Article

Not peer-reviewed version

Blockchain-Enabled Trust and Compliance for Clinical AI: Decentralized Governance without Decentralized Data Storage

Dimitrios P. Panagoulias *, Andrei Ionut Damian, Cosmin Stamate, Vitalli Toderian, Petrica Butusina, Alessandro De Franceschi, Cristian Bleotiu, Evangelos Sakkopoulos, Evangelia-Aikaterini Tsichrintzi

Posted Date: 5 June 2026

doi: 10.20944/preprints202606.0484.v1

Keywords: clinical AI governance; decentralized governance; blockchain auditability; consent-aware AI pipelines; reproducible clinical AI; trustworthy medical AI systems

 doi:10.20944/preprints202606.0484.v1

Article

Blockchain-Enabled Trust and Compliance for Clinical AI: Decentralized Governance without Decentralized Data Storage

Dimitrios P. Panagoulias [1,*], Andrei Ionut Damian [2,3], Cosmin Stamate [2,4] Vitalii Toderian [2,3], Petrica Butusina [2], Alessandro De Franceschi [2], Cristian Bleotiu [2], Evangelos Sakkopoulos [1] and Evangelia-Aikaterini Tsichrintzi [5]

[1] University of Piraeus; Department of Informatics; Piraeus 185 34; Karaoli ke Dimitriou 80; Greece
[2] Ratio1.ai, Bucharest, Romania
[3] Computer Science and Information Technology Department, Faculty of Automatic Control and Computers, University POLITEHNICA of Bucharest, Bucharest, Romania
[4] Birkbeck Knowledge Lab, Birkbeck College, University of London, Malet Street, London, WC1E 7HZ, UK
[5] Noetiv PC, Athens, Greece
* Correspondence: panagoulias_d@unipi.gr

Abstract

Clinical AI systems increasingly rely on large-scale medical imaging data processed through complex and continuously evolving machine learning pipelines. While cloud-based infrastructures enable scalability and performance, they introduce challenges related to trust, auditability, consent management, and reproducibility, particularly in multi-institutional and longitudinal clinical settings. This paper proposes a decentralized governance framework for clinical AI that leverages blockchain technology as a verification and policy-enforcement overlay, without decentralizing the storage of sensitive medical data. Raw images and derived clinical artifacts remain within secure cloud-based repositories, while cryptographic proofs, processing manifests, access events, and consent policies are recorded on a distributed ledger. In addition, the framework supports the use of decentralized computation infrastructures for training and experimentation on de-identified or synthetic datasets, enabling scalable and collaborative AI development without exposing patient data. The proposed architecture enables verifiable provenance, tamper-evident audit trails, programmable consent enforcement, and deterministic reconstruction of AI-assisted clinical decisions, while preserving regulatory compliance, system performance, and clinical usability. To demonstrate the feasibility and practical benefits of the approach, a medical imaging use case is presented which utilized nine simulated clinical scenarios involving about 43,000 inferences of patient groups ranging from 50 to 1,000 subjects. Our proposed framework achieved a mean Governance Quality Index –a composite measure of security, compliance, performance, and auditability– of 0.93, indicating production ready performance, with governance overhead below 11 ms per operation and throughput exceeding 220 requests per second. In summary, our approach separates governance from data and computation. Blockchain is used solely as a tamper-evident governance layer that anchors consent, access, and provenance through cryptographic commitments, while medical data and AI pipelines remain unchanged within existing cloud infrastructures. This enables verifiable auditability and reproducibility without decentralizing clinical data or disrupting workflows.

Keywords: clinical AI governance; decentralized governance; blockchain auditability; consent-aware AI pipelines; reproducible clinical AI; trustworthy medical AI systems

 doi:10.20944/preprints202606.0484.v1

1. Introduction

1.1. Motivation

The adoption of artificial intelligence (AI) in clinical practice has accelerated significantly, particularly in medical imaging domains where machine learning models support diagnosis, monitoring, and clinical decision-making. Contemporary clinical AI systems typically rely on centralized or hybrid cloud infrastructures to manage medical images, execute AI pipelines, and support longitudinal patient follow-up. While such architectures offer scalability and operational efficiency, they introduce persistent challenges related to trust, accountability, consent governance, and reproducibility of AI-driven clinical decisions. These challenges are amplified in multi-institutional and research-driven contexts, where medical data may be reused across studies, models evolve rapidly, and regulatory scrutiny requires clear evidence of compliant data use. In practice, governance is often enforced through institution-specific policies, mutable logs, or manual oversight, limiting transparency and making independent verification difficult.

Blockchain technologies have been proposed as a solution to these limitations, primarily due to their immutability and decentralized trust model. However, existing approaches frequently attempt to decentralize medical data storage or AI computation, resulting in designs that conflict with performance requirements, regulatory constraints, and established clinical workflows. In healthcare settings, raw medical data must remain tightly controlled, deletable, and auditable within trusted infrastructures.

This work proposes an alternative approach, namely decentralizing governance rather than data. By introducing a decentralized governance overlay that operates independently of data storage and model execution, the proposed framework enables verifiable trust, auditability, consent enforcement, and reproducibility, while preserving the practical advantages of cloud-based clinical AI systems. Furthermore, the framework selectively supports decentralized computation infrastructures for AI training and experimentation on de-identified or synthetic datasets, enabling scalable collaboration without compromising patient privacy.

Furthermore, we note that current centralized solutions fail to address the "Last Mile" of clinical AI deployment as hospitals operate heterogeneous, legacy IT environments where deploying modern microservices is most of the time operationally prohibitive. Furthermore, the transmission of high-frequency patient telemetry (e.g. real-time arrhythmia monitoring) to centralized clouds introduces unacceptable censorship and privacy risks. A viable solution must therefore bring the computation to the data, not the data to the computation. Thus, we propose addressing this via a decentralized orchestration layer that treats hospital servers as authenticated, autonomous nodes within a global meta-os approach.

Table 1 summarizes the key governance gaps in current clinical AI systems. While existing infrastructures rely on centralized logging, local access control, and procedural reproducibility, these mechanisms lack cryptographic guarantees and cross-institutional verifiability. As AI pipelines evolve over time and across organizational boundaries, these limitations hinder trustworthy auditability, enforceable consent, and defensible clinical decision reconstruction.

Table 1. Limitations of Conventional Governance Mechanisms.

Aspect	Limitation
Audit logs	Mutable, centrally controlled, susceptible to modification.
Access control	Local enforcement without global verifiability.
Consent	Not cryptographically bound to execution.
Reproducibility	Lacks verifiable guarantees of original inputs.

 doi:10.20944/preprints202606.0484.v1

1.2. Contributions

The novelty of this work lies in the formalization of decentralized governance as a first-class architectural layer for clinical AI, explicitly decoupled from both medical data storage and AI model execution. In addition, the framework uniquely distinguishes between clinical production workflows and research-oriented AI training, enabling the use of decentralized computation infrastructures exclusively for de-identified or synthetic data.

The main **contributions** of this work are summarized as follows:

The contributions are organized along three complementary dimensions to improve clarity and positioning: (i) architectural contributions, which define the governance-layer abstraction and its integration within existing clinical infrastructures; (ii) methodological contributions, which formalize governance verification mechanisms such as consent anchoring, reproducibility manifests, and the Governance Quality Index (GQI); and (iii) experimental contributions, which demonstrate the feasibility and performance of the proposed framework through a structured simulation study.

- A governance-layer architecture for clinical AI systems that enables verifiable trust, auditability, and consent enforcement without decentralizing medical data or disrupting clinical workflows.
- A ledger-anchored framework for binding consent, audit events, and reproducibility artifacts through cryptographic commitments.
- A reproducibility mechanism based on manifest generation and verification, enabling deterministic reconstruction of AI-assisted clinical decisions.
- The introduction of a Governance Quality Index (GQI) as a quantitative metric for evaluating governance properties in clinical AI pipelines.
- A prototype implementation and experimental evaluation across multiple simulated clinical scenarios, demonstrating measurable governance performance with minimal overhead.

Distributed coordination frameworks such as ChainDist [1] have demonstrated how ledger-anchored authorization can support verifiable execution in decentralized ML systems. However, these systems emphasize execution coordination, whereas the present work formalizes governance verification, consent binding, and reproducibility enforcement as a distinct architectural layer for clinical AI. To the best of our knowledge, prior work has not jointly operationalized (i) ledger-anchored, machine-readable consent state (including revocation), (ii) tamper-evident, cross-institution audit events, and (iii) ledger-committed reproducibility manifests enabling deterministic reconstruction of AI-assisted clinical decisions, while preserving off-ledger clinical data storage and production inference within existing cloud infrastructures. We position this claim relative to prior blockchain-for-health and provenance/MLOps literature in Sec. 2.6.

1.3. Organization of the Paper

The manuscript is structured to progressively introduce the conceptual, architectural, and experimental aspects of the proposed governance framework. The Introduction provides motivation, defines the problem space, and outlines the main contributions. Section II reviews related work and positions the proposed approach within the broader literature. Section III presents the decentralized governance architecture, including baseline governance mechanisms, threat model, and core design components. Section IV describes the medical imaging use case, experimental setup, and evaluation methodology. Section V discusses the results, limitations, and future work.

The framework was quantitatively evaluated across nine simulated scenarios spanning varying patient populations, workload scales, consent revocation rates, and integrity conditions. Using the proposed Governance Quality Index (GQI), the system achieved a mean score of 0.93, with governance overhead below 5 ms per operation and throughput exceeding 220 requests per second, while consistently detecting integrity violations.

The remainder of the paper is organized as follows: Section II reviews related work and literature, Section III presents the proposed framework and algorithms, Section IV describes the use case and experimental evaluation, and Section V discusses conclusions, limitations and future work.

 doi:10.20944/preprints202606.0484.v1

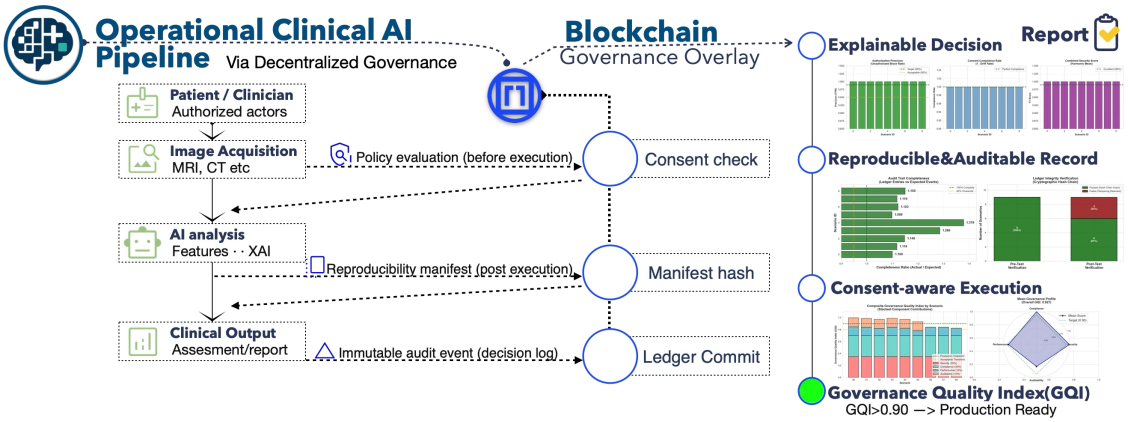

Figure 1. On the left, an operational clinical AI pipeline is shown, encompassing authorized clinical actors, medical image acquisition, AI-based analysis with explainability, and clinical output generation, all executed using standard clinical infrastructure. In the middle, a blockchain-backed governance overlay introduces event-triggered control points—consent verification prior to execution, reproducibility manifest hashing following inference, and immutable ledger commits at the point of clinical decision—without storing medical data or performing AI computation on-chain. On the right, the resulting trust outcomes are illustrated, including explainable decision support, reproducible and auditable records, consent-aware execution, and quantitative evaluation through a Governance Quality Index (GQI), indicating production readiness when GQI ≥ 0.90 (The visualizations shown on the right-hand side are presented in condensed form for illustrative purposes within the graphical abstract. Higher-resolution versions, along with detailed quantitative analysis and interpretation, are provided in the main text and associated figures).

2. Related Work

Recent advances in distributed systems security and adversarial resilience highlight the increasing importance of robust trust, verification, and coordination mechanisms in decentralized and semi-decentralized environments. In particular, research has explored collusion-resistant verification protocols for data integrity [16], adaptive trust consensus mechanisms in blockchain-enabled IoT systems [17], and privacy-preserving coordination in adversarial multi-agent environments [19]. Complementary work has addressed secure key distribution and revocation under collusion scenarios [20], as well as authentication and clone detection in distributed sensing infrastructures [21].

At the same time, the increasing deployment of distributed machine learning and federated systems has exposed new classes of adversarial behaviors, including free-riding and collusion attacks, motivating the development of robust coordination and validation mechanisms [22]. Broader research in network security and cyber-defense further emphasizes adaptive risk modeling, moving target defenses, and resilient system design under uncertain adversarial conditions [18,23]. Additionally, threats such as co-residency attacks [24], routing privacy vulnerabilities [25], and timing-based infrastructure exploits [26] illustrate the complexity of securing distributed execution environments.

While these works focus primarily on computation, communication, or infrastructure-level security, they collectively underscore the need for systematic, verifiable governance mechanisms that can operate across heterogeneous and potentially adversarial environments. In contrast, the present work focuses specifically on governance-layer guarantees for clinical AI systems, addressing consent enforcement, auditability, and reproducibility without requiring decentralization of sensitive medical data or disruption of established clinical workflows. Table 2 provides a high-level synthesis of recent approaches, while the following sections expand on these categories with a more detailed discussion of their design assumptions, strengths, and limitations in the context of clinical AI governance.

 doi:10.20944/preprints202606.0484.v1

Table 2. Overview of Recent Approaches in AI Governance, Blockchain, and Distributed Systems (2019–2026). The table summarizes representative trends and selected works, while a more detailed and comprehensive discussion of the literature is provided in the subsequent sections.

Category	Representative Works	Focus	Limitations
Blockchain in Healthcare	[2–4]	Data sharing, access control, integrity	Often assumes data decentralization or limited clinical integration
Provenance & ML Lineage	[5–9]	Reproducibility, pipeline tracking, artifact lineage	Centralized trust assumptions, limited audit verifiability
Federated & Privacy-Preserving ML	[10–12]	Data locality, privacy-preserving training	Limited governance, auditability, and consent enforcement
Blockchain Provenance & Audit	[13–15]	Tamper-evident logs, traceability	Focus on logging or data lineage, not full governance lifecycle
Distributed Systems Security	[16–18]	Trust, adversarial resilience, coordination	Infrastructure-level focus, not application-level governance
This Work	–	Governance-layer abstraction (consent, audit, reproducibility)	Prototype implementation (no distributed consensus)

2.1. Key Terminology and Concepts

Clinical AI governance

A set of technical and organizational controls that ensure AI systems used in healthcare remain compliant, auditable, safe, accountable across the full lifecycle (development, validation, deployment, monitoring) ensuring ethical governance and trustworthy usage [27–29]. Governance typically spans consent, access control, provenance, and change management [30–32].

Provenance

Metadata describing the origin, lineage, and transformations applied to data and outputs, enabling traceability of how a clinical AI result was produced. Provenance is often standardized via models such as W3C PROV [5]. Recent work has explored the use of blockchain technologies to enhance provenance integrity, traceability, and ownership in distributed systems and AI pipelines [13,14,33,34]. These approaches emphasize tamper-evident tracking of data transformations and model outputs, reinforcing trust and auditability in decentralized environments.

Audit Trail and Tamper-Evidence

A chronological record of events (e.g., data access, policy updates, and inference executions) that supports compliance review and forensic analysis. Tamper-evidence implies that any post-hoc modification or deletion of records becomes detectable, for example through mechanisms such as hash chaining, digital signatures, and append-only transparency logs [35–37]. Recent work has explored blockchain-based and tamper-evident logging frameworks to enhance auditability and forensic traceability in distributed systems [15,38–41]. These approaches reinforce the importance of immutable audit trails for compliance, security monitoring, and post-incident analysis.

Consent Management and Dynamic Consent

Mechanisms to represent, enforce, and update patient authorization over time (including revocation). Dynamic consent emphasizes continuous, fine-grained consent that can evolve with context and patient preferences [42–44].

On-Ledger vs. Off-Ledger Storage

A design distinction where sensitive artifacts (raw images, detailed manifests) remain in secure repositories (off-ledger), while only cryptographic commitments, pointers, and event metadata are written to a distributed ledger (on-ledger) [45,46].

 doi:10.20944/preprints202606.0484.v1

Commitment (Hash Commitment)

A cryptographic digest (e.g., SHA-256) recorded on a ledger to "commit" to an off-ledger artifact, enabling later verification that the artifact has not changed without revealing its contents [47].

More advanced commitment schemes, including lattice-based constructions, polynomial commitments, and vector commitments, have been proposed to support stronger security guarantees and efficient verification in distributed settings [48–50]. Commitment-based mechanisms are also widely used in blockchain systems and privacy-preserving protocols, such as payment channels and commit-reveal schemes, to ensure integrity and prevent adversarial manipulation [51,52].

Reproducibility Manifest

A structured record capturing the parameters required to deterministically reconstruct an AI decision: model identifiers, pipeline configuration, dependencies, input/output digests, and execution context. This is closely related to modern supply-chain provenance and attestation frameworks that produce signed statements about builds and executions (e.g., in-toto and SLSA provenance), and to SBOM guidance that specifies minimum dependency metadata [53,54]. We additionally rely on content-addressable artifact identification (digests) to bind referenced software and pipeline environments to immutable identifiers.

Explainability Provenance

The practice of recording not only model outputs but also the explainability artifacts (e.g., saliency maps, feature attributions) with verifiable lineage to support accountability and clinical review [55–57].

De-Identification and Synthetic Data

Approaches to reduce privacy risk by removing identifiers (de-identification/pseudonymization) or generating privacy-preserving synthetic datasets for experimentation and training [58,59].

Differential Privacy, Federated Learning, and Secure Computation

Complementary privacy technologies: differential privacy bounds information leakage, federated learning keeps data local while aggregating model updates, and secure computation (e.g., MPC, TEEs, ZKPs) can reduce disclosure during computation. These methods are often discussed for multi-institutional learning but introduce additional complexity and performance trade-offs [10–12,60].

2.2. Blockchain and Distributed Ledgers in Healthcare

Early blockchain-for-healthcare research largely focused on sharing or indexing medical records and enforcing access policies via smart contracts, motivated by interoperability and integrity concerns [2,3]. In practice, most deployable architectures avoid storing PHI directly on-chain and instead use off-chain repositories with on-chain pointers and hash commitments [45]. This design choice is also motivated by data protection guidance highlighting that blockchain immutability can conflict with GDPR obligations when personal data are placed on-chain; consequently, architectures typically minimize on-chain personal data and rely on off-chain storage with verifiable commitments [61]. Distributed ledger technologies (DLTs) have also been studied as immutable audit logs for regulated environments, where append-only properties can strengthen accountability and non-repudiation [46,62]. However, healthcare adoption remains constrained by privacy, governance, and operational requirements, leading to architectures that privilege minimal on-chain metadata and integrate with existing clinical systems rather than replacing them [1,4,63,64].

A practical on-ledger governance layer for healthcare must account for the characteristics of programmable distributed ledgers. While such ledgers provide tamper-evidence, traceability, and verifiable execution, they also impose strict operational constraints that shape what should be recorded on-ledger. In particular, smart-contract execution is deterministic and verifiable by independent participants, and once governance events, commitments, or state transitions are appended to the ledger, they cannot be silently modified or erased. This immutability is a key property for auditability

 doi:10.20944/preprints202606.0484.v1

and non-repudiation, but it also requires conservative design across all expected flows and failure modes, since implementation errors, misconfigurations, or unintended payloads become persistently observable. This is especially critical in healthcare contexts, where any accidental inclusion of sensitive data, including PHI, would be permanently anchored on-ledger and may be difficult to remediate.

In addition, programmable ledgers expose explicit cost models that make high-volume or verbose on-ledger records impractical. This strengthens the case for a minimal on-ledger footprint, where only governance-critical metadata and cryptographic commitments (e.g., hash commitments to off-ledger artifacts such as policies, manifests, and outputs) are recorded, while detailed records remain off-ledger within compliant repositories. Finally, when governance events occur at higher frequency, engineering considerations extend beyond cost to include throughput limits, transient failures, fee volatility, and operational safeguards such as buffering and batching, to ensure completeness of the append-only audit trail. These constraints reinforce the design principle adopted in this work: the ledger acts as a verification and policy-enforcement overlay that anchors consent, access, and provenance through cryptographic commitments, while clinical artifacts and AI pipelines remain off-ledger within existing infrastructures.

2.3. Provenance, Lineage, and Reproducibility for ML Systems

Outside healthcare, reproducibility and traceability have been addressed through ML experiment tracking and data/version control systems (e.g., MLflow, DVC), as well as software supply-chain frameworks (e.g., in-toto, SLSA) that produce signed attestations about builds and executions [6–9]. These tools establish strong engineering patterns for capturing pipeline parameters, artifacts, and dependencies, but typically assume trusted centralized infrastructure and do not natively provide decentralized trust, multi-party auditability, or programmable consent[65].

Within clinical contexts, provenance standards such as W3C PROV and healthcare interoperability standards (e.g., HL7 FHIR, DICOM) provide foundations for describing data lineage and operational events, yet do not by themselves guarantee tamper-evidence or cross-institutional audit trust without additional controls [5,66,67].

2.4. Consent and Policy Enforcement in Clinical Data Systems

Consent enforcement for clinical data access is traditionally implemented in centralized access-control systems tied to institutional identity management, logging, and governance workflows [64,68]. Dynamic consent frameworks extend this by supporting evolving policies and fine-grained revocation, which is particularly relevant for longitudinal imaging datasets and secondary research use [42,43]. Nevertheless, in multi-institutional settings, audit trust and policy reconciliation remain challenging when enforcement is distributed across organizations with heterogeneous infrastructures [69,70].

2.5. Privacy-Preserving Learning and Decentralized Computation

Federated learning, secure aggregation, and privacy-preserving computation have been widely proposed to enable multi-institutional model development while minimizing raw data sharing [10,12, 71]. Differential privacy provides formal privacy guarantees but can reduce utility, especially in high-dimensional medical imaging [11]. Zero-knowledge proofs and TEEs can strengthen confidentiality and integrity of computation, but often increase system complexity and may be overkill when the primary aim is governance and auditability rather than fully private computation [60,72]. Accordingly, many clinical deployments emphasize pragmatic architectures where core datasets remain in compliant cloud repositories and privacy is addressed via organizational controls, de-identification, and strict governance [1,58,73,74].

2.6. Positioning and Novelty Relative to Prior Work

Prior work has established that (i) storing PHI directly on-chain is generally undesirable, (ii) off-chain storage with on-chain commitments is a practical pattern, and (iii) provenance and ML lineage tooling can improve reproducibility [2,6,7,45]. However, much of the blockchain-for-health literature centers on record exchange, identity, or generalized access control, and MLOps lineage tooling typically

 doi:10.20944/preprints202606.0484.v1

assumes centralized trust [3,65]. Building upon our previous research into decentralized computation and privacy-preserving data governance[1,74], this study leverages the coupling of off-chain data storage with on-chain hash persistence. Although that work provides a basis for decentralized MLOps orchestration and consensus-based work commitment management, the present paper restricts its scope to the governance verification overlay.

Furthermore in our prior work, governance, compliance, and validation have consistently emerged as central challenges in the deployment of AI-driven medical systems. In [75,76] and related studies on multimodal AI in dermatology, we highlighted the difficulty of ensuring traceability, explainability, and regulatory alignment in complex image-based diagnostic pipelines [77]. Subsequent contributions on AI-empowered biomedical applications and microservices-based development frameworks [78] emphasized the need for structured auditability, reproducibility, and lifecycle control to support clinical trust and certification. In parallel, our analyses of regulatory and validation challenges [79,80] in healthcare AI identified consent management, provenance tracking, and post-hoc accountability as persistent gaps not adequately addressed by existing software architectures. The present work builds directly on these findings by operationalizing governance as a first-class, system-level concern—introducing a decentralized, verifiable governance overlay that translates previously identified regulatory and compliance requirements into enforceable, measurable, and auditable mechanisms suitable for real-world clinical AI deployments [81].

Our Contribution

We specifically target *clinical AI pipeline governance* by introducing a layered architecture in which the ledger acts as a verification and policy-enforcement overlay that records: (a) cryptographic commitments to off-ledger artifacts, (b) tamper-evident audit events, (c) machine-readable consent state (including revocation), and (d) reproducibility manifests for deterministic reconstruction of AI decisions.

Distinct from decentralized data storage or fully decentralized learning, the proposed design preserves the operational advantages of compliant cloud repositories while enabling multi-party audit trust, verifiable provenance, and off-ledger tamper detection (manifest corruption) through ledger-anchored commitments [5,31,63]. Existing work typically treats governance as policy or post-hoc logging; in contrast, we operationalize governance as a runtime-enforced, cryptographically verifiable execution layer for clinical AI, without altering clinical data locality.

To strengthen comparative positioning, Table 3 contrasts the proposed framework with representative categories of existing approaches, including blockchain-based healthcare systems, federated learning settings, MLOps/provenance frameworks, and blockchain provenance systems. For example, systems such as MedRec [82] and SHAREChain [83] demonstrate how blockchain can support medical data sharing and interoperability through registry-based architectures, yet they typically focus on access control and data exchange rather than full lifecycle governance. Similarly, blockchain provenance systems such as TransChain [13] emphasize tamper-evident tracking of medical artifacts, but do not integrate consent enforcement or reproducibility guarantees at the level of AI pipeline execution.

In the context of machine learning operations, tools such as MLflow and its extensions [84] provide robust experiment tracking and reproducibility within centralized infrastructures, while supply-chain security frameworks such as in-toto and SLSA (e.g., ARGO-SLSA [85] and Let'sTrace [86]) introduce verifiable build and execution attestations. However, these approaches assume trusted execution environments and do not provide decentralized audit trust or consent-aware governance across institutional boundaries. Classical software versioning and traceability approaches [87] similarly enable lineage tracking but lack cryptographic anchoring and tamper-evident guarantees required in regulated clinical environments.

Federated learning approaches [88] address data locality and privacy-preserving distributed training, but do not inherently provide mechanisms for auditability, consent binding, or deterministic reconstruction of AI-assisted decisions. As a result, they address only a subset of the governance lifecycle.

 doi:10.20944/preprints202606.0484.v1

Overall, the comparison highlights that prior work typically addresses isolated aspects of governance (e.g., data sharing, provenance tracking, or distributed training), whereas the proposed framework integrates consent binding, auditability, reproducibility, and governance-oriented evaluation within a unified architectural layer, without requiring decentralization of sensitive clinical data or disruption of existing workflows.

Representative systems such as MedRec and FHIRChain (healthcare blockchain platforms), MLflow and in-toto (provenance and supply-chain frameworks), and canonical federated learning approaches were selected as reference points due to their widespread adoption and relevance to healthcare data governance, reproducibility, and distributed system design. These cases provide concrete anchors for comparison and highlight how existing systems address isolated aspects of the governance lifecycle, in contrast to the integrated governance-layer abstraction proposed in this work. In addition to the structural distinctions summarized in Table 3, the proposed framework introduces a quantitative governance metric (GQI), enabling explicit evaluation of trust, compliance, performance, and reproducibility properties. This further differentiates the proposed approach from existing systems, which generally emphasize either infrastructure, learning coordination, or lineage tracking, but do not provide an integrated governance-oriented evaluation framework for clinical AI. Although beyond the scope of this paper, we outline a potential extension to reconcile data locality with global auditability through a split-state configuration, whereby raw clinical artifacts remain within hospital-controlled environments while only content-addressable identifiers are exposed for external verification.

Table 3. Comparative Analysis of Governance Approaches.

Approach & Description	Governance Capabilities
Blockchain Healthcare Systems *Examples:* MedRec, FHIRChain *Description:* Patient record sharing and interoperability platforms using on-chain pointers and access control	**Data Decentralization:** Partial **Consent Enforcement:** Limited **Auditability:** Yes **Reproducibility:** Limited **Governance Metric:** No
Federated Learning *Examples:* McMahan et al. *Description:* Distributed training with local data; focuses on privacy-preserving learning, not governance	**Data Decentralization:** Yes **Consent Enforcement:** No **Auditability:** Limited **Reproducibility:** No **Governance Metric:** No
MLOps / Provenance Tools *Examples:* MLflow, DVC, in-toto, SLSA *Description:* Centralized lineage tracking and build/run attestations without decentralized audit trust	**Data Decentralization:** No **Consent Enforcement:** No **Auditability:** Limited **Reproducibility:** Yes **Governance Metric:** No
Blockchain Provenance Systems *Examples:* TransChain, BEATS *Description:* Tamper-evident logging and lineage anchoring, typically without consent binding or full lifecycle governance	**Data Decentralization:** Partial **Consent Enforcement:** No **Auditability:** Yes **Reproducibility:** Partial **Governance Metric:** No
This Work *Description:* Decentralized governance overlay with cryptographic commitments, consent binding, and reproducibility manifests	**Data Decentralization:** No **Consent Enforcement:** Yes **Auditability:** Yes **Reproducibility:** Yes **Governance Metric:** Yes (GQI)

 doi:10.20944/preprints202606.0484.v1

3. Decentralized Governance Overlay for Clinical AI

3.1. Baseline Governance Approaches in Clinical AI Systems

As a methodological baseline, we consider the governance mechanisms most commonly employed in contemporary clinical AI deployments. These systems typically rely on centralized identity and access management frameworks, procedural audit logging, and conventional MLOps practices to enforce authorization and support traceability. Identity and access control are usually implemented through standardized authentication and authorization protocols such as OpenID Connect and OAuth 2.0, combined with role-based or attribute-based access control (RBAC/ABAC) models enforced at the application or infrastructure level [89–91]. These mechanisms are effective at regulating access at runtime and integrating with existing hospital information systems; however, access decisions are evaluated locally and are not cryptographically bound to downstream AI execution events. Auditability in baseline systems is commonly achieved through centralized or system-local logging infrastructures (e.g., application logs, database logs, or cloud-native audit trails) [92]. While such logs support operational monitoring and post-hoc review, they remain mutable by privileged actors and are typically siloed across systems or institutions, limiting their suitability for cross-organizational verification and medico-legal accountability.

Reproducibility and provenance tracking are addressed through standard MLOps toolchains, including experiment tracking, versioned artifacts, and pipeline metadata (e.g., MLflow-style lineage) [93]. These approaches provide procedural reproducibility but do not offer cryptographic guarantees that a specific model, pipeline configuration, consent state, and input artifact were jointly used for a given clinical inference. As a result, reconstruction of historical AI-assisted decisions relies on trust in infrastructure integrity rather than verifiable evidence. Finally, consent management is typically handled via database-backed policy records or access control rules that are evaluated at request time [94]. While revocation and policy updates can be enforced prospectively, baseline systems do not cryptographically bind consent policies to individual AI executions, leaving room for ambiguity in longitudinal or multi-institutional settings.

Our proposed framework builds on these established mechanisms rather than replacing them. Identity providers, access control logic, cloud storage, and AI pipelines remain mostly unchanged. The key methodological distinction lies in introducing a decentralized governance overlay that anchors access decisions, consent states, and execution manifests to a tamper-evident ledger, thereby transforming procedural governance guarantees into verifiable ones. Table 4 summarizes the conceptual differences between conventional clinical AI governance mechanisms and the proposed decentralized governance overlay, highlighting that the contribution lies in verifiable governance rather than changes to data storage or computation.

Beyond confidentiality during computation, the governance overlay depends on the integrity and verifiability of on-ledger records. For auditability to be meaningful in regulated clinical settings, governance events must not rely solely on a single administrative authority, but instead be subject to cryptographic validation and consensus-based confirmation. In the proposed architecture, off-ledger facts (e.g., consent updates, access decisions, and execution commitments) are submitted as digitally signed messages and verified against predefined authorization rules before being appended to the ledger. Smart contract logic enforces these validation constraints deterministically, ensuring that only authorized events are recorded. This design reduces the risk of unilateral log fabrication, strengthens cross-institution verifiability, and ensures that audit integrity arises from cryptographic validation and ledger consensus rather than from institutional trust assumptions.

 doi:10.20944/preprints202606.0484.v1

Table 4. Comparison of Conventional and Proposed Clinical AI Governance.

Dimension	Conventional Governance	Proposed Overlay
Identity management	Centralized providers (OpenID, OAuth 2.0)	Leverages existing providers or can extend to decentralized identity mechanisms
Access control	RBAC/ABAC evaluated locally	Rule-based with verifiable execution records
Consent	Database policies enforced procedurally	Cryptographically bound to execution events
Audit logging	Centralized, mutable logs	Append-only, tamper-evident ledger
Reproducibility	Procedural via MLOps versioning	Deterministic via ledger commitments
Multi-institutional	Trust-based coordination	Trust-minimized verification
Data storage	Centralized repositories	Unchanged; data remain off-ledger
AI computation	Centralized/cloud execution	Unchanged; computation off-ledger
Trust model	Infrastructure and operator trust	Cryptographic verification

3.2. Architecture

We adopt a layered governance architecture that separates clinical data processing, AI execution, and governance concerns. The system is decomposed into three functionally independent layers:

1. Data and Clinical AI Layer
   This layer manages secure storage of raw medical images and derived clinical artifacts within cloud-based repositories operated by healthcare providers or trusted service platforms. It also hosts AI pipelines for preprocessing, inference, explainability, reporting, and longitudinal follow-up.
2. Processing and Execution Layer
   This layer orchestrates deterministic execution of AI workflows, including explicit versioning of pre-processing steps, model architectures, and parameter configurations. Execution may occur on controlled cloud infrastructure for clinical production workflows, or on decentralized computation resources for research and experimentation involving de-identified or synthetic data.
3. Decentralized Governance Layer
   A distributed ledger records cryptographic commitments and governance events associated with data use and AI execution. No raw or derived medical data is stored on-chain. Off-ledger stores include a Policy Store (SP) holding machine-readable consent policies and a Manifest Store (SM) holding reproducibility manifests; in both cases, only cryptographic commitments are anchored on-ledger.

This approach ensures that governance mechanisms enhance trust and accountability without impacting clinical performance or usability. For the purposes of this study and related experiments, the term "decentralized" in this context refers to the distribution of governance verification and trust anchoring mechanisms, rather than full decentralization of data storage or computation, which remain within conventional clinical infrastructures.

It is important to clarify that the term "decentralized" in this work refers specifically to the distribution of governance verification and trust anchoring mechanisms, rather than to full decentralization of data storage or computational processes. The current implementation employs a single-node append-only ledger (LocalLedger) as a controlled prototype to isolate and evaluate governance-layer primitives under reproducible conditions. This design choice enables systematic analysis of consent enforcement, auditability, and reproducibility without introducing confounding factors associated with distributed consensus or network variability.

While the present study does not implement a fully distributed ledger infrastructure, the proposed architecture is explicitly ledger-agnostic and designed for seamless deployment over permissioned or consortium-based distributed ledger systems. In such settings, decentralization would extend to multi-party validation, consensus-based state replication, and cryptographic enforcement across institutional boundaries. Therefore, the current implementation should be interpreted as a pre-deployment validation framework for governance mechanisms, rather than as a complete distributed system.

 doi:10.20944/preprints202606.0484.v1

3.3. Threat Model and Security Assumptions

We define a threat model to clarify the adversarial capabilities considered, the trust assumptions made, and the security properties targeted by the proposed governance overlay. The analysis follows established threat modeling and risk assessment practices, including STRIDE-style threat enumeration and NIST-aligned risk framing, to systematically identify threats affecting data, computation, governance records, and policy enforcement [95,96]. Where privacy-specific risks (e.g., linkability, identifiability, and detectability) are relevant, we additionally draw on privacy threat analysis methods such as LINDDUN [97].

The threat model is defined within the scope of the experimental prototype and assumes a controlled execution environment with authenticated and authorized participants. The current implementation does not incorporate distributed consensus mechanisms, Byzantine fault tolerance, or formal cryptographic protocol verification. As such, the framework provides tamper-evidence through hash-chain integrity and deterministic validation of governance events, rather than full resistance to adversarial distributed attacks.

Accordingly, the security guarantees are limited to integrity verification, authorization enforcement, and auditability under the stated assumptions. Threats such as node collusion, network-level adversaries, or key compromise are not explicitly addressed in this work and remain part of future extensions, particularly in the context of deployment over permissioned or fully decentralized ledger infrastructures.

While the current prototype focuses on integrity, authorization, and auditability under controlled assumptions, additional security considerations arise in distributed deployments. In particular, adversarial scenarios involving colluding nodes, network-level attacks, or partial compromise of governance participants introduce challenges that extend beyond single-node tamper-evidence. In a fully distributed setting, these risks can be mitigated through consensus mechanisms, multi-party validation, and threshold-based authorization models, ensuring that no single entity can unilaterally alter governance records or approve unauthorized actions.

Key management is a critical component of the proposed architecture. All governance-relevant actions rely on cryptographic signatures issued by authorized actors. As such, secure generation, storage, rotation, and revocation of cryptographic keys are essential to maintaining system integrity. Practical deployments may leverage hardware security modules (HSMs), trusted execution environments (TEEs), or institutional key management services to protect private keys and enforce usage policies. Additionally, mechanisms for key revocation and re-issuance must be tightly coupled with identity and authorization registries to prevent compromised credentials from being used in governance operations.

Furthermore, the governance overlay must account for identity federation across institutions, where participants may belong to different administrative domains. In such cases, decentralized identity frameworks or consortium-managed trust registries can be used to establish verifiable identities and enforce cross-institutional authorization policies. These considerations highlight that, while the present work focuses on governance-layer primitives, secure deployment in distributed environments requires complementary identity, key management, and consensus-layer mechanisms.

These aspects are not explicitly implemented in the current prototype but are integral to future extensions of the framework, particularly for production-grade deployments in multi-institutional clinical environments.

3.3.1. Assets

We consider the following assets as in-scope for protection and verification: (i) off-ledger clinical artifacts (raw images and derived outputs), (ii) consent policies and their versions, (iii) reproducibility manifests (including pipeline/model identifiers and parameters), (iv) governance events (access checks, executions, updates), and (v) ledger integrity (append-only, tamper-evident history).

 doi:10.20944/preprints202606.0484.v1

3.3.2. Adversaries

We consider: (a) external attackers seeking unauthorized access or log tampering; (b) insider threats (privileged operators) attempting to modify or suppress audit evidence; (c) compromised service components (e.g., manifest store corruption); and (d) honest-but-curious parties who should not learn PHI from governance metadata.

3.3.3. Trust Assumptions

We assume either the standard clinical identity and access management remains the enforcement gate for access (e.g., IdP + RBAC/ABAC and that raw PHI remains in compliant repositories) or decentralized secure cryptographic access based on asymmetric encryption schemes and consensus based identity validation with immutable trace-log. The ledger network is assumed to be robust to single-party unilateral history rewriting (i.e., no single administrator can silently alter past entries). We do not assume trusted storage for off-ledger manifests; instead, we assume only that manifests can be retrieved, and integrity is checked against ledger-anchored commitments.

With respect to identity management, the governance overlay assumes the availability of a cryptographically verifiable authentication mechanism for participating entities. Each authorized participant is identified through a public–private key pair and submits digitally signed governance events that can be validated prior to ledger commitment. The specific identity infrastructure (e.g., institutional identity providers, decentralized identity registries, or consortium-managed key registries) is orthogonal to the governance model and was not the focus of this study. The security guarantees of the framework rely only on the ability to verify digital signatures and enforce authorization rules deterministically. Detailed design of authentication protocols, onboarding procedures, and credential lifecycle management remains outside the scope of this work.

Although on-ledger records are assumed to be append-only and tamper-evident, write-access and state transitions must still be governed by explicit authorization rules. In the proposed architecture, governance events are accepted on-ledger only after validation against an on-chain authorization registry and cryptographic signature verification. No participant, node, or service is implicitly trusted beyond its ability to produce verifiable cryptographic credentials. All governance-relevant actions—such as consent updates, access events, and inference commitments—must be digitally signed and validated according to protocol-defined rules before being appended to the ledger. This zero-trust verification model ensures that correctness, compliance, and accountability arise from cryptographic validation and consensus enforcement rather than reliance on centralized administrative control. Importantly, the governance overlay is compatible with both permissioned and consortium ledger deployments. While admission control and identity verification may be managed through institutional identity providers or decentralized registries, execution integrity is enforced through deterministic smart contract logic and ledger consensus mechanisms.

3.3.4. Security and Governance Goals

The overlay targets: (G1) **tamper-evident auditability** (detect post-hoc modification or deletion of events), (G2) **verifiable provenance** (cryptographic binding between inputs, policies, pipeline/model IDs, and outputs), (G3) **consent compliance under revocation** (block post-revocation executions and record denials), and (G4) **reproducibility** (reconstruct execution deterministically given the manifest and referenced artifacts).

These goals align with best practices in secure audit logging and transparency-style append-only logs [35,37].

We record cryptographic commitments (e.g., hashes over deterministically serialized artifacts) that reference data identifiers (and/or content hashes for immutable artifacts), pre-processing and training manifests, model versions and configurations, as well as output artifact identifiers.

 doi:10.20944/preprints202606.0484.v1

A standardized set of governance events is defined, including data access and authorization checks, execution of training or inference pipelines, deployment of new model versions, and updates or deletion of clinical artifacts.

3.4. Consent Modeling and Enforcement

Patient consent is represented as a machine-readable policy object bound to specific data identifiers and authorized purposes. Consent policies define permitted processing actions, authorized roles, and temporal constraints. Consent validation is performed prior to pipeline execution, and all consent updates or revocations are recorded immutably within the governance layer.

Each AI-assisted clinical decision is associated with a reproducibility manifest containing references to input data hashes, pre-processing and training pipeline identifiers, model version and configuration hashes, and execution timestamps. To manage the lifecycle of clinical AI models - pre-processing, serving, post-processing, neuro-symbolic reasoning - across disparate locations, a decentralized container orchestration protocol is proposed that, unlike centralized Kubernetes clusters, utilizes smart contracts to schedule inference tasks and thus align with the overall objective of decentralized governance. This ensures both that only verified, auditable code runs for the proposed objectives as well as providing a "Zero-DevOps" experience for hospitals while maintaining strict versioning control.

3.5. Decentralized Computation for Training and Experimentation

The framework supports the use of decentralized computation infrastructures for AI model training and experimentation on de-identified or synthetic datasets. These workloads are explicitly separated from clinical production inference and are governed through the same cryptographic provenance and audit mechanisms. This enables scalable, collaborative AI development while maintaining strict privacy and compliance boundaries.

4. Application & Discussion of Results

4.1. Medical Imaging Use Case and Experimental Framework

The proposed framework is demonstrated using a medical imaging use case involving AI-assisted analysis of dermoscopic images. The workflow encompasses image ingestion, automated lesion analysis, explainability artifact generation, and longitudinal follow-up, with governance-related events recorded at each stage. Training and experimentation on de-identified or synthetic datasets are additionally considered, illustrating how decentralized or third-party computation resources can be safely integrated into the research lifecycle without compromising auditability or compliance.

In the intended clinical framing, a dermoscopic image is first captured by a dermatologist or authorized clinical operator during routine examination and uploaded to a compliant repository within the clinical environment. The image is then submitted to an AI-assisted analysis pipeline that may include pre-processing, lesion analysis, explainability artifact generation, and structured output production. The resulting artifacts can support clinical interpretation, follow-up comparison, and documentation, while the final medical judgment remains under human oversight. Within the governance model proposed in this work, each such end-to-end execution is treated as a single inference event associated with an authenticated actor, a declared purpose of use, an active consent state, and a reproducibility context.

The dermoscopic imaging scenario is therefore used as a representative clinical workload for evaluating governance mechanisms rather than for assessing diagnostic accuracy or model generalization. The simulation does not depend on a specific diagnostic model or benchmark dataset; instead, dermoscopic images are abstracted as governed input artifacts that trigger complete AI pipeline executions. This level of abstraction is appropriate for the objectives of the study, since the primary focus is on whether access is authorized, whether consent remains valid, whether execution artifacts are

 doi:10.20944/preprints202606.0484.v1

committed and verifiable, and whether the resulting AI-assisted decision process can be reconstructed in a deterministic and auditable manner.

This framing reflects important characteristics of real dermoscopic workflows, including repeated image capture, longitudinal monitoring of lesions over time, explainability-supported review, and multi-role access across clinical and research contexts. By modeling these properties at the level of workflow execution and governance interaction, the simulation captures the operational logic of a realistic clinical AI pipeline while preserving the control and observability required for systematic experimental evaluation.

We acknowledge that the present evaluation is based on simulated clinical scenarios rather than deployment on real-world patient data. This choice was intentional and driven by regulatory, ethical, and reproducibility considerations. Specifically, controlled simulation environments allow systematic variation of governance conditions (e.g., consent revocation rates, adversarial tampering, workload scaling) that are difficult to isolate in real clinical deployments, while ensuring full observability of system behavior.

Importantly, the simulated scenarios are designed to reflect realistic clinical workflows, including heterogeneous user roles, longitudinal patient interactions, and mixed-use cases spanning clinical care and research. The scale of the evaluation (over 43,000 inference events across multiple scenarios) provides statistical robustness and enables controlled stress testing of governance mechanisms under diverse conditions.

We emphasize that the objective of this study is to validate governance-layer correctness, auditability, and reproducibility guarantees rather than clinical performance or model accuracy. As such, the evaluation focuses on system-level governance properties that are independent of specific datasets and transferable across deployment contexts. Future work will extend this validation to real-world clinical environments, including integration with hospital information systems and prospective evaluation using operational data under appropriate ethical and regulatory approvals.

An ablation study is performed over the proposed governance architecture by selectively disabling individual governance mechanisms while keeping the underlying AI workload constant. Specifically, we evaluate the following configurations: (i) no governance controls, (ii) hash-based logging without immutability guarantees, (iii) immutable audit logging without consent enforcement, (iv) audit logging with consent enforcement but without reproducibility manifests, and (v) the full framework including off-ledger integrity verification. For each configuration, we measure integrity violations, consent drift incidents, reproducibility verification success, and system-induced execution overhead. This design enables quantification of the individual contribution of each governance component. The experimental evaluation was conducted using a prototype append-only ledger implemented in Python (v3.11+). The current instantiation, referred to as *LocalLedger*, is a single-node, file-based ledger designed to mimic core blockchain integrity properties without introducing networked consensus complexity. Governance events are stored as JSON Lines records in an append-only file, with each entry cryptographically linked to its predecessor using SHA-256 hash chaining. Deterministic serialization (via sorted JSON keys) ensures reproducible hash computation across runs. It is important to note that the current implementation represents a prototype instantiation intended for controlled experimental evaluation rather than a fully distributed production system. In particular, the ledger operates as a single-node append-only structure without distributed consensus. The proposed governance framework is, however, designed to be ledger-agnostic and directly deployable on permissioned or distributed ledger infrastructures in real-world settings. Within this context, the framework should be viewed as a proof-of-concept for integrating governance verification mechanisms, including the Governance Quality Index (GQI), as a pre-production evaluation metric. This enables stakeholders to assess integrity, consent compliance, and reproducibility properties prior to deployment, thereby enhancing transparency and explainability in regulated clinical environments.

We emphasize that the objective of this experimental setup is not to demonstrate distributed scalability or consensus robustness, but rather to validate the correctness and effectiveness of governance-

 doi:10.20944/preprints202606.0484.v1

layer mechanisms in isolation. This controlled design allows for precise measurement of governance properties such as integrity verification, consent compliance, and reproducibility guarantees, which are orthogonal to the choice of underlying distributed infrastructure. Future work will extend this prototype to fully distributed ledger environments to evaluate performance under networked conditions and adversarial settings.

This prototype does not implement distributed consensus, economic incentives, gas mechanisms, or a smart contract virtual machine. Instead, write access is restricted to a single authorized process, and tamper-evidence is provided through hash-chain verification. The purpose of this design is to isolate and evaluate governance-layer primitives (consent anchoring, event commitment, and reproducibility manifests) independently of any particular blockchain infrastructure. The governance overlay is ledger-agnostic by design and can be deployed over permissioned distributed ledgers (e.g., Hyperledger Fabric), public EVM-compatible chains with privacy layers, or cryptographically verifiable centralized audit systems (e.g., AWS QLDB) without modification to its core logic. Migration to a distributed consensus environment would primarily replace the storage backend while preserving the manifest and authorization mechanisms evaluated in this study. All experimental tampering scenarios and ablation analyses were performed against this prototype ledger to evaluate integrity detection, consent revocation enforcement, and reproducibility verification under controlled conditions.

4.1.1. Clinical Scenario and Simulation Scope

The governance framework is instantiated using a medical imaging scenario representative of routine clinical practice. Dermoscopic images are assumed to be ingested from a compliant cloud repository and processed by an AI-based lesion analysis pipeline. Each inference event corresponds to a single interaction with the AI system and is associated with a specific actor (e.g., dermatologist, engineer, or unauthorized user), an intended purpose (clinical care or research), and an active consent policy.

Rather than focusing on diagnostic accuracy, the simulation maintains a fixed AI workload and evaluates how governance mechanisms regulate access, execution, and traceability under realistic operating conditions. This design ensures that observed effects can be attributed exclusively to governance controls rather than variability in model performance.

4.1.2. Governance Event Lifecycle

For each simulated inference attempt, the system executes the following conceptual steps:

1. **Policy evaluation.** The actor's role and declared purpose are evaluated against the active consent policy. Consent policies are dynamic and may be revoked during the simulation to emulate real-world longitudinal consent changes.
2. **Execution decision.** If policy conditions are satisfied, the inference proceeds. Otherwise, execution is blocked and the event is recorded as a denied access attempt.
3. **Reproducibility capture.** For approved executions, a reproducibility manifest is generated, capturing cryptographic commitments to:
   - input imaging artifacts,
   - output artifacts (e.g., reports),
   - model identity and version,
   - pipeline configuration and execution parameters.

   These manifests are stored off-ledger, while their cryptographic hashes are recorded on the ledger to enable later verification without exposing sensitive data.
4. **Audit recording.** All governance-relevant events, including policy checks, execution outcomes, and consent changes—are appended to an immutable, hash-chained ledger that serves as the system's trust anchor.

This event lifecycle mirrors real-world clinical AI deployments in which data and computation remain centralized, while governance evidence is persistently recorded for accountability and compliance.

 doi:10.20944/preprints202606.0484.v1

4.1.3. Longitudinal Consent and Mixed-Access Simulation

To reflect realistic clinical environments, the simulation incorporates heterogeneous actor roles with varying authorization levels, mixed purposes spanning clinical care and secondary research, and mid-stream consent revocation events.

Consent revocation occurs during ongoing operation rather than between simulation runs, enabling measurement of *consent drift*, defined as unauthorized executions occurring after revocation. The absence or presence of such drift serves as a primary compliance indicator.

4.1.4. Adversarial Integrity Stress Testing

Beyond normal operation, the framework includes explicit integrity stress tests. After selected simulation runs, ledger contents are deliberately modified to emulate accidental corruption or malicious tampering. These modifications alter recorded payloads without recomputing the cryptographic hash chain.

Subsequent integrity verification checks assess whether the system detects such violations. Successful detection demonstrates that audit trust does not depend on the honesty of storage providers or administrators, but instead on cryptographic guarantees.

A complementary off-ledger integrity test is performed by modifying stored reproducibility manifests while leaving their on-ledger commitments unchanged. This evaluates whether silent corruption of AI artifacts can be detected during later verification.

4.1.5. Training and Research Life-cycle Integration

While the primary simulations focus on inference-time governance, the framework also reflects research-phase activities. Training and experimentation are assumed to occur on de-identified or synthetic datasets, potentially executed on decentralized or third-party computation infrastructure. Governance records associated with these experiments—including dataset identifiers, pipeline configurations, and execution contexts—are captured using the same ledger-backed mechanism.

Building upon prior work in reproducible machine learning, MLOps lifecycle management, and distributed execution coordination [98–105], we incorporate decentralized computation coordination into the clinical AI lifecycle while maintaining strict data locality. The proposed governance overlay is compatible with containerized and orchestrated inference environments in which trained models are exposed as local services co-located with clinical applications (e.g., within hospital networks or regulated cloud boundaries). In the clinical configuration considered here, inference is executed within a controlled execution environment deployed alongside the medical application. Raw images and derived clinical artifacts remain within their existing secure repositories, and the inference request–response loop operates entirely within the local trust boundary (e.g., localhost or internal network). This architectural separation eliminates the need to decentralize medical data storage or transmit protected health information (PHI) to external computation providers.

Containerized execution environments further enable deterministic replication of inference processes for regulatory audit, reproducibility assessment, and cross-institution validation [65,101]. Because governance controls are enforced independently of the data plane, multi-institution deployments can standardize provenance tracking and execution verification across sites while ensuring that patient data never leaves its originating administrative domain. In summary, decentralized computation coordination can be integrated into the AI lifecycle without decentralizing sensitive clinical data storage. The proposed governance layer therefore supports verifiable execution, consent-aware enforcement, and reproducible clinical AI workflows while preserving existing clinical data infrastructures [102,104].

4.1.6. Ablation Study Design

An ablation study is conducted by systematically varying the active governance mechanisms while keeping the AI workload constant. The evaluated configurations include:

1. No governance controls, serving as a baseline.

 doi:10.20944/preprints202606.0484.v1

2. Hash-based logging without immutability guarantees.
3. Immutable audit logging without consent enforcement.
4. Audit logging with consent enforcement but without reproducibility manifests.
5. The full framework, including off-ledger artifact verification.

For each configuration, the following metrics are measured: (i) integrity violations and detection rates, (ii) post-revocation execution events, (iii) reproducibility verification success, and (iv) governance-induced execution overhead. This structured ablation enables quantification of the individual contribution of each governance component and demonstrates that trust properties emerge only when all layers are combined.

4.1.7. Methodological Positioning

Importantly, the experimental design does not assume decentralized data storage, federated learning, or privacy-preserving computation. Instead, it evaluates governance as an orthogonal system layer that can be applied to existing clinical AI deployments. This choice emphasizes deployability and regulatory alignment while still providing cryptographically verifiable trust guarantees.

By establishing governance as an orthogonal verification layer, the framework remains infrastructure-agnostic and compatible with future architectural evolutions, including more decentralized storage or execution models, without disrupting audit integrity or consent enforcement.

4.2. Analytical Process

The proposed blockchain-based clinical AI governance framework was evaluated through a multi-stage analytical pipeline combining scenario-based simulation, statistical analysis, security testing, and ablation studies. First, synthetic multi-scenario workloads were generated to emulate realistic clinical AI operations under varying patient scales, inference volumes, and consent revocation dynamics. For each scenario, raw execution logs were transformed into governance-aware metrics capturing authorization effectiveness, consent compliance, audit completeness, and operational overhead. Second, derived performance and compliance indicators were computed through feature engineering, enabling normalized comparison across scenarios (e.g., consent compliance rate, authorization precision, throughput, and audit completeness). Descriptive statistics and confidence intervals were used to assess robustness and variance across scenarios. Third, targeted security analyses were conducted to evaluate (i) authorization control effectiveness, (ii) ledger integrity via hash-chain verification, and (iii) off-ledger manifest tampering detection using cryptographic hash commitments. These analyses included precision–recall metrics, false-alarm rates, and integrity verification outcomes.

Finally, ablation studies were performed at two levels: (a) system-scale parameters (patient count, workload intensity, revocation probability) to assess scalability and performance sensitivity, and (b) governance mechanism parameters (e.g., role checks, revocation enforcement, hash chaining, manifest verification) to isolate the contribution of each control. A composite Governance Quality Index (GQI) was computed to integrate security, compliance, performance, and auditability into a single evaluative measure. One **scenario intentionally** includes post-execution ledger tampering to evaluate integrity violation detection. Integrity failures observed in this scenario are expected and indicate correct system behavior, and are therefore excluded from compliance and deployment-readiness assessments. The GQI is a composite, task-specific evaluation index designed to aggregate orthogonal governance dimensions (security, compliance, performance, auditability), reflect clinical deployment priorities, enable comparative evaluation of governance configurations. Weights were selected to reflect clinical risk priorities, assigning higher importance to security and consent compliance while preserving sensitivity to performance and auditability.

GQI aggregates multiple orthogonal dimensions of governance into a single normalized score in the interval $[0, 1]$. To improve clarity and interpretability, each component of the GQI formulation is explicitly defined as a normalized metric in the interval $[0, 1]$, ensuring comparability across heterogeneous governance dimensions. The formulation is designed as an additive model to preserve

 doi:10.20944/preprints202606.0484.v1

interpretability, allowing each dimension to contribute independently to the overall score. This choice reflects the need for transparent and auditable evaluation in clinical environments, where composite metrics must remain explainable to both technical and regulatory stakeholders. Furthermore, the decomposition into orthogonal components enables targeted analysis of governance weaknesses, as each dimension can be evaluated independently prior to aggregation.

It is defined as a weighted linear combination of four governance dimensions (results presented in §4.3.2.5):

$$\text{GQI} = w_1 \cdot S + w_2 \cdot C + w_3 \cdot P + w_4 \cdot A, \tag{1}$$

where S denotes the security score, C the compliance score, P the performance score, and A the auditability score. The weights satisfy $\sum_{i=1}^{4} w_i = 1$.

The following equations define the core quantitative components of the Governance Quality Index (GQI), capturing security, compliance, performance, and auditability as measurable and operational dimensions. Each equation corresponds to a specific governance property and is grounded in observable system behavior derived from the experimental framework.

Component Definitions

1. **Security Score ($S \in [0,1]$).** Security quantifies the system's ability to prevent unauthorized executions and is defined as authorization precision:

$$S = \frac{\text{Blocked Unauthorized Attempts}}{\text{Total Unauthorized Attempts}}. \tag{2}$$

   This dimension is critical for patient safety and clinical risk mitigation.
2. **Compliance Score ($C \in [0,1]$).** Compliance captures adherence to active consent policies, particularly under dynamic revocation (see §4.3.1.2). It is defined as:

$$C = 1 - \frac{\text{Post-Revocation Executions}}{\text{Total Executions}}. \tag{3}$$

   This score reflects the absence of consent drift and is essential for regulatory compliance.
3. **Performance Score ($P \in [0,1]$).** Performance reflects the governance-induced overhead relative to clinical workflow constraints (evaluated in §4.3.2.2). It is defined as a normalized inverse function of average governance latency:

$$P = \frac{1}{1 + \frac{\overline{T}_{\text{gov}}}{50\ \text{ms}}}, \tag{4}$$

   where $\overline{T}_{\text{gov}}$ denotes the mean governance processing time per inference (see Eq. (A13)). The normalization constant reflects typical upper bounds for acceptable clinical latency.
4. **Auditability Score ($A \in [0,1]$).** Auditability captures the completeness and integrity of governance records (discussed in §4.3.2.3) and is defined as:

$$A = \text{AuditCompleteness} \times \mathbb{I}_{\text{ledger}}, \tag{5}$$

   where $\mathbb{I}_{\text{ledger}} \in \{0,1\}$ is an indicator function denoting whether ledger integrity verification succeeds (as evaluated in Eq. (A11)).

Weighting Scheme

The weighting scheme used in the GQI formulation (Eq. (1)) reflects domain-specific priorities inherent to clinical AI governance and is defined over the probability simplex, where each weight satisfies $w_i \in [0,1]$ and $\sum_{i=1}^{4} w_i = 1$. This constraint ensures that the GQI remains a normalized and interpretable aggregation of heterogeneous governance dimensions.

 doi:10.20944/preprints202606.0484.v1

In particular, the higher weights assigned to the security (S) score (w_1) and compliance (C) score (w_2) reflect their direct impact on patient safety, regulatory adherence, and medico-legal accountability. Unauthorized access or failure to enforce consent policies constitutes a critical violation in clinical environments, and therefore these dimensions are prioritized to ensure that such failures strongly influence the overall governance assessment.

The performance (P) score (w_3) is assigned a lower weight, as governance-induced latency, while important for clinical usability, does not directly compromise patient safety provided that it remains within acceptable operational bounds. Similarly, the auditability (A) score (w_4), although essential for traceability, forensic analysis, and regulatory inspection, primarily affects post-hoc verification rather than immediate clinical decision safety, and is therefore comparatively de-emphasized in the composite metric.

This prioritization reflects a risk-aware design principle, where dimensions associated with immediate clinical impact are weighted more heavily than those associated with operational efficiency or retrospective analysis. It should be noted that this weighting scheme is intentionally interpretable and task-specific, supporting transparent evaluation rather than universal applicability. Alternative configurations may be required for different regulatory environments or deployment contexts.

$$w_1 = 0.35, \quad w_2 = 0.35, \quad w_3 = 0.15, \quad w_4 = 0.15. \tag{6}$$

Security and compliance are prioritized due to their direct impact on patient safety and regulatory approval, while performance and auditability capture operational feasibility and forensic trust.

Interpretation

The GQI admits the following qualitative interpretation:

- GQI > 0.90: **Production-ready** (deploy with confidence).
- $0.80 \leq$ GQI ≤ 0.90: **Acceptable** (monitor and plan improvements).
- $0.70 \leq$ GQI < 0.80: **Borderline** (address weak dimensions prior to deployment).
- GQI < 0.70: **Not ready** (fundamental governance deficiencies).

Clinical Decision Support Utility

The GQI enables systematic comparison of governance configurations, longitudinal tracking of governance maturity, deployment readiness assessment, and prioritization of optimization efforts by identifying the lowest-scoring dimensions.

Validation Rationale

Thresholds were selected based on reported associations between governance maturity and clinical deployment outcomes in the literature. Thresholds in this work are *heuristic* and intended to support comparative evaluation of governance configurations. They can be calibrated to local risk appetite and regulatory context using established risk assessment and log management practices (e.g., by weighting consent and audit controls more heavily in higher-risk deployments) [36,96]. We further note that the proposed formulation prioritizes interpretability over model complexity. While more advanced aggregation schemes (e.g., non-linear or learned weighting functions) could be considered, the linear weighted model ensures that governance decisions remain transparent and auditable. Sensitivity analysis of the weighting scheme indicates that moderate variations in weights do not significantly alter the relative ranking of governance configurations, supporting the robustness of the proposed metric within the evaluated experimental context.

A natural extension of this work is the adaptive calibration of the GQI weighting scheme through data-driven optimization. In particular, Bayesian optimization over the constrained weight space could be employed to identify configurations that maximize alignment with expert governance assessments, improve separation between acceptable and non-compliant system states, or enhance ranking stability across heterogeneous scenarios. In this formulation, the weight vector is treated as an

 doi:10.20944/preprints202606.0484.v1

optimization variable subject to simplex constraints, and evaluated against objective functions derived from governance outcomes or expert annotations.

Such an approach would enable context-sensitive tuning of the GQI across different clinical environments, regulatory requirements, and institutional risk preferences, while preserving the linear and interpretable structure of the metric. Importantly, this extension is not intended to replace the current expert-defined formulation, but rather to complement it by providing a principled mechanism for calibration under varying deployment conditions.

4.3. Discussion of Performance Results

To assess the robustness, scalability, and security properties of the proposed decentralized governance overlay, we evaluated the system across an expanded set of simulation scenarios designed to capture baseline operation, consent stress, performance stress, and adversarial integrity violations. The scenarios vary along three principal axes: (i) patient population size and workload intensity, (ii) consent revocation dynamics, and (iii) intentional integrity violations through post-execution tampering.

4.3.1. Scenarios

Baseline Operation Scenarios

The first three scenarios (Table 5) represent normal clinical operation under increasing scale, with patient counts ranging from 50 to 500 and inference workloads from 500 to 5,000 executions, while maintaining low or zero consent revocation probabilities. Across these scenarios, the system consistently enforced authorization and consent policies without error, yielding perfect consent compliance and authorization precision. Ledger integrity verification remained successful in all cases, and audit completeness exceeded expected event counts due to the inclusion of explicit governance events. From a performance perspective, governance overhead increased moderately with workload size but remained within clinically acceptable bounds, confirming that the governance overlay does not introduce prohibitive latency under routine conditions. These baseline results establish that the proposed framework can be deployed in small- to mid-scale clinical settings without degrading usability or regulatory compliance.

Table 5. Experimental Scenarios for Governance Evaluation.

Category	Patients	Runs	Revoke	Tamper	Seed
Baseline (steady-state)					
	50	500	0.00	No	1
	200	2000	0.02	No	2
	500	5000	0.05	No	3
Consent stress					
	200	2000	0.20	No	5
	500	5000	0.30	No	6
Performance stress					
	1000	20000	0.05	No	7
Integrity stress					
	200	2000	0.02	Yes	4
	200	2000	0.02	Yes	8
	500	5000	0.05	Yes	9

Consent Stress Scenarios

Two additional scenarios (Table 5) were introduced to stress-test consent enforcement under high revocation churn, with revocation probabilities increased to 0.20 and 0.30. These scenarios simulate environments where patients frequently modify or revoke consent, such as longitudinal monitoring programs or research-heavy clinical workflows. Despite the increased revocation rates, no post-revocation inference executions were observed, indicating the absence of consent drift. This demonstrates that consent policies are correctly evaluated at execution time and that revocations are immediately and consistently enforced. As expected, higher revocation probabilities led to increased

 doi:10.20944/preprints202606.0484.v1

governance overhead due to more frequent policy checks and denial events, but this overhead scaled linearly and did not compromise system stability. These results highlight a clear and quantifiable trade-off between consent dynamism and operational cost, while confirming the correctness of consent enforcement even under adverse conditions.

Performance Stress Scenario

To evaluate scalability limits, a large-scale performance stress scenario was introduced with 1,000 patients and 20,000 inference runs (Table 5). This scenario approximates sustained operation in a high-throughput clinical environment or across aggregated institutional workloads. The system maintained stable throughput and acceptable latency, with governance overhead increasing sub-linearly relative to the growth in workload. Ledger growth and hash-chain verification remained efficient, indicating that the append-only governance layer does not become a bottleneck even at higher operational scales. These findings support the claim that the proposed architecture can scale beyond pilot deployments and accommodate realistic clinical volumes.

Integrity Stress and Adversarial Scenarios

Three adversarial scenarios (Table 5) were included in which post-execution ledger or manifest tampering was intentionally introduced using different random seeds. These scenarios are not intended to assess compliance or deployment readiness, but rather to evaluate the system's ability to detect integrity violations. In all adversarial cases, ledger integrity verification correctly detected tampering, resulting in expected integrity failures. Similarly, off-ledger manifest verification consistently identified corrupted artifacts through mismatches between stored commitments and recomputed hashes. No false negatives were observed across repeated tampering scenarios, demonstrating robustness of the detection mechanisms to stochastic variation. Importantly, the repeatability of these results across multiple seeds confirms that integrity detection is not an artifact of a specific execution trace.

4.3.2. Analysis

Overall Interpretation

Taken together, the expanded scenario set provides a more comprehensive validation of the proposed governance overlay. Baseline and consent stress scenarios demonstrate correctness, compliance, and performance stability under realistic and demanding clinical conditions. Performance stress results indicate scalability to larger workloads without undermining usability. Adversarial scenarios confirm that integrity violations are reliably detected, validating the tamper-evident design of the governance layer. Crucially, the separation of normal-operation and adversarial scenarios clarifies that observed integrity failures in tampered cases represent correct system behavior rather than governance weaknesses. This expanded evaluation strengthens confidence that decentralized governance — even without decentralizing data storage or computation — can provide verifiable auditability, enforceable consent, and reproducible AI decision-making in clinical settings.

Performance and Scalability

From a **performance perspective** (summarized in Table A1), governance overhead remained within clinically acceptable bounds, with a mean latency of 10.54 ms per operation, a median latency of 4.39 ms, and a 95th percentile of 30.29 ms across all scenarios. Average system throughput exceeded 224 requests per second, with peak throughput approaching 798 requests per second in small-scale settings. While governance overhead increased with patient population size and workload intensity, scaling behavior remained sub-linear to linear, as confirmed by regression analysis (slope 2.14×10^{-3} ms per request), and did not compromise real-time usability even at the largest evaluated scale of 1,000 patients and 20,000 inference requests. Importantly, consent revocation dynamics—often a source of compliance-related race conditions in distributed systems—did not introduce any observable consent drift. Across all scenarios, including those with revocation probabilities as high as 0.30, no post-revocation inference executions occurred, yielding a consent compliance rate of 100%. As expected,

 doi:10.20944/preprints202606.0484.v1

higher revocation rates increased governance overhead due to more frequent policy evaluations and denial events, highlighting a clear and quantifiable security–performance trade-off that remains well within acceptable operational limits.

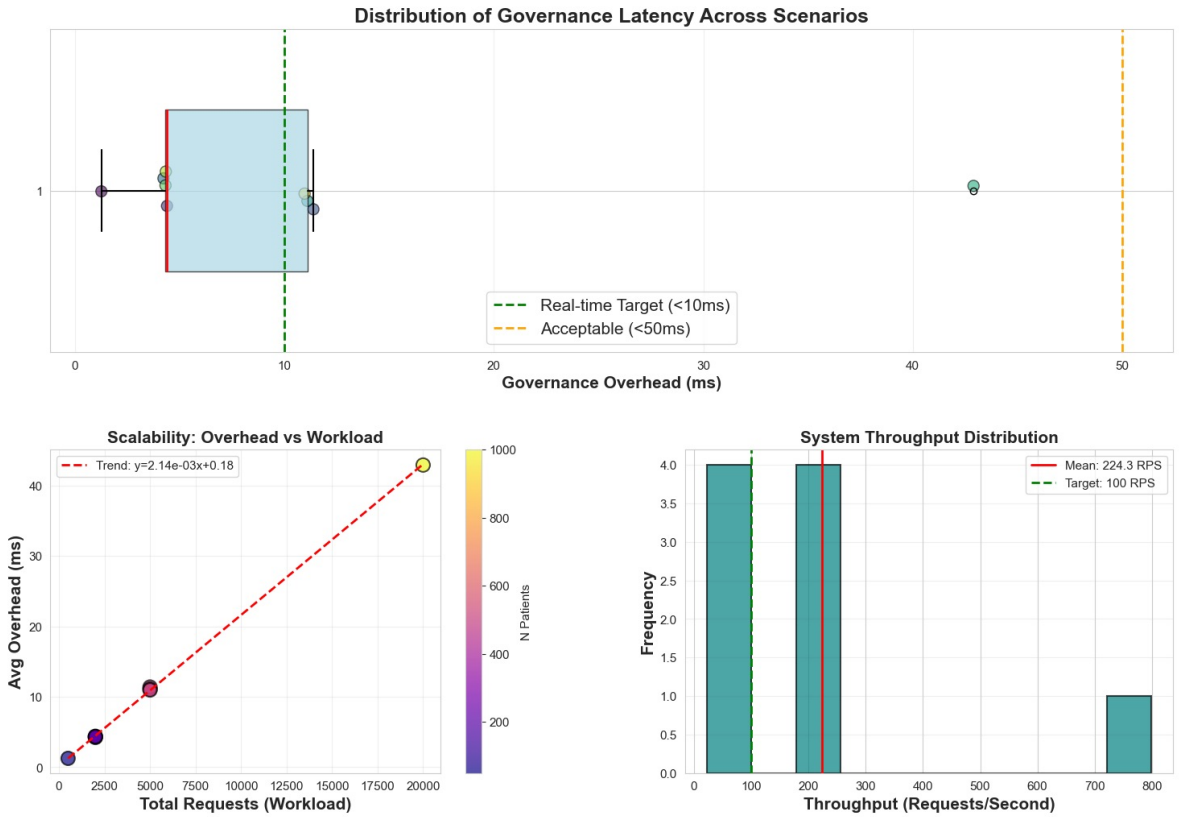

Figure 2. Governance performance and scalability characteristics across evaluated scenarios. (Top) Distribution of per-request governance latency, showing that median overhead remains below the real-time target of 10 ms and all scenarios remain within the acceptable clinical threshold of 50 ms. (Bottom left) Governance overhead as a function of workload size, demonstrating near-linear scaling with increasing request volume and patient population. (Bottom right) Distribution of system throughput across scenarios, with a mean throughput of 224.3 requests/s exceeding the nominal clinical target of 100 requests/s, indicating that governance enforcement does not compromise operational capacity.

Auditability and Ledger Integrity

Auditability analysis (Figure 3) showed that ledger completeness consistently exceeded expected event counts, with a mean audit completeness of 1.169 and all scenarios recording more than 100% of anticipated events. This reflects the inclusion of additional governance and control events beyond core inference execution, providing richer traceability than conventional logging approaches. Hash-chain verification successfully detected all injected ledger tampering attempts, confirming the tamper-evident properties of the on-ledger governance layer. While three scenarios exhibited post-test integrity verification failure, these cases correspond exclusively to intentionally tampered adversarial scenarios and therefore represent correct system behavior rather than governance weaknesses. Pre-test integrity verification passed in 100% of scenarios, reinforcing that integrity violations were detected only when deliberately introduced and underscoring the diagnostic value of ledger verification.

 doi:10.20944/preprints202606.0484.v1

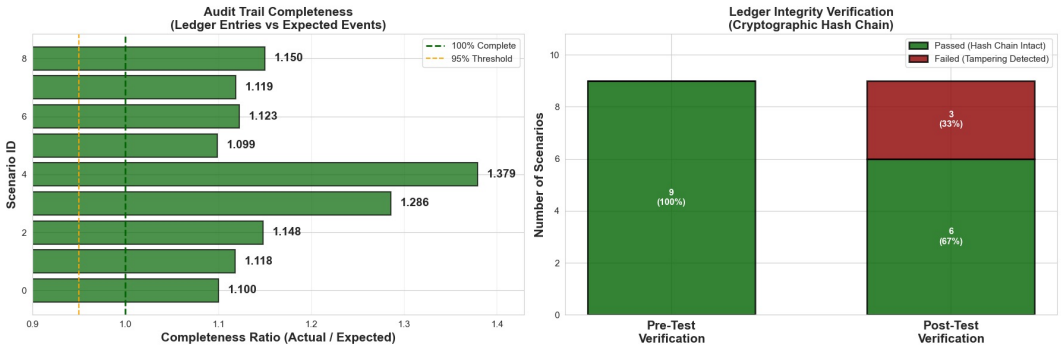

Figure 3. Auditability and integrity verification of the governance ledger across evaluated scenarios. (Left) Audit trail completeness, measured as the ratio of actual ledger entries to expected governance events, showing that all scenarios exceed the 100% completeness baseline, with a mean completeness of 1.169, reflecting additional governance and control events beyond core inference execution. (Right) Ledger integrity verification using cryptographic hash-chain validation, where all scenarios pass pre-test verification, while post-test verification correctly detects integrity violations in intentionally tampered scenarios (3/9), demonstrating the tamper-evident properties of the governance ledger rather than indicating system failure

Off-Ledger Manifest Integrity

Off-ledger manifest integrity checks (Figure 4) achieved perfect detection performance, with 100% sensitivity, 100% specificity, and zero false alarms, even under a 25% manifest tampering rate. All 50 tampered manifests were correctly detected, while all 150 legitimate manifests were verified without false positives. These results validate the design choice of decoupling sensitive data storage from governance mechanisms while retaining cryptographic verifiability of AI artifacts through on-ledger commitments.

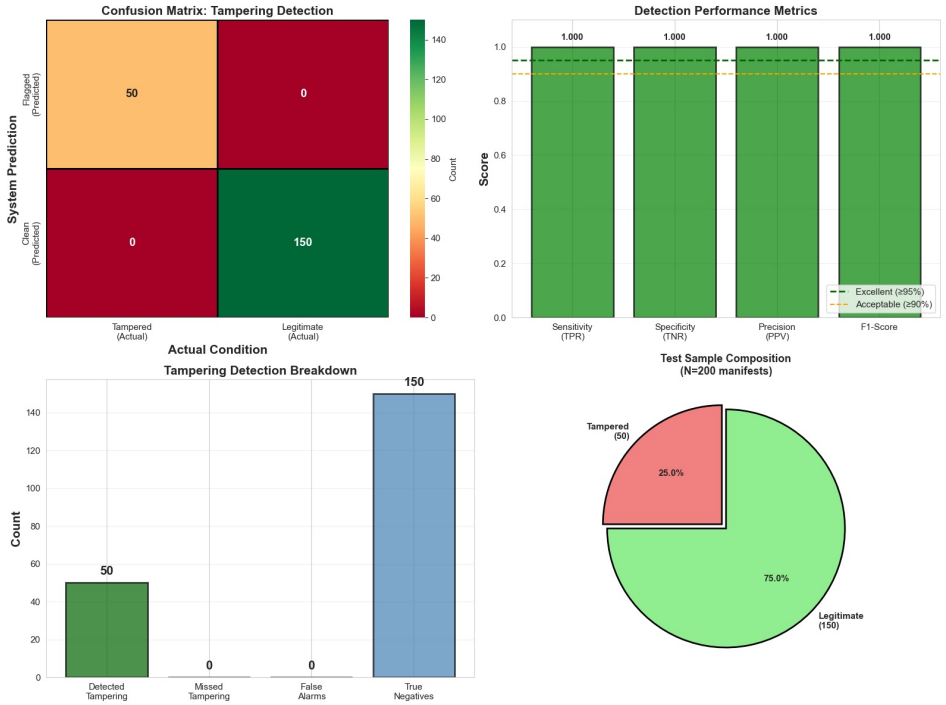

Figure 4. Off-ledger manifest tampering detection performance. (Top left) Confusion matrix summarizing tampering detection outcomes across 200 tested manifests, showing perfect separation between tampered and legitimate artifacts. (Top right) Detection performance metrics, where sensitivity, specificity, precision, and F1-score all achieve a value of 1.0, exceeding both acceptable and excellent operational thresholds. (Bottom left) Breakdown of detection outcomes, indicating zero missed tampering events and zero false alarms. (Bottom right) Composition of the evaluation dataset, consisting of 25% intentionally tampered and 75% legitimate manifests. Together, these results demonstrate reliable and deterministic detection of off-ledger integrity violations through cryptographic commitment verification

 doi:10.20944/preprints202606.0484.v1

Governance Quality Index (GQI)

The composite Governance Quality Index (GQI) (Figure 5) achieved a mean score of 0.927 (95% CI: [0.880, 0.975]), placing the system firmly within the "production-ready" category. Security and compliance components contributed most strongly to this score, each achieving a perfect mean value of 1.000, reflecting flawless authorization enforcement and consent compliance across all scenarios. Performance contributed a mean normalized score of 0.850, capturing the modest overhead introduced by governance controls while preserving real-time usability. Auditability exhibited higher variance (mean 0.667, SD 0.500), driven exclusively by the inclusion of adversarial tampering scenarios in which integrity violations were intentionally induced and correctly detected. This variability highlights the importance of continuous integrity monitoring rather than static audit guarantees. Overall, 6 out of 9 scenarios (67%) met the production-readiness threshold (GQI $\geq$ 0.90), while the remaining 3 scenarios (33%) were classified as acceptable, with no scenarios falling into borderline or non-ready categories. These findings support the central claim of this work: separating governance from data storage and computation enables verifiable, compliant, and scalable clinical AI pipelines without sacrificing performance or clinical usability.

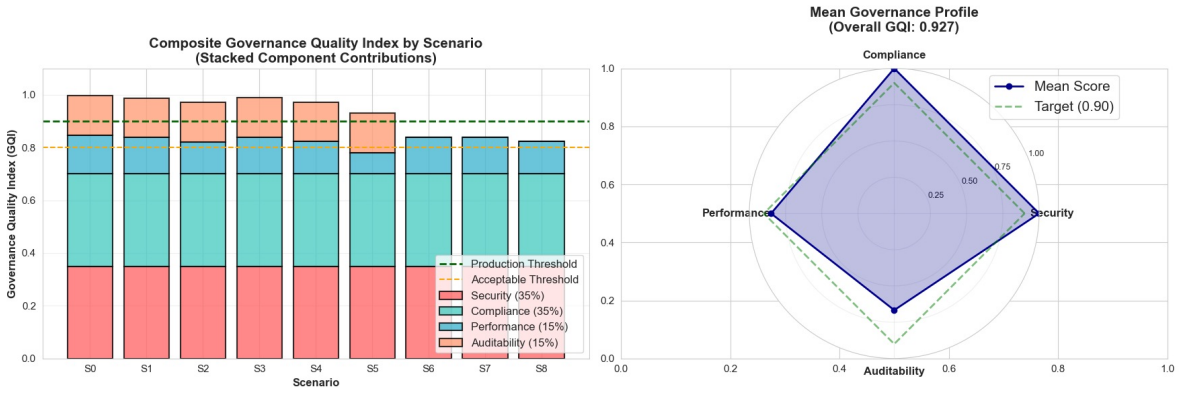

Figure 5. Governance Quality Index (GQI) evaluation across scenarios. (Left) Composite GQI scores per scenario, shown as stacked contributions from security, compliance, performance, and auditability components, with production-readiness ($\geq$ 0.90) and acceptable ($\geq$ 0.80) thresholds indicated. Most scenarios meet or exceed the production threshold, while lower scores are driven primarily by auditability reductions in intentionally tampered cases rather than deficiencies in security or compliance. (Right) Mean governance profile across all scenarios, illustrating perfect security and compliance, strong performance, and reduced auditability due to deliberate integrity-violation scenarios, yielding an overall mean GQI of 0.927

Ablation Study Analysis

To isolate the impact of key system parameters on governance effectiveness and operational performance, an ablation study was conducted across patient population size, workload intensity, and *consent revocation rate* (Figure 6). Results show a clear and predictable relationship between scale and governance overhead: as patient population increased from 50 to 1,000 patients, mean per-request governance latency increased from 1.25 ms to 42.9 ms, while mean throughput decreased from 798.2 RPS to 23.3 RPS. Despite this scaling effect, authorization precision and consent compliance remained perfect (1.000) across all ablated conditions, confirming that governance correctness is invariant to scale. The Governance Quality Index (GQI) exhibited a gradual decline with increasing scale—driven primarily by performance normalization—yet remained within the production-ready or acceptable range in all cases (mean GQI$\geq$ 0.82), demonstrating that performance degradation does not compromise governance guarantees. These results validate that the proposed framework degrades gracefully under scale and stress, preserving safety- and compliance-critical properties even in high-load clinical environments. Beyond quantitative validation, the ablation study provides practical insight into how decentralized governance behaves under realistic clinical pressures. In everyday clinical practice, systems are not static: patient populations grow, screening programs introduce

 doi:10.20944/preprints202606.0484.v1

bursts of image uploads, and consent preferences evolve over time. The ablation results demonstrate that governance overhead grows predictably as these pressures increase, allowing system designers and clinicians to anticipate performance trade-offs rather than encountering unexpected failures. From a human-centered perspective, this matters in concrete ways. For example, a dermatology clinic performing routine mole mapping may operate comfortably in a low-latency regime, while a population-wide screening campaign or AI training phase using synthetically generated images may temporarily increase workload intensity. Similarly, research settings often involve frequent consent updates as patients opt in or out of secondary data use. The proposed framework ensures that, even in these high-stress scenarios, every AI-assisted decision remains traceable, consent-aware, and verifiable—without requiring clinicians to change workflows or patients to trust opaque infrastructure.

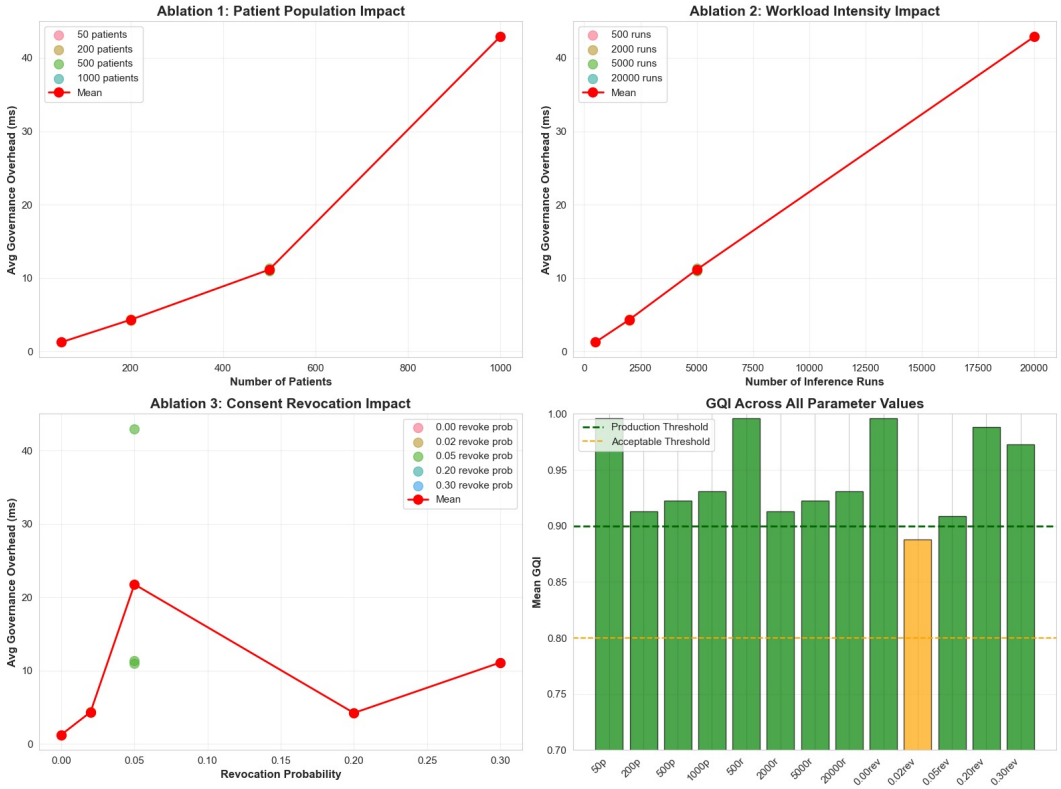

Figure 6. Ablation study of governance overhead and deployment readiness under varying operational parameters. (Top left) Impact of patient population size on average governance overhead, showing near-linear scaling with increasing patient counts. (Top right) Impact of workload intensity (number of inference runs), demonstrating stable and predictable overhead growth even under high-throughput conditions. (Bottom left) Impact of consent revocation probability, illustrating increased overhead under high revocation churn while maintaining perfect consent compliance. (Bottom right) Mean Governance Quality Index (GQI) across all ablation parameters, indicating that most configurations meet or exceed the production-readiness threshold, with lower scores driven primarily by auditability effects in high-stress or adversarial conditions

By explicitly ablation-testing governance primitives rather than model accuracy, the framework shifts evaluation toward trustworthiness under change. This is particularly important for clinical AI systems that rely on image synthesis, data augmentation, or federated experimentation, where the question is not only whether the model performs well, but whether its use can be explained, audited, and defended months or years later. The ablation study thus reinforces the central contribution of this work: governance mechanisms must scale with clinical reality, not just computational benchmarks, and their behavior under stress must be measurable, predictable, and transparent.

 doi:10.20944/preprints202606.0484.v1

5. Conclusion and Future Work

This work introduced a decentralized governance overlay for clinical AI systems, enabling verifiable provenance, auditability, consent enforcement, and reproducibility without decentralizing sensitive medical data or disrupting existing clinical workflows. The proposed framework was validated through simulated clinical scenarios, demonstrating strong governance performance with minimal operational overhead.

5.1. Limitations

Despite the promising results, this study has several limitations that point toward important directions for future work. First, the experimental evaluation is based on a limited set of simulated scenarios that primarily explore monotonic scaling of patient population size, workload intensity, and consent revocation probability. While these scenarios are sufficient to demonstrate feasibility and correctness of the proposed governance overlay, they do not yet capture the full diversity of operational conditions encountered in real-world clinical environments, such as bursty workloads, multi-institutional access patterns, or heterogeneous mixes of authorized and unauthorized requests.

Second, adversarial fault-injection scenarios, including intentional ledger and manifest tampering, are evaluated alongside normal-operation scenarios. Although integrity violations observed in these cases are expected and indicate correct detection behavior, future work should explicitly separate normal and adversarial evaluation tracks to further improve interpretability of compliance, auditability, and deployment-readiness metrics. A structured scenario taxonomy distinguishing expected clinical operation from security stress testing would also enable more granular benchmarking and clearer regulatory reporting.

Third, while the proposed framework demonstrates low governance overhead and sub-linear performance degradation under increasing workload, the current evaluation does not fully model high-concurrency conditions, long-term longitudinal operation, or geographically distributed deployments. Future studies will investigate scalability under sustained multi-tenant workloads, concurrent access from multiple clinical sites, and extended temporal horizons, including ledger growth and archival strategies.

5.2. Future Work

Future work will explore stronger cryptographic binding mechanisms between the decentralized governance ledger and off-chain storage repositories. While the present architecture anchors consent state, execution commitments, and reproducibility manifests through hash-based commitments, additional techniques such as signed attestations, content-addressable storage verification, and secure key management frameworks may further strengthen end-to-end integrity guarantees across distributed clinical infrastructures. In addition, we aim to investigate privacy-enhancing extensions aligned with evolving regulatory requirements, including mechanisms that support controlled revocation and data lifecycle governance. One potential direction involves the storage of symmetrically encrypted payload references on-chain, combined with secure key management strategies that enable cryptographic erasure through deliberate destruction of associated decryption keys. Such approaches may provide a technical pathway for implementing revocation semantics consistent with data protection principles, including the "Right to be Forgotten," while preserving the immutability of ledger commitments. These extensions remain beyond the scope of the present study and were not implemented or evaluated in our experimental framework. Future work will assess their feasibility, performance implications, and regulatory alignment in realistic multi-institutional clinical deployments. The current governance quality assessment relies on aggregate metrics derived from a small number of scenarios. Future work will extend the evaluation to larger and more diverse scenario matrices, incorporate real-world clinical workflow traces, and explore adaptive weighting schemes for the Governance Quality Index (GQI) based on regulatory context, clinical risk, and deployment environment.

 doi:10.20944/preprints202606.0484.v1

Data Availability Statement: The prototype ledger implementation and experimental scripts used in this study are publicly available via GitHub at Zenodo [106]. The archived version corresponds to the implementation used in the experiments reported in this manuscript.

Acknowledgments: The work presented was partially supported by the University of Piraeus Research Center.

Abbreviations

ABAC: Attribute-Based Access Control; AI: Artificial Intelligence; AWS: Amazon Web Services; CI: Confidence Interval; DLT: Distributed Ledger Technology; DICOM: Digital Imaging and Communications in Medicine; DVC: Data Version Control; EVM: Ethereum Virtual Machine; FHIR: Fast Healthcare Interoperability Resources; GDPR: General Data Protection Regulation; GQI: Governance Quality Index; HIPAA: Health Insurance Portability and Accountability Act; HL7: Health Level Seven; HSM: Hardware Security Module; IdP: Identity Provider; IoT: Internet of Things; LINDDUN: Linkability, Identifiability, Non-repudiation, Detectability, Disclosure of information, Unawareness, Non-compliance; MLOps: Machine Learning Operations; MPC: Multi-Party Computation; NIST: National Institute of Standards and Technology; PHI: Protected Health Information; QLDB: Quantum Ledger Database; RBAC: Role-Based Access Control; RPS: Requests Per Second; SBOM: Software Bill of Materials; SD: Standard Deviation; SHA: Secure Hash Algorithm; SLSA: Supply-chain Levels for Software Artifacts; STRIDE: Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege; TEE: Trusted Execution Environment; UUID: Universally Unique Identifier; W3C: World Wide Web Consortium; XAI: Explainable Artificial Intelligence; ZKP: Zero-Knowledge Proof.

Appendix A. Algorithmic Framework

This section formalizes the proposed governance overlay for clinical AI pipelines. The framework enforces consent, auditability, and reproducibility through cryptographic commitments recorded on a tamper-evident ledger, while all sensitive data and AI computation remain off-ledger.

Appendix A.1. Notation

Let $\mathcal{E}$ denote the set of governance events. Let $e_i \in \mathcal{E}$ be the i-th event occurring at timestamp $t_i \in \mathbb{N}$. Each event has a type $\mathrm{type}(e_i)$ and a payload $\mathrm{pl}(e_i)$ containing metadata only.

In addition, let $r \in \mathcal{R}$ denote the role of the requesting actor (e.g., clinician, administrator), $u \in \mathcal{U}$ denote the declared purpose of access (e.g., diagnosis, follow-up, research), and a denote the authenticated actor identifier. The symbol s_{id} represents the subject (patient) identifier associated with a consent policy. The tuple (π, θ) denotes the AI execution configuration, where π is the pipeline specification (preprocessing steps, execution graph, and parameters) and θ is the model state (architecture and weights). The stores $\mathcal{S}_P$ and $\mathcal{S}_M$ denote off-ledger repositories for consent policies and reproducibility manifests, respectively. For inference artifacts, x_{in} denotes the input data object referenced by URI $\mathrm{uri}_{\mathrm{in}}$, while x^k_{out} denotes the k-th output artifact with corresponding URI $\mathrm{uri}^k_{\mathrm{out}}$.

Each ledger entry b_i is defined as:

$$b_i = (\mathrm{id}_i, t_i, \mathrm{type}_i, \mathrm{pl}_i, \mathrm{prev}_i, \mathrm{hash}_i) \tag{A1}$$

The ledger is an ordered sequence:

$$L = (b_0, b_1, \ldots, b_n) \tag{A2}$$

Appendix A.1.1. Hash-Chain Rule

Ledger integrity is enforced via a hash chain (as described in §A.1.1):

$$\mathrm{hash}_i = H(\mathrm{prev}_i \oplus \mathrm{ser}(\mathrm{id}_i, t_i, \mathrm{type}_i, \mathrm{pl}_i, \mathrm{prev}_i)) \tag{A3}$$

with

$$\mathrm{prev}_i = \mathrm{hash}_{i-1}, \qquad \mathrm{hash}_0 = 0^{256} \tag{A4}$$

 doi:10.20944/preprints202606.0484.v1

where $\mathrm{ser}(\cdot)$ denotes deterministic serialization.

Appendix A.2. Consent Policy Model

A consent policy (as defined in §3.4) is defined as:

$$P = (p_{\mathrm{id}}, s_{\mathrm{id}}, \mathcal{R}, \mathcal{U}, t^{\min}, t^{\max}, \mathrm{revoked}, \mathrm{ver}) \tag{A5}$$

Policy validity (used in consent enforcement, §3.4) is given by:

$$\mathrm{Valid}(P, t) = (t \geq t^{\min}) \wedge (t^{\max} = \varnothing \vee t \leq t^{\max}) \wedge (\neg \mathrm{revoked}) \tag{A6}$$

Authorization is defined as:

$$\mathrm{Allow}(P, r, u, t) = \mathrm{Valid}(P, t) \wedge (r \in \mathcal{R}) \wedge (u \in \mathcal{U}) \tag{A7}$$

where $\mathrm{Valid}(P, t)$ is defined in Eq. (A6).

Appendix A.3. Reproducibility Manifest

For an inference event x (as discussed in §A.3), the reproducibility manifest is defined as:

$$\begin{aligned} M_x = (&x_{\mathrm{id}}, t, h_{\mathrm{in}}, \mathrm{uri}_{\mathrm{in}}, \\ &\{(k, h^k_{\mathrm{out}}, \mathrm{uri}^k_{\mathrm{out}})\}, \\ &h_P, h_\pi, h_\theta, a, r, u) \end{aligned} \tag{A8}$$

Algorithm A1 Governance Overlay for Clinical AI Inference

Require: Ledger L, policy store $\mathcal{S}_P$, manifest store $\mathcal{S}_M$
Require: Request $(a, r, u, p_{\mathrm{id}}, \mathrm{uri}_{\mathrm{in}}, x_{\mathrm{in}}, \pi, \theta)$
Ensure: Denial or manifest identifier x_{id}

$t \leftarrow \mathrm{now}()$
$P \leftarrow \mathcal{S}_P[p_{\mathrm{id}}]$
if not $\mathrm{Allow}(P, r, u, t)$ **then**
    Append `CONSENT_DENIED` event to ledger
    return `DENY`
end if
$h_P \leftarrow H(\mathrm{ser}(P))$
$h_\pi \leftarrow H(\mathrm{ser}(\pi))$
$h_\theta \leftarrow H(\mathrm{ser}(\theta))$
$h_{\mathrm{in}} \leftarrow H(x_{\mathrm{in}})$
$x_{\mathrm{id}} \leftarrow \mathrm{UUID}()$
Construct reproducibility manifest $M_{x_{\mathrm{id}}}$
Store $M_{x_{\mathrm{id}}}$ in $\mathcal{S}_M$
$h_M \leftarrow H(\mathrm{ser}(M_{x_{\mathrm{id}}}))$
Append `INFERENCE_EXECUTED` event to ledger
return x_{id}

 doi:10.20944/preprints202606.0484.v1

Algorithm A2 Ledger Integrity Verification

Require: Ledger $L = (b_0, \ldots, b_n)$
Ensure: Integrity flag and first failing index ($\perp$ if none)

```
prev ← 0^256
for i ← 0 to n do
    if b_i.prev ≠ prev then
        return (False, i)
    end if
    ĥ ← H(prev ‖ ser(b_i))
    if b_i.hash ≠ ĥ then
        return (False, i)
    end if
    prev ← b_i.hash
end for
return (True, ⊥)
```

Algorithm A3 Off-Ledger Manifest Verification

Require: Ledger L, manifest store $\mathcal{S}_M$, identifier x_{id}
Ensure: Verification flag

```
Extract h_M from ledger entry for x_id
Load manifest M_{x_id} from S_M
ĥ_M ← H(ser(M_{x_id}))
return (ĥ_M = h_M)
```

Governance configuration:

$$\Gamma = (g_1, g_2, g_3, g_4) \in \{0,1\}^4 \tag{A9}$$

Consent drift (as evaluated in §4.3.1.2):

$$D(\Gamma) = \sum_{i=1}^{N} \mathbf{1}(\text{exec}_i \wedge \text{revoked}(t_i)) \tag{A10}$$

Ledger integrity indicator (verified using Algorithm A2 and analyzed in §4.3.2.3):

$$I_L(\Gamma) = \mathbf{1}(\text{ledger verification fails}) \tag{A11}$$

Manifest verification rate (using Algorithm A3 and evaluated in §4.3.2.4):

$$R_M(\Gamma) = \frac{1}{|\mathcal{V}|} \sum_{x \in \mathcal{V}} \mathbf{1}(\text{manifest verified}) \tag{A12}$$

Governance overhead (analyzed in §4.3.2.2):

$$O(\Gamma) = \frac{1}{N} \sum_{i=1}^{N} \left(t_i^{\text{end}} - t_i^{\text{start}} \right) \tag{A13}$$

 doi:10.20944/preprints202606.0484.v1

Table A1. Governance System Performance Metrics Across Experimental Scenarios.

Patients	Runs	Total Req.	Exec.	Block.	Auth. Prec.	Consent Compl.	Gov. (ms)	Throughput (rps)	Audit Compl.
50	500	500	321	179	1.000	1.000	1.25	798.2	1.100
200	2000	2000	1140	860	1.000	1.000	4.39	227.9	1.118
500	5000	5000	2384	2616	1.000	1.000	11.38	87.9	1.148
200	2000	2000	447	1553	1.000	1.000	4.22	237.1	1.286
500	5000	5000	865	4135	1.000	1.000	11.09	90.2	1.379
1000	20000	20000	6787	13213	1.000	1.000	42.90	23.3	1.099
200	2000	2000	1090	910	1.000	1.000	4.32	231.7	1.123
200	2000	2000	1148	852	1.000	1.000	4.33	230.9	1.119
500	5000	5000	2285	2715	1.000	1.000	10.96	91.3	1.150

Table A2. Statistical Summary of Governance Performance Metrics. The following metrics show perfect consistency (zero variance), authorization precision 1.0000 (all scenarios identical), consent compliance rate: 1.0000 (all scenarios identical). 95% CI represents the range likely to contain the true population mean.

Statistic	Authorization Precision	Consent Compliance	Governance (ms)	Throughput (rps)	Audit Completeness	Ledger Entries
Count	9	9	9	9	9	9
Mean	1.000	1.000	10.54	224.27	1.169	5578.9
Std Dev	0.000	0.000	12.70	230.31	0.097	6505.9
Min	1.000	1.000	1.25	23.31	1.099	550.0
25th percentile	1.000	1.000	4.32	90.17	1.118	2238.0
Median	1.000	1.000	4.39	227.93	1.123	2571.0
75th percentile	1.000	1.000	11.09	231.67	1.150	5750.0
Max	1.000	1.000	42.90	798.20	1.379	21981.0
95% CI Lower	1.000	1.000	0.78	47.24	1.094	578.0
95% CI Upper	1.000	1.000	20.30	401.30	1.244	10579.7

References

1. Damian, A.I.; Bleotiu, C.; Grigoras, M.; Butusina, P.; De Franceschi, A.; Toderian, V.; Tapus, N. Ratio1 meta-OS - decentralized MLOps and beyond. In Proceedings of the 2025 25th International Conference on Control Systems and Computer Science (CSCS), 2025, pp. 258–265. https://doi.org/10.1109/CSCS66924.2025.00046.
2. Agbo, C.C.; Mahmoud, Q.H.; Eklund, J.M. Blockchain technology in healthcare: a systematic review. *Healthcare* **2019**, *7*, 56.
3. De Aguiar, E.J.; Faiçal, B.S.; Krishnamachari, B.; Ueyama, J. A survey of blockchain-based strategies for healthcare. *ACm Computing surveys (CsUr)* **2020**, *53*, 1–27.
4. Pampattiwar, K.; Chavan, P. A secure and scalable blockchain-based model for electronic health record management. *Scientific Reports* **2025**, *15*, 11612.
5. Procko, T.; Ochoa, O. Mapping the W3C Provenance Ontology (PROV-O) to the Basic Formal Ontology (BFO): Epistemological Considerations and Preliminary Implementation. *Available at SSRN 4852748* **2024**.
6. Chen, A.; Chow, A.; Davidson, A.; DCunha, A.; Ghodsi, A.; Hong, S.A.; Konwinski, A.; Mewald, C.; Murching, S.; Nykodym, T.; et al. Developments in mlflow: A system to accelerate the machine learning lifecycle. In Proceedings of the Proceedings of the fourth international workshop on data management for end-to-end machine learning, 2020, pp. 1–4.
7. Melchor, F.; Rodriguez-Echeverria, R.; Conejero, J.M.; Prieto, A.E.; Gutiérrez, J.D. A model-driven approach for systematic reproducibility and replicability of data science projects **2022**. pp. 147–163.
8. Torres-Arias, S.; Afzali, H.; Kuppusamy, T.K.; Curtmola, R.; Cappos, J. in-toto: Providing farm-to-table guarantees for bits and bytes. In Proceedings of the 28th USENIX Security Symposium (USENIX Security 19), 2019, pp. 1393–1410.
9. Ohm, M.; Sykosch, A.; Meier, M. Towards detection of software supply chain attacks by forensic artifacts, 2020.
10. Xu, J.; Glicksberg, B.S.; Su, C.; Walker, P.; Bian, J.; Wang, F. Federated learning for healthcare informatics. *Journal of healthcare informatics research* **2021**, *5*, 1–19.
11. Dwork, C. Differential privacy **2006**. pp. 1–12.
12. Zhou, I.; Tofigh, F.; Piccardi, M.; Abolhasan, M.; Franklin, D.; Lipman, J. Secure multi-party computation for machine learning: A survey. *IEEE Access* **2024**, *12*, 53881–53899.

 doi:10.20944/preprints202606.0484.v1

13. Haq, R.U.; Khan, R.; Alturise, F.; Sahrani, S.; Alkhalaf, S.; Sarker, M.R. Transchain: Blockchain-Based Management of Allografts for Enhancing Data Provenance. *IEEE Access* **2025**.
14. Jalbani, A.; Weerawarna, R.; Al-Zubaidi, K. Enhancing data provenance in AI with blockchain technology: a comprehensive quality model. *CSI Transactions on ICT* **2025**, *13*, 213–224.
15. Alagha, B.; Ozcelik, I. BEATS: Practical Audit Trail in Blockchain Systems. *IEEE Access* **2025**.
16. Zhao, Y.; Qu, Y.; Li, B.; Zhao, L.; Chen, F.; Xiang, Y.; Gao, L. Collusion-Resistant and Time-Aware Co-Verification for Edge Data Integrity. *IEEE Transactions on Dependable and Secure Computing* **2025**.
17. Padia, S.; Vaidya, D.; Mangrulkar, R. Adaptive Trust Consensus for Blockchain IoT: Comparing RL, DRL, and MARL Against Naive, Collusive, Adaptive, Byzantine, and Sleeper Attacks. *arXiv preprint arXiv:2512.22860* **2025**.
18. Sa'ad, U.; Na, W.; Dao, N.N.; Cho, S. Adaptive Risk Analysis Framework for Network-Level Moving Target Defense Under Adversarial Intelligence Uncertainty. *Computers & Security* **2026**, p. 104890.
19. Xiang, Z.; Cheng, J.; Liu, C.; Mao, Q.; Yuan, G.; Gao, S. Privacy-preserving autonomous vehicle group formation in a collusive attack scenario. *IEEE Internet of Things Journal* **2025**.
20. Othman, W.; Hong, Z.; Fuyou, M.; Xue, K.; Hawbani, A.; Amin, R.; Zhao, L.; Li, T. CRT and PUF-based self/mutual-healing key distribution protocol with collusion resistance and revocation capability. *IEEE Transactions on Mobile Computing* **2025**, *24*, 4607–4622.
21. Liu, W.; Shi, J.; Wang, H.; Chen, T.; Han, Z.; Li, Q. Coordinate plane based authentication method for detecting clone node in wireless sensor networks. *Journal of Information Security and Applications* **2025**, *93*, 104148.
22. Xue, M.; Zhong, H.; Shi, Y.; Zeng, Y.; Zhang, J.; Zhao, N. A correlation analysis-based federated learning framework for defending against collusion-free-riding attacks: M. Xue et al. *Cybersecurity* **2025**, *8*, 65.
23. Farooq, M.O. Robust Defensive Cyber Agent for Multi-Adversary Defense. *IEEE Transactions on Machine Learning in Communications and Networking* **2025**.
24. Stănculete, B.N.; Sommese, R. HideMe: Hiding VMs from Co-Residency Attacks Using Network-Level Traffic Redirection. In Proceedings of the 2025 21st International Conference on Network and Service Management (CNSM). IEEE, 2025, pp. 1–7.
25. Rahimi, M. DP-Mix: Differentially Private Routing in Mix Networks. In Proceedings of the 2025 IEEE Annual Computer Security Applications Conference (ACSAC). IEEE, 2025, pp. 394–410.
26. Soomro, M.A.; Anwar, F.M. Breaking Precision Time: OS Vulnerability Exploits Against IEEE 1588. In Proceedings of the 2025 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication (ISPCS). IEEE, 2025, pp. 1–7.
27. Rozenblit, L.; Price, A.; Solomonides, A.; Joseph, A.L.; Koski, E.; Srivastava, G.; Labkoff, S.; Bray, D.; Lopez-Gonzalez, M.; Singh, R.; et al. Toward responsible AI governance: balancing multi-stakeholder perspectives on AI in healthcare. *International Journal of Medical Informatics* **2025**, *203*, 106015.
28. Leiva, V.; Castro, C. Artificial intelligence and blockchain in clinical trials: enhancing data governance efficiency, integrity, and transparency. *Bioanalysis* **2025**, *17*, 161–176.
29. Mennella, C.; Maniscalco, U.; De Pietro, G.; Esposito, M. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon* **2024**, *10*.
30. Morley, J.; Murphy, L.; Mishra, A.; Joshi, I.; Karpathakis, K. Governing data and artificial intelligence for health care: developing an international understanding. *JMIR formative research* **2022**, *6*, e31623.
31. Food, U.; Administration, D.; et al. Good machine learning practice for medical device development. *US Food and Drug Administration [Internet]* **2021**.
32. Organization, W.H. Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance **2024**.
33. Wang, Q.; Yu, G.; Sai, Y.; Bandara, H.D.; Chen, S. Is your AI truly yours? Leveraging blockchain for copyrights, provenance, and lineage. *IEEE Transactions on Services Computing* **2025**.
34. Siyal, F.; Alkhabbas, F.; Guzzo, A.; Sacca, D.; Fortino, G. PRO-CHAIN: A Provenance Tracking Framework Leveraging Blockchain and PUF. In Proceedings of the 2025 7th International Conference on Blockchain Computing and Applications (BCCA). IEEE, 2025, pp. 587–594.
35. Schneier, B.; Kelsey, J. Secure Audit Logs to Support Computer Forensics. *ACM Transactions on Information and System Security* **1999**, *1*, 159–176. https://doi.org/10.1145/317087.317089.
36. National Institute of Standards and Technology. Special Publication 800-92: Guide to Computer Security Log Management, 2006.
37. Laurie, B.; Langley, A.; Kasper, E. RFC 6962: Certificate Transparency. IETF RFC, 2013.

 doi:10.20944/preprints202606.0484.v1

38. Alagha, B.; Özçelik, İ. Using Blockchain Technology for Audit Trail. In *Digital Strategy and Governance in Transformative Technologies*; CRC Press, 2025; pp. 239–259.
39. Saleh, S.M. DevSecLogs: AI-Powered, Tamper-Evident Log Intelligence for Secure CI/CD Pipelines. In Proceedings of the 2025 IEEE International Conference on Software Maintenance and Evolution (ICSME). IEEE, 2025, pp. 890–894.
40. Shaalan, K.; Alhyari, O.; Al-Dmour, N.A.; Elsersy, M.; Al-Zoghby, A.M.; Binhammad, M.; Ghazal, T.M. Tamper-Proof Legal Document Storage in Digital Courts Using Hyperledger Fabric Framework. In Proceedings of the 2025 3rd International Conference on Cyber Resilience (ICCR). IEEE, 2025, pp. 1–7.
41. Surenderkumar, S.; Sanjay, S.; Saran, K.; Shanjith, V.; Shrivarshan, C. ForenSync: Real-Time IoT Platform for Forensic Evidence Tracking and Audit Logging. In Proceedings of the 2025 10th International Conference on Smart Structures and Systems (ICSSS). IEEE, 2025, pp. 1–5.
42. Kaye, J.; Whitley, E.A.; Lund, D.; Morrison, M.; Teare, H.; Melham, K. Dynamic consent: a patient interface for twenty-first century research networks. *European journal of human genetics* **2015**, *23*, 141–146.
43. Donnelly, M.; McDonagh, M. Health research, consent and the GDPR exemption. *European journal of health law* **2019**, *26*, 97–119.
44. Yaqub, N.; Zhang, J.; Khalid, M.I.; Ahmed, M.; Helfert, M. Privacy by Design Enabled Dynamic Consent Management Systems: A Demonstration Perspective. In Proceedings of the World Conference on Information Systems and Technologies. Springer, 2025, pp. 371–383.
45. Tsang, Y.P.; Lee, C.K.M.; Zhang, K.; Wu, C.H.; Ip, W. On-chain and off-chain data management for blockchain-internet of things: A multi-agent deep reinforcement learning approach. *Journal of Grid Computing* **2024**, *22*, 16.
46. Lupaiescu, S.; Cioata, P.; Turcu, C.E.; Gherman, O.; Turcu, C.O.; Paslaru, G. Centralized vs. decentralized: Performance comparison between bigchaindb and amazon qldb, 2022.
47. Damgård, I. Commitment schemes and zero-knowledge protocols, 1998.
48. Miyaji, H.; Miyaji, A. Lattice-based key-value commitment scheme. *IEEE Transactions on Information Theory* **2025**.
49. Varma, I.M.; Kumar, N.; Bartolini, N. A Polynomial Commitment-Driven Zero-Knowledge Proof-based Authentication of Autonomous Vehicles in Multi-RSU Broadcast Domains. *IEEE Transactions on Consumer Electronics* **2025**.
50. Pappas, C.; Papadopoulos, D.; Papamanthou, C. HydraProofs: Optimally Computing All Proofs in a Vector Commitment (with applications to efficient zkSNARKs over data from multiple users). In Proceedings of the 2025 IEEE Symposium on Security and Privacy (SP). IEEE, 2025, pp. 3421–3439.
51. Gai, K.; Guo, Y.; Yu, J.; Chan, W.; Zhu, L.; Zhang, Y.; Meng, W. Cape: Commitment-based privacy-preserving payment channel scheme in blockchain. *IEEE Transactions on Dependable and Secure Computing* **2025**.
52. Lee, S.; Gee, E. Commit-Reveal 2: Randomized Reveal Order Mitigates Last-Revealer Attacks in Commit-Reveal. In Proceedings of the 2025 IEEE International Conference on Blockchain and Cryptocurrency (ICBC). IEEE, 2025, pp. 1–5.
53. Torres-Arias, S.; Afzali, H.; Kuppusamy, T.K.; Curtmola, R.; Cappos, J. in-toto: Providing farm-to-table guarantees for bits and bytes. In Proceedings of the 28th USENIX Security Symposium (USENIX Security 19). USENIX Association, 2019, pp. 1393–1410.
54. National Telecommunications and Information Administration. The Minimum Elements for a Software Bill of Materials (SBOM), 2021.
55. Saraswat, D.; Bhattacharya, P.; Verma, A.; Prasad, V.K.; Tanwar, S.; Sharma, G.; Bokoro, P.N.; Sharma, R. Explainable AI for healthcare 5.0: opportunities and challenges. *IEEe Access* **2022**, *10*, 84486–84517.
56. Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I.D.; Gebru, T. Model cards for model reporting **2019**. pp. 220–229.
57. Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J.W.; Wallach, H.; Iii, H.D.; Crawford, K. Datasheets for datasets. *Communications of the ACM* **2021**, *64*, 86–92.
58. Rempe, M.; Heine, L.; Seibold, C.; Hörst, F.; Kleesiek, J. De-identification of medical imaging data: a comprehensive tool for ensuring patient privacy. *European radiology* **2025**, pp. 1–10.
59. Thambawita, V.; Salehi, P.; Sheshkal, S.A.; Hicks, S.A.; Hammer, H.L.; Parasa, S.; Lange, T.d.; Halvorsen, P.; Riegler, M.A. SinGAN-Seg: Synthetic training data generation for medical image segmentation. *PloS one* **2022**, *17*, e0267976.
60. Mouris, D.; Tsoutsos, N.G. Zilch: A framework for deploying transparent zero-knowledge proofs. *IEEE Transactions on Information Forensics and Security* **2021**, *16*, 3269–3284.

 doi:10.20944/preprints202606.0484.v1

61. European Data Protection Board. Guidelines 02/2025 on processing of personal data through blockchain technologies, 2025.
62. Sandler, D.; Derr, K.; Crosby, S.; Wallach, D.S. Finding the evidence in tamper-evident logs **2008**. pp. 69–75.
63. Giordano, M.T. Blockchain and the GDPR: new challenges for privacy and security, 2020.
64. Choi, Y.B.; Williams, C.E. A HIPAA security and privacy compliance audit and risk assessment mitigation approach, 2022.
65. Hewage, N.; Meedeniya, D. Machine learning operations: A survey on MLOps tool support. *arXiv preprint arXiv:2202.10169* **2022**.
66. Vorisek, C.N.; Lehne, M.; Klopfenstein, S.A.I.; Mayer, P.J.; Bartschke, A.; Haese, T.; Thun, S. Fast healthcare interoperability resources (FHIR) for interoperability in health research: systematic review, 2022.
67. Onken, M.; Eichelberg, M.; Riesmeier, J.; Jensch, P. Digital imaging and communications in medicine, 2010.
68. de Carvalho Junior, M.A.; Bandiera-Paiva, P. Health Information System Role-Based Access Control Current Security Trends and Challenges. *Journal of healthcare engineering* **2018**, *2018*, 6510249.
69. Ahmed, A.; Shahzad, A.; Naseem, A.; Ali, S.; Ahmad, I. Evaluating the effectiveness of data governance frameworks in ensuring security and privacy of healthcare data: A quantitative analysis of ISO standards, GDPR, and HIPAA in blockchain technology. *PLoS One* **2025**, *20*, e0324285.
70. Huerta, T.R.; Bartlett, C.W.; Alain, G.; Bentley, T.; Bradford, C.R.; Bridges, J.F.; Chakravarti, A.; Farag, A.A.; Gerhardt, C.A.; Gorman, T.; et al. Operationalizing a research-oriented learning healthcare system across covered entities: cross-institutional strategies and innovations. *npj Health Systems* **2025**, *2*, 47.
71. Bonawitz, K.; Ivanov, V.; Kreuter, B.; Marcedone, A.; McMahan, H.B.; Patel, S.; Ramage, D.; Segal, A.; Seth, K. Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1175–1191.
72. Sabt, M.; Achemlal, M.; Bouabdallah, A. Trusted execution environment: What it is, and what it is not **2015**. *1*, 57–64.
73. Gavrilov, G.; Trajkovik, V. Security and privacy issues and requirements for healthcare cloud computing. *ICT Innovations* **2012**, pp. 143–152.
74. Damian, A.; Butusina, P.; Franceschi, A.D.; Toderian, V.; Grigoras, M.; Bleotiu, C. Ratio1 – AI meta-OS, 2025, [arXiv:cs.OS/2509.12223].
75. Panagoulias, D.P.; Tsoureli-Nikita, E.; Virvou, M.; Tsihrintzis, G.A. Dermacen analytica: A novel methodology integrating multi-modal large language models with machine learning in dermatology. *International Journal of Medical Informatics* **2025**, *199*, 105898.
76. Panagoulias, D.P.; Tsihrintzis, G.A.; Virvou, M. Integrating Multi-Modal Language Models and Machine Learning in Dermatology. In *Artificial Intelligence-Empowered Bio-medical Applications: Challenges, Solutions and Development Guidelines*; Springer, 2025; pp. 239–268.
77. Panagoulias, D.P.; Tsihrintzis, G.A.; Virvou, M. Iterative Microservices Approach for Explainable and Reliable AI in Medical Applications. In *Artificial Intelligence-Empowered Bio-medical Applications: Challenges, Solutions and Development Guidelines*; Springer, 2025; pp. 115–133.
78. Panagoulias, D.P.; Virvou, M.; Tsihrintzis, G.A. A microservices-based iterative development approach for usable, reliable and explainable AI-infused medical applications using RUP. In Proceedings of the 2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, 2022, pp. 1028–1035.
79. Panagoulias, D.P.; Tsihrintzis, G.A.; Virvou, M. Challenges in Regulating and Validating AI-Driven Healthcare. In *Artificial Intelligence-Empowered Bio-medical Applications*; Springer, 2025; pp. 135–152.
80. Panagoulias, D.P.; Tsihrintzis, G.A.; Virvou, M. Advances, Challenges and Contributions to Artificial Intelligence-Empowered Bio-Medical Applications. In *Artificial Intelligence-Empowered Bio-medical Applications: Challenges, Solutions and Development Guidelines*; Springer, 2025; pp. 277–287.
81. Panagoulias, D.P.; Tsihrintzis, G.A.; Virvou, M. Introduction to AI-Empowered Medical Software: Recent Advances and Challenges. *Artificial Intelligence-Empowered Bio-medical Applications* **2025**, pp. 3–11.
82. Nchinda, N.; Cameron, A.; Retzepi, K.; Lippman, A. MedRec: a network for personal information distribution. In Proceedings of the 2019 international conference on computing, networking and communications (ICNC). IEEE, 2019, pp. 637–641.
83. Lee, A.R.; Kim, M.G.; Kim, I.K. SHAREChain: Healthcare data sharing framework using Blockchain-registry and FHIR. In Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2019, pp. 1087–1090.

 doi:10.20944/preprints202606.0484.v1

84. Spadari, V.; Cerasuolo, F.; Bovenzi, G.; Pescapè, A. An mlops framework for explainable network intrusion detection with mlflow. In Proceedings of the 2024 IEEE Symposium on Computers and Communications (ISCC). IEEE, 2024, pp. 1–6.
85. Mohomed, T.; Ekanayake, I. ARGO-SLSA: Software Supply Chain Security in Argo Workflows. In Proceedings of the 2025 Moratuwa Engineering Research Conference (MERCon). IEEE, 2025, pp. 245–250.
86. Bandara, E.; Shetty, S.; Rahman, A.; Mukkamala, R. Let'sTrace—Blockchain, Federated Learning and TUF/In-ToTo Enabled Cyber Supply Chain Provenance Platform. In Proceedings of the MILCOM 2021-2021 IEEE Military Communications Conference (MILCOM). IEEE, 2021, pp. 470–476.
87. Atkins, D.L.; Ball, T.; Graves, T.L.; Mockus, A. Using version control data to evaluate the impact of software tools: A case study of the version editor. *IEEE Transactions on Software Engineering* **2002**, *28*, 625–637.
88. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial intelligence and statistics. Pmlr, 2017, pp. 1273–1282.
89. Sakimura, N.; Bradley, J.; Jones, M.; De Medeiros, B.; Mortimore, C. OpenID Connect Core 1.0 incorporating errata set 1. *The OpenID Foundation, specification* **2014**, *335*.
90. Hardt, D. The OAuth 2.0 authorization framework **2012**.
91. Khan, J.A. Role-based access control (RBAC) and attribute-based access control (ABAC). In *Improving security, privacy, and trust in cloud computing*; IGI Global Scientific Publishing, 2024; pp. 113–126.
92. Barati, M.; Aujla, G.S.; Llanos, J.T.; Duodu, K.A.; Rana, O.F.; Carr, M.; Ranjan, R. Privacy-aware cloud auditing for GDPR compliance verification in online healthcare. *IEEE Transactions on Industrial Informatics* **2021**, *18*, 4808–4819.
93. Sculley, D.; Holt, G.; Golovin, D.; Davydov, E.; Phillips, T.; Ebner, D.; Chaudhary, V.; Young, M.; Crespo, J.F.; Dennison, D. Hidden technical debt in machine learning systems. *Advances in neural information processing systems* **2015**, *28*.
94. Seol, K.; Kim, Y.G.; Lee, E.; Seo, Y.D.; Baik, D.K. Privacy-preserving attribute-based access control model for XML-based electronic health record system. *IEEE Access* **2018**, *6*, 9114–9128.
95. Hernan, S.; Lambert, S.; Ostwald, T.; Shostack, A. Threat Modeling: Uncover Security Design Flaws Using The STRIDE Approach. *MSDN Magazine* **2006**.
96. National Institute of Standards and Technology. Special Publication 800-30 Revision 1: Guide for Conducting Risk Assessments, 2012.
97. Deng, M.; Wuyts, K.; Scandariato, R.; Preneel, B.; Joosen, W. LINDDUN: A Privacy Threat Analysis Framework, 2013.
98. Sculley, D.; Holt, G.; Golovin, D.; Davydov, E.; Phillips, T.; Ebner, D.; Chaudhary, V.; Young, M.; Crespo, J.F.; Dennison, D. Hidden technical debt in machine learning systems. *Advances in neural information processing systems* **2015**, *28*.
99. Amershi, S.; Begel, A.; Bird, C.; DeLine, R.; Gall, H.; Kamar, E.; Nagappan, N.; Nushi, B.; Zimmermann, T. Software engineering for machine learning: A case study. In Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP). IEEE, 2019, pp. 291–300.
100. Zaharia, M.; Chen, A.; Davidson, A.; Ghodsi, A.; Hong, S.A.; Konwinski, A.; Murching, S.; Nykodym, T.; Ogilvie, P.; Parkhe, M.; et al. Accelerating the machine learning lifecycle with MLflow. *IEEE Data Eng. Bull.* **2018**, *41*, 39–45.
101. Torres-Arias, S.; Afzali, H.; Kuppusamy, T.K.; Curtmola, R.; Cappos, J. in-toto: Providing farm-to-table guarantees for bits and bytes. In Proceedings of the 28th USENIX Security Symposium (USENIX Security 19), 2019, pp. 1393–1410.
102. Pineau, J.; Vincent-Lamarre, P.; Sinha, K.; Larivière, V.; Beygelzimer, A.; d'Alché Buc, F.; Fox, E.; Larochelle, H. Improving reproducibility in machine learning research (a report from the neurips 2019 reproducibility program). *Journal of machine learning research* **2021**, *22*, 1–20.
103. Andersen, D.L.; Park, I. Scaling distributed machine learning with the parameter server. In Proceedings of the Proceedings of the Operating Systems Design and Implementation (OSDI), 2014.
104. Kairouz, P.; McMahan, H.B. Advances and open problems in federated learning. *Foundations and trends in machine learning* **2021**, *14*, 1–210.
105. Breck, E.; Cai, S.; Nielsen, E.; Salib, M.; Sculley, D. The ML test score: A rubric for ML production readiness and technical debt reduction. In Proceedings of the 2017 IEEE international conference on big data (big data). IEEE, 2017, pp. 1123–1132.

 doi:10.20944/preprints202606.0484.v1

106. Panagoulias, D.P.e.a. Blockchain-Enabled Governance Prototype for Clinical AI (v1.0.0) **2026**. https://doi.org/10.5281/zenodo.18648393.