Original Article

Learning Medication-Safety Lessons across Hospitals while Every Record Stays Home

Sara Ben Youssef[1]*, Amal Trabelsi[1], Karim Boudiaf[2], Nabil Jebali[3]

[1]Department of Federated Learning and Medication Safety, Faculty of Pharmacy, University of Tunis, Tunis, Tunisia. [2]Department of Privacy-Preserving Pharmacy AI, Faculty of Pharmacy, University of Sousse, Sousse, Tunisia. [3]Department of Cross-Hospital AI Learning, Faculty of Pharmacy, University of Sfax, Sfax, Tunisia.

Abstract

Hospitals could learn from one another's medication-related risks without transferring patient-level records into a common repository. However, keeping records local does not by itself establish privacy, security, interoperability, fairness, clinical usefulness, or accountable decision use. This article develops an original, non-empirical federated medication-safety architecture for collaboration among heterogeneous hospitals. It defines a medication-safety lesson as a proposed, versioned package linking a bounded clinical task with locally derived model information, provenance, uncertainty, validation evidence, workflow purpose, and explicit limits on use. The architecture separates six functions: local data stewardship, local training, protected update exchange, shared candidate-model construction, site-specific qualification, and consortium governance. Raw records remain within participating institutions; only authorized computational updates and contextual metadata enter the shared process. A resulting shared model remains a candidate until each hospital assesses local calibration, medication-safety relevance, subgroup performance, workflow burden, technical compatibility, and professional oversight requirements. Separate controls address statistical and semantic heterogeneity, privacy leakage, malicious or unreliable updates, inequitable performance, institutional burden, and unequal distribution of benefits. Evaluation must extend beyond aggregate predictive performance to credible local comparators, site-separated results, workflow consequences, human–AI interaction, attack resistance, equity, change control, and prospective validation. The architecture does not establish that federated learning is inherently safer, more private, more equitable, or more clinically effective than centralized or local alternatives. Its original contribution is a proposed boundary-governed approach in which collaboration creates locally contestable medication-safety candidates rather than transferable clinical authority.

Keywords: Digital pharmacy, Artificial intelligence, Pharmacy practice, Medication safety, Human–AI collaboration, Clinical decision support

Introduction

Medication-safety learning is often constrained by the fragmentation of hospital data. Individual institutions may observe only part of the variation in prescribing, dispensing, administration, monitoring, and adverse-event patterns needed to improve a model. Federated learning offers a mechanism for multi-institutional healthcare modelling without pooling raw patient records, although its use remains conditional on privacy, security, and institutional controls [1-3]. Record locality therefore describes where patient-level data remain; it does not establish that computational updates are non-identifying, that participating systems are secure, or that a shared model is clinically appropriate.

Structured electronic-health-record applications demonstrate considerable technical activity, but their designs, outcomes, comparators, and clinical evaluation remain heterogeneous [4]. This matters for medication safety because a technically transferable model may still represent different prescribing conventions, formularies, measurement practices, patient populations, or workflow priorities at different hospitals. A common optimization objective can conceal disagreement about what constitutes an error, what outcome should be predicted, when intervention remains possible, and which professional is expected to act.

Medication-error risk prioritization illustrates the distinction between prediction and pharmacy practice. A model may help direct pharmacist attention toward prescriptions with elevated error risk, but evidence obtained in one workflow does not establish that the same signal will remain calibrated, useful, or safe elsewhere [5]. Neither an aggregated parameter set nor a high-performing risk score is equivalent to a medication-safety lesson. The latter requires interpretable links among the clinical task, local context, model version, supporting evidence, uncertainty, intended user, and permissible action.

This article proposes a federated medication-safety architecture rather than a new predictive algorithm. Its central claim is that cross-hospital learning requires coordinated boundaries around what is trained locally, what is shared, how updates are admitted, how a shared candidate is evaluated, who may authorize local use, and how privacy, security, equity, and institutional responsibility are

monitored. The architecture does not presume that federation is preferable for every task. A local model, conventional multicentre study, trusted research environment, or centralized analysis may be more appropriate when lawful data sharing is feasible, heterogeneity overwhelms shared learning, or federation adds complexity without credible medication-safety value.

Medication-Safety Learning across Heterogeneous Hospitals

Federated learning can train predictive models from institutionally distributed electronic records while source data remain locally controlled [6]. For medication safety, however, the learnable object must be narrower than "hospital data." Each collaboration requires a specified population, medication-related target, prediction or detection time, outcome definition, intended professional user, and action that remains possible after an output is produced. Without these elements, hospitals may optimize superficially similar tasks that represent different clinical questions.

Multi-institutional studies have demonstrated that models can be trained across geographically or organizationally separated clinical datasets without creating one pooled patient-level repository [7-9]. They also show why site-separated assessment remains necessary. Hospitals differ in case mix, prevalence, equipment, data completeness, and clinical processes. These differences may be even more consequential for medication systems, where local formularies, substitution rules, order sets, pharmacy staffing, prescribing privileges, laboratory reporting, and medication-administration records influence both the input and the meaning of an outcome.

Collaboration may be especially attractive for uncommon but consequential medication-related events because one institution may have insufficient examples for stable learning. Evidence from federated rare-event applications supports this rationale at a mechanism level, but it does not prove that federation will improve detection of rare medication harm [9]. A participating hospital must therefore retain a credible local-only comparator and determine whether shared learning adds value without degrading locally important performance. The intended pharmacy action must also remain explicit because prioritizing review is different from recommending therapy, verifying a prescription, or changing an order [5].

This article proposes the medication-safety lesson as the unit connecting learning with responsible use. It consists of a bounded task definition, an authorized update or model representation, provenance, software and model version, local validation results, uncertainty, relevant subgroup information, workflow purpose, and a decision boundary. The construct responds to reported variation in structured-record federation but is not itself validated by that literature [4]. It separates a patient record from a local model, a model update from a lesson, a shared candidate from an approved local system, and decision support from professional medication authority.

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for definitions, components, relationships, and intended uses for the article Learning Medication-Safety Lessons across Hospitals while Every Record Stays Home.

Table 1. Definitions, components, relationships, and intended uses for the article Learning Medication-Safety Lessons across Hospitals while Every Record Stays Home.

Component or scientific dimension	Problem addressed	Inputs or determinants	Proposed mechanism or relationship	Expected contribution	Evidence required	Failure or uncertainty risk	Boundary statement
Local medication record	Central pooling may be unlawful, unacceptable, or unnecessary.	Prescriptions, administrations, laboratory results, diagnoses, outcomes, and workflow data.	Records remain within hospital control while approved computation occurs locally [1, 3, 6].	Enables collaboration without routine patient-level transfer.	Data provenance, lawful-use basis, quality assessment, and access controls.	Local records may still be incomplete, biased, or exposed through weak systems.	Record locality does not prove privacy or data quality.
Local medication-safety task	Hospitals may use the same label for different clinical problems.	Population, target, time horizon, outcome definition, user, and intended action.	**Proposed:** every collaboration begins with a task contract tied to a pharmacy workflow [5].	Aligns learning with an identifiable medication-safety purpose.	Clinical adjudication and workflow-specific validation.	Outcome labels may reflect local policy rather than preventable harm.	A prediction target is not a medication decision.
Local model	Shared models may obscure hospital-specific baselines.	Local task data, preprocessing, algorithm, and validation design.	Training and internal validation occur inside the institutional environment [6].	Preserves a local comparator and identifies site-specific limitations.	Calibration, discrimination, error analysis, and temporal testing.	Poor local data may yield an unreliable update.	Local performance does not establish external usefulness.

Authorized model update	Unrestricted exchange may disclose information or accept low-quality contributions.	Parameters, gradients, statistics, metadata, and validation summaries.	**Proposed:** only authenticated, policy-compliant update packages enter aggregation.	Restricts shared information to task-relevant artifacts.	Leakage testing, integrity checks, provenance, and authorization.	Updates may reveal information or contain malicious or incompatible content.	An update is not a complete medication-safety lesson.
Medication-safety lesson	Model parameters lack clinical context and use boundaries.	Task contract, update, provenance, uncertainty, subgroup evidence, and intended use.	**Proposed original construct:** package computational learning with contextual evidence and limits.	Makes shared learning traceable, contestable, and locally interpretable.	Empirical testing of completeness, usability, and transferability.	Required metadata may be unavailable or misleading.	The construct is proposed and not empirically validated.
Shared candidate model	Aggregation can be mistaken for approval.	Admitted local updates and aggregation rules.	Accepted updates create a versioned candidate for site-level reassessment [7, 8].	Supports collaborative learning while retaining local authority.	Site-separated evaluation against local and non-AI comparators.	Mean gains may conceal serious local deterioration.	A candidate cannot enter workflow automatically.
Qualified decision-support output	A model output may be confused with professional action.	Locally qualified model, current patient information, and workflow context.	**Proposed:** output may prioritize or inform authorized review after local approval.	Supports human–AI collaboration without transferring professional authority.	Prospective workflow, usability, workload, and safety evaluation.	Automation bias, alert burden, delay, or inappropriate reliance.	The model does not prescribe, verify, dispense, administer, or alter therapy.

Proposed Federated Collaboration Architecture

The proposed architecture contains six connected planes. The local data plane holds medication, laboratory, diagnostic, administration, outcome, and workflow records under hospital control. The local learning plane performs task-specific preparation, training, internal validation, pharmacist review, and update packaging. Local computation can be combined with formal privacy protections, but privacy parameters and their utility consequences must be disclosed rather than assumed [10]. Hospital-side federated computation can also be supported through on-premises infrastructure, although feasibility in one clinical application does not establish suitability for medication-safety work [11].

The protected exchange plane authenticates participants and governs what may leave each hospital. Its controls include update provenance, task compatibility, software version, integrity checking, encryption, transmission logging, and anomaly screening. Cryptographic protected computation can reduce exposure during distributed analysis, but it creates computational and coordination requirements and cannot be described as protection from every threat [12].

Figure 1 presents the original conceptual synthesis for federated medication-safety learning architecture, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

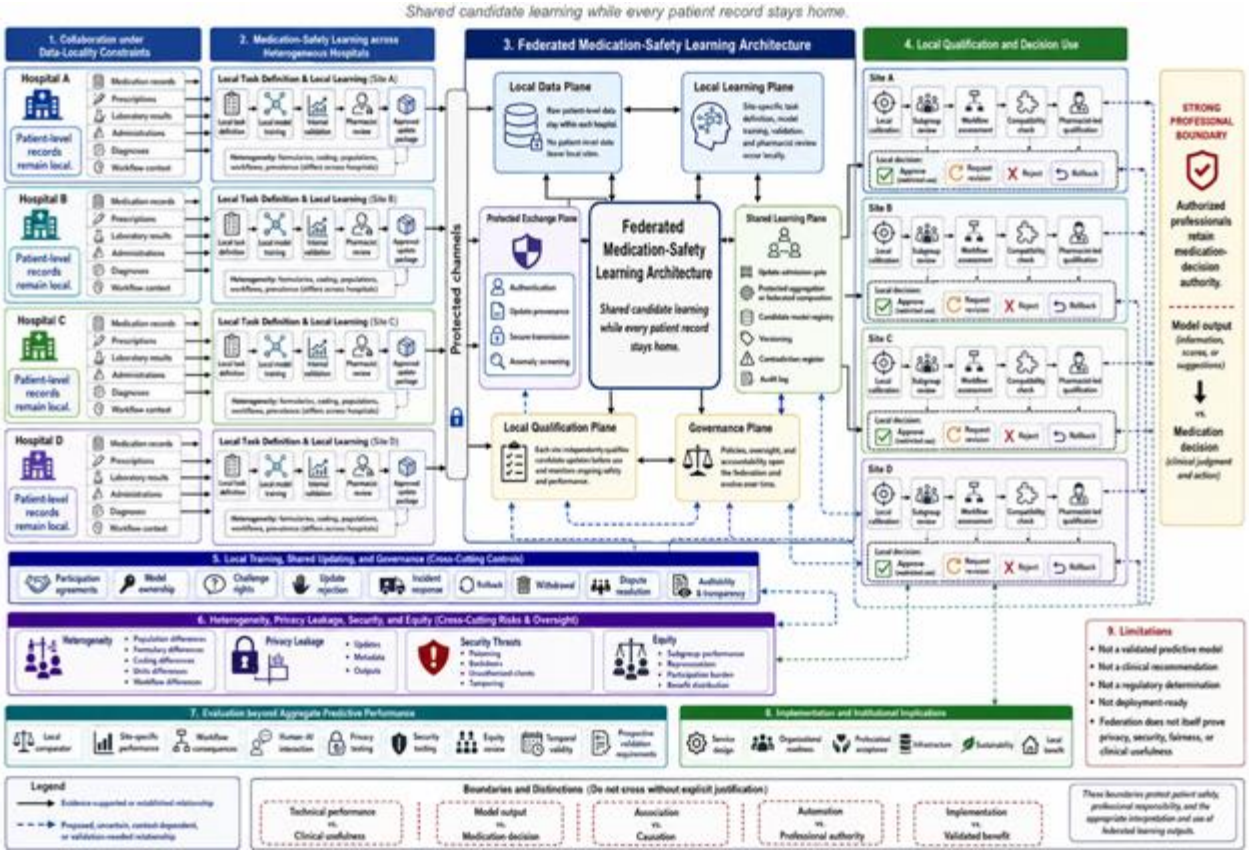

Figure 1. Federated Medication-Safety Learning Architecture

The shared learning plane applies approved aggregation rules, records participation, versions candidate models, and retains an auditable account of accepted, rejected, and anomalous updates. Health-specific federated platforms demonstrate that collaboration requires orchestration, identity, controlled execution, and reproducible analysis rather than aggregation alone [13]. The architecture adds proposed medication-safety controls: a contradiction register for incompatible local findings, an update-admission decision, and a candidate registry that prevents an aggregated model from being represented as clinically approved.

The local qualification plane returns each candidate to participating hospitals. Qualification examines local calibration, task validity, subgroup performance, workflow burden, technical compatibility, and the ability of pharmacists or other authorized users to challenge an output. Approval is therefore local and conditional. A hospital may restrict use, reject a candidate, request revision, retain its local model, or withdraw from the update cycle. This design rejects the assumption that participation creates an obligation to deploy.

The sixth plane is governance, which spans the complete architecture. It determines participation authority, model ownership, update rules, audit access, incident responsibility, rollback, withdrawal, dispute resolution, and allocation of burdens and benefits. The six-plane design is proposed. Existing work supports the need to separate privacy, security, infrastructure, and evaluation, but it does not validate this medication-safety arrangement [2].

Table 2 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for evaluation criteria, evidence requirements, and failure modes for the article Learning Medication-Safety Lessons across Hospitals while Every Record Stays Home.

Table 2. Evaluation criteria, evidence requirements, and failure modes for the article Learning Medication-Safety Lessons across Hospitals while Every Record Stays Home.

Architecture layer or process stage	Core function	Information flow	Interaction with other components	Evaluation criterion	Potential failure mode	Human or organizational responsibility	Readiness boundary
Local data plane	Preserve institutional control over	Records remain local; approved	Supplies the local learning plane.	Provenance, completeness,	Missingness, coding mismatch,	Hospital data steward and medication-safety lead.	No training when task data cannot

	patient-level records.	computations access specified variables.		semantic validity, lawful access.	biased care-process data.		be interpreted reliably.
Local learning plane	Train and internally test the bounded task.	Local data produce a model, validation results, and update dossier.	Feeds protected exchange and retains a local comparator [6, 11].	Calibration, error analysis, temporal validity, reproducibility.	Overfitting, label error, unstable estimates, inappropriate target.	Local technical team and pharmacist reviewer.	Internal performance alone cannot authorize shared or clinical use.
Protected exchange plane	Restrict and authenticate shared artifacts.	Authorized updates and metadata leave the institution.	Connects local learning with shared computation [10, 12].	Leakage resistance, integrity, identity, policy compliance.	Inference, tampering, incompatible updates, unauthorized clients.	Privacy, security, and local authorization functions.	Locality or encryption alone cannot establish adequate protection.
Shared learning plane	Aggregate admitted contributions and create a versioned candidate.	Accepted updates enter protected computation; a candidate returns to hospitals.	Depends on orchestration, logging, and governance [13].	Reproducibility, contribution traceability, anomaly detection.	Poisoned aggregation, dominant-site influence, unrecorded changes.	Federated coordinator under consortium authority.	Aggregation produces a candidate, not an approved intervention.
Local qualification plane	Test whether the candidate is suitable for a participating hospital.	Candidate model enters local calibration, safety, workflow, and subgroup assessment.	May approve, restrict, reject, or request revision.	Local value relative to credible comparators and acceptable workflow consequences.	Aggregate improvement with local harm or excessive workload.	Hospital governance body and authorized clinical users.	No local use without task-specific qualification.
Governance plane	Allocate authority and manage lifecycle decisions.	Policies, audit records, incidents, challenges, and decisions flow across all planes.	Can suspend training, reject updates, roll back, or permit withdrawal.	Clear ownership, challenge rights, response authority, and documentation.	Diffuse accountability, delayed response, coercive participation.	Consortium board and participating institutions.	Technical operability cannot substitute for accountable governance.
Professional decision boundary	Preserve human authority for medication-related action.	Qualified output informs, but does not execute, the medication decision.	Connects local qualification with clinical workflow.	Appropriate interpretation, challenge, documentation, and override.	Automation bias, authority displacement, unreviewed action.	Authorized pharmacist, prescriber, or clinical professional.	No model output independently prescribes or changes therapy.

Local Training, Shared Updating, and Governance

Local training begins with an institution-specific task contract. The hospital must identify the medication-related target, clinical time point, eligible population, data period, intended user, available action, and unacceptable error modes. Preprocessing and outcome construction remain locally inspectable. The resulting update dossier should contain model and software versions, task compatibility, local validation, uncertainty, subgroup results, and known limitations. These requirements are proposed controls rather than a validated minimum dataset.

Shared updating requires governance before computation. Healthcare federated-learning governance remains underdeveloped across organizational, legal, data, technical, and accountability dimensions [14]. Participation agreements should therefore specify who may submit an update, which aggregation rule applies, who can inspect logs, how intellectual contributions are recognized, when a round may be paused, and how a hospital can challenge or withdraw. Model ownership must be separated from authority to approve clinical use.

The viability of a federation also depends on institutional relationships. Experience from healthcare federated initiatives indicates that trust, legal agreements, multidisciplinary cooperation, resources, and allocation of power can be more decisive than the learning algorithm itself [15]. A hospital with limited technical capacity may contribute important medication-safety experience while having less influence over architecture choices. Governance must prevent technical capacity from becoming the sole basis for authority or benefit.

Responsibility extends across the lifecycle. Responsible healthcare machine learning requires explicit stakeholders, failure detection, monitoring, and attention to patient harm rather than reliance on predictive performance alone [16]. In the proposed architecture, the participating hospital remains responsible for its local data, task definition, validation, workflow decision, and response to patient-facing risk. The coordinator is responsible for authenticated orchestration, versioning, aggregation integrity, and audit records but does not acquire clinical authority. Pharmacist medication-safety leadership is responsible for defining relevant errors, assessing review burden, challenging unsuitable outputs, and preserving escalation pathways.

Transparency and effectiveness must remain open to structured challenge [17]. Each shared update should therefore be contestable: another hospital may identify incompatible coding, a contradictory local effect, an inequitable subgroup pattern, or an unacceptable workflow consequence. Proposed responses include rejection of an update, restriction of a candidate, revision of the task contract, suspension of a training round, rollback to a prior version, and institutional withdrawal. These mechanisms are not evidence that harm will be prevented. They define the responsibilities and evidence that would need to be tested before a federation could be considered trustworthy for a specific medication-safety task.

Heterogeneity, Privacy Leakage, Security, and Equity

Hospital heterogeneity is not one technical defect. It includes differences in patient populations, formularies, prescribing policies, coding systems, laboratory units, missingness, medication-event prevalence, outcome ascertainment, and workflow. Hospitals may also expose different feature sets or differently organized machine-learning-ready data views, making ordinary parameter aggregation unreliable [18]. A federation must therefore document which variables and clinical concepts are comparable, transformed, absent, or locally specific.

Statistical alignment is equally important. Covariate shift may cause a shared model to fit one participating hospital poorly even when federation-wide performance appears acceptable [19]. Reweighting or personalization may mitigate particular forms of mismatch, but no single method resolves differences in labels, interventions, professional practice, or causal pathways. The proposed architecture consequently requires site-separated results and permits a hospital to retain or restore a local model.

Keeping records at home does not eliminate disclosure or attack risk. Information may be inferred from updates, parameters, outputs, logs, or repeated interactions. Federated systems may also be exposed to poisoning, backdoors, unauthorized clients, tampering, compromised coordinators, or denial of service [20]. Privacy and security must therefore be evaluated separately: privacy concerns what information can be learned about people or institutions, whereas security concerns whether the learning process, participants, and resulting artifacts remain trustworthy.

Equity also has patient-level and institutional dimensions. Federated fairness methods show that global optimization may distribute performance unevenly across population groups or participating clients [21]. The proposed equity gate therefore examines subgroup performance, representation, hospital influence, participation burden, and access to resulting benefits. Mathematical parity alone cannot establish health equity, and no hospital should be presumed to benefit merely because it contributes data-derived updates. These controls remain task-specific proposals requiring empirical and normative evaluation.

Evaluation beyond Aggregate Predictive Performance

A federated medication-safety candidate should be evaluated at the level at which harm or benefit could occur. Early clinical evaluation of artificial-intelligence decision support requires attention to the setting, intended users, workflow, human–AI interaction, system failures, and changes made during use [22]. Retrospective discrimination is therefore only one component of an evaluation program.

Transparent reporting should identify the target population, data sources, preprocessing, missing-data handling, model development, validation design, intended use, performance measures, and limitations [23]. Results should be presented separately for each hospital and clinically relevant subgroup. Aggregate results may be included, but they cannot substitute for local calibration, error analysis, or comparison with a credible site-only alternative.

Federated benchmarking provides a mechanism for moving a model to locally held evaluation data without transferring those records [24]. For medication safety, this approach should compare the shared candidate with local models, existing clinical rules, routine pharmacy review, and appropriate non-AI baselines. Evaluation should also test update compatibility, temporal drift, workload, alert burden, disagreement handling, privacy leakage, attack resistance, and rollback procedures.

Prospective protocols should prespecify intended users, input requirements, permissible actions, outcomes, failure handling, and model-change procedures [25]. Progression should be blocked when a candidate produces clinically material local deterioration, unacceptable subgroup disparities, unmanageable work redistribution, unresolved security weaknesses, or unclear responsibility. These are proposed transition conditions rather than universal thresholds. Their meaning must be defined for the medication-safety task before evaluation begins.

Implementation and Institutional Implications

Implementation is not completed when hospitals connect to an aggregation service. Complex health technologies may be rejected, abandoned, fail to scale, or become unsustainable after initial adoption [26]. A federation therefore requires continuing resources for local data mapping, technical maintenance, pharmacy review, governance meetings, incident management, requalification, and participant support.

The architecture should be implemented as a clinical service rather than an isolated model-delivery pipeline [27]. Its design must specify how an output enters pharmacy work, who receives it, what action is expected, how disagreement is

documented, and how users challenge or override the system. Infrastructure, professional acceptance, regulation, organizational readiness, and available resources may facilitate or obstruct implementation [28]. These conditions can differ substantially between hospitals and should not be treated as secondary to algorithm selection.

Technical performance does not establish clinical impact, generalizability, workflow value, or safe use [29]. Participation should continue only when each hospital can identify a plausible local benefit, maintain appropriate professional oversight, meet the consortium's privacy and security requirements, and exercise meaningful rights to challenge, restrict, pause, or withdraw.

Q1-Journal-Publication-Ready Figure Prompt

Create a publication-grade conceptual figure titled "Local Training, Shared Updating, Privacy, Heterogeneity, Security, and Equity" for the article "Learning Medication-Safety Lessons across Hospitals while Every Record Stays Home." Build the composition around one unmistakable central visual anchor representing the article's core proposed architecture, framework, model, theory, review synthesis, or decision pathway. Organize the surrounding modules according to these manuscript sections: Introduction: Collaboration under Data-Locality Constraints; Medication-Safety Learning across Heterogeneous Hospitals; Proposed Federated Collaboration Architecture; Local Training, Shared Updating, and Governance; Heterogeneity, Privacy Leakage, Security, and Equity; Evaluation beyond Aggregate Predictive Performance; Implementation and Institutional Implications; Limitations. Show directional relationships with solid arrows only where the selected evidence supports the connection and dashed arrows where relationships are proposed, uncertain, context-dependent, or require validation. Include explicit boundary markers separating model output from medication decision, technical performance from clinical usefulness, association from causation, automation from professional authority, and implementation from validated benefit where relevant. Use abstract scientific icons and labelled modules only. Do not use corporate logos, journal logos, software interfaces, copied scientific figures, copyrighted imagery, stock photography, approval stamps, or certification marks. Do not invent numerical values, effect sizes, performance scores, clinical outcomes, regulatory thresholds, or readiness scores. Use a clean white background, precise geometric alignment, professional sans-serif typography, thin uniform borders, strong visual hierarchy, a restrained colorblind-safe palette, and sufficient label size for journal-column reduction and 300 dpi export.

Exact Placement

Section "Implementation and Institutional Implications," immediately after the corresponding in-text callout.

Table 3 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for governance responsibilities, implementation conditions, and monitoring requirements for the article Learning Medication-Safety Lessons across Hospitals while Every Record Stays Home.

Table 3. Governance responsibilities, implementation conditions, and monitoring requirements for the article Learning Medication-Safety Lessons across Hospitals while Every Record Stays Home.

Governance domain	Responsible actor	Required authority	Documentation requirement	Monitoring indicator	Escalation trigger	Corrective action	Accountability boundary
Local task and data governance	Participating hospital and data steward	Approve lawful local use, task definition, access, and data mapping.	Provenance, variable definitions, data period, quality limitations, and change notices [14].	Missingness, coding changes, outcome stability, and mapping failures.	Material change in data-generating process or lawful-use basis.	Pause training, remap data, revise the task, or withdraw the update.	The hospital remains accountable for local data and task validity.
Medication-safety governance	Pharmacist medication-safety lead	Define the medication-related problem and permissible workflow use.	Error definition, intended user, action pathway, review burden, and escalation plan [5].	Output relevance, review burden, disagreement, and override patterns.	Unsafe recommendation, unclear action, or loss of professional challenge.	Restrict or stop use and revise the task contract.	Model output cannot acquire medication-decision authority.
Federated coordination	Consortium coordinator	Authenticate clients, orchestrate rounds, version candidates, and preserve logs.	Participation record, aggregation configuration, accepted and rejected updates, and model version [13].	Failed rounds, anomalous updates, version inconsistency, and unavailable audit records.	Unexplained update, integrity failure, or untraceable model change.	Reject the update, suspend aggregation, investigate, or roll back.	Coordination authority does not include local clinical approval.
Privacy and security	Independent privacy and security functions	Define threat models, controls, testing, and incident response.	Privacy parameters, cryptographic configuration, access logs, tests, and incident records [20].	Leakage signals, unauthorized access, poisoning indicators, and compromised clients.	Risk outside the approved threat model or confirmed compromise.	Isolate clients, suspend exchange, notify participants, remediate, or roll back.	Record locality is not evidence of complete privacy or security.

Equity oversight	Consortium and local equity reviewers	Examine patient- and hospital-level burdens, influence, and benefits.	Representation, subgroup results, participation resources, and benefit-distribution rationale [21].	Subgroup deterioration, persistent underrepresentation, or unequal institutional burden.	Unexplained inequity or participation without credible local benefit.	Restrict use, revise objectives, support affected sites, or pause participation.	Statistical fairness does not independently establish health equity.
Local qualification	Hospital governance body and authorized clinical users	Approve, restrict, reject, or withdraw a shared candidate.	Local comparator results, calibration, workflow assessment, subgroup analysis, and approval conditions [22].	Local performance, workload, override, safety signals, and drift.	Material local deterioration or unacceptable workflow consequence.	Reject, restrict, requalify, or restore a prior local version.	Consortium acceptance cannot substitute for local authorization.
Change and lifecycle control	Consortium board and participating hospitals	Approve update policies, suspension, rollback, and withdrawal.	Change requests, rationale, validation results, release notes, and rollback history [17].	Update frequency, unresolved challenges, recurring incidents, and requalification status.	Unvalidated material change or unresolved cross-hospital contradiction.	Delay release, request evidence, revert the candidate, or terminate the round.	A new version is a new candidate until locally requalified.
Sustainability and participation	Institutional leadership and consortium governance	Commit resources, negotiate responsibilities, and exercise exit rights.	Resource plan, agreements, service dependencies, benefit rationale, and exit procedure [15, 26].	Staffing, maintenance burden, participation continuity, and local benefit.	Unsustainable workload, loss of trust, or inability to meet obligations.	Renegotiate scope, provide support, reduce participation, or withdraw.	Continued participation must not be assumed or coerced.

Limitations

The architecture is constrained by the quality and meaning of hospital records. Electronic records may encode selection, measurement, missingness, and care-process bias; fairness also depends on explicit objectives and the inequities embedded in healthcare delivery. Proxy targets can appear technically useful while distributing benefit unequally [30-32].

Methodological leakage, inadequate comparators, weak external validation, and selective reporting can produce apparently strong but non-transportable models [33]. The architecture does not identify a universally optimal algorithm, aggregation rule, privacy parameter, fairness definition, or governance structure. It also does not establish that federation is preferable to local or centralized alternatives. Medication-specific prospective evaluation, privacy and attack testing, human-factors research, equity assessment, economic analysis, and longitudinal governance studies remain necessary before decision readiness can be considered.

Conclusion

Learning medication-safety lessons across hospitals while every record stays home requires more than distributed model training. This article proposes an architecture in which hospitals define bounded medication-safety tasks, retain control of patient-level records, share only authorized computational artifacts, create versioned shared candidates, and preserve local authority over qualification and use.

The contribution separates six functions: local data stewardship, local learning, protected exchange, shared updating, local qualification, and governance. It also distinguishes heterogeneity from privacy leakage, privacy from security, statistical fairness from health equity, predictive performance from workflow value, and model output from professional medication authority. These distinctions prevent record locality or aggregate performance from being treated as evidence of safety.

The proposed medication-safety lesson connects computational learning with provenance, uncertainty, local evidence, intended use, and an explicit decision boundary. Its purpose is to make shared learning traceable and contestable rather than to make clinical authority transferable. Hospitals must remain able to reject updates, restrict candidates, challenge shared conclusions, roll back versions, or withdraw.

The architecture is a conceptual contribution requiring empirical validation. It does not establish clinical benefit, regulatory acceptance, universal applicability, or deployment readiness. Its value lies in defining the scientific, professional, organizational, security, equity, and governance conditions under which federated collaboration could be evaluated responsibly for a specific medication-safety purpose.

ACKNOWLEDGMENTS: None
CONFLICT OF INTEREST: None
FINANCIAL SUPPORT: None
ETHICS STATEMENT: None

References

1. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med. 2020;3(1):119. doi:10.1038/s41746-020-00323-1
2. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nat Mach Intell. 2020;2(6):305-11. doi:10.1038/s42256-020-0186-1

3. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. Sci Rep. 2020;10(1):12598. doi:10.1038/s41598-020-69250-1
4. Li S, Liu P, Nascimento GG, Wang X, Leite FRM, Chakraborty B, et al. Federated and distributed learning applications for electronic health records and structured medical data: A scoping review. J Am Med Inform Assoc. 2023;30(12):2041-9. doi:10.1093/jamia/ocad170
5. Corny J, Rajkumar A, Martin O, Dode X, Lajonchère JP, Billuart O, et al. A machine learning-based clinical decision support system to identify prescriptions with a high risk of medication error. J Am Med Inform Assoc. 2020;27(11):1688-94. doi:10.1093/jamia/ocaa154
6. Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W. Federated learning of predictive models from federated electronic health records. Int J Med Inform. 2018;112:59-67. doi:10.1016/j.ijmedinf.2018.01.007
7. Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, Abidin AZ, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat Med. 2021;27(10):1735-43. doi:10.1038/s41591-021-01506-3
8. Dou Q, So TY, Jiang M, Liu Q, Vardhanabhuti V, Kaissis G, et al. Federated deep learning for detecting COVID-19 lung abnormalities in CT: A privacy-preserving multinational validation study. NPJ Digit Med. 2021;4(1):60. doi:10.1038/s41746-021-00431-6
9. Pati S, Baid U, Edwards B, Sheller M, Wang SH, Reina GA, et al. Federated learning enables big data for rare cancer boundary detection. Nat Commun. 2022;13(1):7346. doi:10.1038/s41467-022-33407-5
10. Sadilek A, Liu L, Nguyen D, Kamruzzaman M, Serghiou S, Rader B, et al. Privacy-first health research with federated learning. NPJ Digit Med. 2021;4(1):132. doi:10.1038/s41746-021-00489-2
11. Soltan AAS, Thakur A, Yang J, Chauhan A, D'Cruz LG, Dickson P, et al. A scalable federated learning solution for secondary care using low-cost microcomputing: Privacy-preserving development and evaluation of a COVID-19 screening test in UK hospitals. Lancet Digit Health. 2024;6(2). doi:10.1016/S2589-7500(23)00226-1
12. Froelicher D, Troncoso-Pastoriza JR, Raisaro JL, Cuendet MA, Sousa JS, Cho H, et al. Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption. Nat Commun. 2021;12(1):5910. doi:10.1038/s41467-021-25972-y
13. Mullie L, Afilalo J, Archambault P, Bouchakri R, Brown K, Buckeridge DL, et al. CODA: An open-source platform for federated analysis and machine learning on distributed healthcare data. J Am Med Inform Assoc. 2024;31(3):651-65. doi:10.1093/jamia/ocad235
14. Eden R, Chukwudi I, Bain C, Barbieri S, Callaway L, de Jersey S, et al. A scoping review of the governance of federated learning in healthcare. NPJ Digit Med. 2025;8(1):427. doi:10.1038/s41746-025-01836-3
15. Peltonen LM, Chomutare T. Federated learning's uncomfortable truth: Why human networks matter more than neural networks. J Am Med Inform Assoc. 2026;33(6):1236-40. doi:10.1093/jamia/ocag047
16. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: A roadmap for responsible machine learning for health care. Nat Med. 2019;25(9):1337-40. doi:10.1038/s41591-019-0548-6
17. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ. 2020;368. doi:10.1136/bmj.l6927
18. Thakur A, Molaei S, Nganjimi PC, Liu F, Soltan A, Schwab P, et al. Knowledge abstraction and filtering based federated learning over heterogeneous data views in healthcare. NPJ Digit Med. 2024;7(1):283. doi:10.1038/s41746-024-01272-9
19. Zhu H, Bai J, Li N, Li X, Liu D, Buckeridge DL, et al. FedWeight: Mitigating covariate shift of federated learning on electronic health records data through patients re-weighting. NPJ Digit Med. 2025;8(1):286. doi:10.1038/s41746-025-01661-8
20. Mothukuri V, Parizi RM, Pouriyeh S, Huang Y, Dehghantanha A, Srivastava G. A survey on security and privacy of federated learning. Future Gener Comput Syst. 2021;115:619-40. doi:10.1016/j.future.2020.10.007
21. Zhang F, Shuai Z, Kuang K, Wu F, Zhuang Y, Xiao J. Unified fair federated learning for digital healthcare. Patterns (N Y). 2024;5(1):100907. doi:10.1016/j.patter.2023.100907
22. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924-33. doi:10.1038/s41591-022-01772-9
23. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385. doi:10.1136/bmj-2023-078378
24. Karargyris A, Umeton R, Sheller MJ, Aristizabal A, George J, Wuest A, et al. Federated benchmarking of medical artificial intelligence with MedPerf. Nat Mach Intell. 2023;5(7):799-810. doi:10.1038/s42256-023-00652-2
25. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. Nat Med. 2020;26(9):1351-63. doi:10.1038/s41591-020-1037-7
26. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res. 2017;19(11). doi:10.2196/jmir.8775
27. Shaw J, Agarwal P, Desveaux L, Cornejo Palma D, Stamenova V, Jamieson T, et al. Beyond "implementation": Digital health innovation and service design. NPJ Digit Med. 2018;1(1):48. doi:10.1038/s41746-018-0059-8
28. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence applications in radiology: Hindering and facilitating factors. Eur Radiol. 2020;30(10):5525-32. doi:10.1007/s00330-020-06946-y
29. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195. doi:10.1186/s12916-019-1426-2
30. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. JAMA Intern Med. 2018;178(11):1544-7. doi:10.1001/jamainternmed.2018.3763
31. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med. 2018;169(12):866-72. doi:10.7326/M18-1990
32. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53. doi:10.1126/science.aax2342
33. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat Mach Intell. 2021;3(3):199-217. doi:10.1038/s42256-021-00307-0