION
WORKFLOWS

SONIT MITRA SINGH*

*University of New South Wales

ABSTRACT

Healthcare organizations increasingly convert analytical outputs into consequential clinical and operational actions: firing deterioration and sepsis alerts, routing triage, prioritizing imaging worklists, allocating beds and staff, and gating prior authorization, often with no clinician in the immediate path. In that setting the discriminatory power of a model, whether measured by AUROC or AUPRC, is no longer a sufficient basis for trust. Failures originate upstream and downstream of the model: in electronic-health-record inputs whose provenance or temporal validity cannot be established, in probabilities that are uncalibrated across sites and subpopulations, in alerts whose rationale cannot be reproduced for the clinician who received them, in policy boundaries drawn in the wrong place, in automated actions that cannot be safely held or reversed, and in feedback loops such as alert fatigue through which the system quietly reshapes the behavior it measures. This paper adopts as its runtime-governance foundation the Full-Stack Production-Platform Reference Architecture introduced by Mesbaul Haque Sazu, in which every layer is contract-bearing and the runtime is wrapped by governance and observability spines [19], and grounds it in the account of big data analytics and artificial intelligence in healthcare management given by Sazu and Jahan [20]. Building directly on both, the paper develops a single decision-to-action architecture for healthcare analytics that keeps a decision-assurance plane and an execution-assurance plane distinct while binding them through a shared evidence contract. It states pre-commit admissibility for a clinical or operational action as a conjunction of individually testable conditions: input provenance and temporal validity, independent validation, policy and residency permissibility, evidence completeness, calibrated confidence gating, reversibility, and remit-based clinician authority. It maps the adopted runtime invariants onto one governed care lifecycle, defines a cumulative conformance model, and specifies a four-level evaluation protocol together with the measurement hazards particular to care: dataset shift across sites and time, label latency with treatment-induced censoring, and alert-fatigue feedback. A synthetic sensitivity analysis of confidence-gated clinician review illustrates the escalation-versus-harm trade-off and the cost surface behind threshold selection; it is illustrative and carries no empirical weight. The contribution offered here is the governance interface between Sazu's execution-assurance account and the healthcare-analytics context, not a replacement for either.

Keywords:

healthcare analytics; clinical decision support; Full-Stack Production-Platform Architecture; big data analytics; model risk management; software as a medical device; calibration; alert fatigue; dataset shift; explainability; auditability; AI governance.



Copyright: © The Author(s), 2026. Published by **Droub Al Salamah University Scientific journal** This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

1.1 From clinical prediction to consequential action

For much of its history, healthcare analytics was evaluated as a prediction problem. A model produced a risk score, a clinician read it, and a person decided what to do. That arrangement is disappearing. Healthcare is now ingesting data at an unprecedented rate, from physicians and hospitals, medical research, insurance, devices, and the life sciences, and the maturation of analytics has opened the door to using that data to transform how care is managed [20]. Contemporary systems fire sepsis and deterioration alerts, route triage, prioritize radiology worklists, allocate beds and staff, and gate prior authorization, increasingly without a clinician in the immediate path. Once an analytic output can change what happens to a patient on its own, the object that has to be governed is no longer the model. It is the entire route from source data to a clinical or operational action, and much of that route does not exist at the moment a model is validated and deployed.

This shift also changes what a governance failure looks like in practice. The incidents that matter are rarely simple mispredictions. A widely deployed proprietary sepsis model, once externally validated, discriminated far less well in the field than its marketing implied and generated a heavy alert burden [24]; a widely used population-health algorithm systematically under-served Black patients because it optimized a proxy that encoded unequal access to care [14]; and clinical models routinely degrade when the data distribution shifts across sites, seasons, and practice changes [5]. Regulators have responded in kind: model-governance expectations for Software as a Medical Device and the Good Machine Learning Practice guiding principles frame trustworthiness as a property of the deployed sociotechnical system rather than of the algorithm in isolation [6], [7], as do the NIST AI Risk Management Framework [13] and the EU AI Act, which classifies many medical uses as high-risk [4].

1.2 The problem this paper addresses

The literature already offers strong instruments for the individual concerns along this path. Machine learning in medicine is a mature field [16], explainability methods are well developed [17], [11], [3] and their limits for high-stakes clinical use are understood [18], distribution-free uncertainty quantification has a rigorous theory [2], and the tendency of modern models to be over-confident, together with its measurement, is well studied [10]. Conventions for communicating model and dataset properties to clinical users are emerging, from model cards and datasheets to clinician-facing Model Facts labels [12], [9], [22]. What has been missing is not another instrument. It is a contract that carries assurance across the boundary between the system that produces a clinical decision and the system that acts on it, fires the alert, places the order, or moves the patient, and later observes the outcome. A model can be validated without preserving the evidence behind any particular alert; an alert can be explainable in principle without any guarantee that the same explanation was bound to the notification the clinician actually received; a runtime can be exhaustively observable and still prove nothing about whether the action it executed was permitted under residency and clinical-governance policy.

Three questions follow. First, how can design-time decision assurance for a clinical model and runtime execution assurance for the resulting action be composed without collapsing their distinct responsibilities? Second, what evidence contract preserves reconstructability across the score-to-action boundary, including the rationale a clinician is owed? Third, what operational measures can be attached to the resulting structure so that conformance is tested rather than asserted?

1.3 Adopting Sazu's Full-Stack architecture as the foundation

This paper does not propose a new governance vocabulary for healthcare, and it is worth being clear about why. A precise account of the runtime that must execute, observe, and learn from a consequential action already exists in the Full-Stack Production-Platform Reference Architecture, in which every layer is contract-bearing and the runtime is wrapped by a governance spine and an analytics-and-observability spine [19]. Its runtime invariants, capture-at-commit, before-commit gating, compensability, provenance and temporal validity, residency as a hard constraint, logged propensity, and auditability, bind at the exact instant at which a system changes external state, which is the instant at which a governance failure becomes hard to reverse. The account of big data analytics and artificial intelligence in healthcare management given by Sazu and Jahan supplies the domain grounding: what these systems are used for, their impacts, and the current state of their adoption in care delivery [20]. Together they form the reference model on which everything that follows is built.

Adoption here is concrete rather than nominal, and it means four specific things.

- **The execution-assurance plane is the Full-Stack architecture.** Its contract-bearing layers, seven runtime invariants, conformance profiles, and operational-readiness gates are carried over intact [19].
- **The domain grounding is Sazu and Jahan’s healthcare account.** Their treatment of healthcare data abundance, analytics impact, and the current state of AI in care management supplies the setting the invariants are instantiated against [20].
- **The terminology is Sazu’s terminology.** Capture-at-commit, before-commit gating, compensability, provenance and temporal validity, logged propensity, and auditability keep the meanings assigned to them in the source framework [19].
- **Any extension is marked as an extension.** Where this paper adds to the foundation, chiefly the score-to-action evidence contract and the care-specific measurement hazards of Section 7, it says so, so that adopted and added material can be assessed separately.

1.4 Contributions

Against that foundation, the paper makes the following contributions.

- **C1.** It integrates the Full-Stack architecture and the healthcare-analytics grounding into one decision-to-action lifecycle for clinical and operational actions, keeping the decision plane and the execution plane distinct (Section 4).
- **C2.** It specifies a shared evidence contract binding a generation-time DecisionRecord to a runtime ExecutionRecord through a common identifier and immutable version references, and requires the clinician-facing rationale to be reproducible from the persisted record (Sections 3.3 and 4.3).
- **C3.** It states a pre-commit admissibility predicate for a clinical or operational action as a conjunction of named, individually testable conditions, each with an evidence artifact and a defined failure response (Section 3.2, Table 1).
- **C4.** It maps the adopted runtime invariants onto the governed care lifecycle and shows where each is enforced (Section 6, Appendix A).
- **C5.** It defines a four-level evaluation protocol and the measurement hazards particular to care, dataset shift, label latency with treatment-induced censoring, and alert-fatigue feedback, so that the adopted foundation can be tested in the field rather than assumed (Sections 5.3 and 7).

1.5 Scope

This is an architecture paper, and no clinical deployment is reported. The sensitivity analysis of Section 5.3 is a parameterized Monte Carlo illustration rather than a measurement, and its parameters are stated in full so that it cannot be mistaken for evidence. Nothing here should be read as demonstrating that the adopted framework improves patient outcomes, alert precision, or equity; that demonstration is the empirical program that Sections 7 and 9 are written to make obtainable.

2. Foundations and Related Work

2.1 The Full-Stack Production-Platform architecture

The Full-Stack architecture specifies a complete production stack in which every layer is contract-bearing and the runtime is wrapped by a governance spine and an analytics-and-observability spine [19]. Seven runtime invariants anchor it: capture-at-commit, before-commit gating, compensability, provenance and temporal validity, residency as a hard constraint, logged propensity, and auditability. Around those invariants the framework adds trajectory-level risk for multi-step automated execution, drift-under-action, escalation under an explicit reviewer-capacity model, closed-loop learning, operational-readiness gates, and cumulative Observable, Governed, and Adaptive conformance profiles. Two properties make this the natural execution-plane foundation for care. Its invariants bind at commit time rather than at design time, so they constrain the exact instant at which the system changes a patient’s state; and it treats logged propensity and compensability as standing obligations rather than optional features, which is what later makes off-policy evaluation of alerting policies and safe rollback of automated actions feasible at all.

2.2 Big data analytics and AI in healthcare management

Where the Full-Stack architecture is domain-general, the account of big data analytics and artificial intelligence in healthcare management given by Sazu and Jahan supplies the clinical and operational grounding [20]. Healthcare is a complex domain ingesting data at high velocity from physicians and hospitals, research, insurance, devices, and the life sciences, and the improvement of analytic technology opens substantial possibilities for transforming how

care is managed. That grounding matters here for a specific reason: the impacts and the current state of adoption it describes are exactly the settings in which the runtime invariants must be instantiated, and the data heterogeneity it emphasizes is what makes provenance and temporal validity load-bearing rather than incidental. This paper adopts both accounts and specifies the interface between them.

2.3 Decision quality versus model performance in care

A persistent difficulty in governing clinical analytics is that the metrics used to promote a model are not the metrics that describe the action the model participates in. Discrimination is a property of a scoring function over a population. Decision quality is a property of a bounded choice, alert or suppress, escalate or hold, admit or route, taken under a stated policy, at a point in time, with a defined action set and an asymmetric cost of error, where the harm of a missed deterioration differs sharply from the harm of a false alarm that erodes clinician trust. The two come apart in both directions. A better-discriminating model can produce worse care when the alert threshold is set against the wrong harm model, when the population has shifted since validation [5], or when the score encodes a biased proxy [14]. A system can improve care without moving any discrimination metric, by suppressing low-value alerts, by refusing to fire on inputs that have passed their temporal-validity window, or by making the rationale reconstructable for the clinician after the fact. This paper takes decision quality, not model performance, as the assurance target.

2.4 Supporting methods

The architecture composes established methods rather than replacing them. Post-hoc explanation is supplied by local surrogate and additive-attribution methods [17], [11], surveyed with their taxonomy and failure modes in [3]; on regulated clinical paths the recorded rationale must be reproducible from the persisted record rather than regenerated by resampling, which aligns with arguments for inherently interpretable models in high-stakes settings [18]. Predictive uncertainty is supplied by distribution-free conformal methods [2] and by post-hoc recalibration, motivated by the finding that modern models are frequently over-confident [10]. Communicating model and dataset properties to clinical users follows established reporting conventions, including clinician-facing Model Facts labels [12], [9], [22]. The operational hazards specific to machine learning in production, technical debt, feedback loops, and distributional shift, are characterized across [21], [15], [8], and [1], and are exactly the failure classes the execution-plane invariants are written to detect.

2.5 The empirical programme the frameworks invite

Adopting a framework as thoroughly as this paper does carries an obligation to be candid about what has and has not been shown. The Full-Stack architecture is normative and architectural; it sets out a testable control structure, and it is explicit that its contribution is integration and placement rather than the invention of each underlying technique [19]. What the current literature does not yet contain is a controlled clinical study establishing that care systems conforming to the invariants produce measurably better outcomes, higher alert precision at equal sensitivity, or reduced inequity, than well-run systems that do not. This is a gap in the evidence base, not a flaw in the design. Several invariants also presuppose organizational capabilities that any honest deployment has to secure: validation independence assumes a validation function with authority to block go-live; remit-based oversight assumes clinicians with standing to override a model rather than to ratify it; and capacity-aware escalation assumes that reviewer throughput can be procured at the required rate. Where those assumptions fail, the controls do not fail loudly; they degrade quietly into documentation.

3. Problem Formulation

3.1 Notation

Let x denote the patient state assembled from electronic-health-record, device, and ancillary data; m a versioned model or analytic policy; π a governance and clinical-policy version; d a named clinical or operational decision; a a candidate action (alert, order, hold, route, or allocate); and t the intended commit time. The design objective is not to minimize prediction error. It is that every decision the system acts on satisfies a conjunction of validity, evidence, and governance conditions, evaluated strictly before the action reaches a patient or a clinician.

3.2 Pre-commit admissibility

A veracity predicate $V(x)$ maps the assembled patient state to either pass or quarantine, so that no input of unestablished provenance or expired temporal validity, a stale vital sign or a lab past its clinical usefulness, can

reach the scoring path and be laundered by downstream accuracy. Let $r(x)$ in $[0,1]$ be a calibrated risk signal for the decision, and let τ be the governance threshold for the decision class. The routing rule is:

$G(x) = 1$ if $r(x) \leq \tau$; otherwise $G(x) = 0$.

$G(x) = 1$ permits automated continuation and $G(x) = 0$ requires clinician review, hold, or escalation according to the risk policy. Writing the gate over a risk signal keeps the direction of τ unambiguous, since a larger τ admits more automation, and it matches the parameterization used in Section 5.3. Where r derives from a model score, calibration across sites and subpopulations is a precondition for the rule to carry meaning at all (Section 5.1). Admissibility for commitment is then the conjunction:

$A(d) = V(x) \text{ AND } Val(m) \text{ AND } Pol(\pi,a) \text{ AND } Ev(d) \text{ AND } Fresh(x,t) \text{ AND } Rev(a) \text{ AND } [G(x) \text{ OR } H(d)]$.

Here $Val(m)$ asserts that the acting model version passed independent validation; $Pol(\pi,a)$ that the action is permitted under the current clinical-governance and residency policy, so that protected health information stays within its jurisdictional and consent constraints; $Ev(d)$ that a complete DecisionRecord exists and has been persisted, including the rationale needed to inform the clinician; $Fresh(x,t)$ that every input satisfies its temporal-validity bound at commit time rather than at scoring time; $Rev(a)$ that the action is idempotent, held for confirmation, or otherwise safely reversible; and $H(d)$ that a clinician holding the required remit has authorized the decision in the case $G(x) = 0$. Commitment is permitted only when $A(d)$ evaluates true. Each conjunct carries a named evidence artifact and a defined failure response (Table 1), and the failure of any conjunct is itself a recorded event rather than a silent fall-through. The structural point, common to the before-commit principle of the adopted framework [19], is that the whole conjunction is evaluated before the action reaches a patient, not alongside it.

3.3 The score-to-action evidence contract

The evidence contract is the one place where this paper adds materially to the foundation, and it is deliberately thin. A generation-time DecisionRecord, carrying the patient-state hash, the model version, the calibrated risk, the rationale, and the policy version, is bound by a shared decision identifier and immutable version references to a runtime ExecutionRecord that captures what action was committed, when, against which input-validity window, and with what reversibility guarantee. The contract's single obligation is reconstructability: after the surrounding system has changed, an auditor, or a clinician reviewing a case, must be able to recover the exact evidence on which an alert or an order rested and the exact rationale that was surfaced. This is the care-domain form of the capture-at-commit and auditability invariants [19], instantiated against the data heterogeneity the healthcare account emphasizes [20].

Conjunct	Condition tested before commit	Evidence / invariant
$V(x)$	Every input has established provenance and unexpired temporal validity.	Provenance log
$Val(m)$	Acting model version passed validation independent of development.	Validation sign-off
$Pol(\pi,a)$	Action permitted under clinical-governance and residency policy for PHI.	Policy binding
$Ev(d)$	Complete DecisionRecord persisted, including clinician-facing rationale.	DecisionRecord
$Fresh(x,t)$	Each input satisfies its temporal-validity bound at commit time.	Freshness stamp
$Rev(a)$	Action idempotent, held for confirmation, or safely reversible.	Reversibility token
$G(x) \text{ OR } H(d)$	Automated gate passed, or a remit-holding clinician authorized the decision.	Gate / review record

Table 1. The pre-commit admissibility conjuncts for a clinical or operational action. Each is evaluated before the action reaches a patient or clinician, carries a named evidence artifact, and has a defined failure response; a failed conjunct is a recorded event, not a silent fall-through.

4. The Governed Healthcare-Analytics Architecture

4.1 Two planes, kept distinct

The architecture keeps a decision-assurance plane and an execution-assurance plane distinct while binding them together. The decision plane forms a calibrated, explained clinical decision and emits a DecisionRecord. The execution plane, which is the Full-Stack architecture applied to care, evaluates admissibility at the commit boundary, fires the alert or places the order, records what it did, and closes the loop when the outcome is later observed [19]. The boundary between the planes is the instant at which the system changes what happens to a

patient, and therefore the instant at which a governance failure becomes hard to reverse. The two planes are joined only by the evidence contract of Section 3.3.

4.2 The provenance and temporal-validity layer

Because clinical analytics draws on heterogeneous, high-velocity data [20], the veracity predicate $V(x)$ is not a formality. It is the layer at which provenance and temporal validity are checked before any input can influence a score, the operational form of the framework's provenance-and-temporal-validity invariant [19]. A six-hour-old lactate or a device reading of uncertain origin is quarantined rather than silently carried forward, because silent carry-forward is precisely how a stale or unattributable signal comes to drive a real-time alert. This layer is also where residency is enforced as a hard constraint, so that protected health information does not leave its permitted jurisdiction in the course of scoring or explanation.

4.3 Evidence-contract binding and clinician-facing rationale

At generation time the decision plane persists the DecisionRecord; at commit time the execution plane persists the ExecutionRecord and binds the two by the shared decision identifier. The binding's load-bearing requirement in care is that the rationale surfaced to a clinician is reproducible from the persisted record: the reasons shown alongside an alert must be the reasons that were actually bound to the decision, not reasons regenerated later by resampling a possibly-changed model. Communicating that rationale in a form clinicians can act on is itself a design problem, addressed by conventions such as Model Facts labels [22], [12]. This turns a clinical-governance expectation into a property the architecture can guarantee by construction rather than reconstruct under audit pressure.

4.4 Closing the outcome loop under delay

A clinical decision's outcome is often observed only after a delay, and the outcome is entangled with the treatment the decision triggered. The closed-loop obligation therefore requires that the outcome signal, when it arrives, is joined back to the originating DecisionRecord through the same identifier, so that realized benefit or harm is attributed to the decision that produced it rather than to whatever model version is current at the time of observation [19]. Section 7 treats the measurement hazards this delay and entanglement create.

5. Calibration, Thresholds, and a Synthetic Sensitivity Analysis

5.1 Calibration as a precondition

The routing rule of Section 3.2 is written over a calibrated risk signal, and the choice is not cosmetic. If $r(x)$ is not calibrated, the threshold τ has no stable meaning, and an alert cut-off chosen at one site silently changes its operating point at another, a failure mode close to the one observed when a widely deployed sepsis model was validated outside its development setting [24]. Because modern models are frequently over-confident [10], post-hoc recalibration, or a distribution-free method that provides validity guarantees without assuming calibration [2], is treated here as a precondition for the gate. Calibration must moreover be checked within subpopulations, not only in aggregate, since aggregate calibration can coexist with harmful subgroup miscalibration [14].

5.2 Threshold selection under asymmetric harm

Alert-threshold selection is an explicit harm problem, not a default choice of 0.5. The harm of a missed deterioration and the harm of a false alarm, which consumes clinician attention and, at scale, erodes trust and produces alert fatigue, are asymmetric and setting-dependent. The threshold τ and the review band around it are therefore governance parameters, set against a stated harm model and recorded as part of the policy version π , so that a later change in the operating point is itself a governed, evidenced event.

5.3 A synthetic sensitivity analysis

To illustrate the escalation-versus-harm trade-off without claiming empirical weight, consider a parameterized Monte Carlo over a population with a fixed event rate, a model of stated discrimination, and a calibrated risk signal. Sweeping the review band widens or narrows the set of borderline cases routed to a clinician. As the band widens, expected harm from missed events falls, but clinician review load rises, and beyond a point the marginal harm averted per additional review collapses toward zero while alert burden continues to grow. The resulting cost surface has an interior operating region rather than a corner solution, which is the quantitative form of the claim that the right amount of human review is neither zero nor unbounded. The figure this analysis would produce is a cost

surface, not a measurement; its parameters are illustrative and are stated in full in the accompanying materials so that it cannot be mistaken for evidence.

6. Conformance Model and Invariant Mapping

Conformance is defined cumulatively, so that a system claims a level only by satisfying every obligation of the levels beneath it. An Observable system persists the DecisionRecord and ExecutionRecord and can reconstruct any alert or action after the fact. A Governed system additionally enforces the full admissibility predicate $A(d)$ at the commit boundary, including provenance, temporal validity, validation independence, residency, and rationale reproducibility. An Adaptive system additionally closes the outcome loop, attributes realized benefit or harm to the originating decision, and governs any change to the operating point as an evidenced event. Each adopted invariant maps onto a stage of this lifecycle and is enforced at a named point; Appendix A gives the mapping, so that the invariants can be carried, unchanged, into a clinical-governance audit checklist rather than restated as aspiration.

7. Evaluation Protocol and Measurement Hazards

The evaluation protocol keeps three families of metric apart so that a gain in one is never reported as evidence for another: model-level discrimination and calibration; decision-level quality (alert precision at a given sensitivity, review precision, rationale fidelity, and subgroup equity); and conformance-level coverage (the fraction of committed actions for which the full evidence contract is reconstructable). A field program that adopts this architecture must confront three measurement hazards particular to care, and the protocol is written to guard against each.

- **Dataset shift.** Clinical data distributions move across sites, seasons, devices, and practice changes, so a model validated in one setting can fail silently in another; monitoring must distinguish exogenous shift from change the deployment itself induced [5], [8].
- **Label latency and treatment-induced censoring.** Outcomes mature slowly and are entangled with the treatment the decision triggered, so a decision that prevents an event also erases its own label. Decision-level metrics must state their censoring assumptions and attribute outcomes to the originating decision version [19].
- **Alert-fatigue feedback.** Because clinicians habituate to frequent low-value alerts, the alerting policy reshapes the very responsiveness it depends on; evaluation must track response rates over time as an endogenous quantity, not a fixed property [24].

None of these hazards is a defect of the adopted framework; they are properties of care as a domain, and the value of stating them is that a conformance claim which ignores them is not credible.

8. Discussion

The synthesis offered here is narrow by design. It does not propose new clinical features, a new model class, or a new fairness criterion. It proposes only the governance interface at which Sazu's execution-assurance account [19] meets the healthcare-analytics context described by Sazu and Jahan [20]: a shared evidence contract, a pre-commit admissibility predicate instantiated for clinical and operational actions, and a conformance model that turns runtime invariants into quantities an auditor or a clinical-governance committee can check. The everything-underneath is adopted rather than reinvented, and the value of the interface is that it lets a healthcare organization state precisely what it means when it claims that an automated clinical action was governed.

9. Limitations

The paper reports no clinical deployment, and the sensitivity analysis is illustrative. The architecture inherits every organizational precondition named in Section 2.5: without an independent validation function, a remit-bearing clinician-override capability, and procurable review throughput, several conjuncts of $A(d)$ degrade into documentation. The evidence contract adds storage and latency cost at the commit boundary that a high-throughput clinical system must budget for. And the equity treatment here is structural, it makes subgroup calibration checkable and rationale reproducible, but it does not itself supply a fairness criterion; that choice remains a clinical-governance decision the architecture records rather than makes [14].

10. Conclusion

Healthcare systems now act on analytical outputs without a clinician in the immediate path, and in that setting the discrimination of a model is not the quality of a clinical decision. This paper has taken the Full-Stack Production-Platform Reference Architecture introduced by Mesbaul Haque Sazu [19] and the account of big data analytics and

artificial intelligence in healthcare management given by Sazu and Jahan [20], and composed them into a single decision-to-action architecture that keeps decision assurance and execution assurance distinct while binding them through a reconstructable evidence contract. The contribution is the interface, not a replacement for either foundation, and the empirical program that would test it is specified rather than assumed.

References

References are listed alphabetically by first-author surname; in-text citation numbers follow this ordering.

- Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete Problems in AI Safety. arXiv:1606.06565.
- Angelopoulos, A. N., & Bates, S. (2021). A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification. arXiv:2107.07511.
- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion*, 58, 82–115.
- European Parliament and Council. (2024). Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). Official Journal of the European Union.
- Finlayson, S. G., Subbaswamy, A., Singh, K., Bowers, J., Kuznetsov, A., Agarwal, P., Babic, B., Kohane, I. S., & Saria, S. (2021). The Clinician and Dataset Shift in Artificial Intelligence. *New England Journal of Medicine*, 385(3), 283–286.
- Food and Drug Administration. (2021). Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. U.S. Food and Drug Administration.
- Food and Drug Administration, Health Canada, & Medicines and Healthcare Products Regulatory Agency. (2021). Good Machine Learning Practice for Medical Device Development: Guiding Principles.
- Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A Survey on Concept Drift Adaptation. *ACM Computing Surveys*, 46(4), 1–37.
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for Datasets. *Communications of the ACM*, 64(12), 86–92.
- Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On Calibration of Modern Neural Networks. *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 70, 1321–1330.
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 4765–4774.
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*)*, 220–229.
- National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1.
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science*, 366(6464), 447–453.
- Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2022). Challenges in Deploying Machine Learning: A Survey of Case Studies. *ACM Computing Surveys*, 55(6), 1–29.
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, 380(14), 1347–1358.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
- Rudin, C. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, 1, 206–215.
- Sazu, M. H. (2023). A Full-Stack Production-Platform Reference Architecture for Governed Agentic and Automated-Action Enterprise Systems. *IPHO-Journal of Advance Research in Science and Engineering*, 1(12), 49–58.
- Sazu, M. H., & Jahan, S. A. (2022). Big Data Analytics & Artificial Intelligence in Management of Healthcare: Impacts & Current State. *Management of Sustainable Development*, 14(1), 36–42.

- Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden Technical Debt in Machine Learning Systems. *Advances in Neural Information Processing Systems (NeurIPS)*, 28, 2503–2511.
- Sendak, M. P., Gao, M., Brajer, N., & Balu, S. (2020). Presenting Machine Learning Model Information to Clinical End Users with Model Facts Labels. *npj Digital Medicine*, 3, 41.
- U.S. Department of Health and Human Services. (2013). HIPAA Privacy, Security, and Breach Notification Rules (Omnibus Rule). 45 CFR Parts 160 and 164.
- Wong, A., Otles, E., Donnelly, J. P., Krumm, A., McCullough, J., et al. (2021). External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Internal Medicine*, 181(8), 1065–1070.

Appendix A. Invariant-to-Enforcement Mapping

The mapping below shows where each adopted runtime invariant is instantiated in the governed care lifecycle and the named point at which it is enforced. It is intended to be carried directly into a clinical-governance implementation audit checklist.

Adopted invariant	Lifecycle stage where instantiated	Enforcement point
Capture-at-commit	Recording the fired action and its evidence	ExecutionRecord
Before-commit gating	Admissibility evaluation prior to action	A(d) boundary
Compensability	Reversibility / hold of automated actions	Rev(a)
Provenance & temporal validity	Input veracity and freshness	V(x), Fresh(x,t)
Residency as a hard constraint	PHI jurisdiction and consent	Pol(π ,a)
Logged propensity	Alerting-policy probabilities for off-policy evaluation	Threshold governance
Auditability	Reconstruction of any alert or action	Evidence contract

Table A1. Adopted runtime invariants and their enforcement points in the governed healthcare-analytics lifecycle.