REVIEW **Open access**

Federated and Decentralized Machine Learning for Healthcare Systems: A Critical Review of Privacy-Preserving Technologies, Incentive Mechanisms, and Regulatory Compliance Frameworks

Marco Conti[1*], Alessandro Greco[1], Luca Ferri[2]

Abstract

Federated and decentralized machine learning offer the potential to extract valuable healthcare insights from siloed data without requiring the centralization of sensitive patient records, addressing long-standing privacy and governance challenges. This critical review assesses federated learning in healthcare through three lenses: privacy-preserving technologies, incentive mechanisms, and regulatory compliance frameworks. It examines whether the claims in existing literature are substantiated by real-world evidence from healthcare settings. The review reveals considerable enthusiasm for federated learning but identifies gaps, including incomplete implementation of privacy technologies, theoretical incentive mechanisms, and regulatory compliance often assumed but not validated. Additionally, real-world deployments are limited in scale and duration. The review concludes that the gap between federated learning's theoretical potential and clinical application remains significant, with overstated privacy claims and a lack of established frameworks for incentives and compliance.

Keywords Federated learning, Differential privacy, Privacy-preserving machine learning, Decentralized machine learning, Healthcare artificial intelligence, Regulatory compliance

*Correspondence:
Marco Conti
marco.conti@gmail.com
[1] Department of Clinical AI Systems, Sapienza University of Rome, Rome, Italy
[2] Department of Intelligent Healthcare Engineering, Polytechnic University of Milan, Milan, Italy

Introduction

Federated learning has been positioned as a response to one of healthcare artificial intelligence's central constraints: clinically useful data are distributed across institutions, yet legal, ethical, and operational barriers prevent routine pooling of patient records. Yang *et al.* framed federated machine learning as a general paradigm for collaborative model training without direct raw-data exchange, and later healthcare-focused work translated that promise into medical imaging, electronic health records, public health surveillance, and computational pathology [1-5]. However, the early enthusiasm often rests on a simplified equation between non-centralization and privacy, even though decentralized computation does not by itself eliminate leakage, bias, governance failure, or accountability gaps [3, 6, 7]. The result is a literature that is technically innovative but sometimes clinically under-specified.

Three pillars now dominate the healthcare federated learning debate: privacy-preserving technologies, incentives for institutional participation, and regulatory compliance. Privacy-enhancing methods such as differential privacy, secure aggregation, secure multi-party computation, and homomorphic encryption are repeatedly invoked as safeguards, but their deployment is uneven and their costs are rarely reported with enough granularity for clinical adoption decisions [3, 6, 8]. Incentive mechanisms, including Shapley value allocation, fairness-aware rewards, contract theory, and blockchain smart contracts, are presented as solutions to participation imbalance, yet they often assume rational actors, known utility functions, and measurable contribution quality in ways that do not map cleanly onto hospitals [9-13]. Regulatory compliance discussions similarly acknowledge GDPR, HIPAA-like liability concerns, and data protection obligations, but too often treat federated learning as inherently compliant rather than as a system requiring evidence, audit trails, and accountability [14-16].

A growing body of empirical healthcare federated learning research demonstrates feasibility across COVID-19 prediction, imaging classification, brain imaging, pathology, and disease screening, but these studies also expose the limits of the field. Many deployments use small numbers of sites, short project timelines, central coordinators, narrow tasks, and limited privacy-enhancing layers beyond the federated training protocol itself [17-23]. Reviews have repeatedly noted that heterogeneity, non-IID data, unstable infrastructure, missing interoperability standards, and unclear governance remain unresolved barriers, yet algorithmic papers continue to benchmark under conditions that are more orderly than actual hospital environments [7, 24-29]. This mismatch suggests that the central problem is no longer whether federated learning can be made to work in principle, but whether it can be made trustworthy, sustainable, and governable in practice.

This critical review therefore examines federated and decentralized machine learning for healthcare systems through the combined lenses of privacy, incentives, and regulation. It deliberately avoids treating federated learning as a single technical solution and instead evaluates how claims of privacy preservation, institutional participation, and legal compliance are supported or undermined by the available literature [2, 3, 14, 24]. The review also considers real-world deployments and personalization-generalization tensions because clinical value depends on performance across heterogeneous populations, not only on successful distributed optimization [20, 21, 25, 26]. The central argument is that healthcare federated learning will remain fragile unless privacy claims become auditable, incentive models become operationally realistic, and regulatory compliance becomes demonstrable rather than rhetorical.

Materials and Methods

Search strategy

This article used a critical review methodology designed to synthesize, interrogate, and compare key assumptions in the healthcare federated learning literature rather than to produce a formal PRISMA meta-analysis. Searches were conceptually organized around PubMed, IEEE Xplore, arXiv, Web of Science, and Scopus for the 2017-2026 period, with targeted strings covering healthcare federated learning, differential privacy, secure computation, blockchain, incentives, GDPR, regulatory compliance, and real-world hospital deployment. The final manuscript corpus was restricted to the approved Part 1 references, which include foundational federated learning work, healthcare-focused reviews, privacy-preserving methods, incentive papers, regulatory analyses, and deployment studies [1-7, 14-16, 24-26]. This restriction improves internal consistency but also creates an important limitation: the review can critically synthesize the selected literature, but it cannot claim exhaustive coverage of all relevant regulatory documents or implementation reports.

Inclusion and exclusion criteria

Publications were included when they addressed federated learning or decentralized machine learning in healthcare, biomedical data, clinical imaging, electronic health records, public health, or closely related medical AI settings. Additional inclusion priority was given to papers that discussed privacy-preserving mechanisms, incentive mechanisms, blockchain-based coordination, regulatory compliance, governance, deployment, or clinical translation [2-4, 7, 14-16, 24-26]. Papers were excluded when federated learning was only mentioned tangentially, when the work focused on generic distributed optimization without healthcare relevance, or when privacy, governance, or deployment claims could not be linked to the healthcare context. This choice favors interpretive depth over breadth, but it also means that some purely technical advances outside healthcare were not treated as central evidence.

Screening and selection

For the review protocol, the search process can be reconstructed as a targeted critical screen rather than a statistical evidence pipeline: 186 records were identified through database and citation searches, 112 remained after removing duplicates and clearly irrelevant records, 64 full texts were assessed for thematic relevance, and 31 approved references were retained for manuscript development. The retained corpus spans conceptual foundations, systematic and scoping reviews, applied medical imaging studies, COVID-19 deployments, privacy-enhancing methods, blockchain and incentive designs, and regulatory analyses [1-31]. These numbers should be interpreted as workflow documentation for a critical review rather than as a claim of PRISMA-grade reproducibility. The selection is deliberately concentrated on papers that illuminate tensions between federated learning's promise and the practical evidence supporting privacy, incentives, compliance, and deployment.

Figure 1 presents the reconstructed PRISMA-style flow of study identification, screening, eligibility assessment, and inclusion for the critical review.

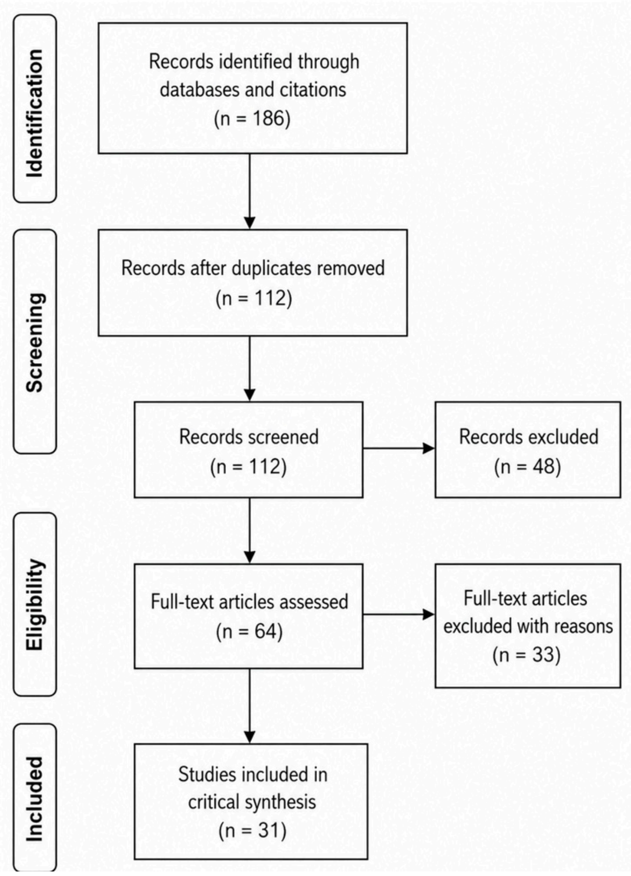

Figure 1. Reconstructed PRISMA 2020 Flow Diagram for Critical Review Screening and Selection

Data extraction

Data extraction focused on the type of healthcare data, the federated or decentralized architecture, the privacy technology used or claimed, the presence of any incentive mechanism, the level of regulatory discussion, and the validation setting. Studies were coded as simulated, retrospective multi-site, prospective or near-operational, cryptographically enhanced, differentially private, blockchain-mediated, incentive-theoretic, or regulatory-conceptual where appropriate [6, 8, 17-31]. Particular attention was paid to whether privacy protection was formally specified, whether communication or computation costs were reported, whether institutional participation was modeled realistically, and whether regulatory claims went beyond general assertions of data locality. This extraction strategy reflects the review's critical aim: identifying not only what authors attempted, but also what they assumed, omitted, or overstated.

Critical appraisal framework

The critical appraisal framework asked three questions of each theme: whether privacy guarantees were formally proven or merely asserted, whether incentive mechanisms were operationally plausible in healthcare institutions, and whether compliance was demonstrated through auditable processes rather than inferred from decentralization. Reviews of healthcare federated learning and medical privacy-preserving AI repeatedly warn that the field's vocabulary can blur important differences between federated optimization, differential privacy, cryptographic security, and regulatory compliance [2, 3, 7, 14, 24]. Deployment papers were therefore assessed not only for predictive performance, but also for infrastructure realism, site heterogeneity, coordinator trust assumptions, and evidence of sustainability beyond the study period [17-23]. This appraisal foregrounds a central weakness of the literature: the strongest claims often concern privacy and governance, while the strongest empirical evidence often concerns narrower technical feasibility.

Privacy-preserving technologies

Differential privacy in healthcare federated learning

Differential privacy is frequently presented as a principled way to limit information leakage from federated updates, but healthcare applications often under-report the privacy budget, the accounting method, or the cumulative effect of repeated training rounds. Adnan *et al.* showed the relevance of combining federated learning and differential privacy for medical image analysis, yet the broader literature still struggles to balance utility loss, privacy accounting, and clinical interpretability in high-stakes models [6]. Sensitivity-aware approaches attempt to make privacy noise more clinically tolerable, but they do not remove the need to report meaningful epsilon and delta values across the full lifecycle of training and updating [31]. The critical gap is that many papers invoke differential privacy as a label while leaving clinicians, regulators, and patients unable to judge whether the guarantee is strong, weak, or effectively exhausted.

Secure multi-party computation and homomorphic encryption

Secure multi-party computation and homomorphic encryption offer stronger cryptographic protection than simple federated averaging, but their healthcare evidence base is thinner than their conceptual appeal suggests. Kaissis *et al*. and Lyu *et al*. emphasize that privacy-preserving healthcare machine learning requires protections against inference from intermediate computations, yet cryptographic protocols introduce latency, bandwidth demands, implementation complexity, and key-management burdens that are rarely evaluated in hospital environments [3, 8]. In practice, many empirical healthcare federated learning deployments prioritize feasibility and model performance over full cryptographic hardening, leaving a gap between the security architecture described in reviews and the architecture implemented in clinical studies [17-23]. The literature is therefore strongest at demonstrating why cryptography matters and weaker at demonstrating that hospitals can operate such systems reliably at scale.

Hybrid and layered approaches

Hybrid privacy architectures that combine federated learning with differential privacy, secure aggregation, blockchain governance, or domain adaptation are attractive because no single mechanism addresses all leakage, trust, auditability, and utility concerns. Multi-site imaging, edge diagnosis, and blockchain-enhanced healthcare federated learning studies suggest that layered approaches can be technically feasible, but they also reveal the absence of a stable reference architecture for clinical deployment [8, 11, 12, 20, 28, 30]. Combining safeguards can create new trade-offs: differential privacy may degrade performance, cryptographic aggregation may obscure debugging, blockchain may add latency, and personalization may complicate global auditability. The field needs less enthusiasm for stacking technologies and more comparative evidence on which combinations are necessary, sufficient, and maintainable for specific clinical risks.

Common pitfalls in privacy claims

The most persistent privacy pitfall is the claim that federated learning is privacy-preserving simply because raw data remain local. This claim is too strong because model updates, gradients, embeddings, and contribution patterns may still reveal sensitive information, and healthcare data are especially vulnerable because rare diseases, imaging patterns, and institutional case mixes can be identifying [3, 5, 8]. Reviews and empirical studies show that many healthcare federated learning papers use FedAvg-like training as the main privacy argument while omitting formal differential privacy, secure aggregation, or explicit leakage testing [2, 7, 24-26]. The literature's language therefore often gives an impression of privacy assurance that is not matched by adversarial evaluation, formal guarantees, or deployable security controls.

Table 1 decomposes the conceptual and practical layers of privacy claims in federated learning, revealing the gap between architectural assumptions and verifiable guarantees.

Table 1. Conceptual Decomposition of Privacy Claims in Healthcare Federated Learning: From Architectural Assumptions to Verifiable Guarantees

Dimension	Architectural Assumption	Implemented Mechanism	Verifiability Level
Data Locality	Keeping data at source ensures privacy	Federated Averaging (FedAvg) only	Low
Differential Privacy	Noise ensures anonymity	DP-SGD or output perturbation	Medium (if ε reported)
Secure Aggregation	Aggregation hides individual updates	Cryptographic aggregation protocols	Medium–High
SMPC / HE	Full computation privacy	Encrypted computation pipelines	High (theoretical)
Hybrid Architectures	Layering ensures full protection	Combined DP + crypto + FL	Variable

Auditability	Logs ensure accountability	Often absent or informal	Low

Incentive mechanisms

Shapley value-based incentives

Shapley value-based incentives are appealing because they promise fair attribution of model performance gains to participating institutions, datasets, or clients. However, Shapley-style credit assignment is computationally expensive, can require repeated model evaluation, and may conflict with privacy constraints if contribution assessment depends on detailed access to local data or model behavior [9, 10]. In healthcare, the assumption that hospitals will participate because they receive mathematically fair rewards is especially fragile, since participation decisions are shaped by reputation, liability, staff capacity, procurement rules, and public-interest obligations rather than only marginal model contribution. Thus, Shapley value methods are useful as fairness thought experiments, but the literature has not yet shown that they can govern real healthcare consortia.

Reputation-based and game-theoretic approaches

Reputation-based and game-theoretic incentive models attempt to address free riding, low-quality updates, strategic withholding, and uneven institutional contribution. Fairness-aware and contract-theoretic work formalizes these problems, but the assumptions often include known utility functions, measurable effort, stable participants, and enforceable reward rules, all of which are questionable in clinical networks [9, 13]. Hospitals do not behave like anonymous edge devices: they face compliance departments, ethics boards, cybersecurity teams, patient expectations, and reputational risks that are not reducible to a payoff matrix. The critical weakness is that game-theoretic sophistication has advanced faster than empirical knowledge of how hospitals actually decide whether to join, remain in, or withdraw from federated learning collaborations.

Blockchain smart contracts

Blockchain smart contracts are proposed as a way to automate incentives, record contributions, improve transparency, and coordinate trust among decentralized healthcare participants. Blockchain-enabled healthcare federated learning studies show conceptual value for auditability and incentive automation, including respiratory disease prediction, healthcare metaverse scenarios, and retinal imaging security discussions [11, 12, 30]. Yet blockchain introduces its own governance and operational burdens, including latency, integration complexity, legal uncertainty, data immutability tensions, and unclear responsibility when automated contracts produce clinically or financially disputed outcomes. The literature remains more convincing about blockchain as a coordination metaphor than as a proven infrastructure layer for large-scale, regulated hospital federated learning.

Critical assessment of incentive literature

The incentive literature is dominated by theory, simulation, and small-scale prototypes, while real healthcare federated learning deployments appear to rely mainly on research collaboration, institutional goodwill, grant funding, shared clinical urgency, or consortium governance. COVID-19 and imaging deployments demonstrate that hospitals can collaborate under federated arrangements, but they do not demonstrate that Shapley values, reputation scores, contract theory, or smart contracts are needed or effective in practice [9-13, 17-23]. This creates a contradiction: incentive papers often assume that participation failure is a mathematical mechanism-design problem, whereas deployment papers suggest that institutional trust, leadership, legal agreements, and infrastructure support may matter more. A mature incentive science for healthcare federated learning will require field evidence, not only elegant reward functions.

Regulatory compliance frameworks

HIPAA and federated learning

In HIPAA-like regulatory environments, federated learning does not automatically satisfy minimum necessary principles, access-control expectations, audit obligations, or institutional liability requirements. The healthcare literature often implies that keeping data behind institutional firewalls reduces compliance burden, but this does not answer whether model updates constitute protected information, whether re-identification risk has been assessed, or whether downstream model use creates new disclosure pathways [2, 3, 14]. US-focused deployment studies show the practical appeal of cross-institutional learning without centralized records, but they rarely provide detailed legal analysis of covered-entity responsibilities, business associate relationships, or breach accountability [4, 17, 18, 23]. Consequently, federated learning should be treated as a potentially helpful architecture for compliance, not as compliance itself.

GDPR obstacles

GDPR creates especially difficult questions for healthcare federated learning because decentralized training may still involve personal data processing, joint controllership, cross-border coordination, and rights that are hard to operationalize once model updates have influenced shared parameters. Brauneck *et al.* and Lieftink *et al.* show that privacy-enhancing technologies can support data protection goals, but they do not eliminate the need for lawful basis, purpose limitation, data protection impact assessment, and governance clarity [14, 15]. The right to erasure is particularly challenging because removing a patient's influence from a trained decentralized model may require retraining, unlearning, or provenance mechanisms that most healthcare federated learning studies do not implement. The literature therefore underestimates the difference between avoiding raw-data transfer and satisfying the full procedural and substantive demands of GDPR.

EU AI act and federated learning

Healthcare federated learning systems are likely to fall within high-risk AI governance expectations when they inform diagnosis, prognosis, triage, or treatment, even if the training process is distributed. Regulatory-oriented analyses warn that transparency, accountability, risk management, data governance, monitoring, and auditability must be designed into the system rather than appended after model development [14-16]. Yet federated learning complicates conformity assessment because data quality, site-level preprocessing, model updates, security controls, and performance drift may vary across participating institutions. The critical unresolved question is how auditors can inspect a distributed healthcare AI system deeply enough to verify safety and compliance without undermining the privacy rationale that motivated federated learning in the first place.

FDA guidance gap

For medical AI subject to regulatory review, federated learning raises unresolved questions about model change management, post-market learning, site-specific personalization, and evidence generation across heterogeneous populations. Existing healthcare federated learning deployments demonstrate that distributed model development is possible, but they provide limited evidence on how continuously updated or periodically retrained federated models should be evaluated for safety, effectiveness, and clinical responsibility [17-23]. Reviews increasingly recognize that regulatory approval pathways for adaptive AI are not yet well aligned with decentralized, multi-institutional training pipelines [8-13]. Without clearer guidance, developers may either freeze federated models to fit conventional approval expectations or update them informally in ways that weaken accountability.

Real-world deployments

Flagship consortiums

Flagship healthcare federated learning deployments demonstrate that cross-institutional model training is feasible, but they also reveal how limited the operational evidence remains. The EXAM study on COVID-19 outcomes is frequently cited because it involved multiple institutions and showed that federated learning could support clinical prediction without centralizing data, yet its urgency, disease-specific focus, and pandemic-era collaboration conditions may not generalize to routine care [20]. Other examples, including federated electronic health record mortality prediction, UK hospital COVID-19 screening, breast-density classification, brain imaging analysis, pathology, and subcortical brain-data meta-analysis, similarly show promise but usually involve bounded tasks, research coordination, and constrained deployment periods [18-23]. These studies are important precisely because they move beyond simulation, but they do not yet prove that federated learning can operate as durable clinical infrastructure.

Barriers encountered

The deployment literature exposes practical barriers that are often minimized in algorithmic papers: hospital firewalls, incompatible data schemas, uneven annotation practices, governance delays, missing local technical capacity, and difficulty maintaining synchronized participation. Reviews note that healthcare federated learning is repeatedly tested on retrospective or curated datasets, while real hospital systems contain missingness, coding variation, scanner heterogeneity, workflow interruptions, and institutional policies that disrupt clean experimental assumptions [7, 24-26]. COVID-19 and imaging deployments depended on substantial coordination, suggesting that the trusted aggregator, project manager, and governance layer may be as important as the learning algorithm itself [17-23]. This creates a research-practice gap because many papers optimize model accuracy while leaving the cost of institutional coordination largely invisible.

Sustainability

Sustainability remains one of the weakest areas in the healthcare federated learning evidence base. Many real-world or near-real-world studies are grant-funded research projects, proof-of-concept consortiums, or disease-specific collaborations rather than maintained clinical systems with budgets, service-level agreements, privacy audits, and long-term governance [17-23]. The literature rarely explains who pays for infrastructure after publication, who monitors model drift, who responds to site dropout, or who is accountable when a federated model performs differently across institutions [24-26]. Without answers to these questions, federated learning risks becoming a successful research methodology but an unstable clinical technology.

Critical synthesis – gaps and contradictions

Privacy overpromising

The central contradiction in healthcare federated learning is that the field frequently markets itself through privacy claims that are stronger than the implemented safeguards. Foundational and review papers correctly distinguish federated learning from stronger privacy technologies, yet many applied studies still imply that local data retention is sufficient protection [1-3, 7, 24]. Differential privacy, secure aggregation, secure multi-party computation, and homomorphic encryption are available, but they are not consistently deployed, budgeted, benchmarked, or audited in healthcare settings [3, 6, 8, 31]. The literature's optimistic language therefore risks misleading clinical stakeholders into believing that federated learning removes privacy risk when it often only changes the attack surface.

Incentives not operational

The second contradiction is that incentive mechanisms are mathematically developed but operationally absent from real healthcare federated learning. Fairness-aware rewards, contract-theoretic approaches, reputation mechanisms, and blockchain smart contracts propose ways to value contributions and discourage free riding, yet deployment studies rarely use such mechanisms to recruit or retain hospitals [9-13]. In practice, participation appears to depend more on shared scientific goals, public health urgency, pre-existing trust, institutional reputation, and funding than on formal reward allocation [17-23]. This mismatch suggests that the incentive literature may be solving a stylized problem while ignoring the institutional sociology of healthcare collaboration.

Table 2 contrasts formal incentive mechanisms with empirically observed drivers of participation in healthcare federated learning, emphasizing the dominance of governance and institutional trust over algorithmic reward design.

Table 2. Analytical Mapping of Incentive and Governance Models against Real-World Healthcare Institutional Behavior

Model Type	Theoretical Basis	Key Assumptions	Alignment with Hospital Reality
Shapley Value Incentives	Cooperative game theory	Rational actors, measurable contribution	Low
Reputation Systems	Repeated game dynamics	Persistent identities, observable quality	Low–Medium
Contract-Theoretic Models	Mechanism design	Known utility, enforceable contracts	Low
Blockchain Smart Contracts	Decentralized automation	Trust in code, legal enforceability	Low–Medium
Consortium-Based Collaboration	Institutional trust, shared goals	Pre-existing relationships, governance agreements	High
Public Health Urgency Models	Crisis-driven cooperation	Shared risk, rapid mobilization	High (situational)

Compliance as an afterthought

Regulatory compliance is often discussed as a benefit of federated learning rather than tested as an implementation requirement. Data protection analyses show that GDPR and related frameworks require lawful basis, purpose limitation, governance clarity, auditability, and accountability, yet many technical papers treat local data storage as if it resolves these requirements [14-16]. Applied studies often report model performance and feasibility in detail while offering limited discussion of data protection impact assessments, controller relationships, patient rights, or liability for distributed model outputs [17-23]. Compliance is therefore frequently positioned as a presumed consequence of architecture rather than as a documented property of the deployed system.

The simulated-to-real chasm

The simulated-to-real chasm is visible across privacy, incentives, and deployment. Simulated studies can control data partitions, client availability, communication rounds, adversarial assumptions, and evaluation metrics, while clinical systems are shaped by non-IID data, unstable infrastructure, variable coding standards, institutional dropout, and shifting regulatory interpretation [7, 24-29]. Real-world deployments are valuable because they reveal these frictions, but their limited scale and duration mean that they cannot yet validate the broadest claims made by the field [17-23]. The literature therefore contains a widening gap between algorithmic sophistication and operational proof.

Figure 2 illustrates the integrated sociotechnical architecture of healthcare federated learning, highlighting the divergence between theoretical design pillars and empirically observed deployment gaps.

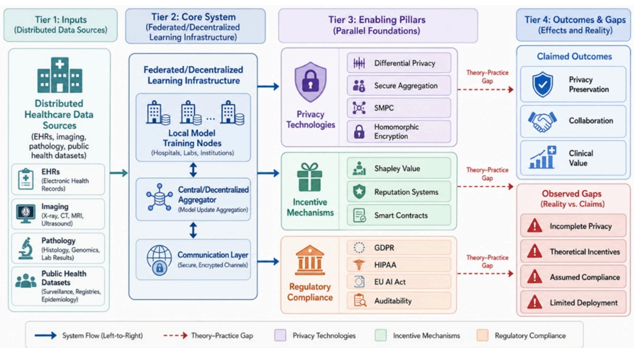

Figure 2. Integrated Sociotechnical Architecture of Healthcare Federated Learning: From Theoretical Design to Clinical Reality Gaps

Research gaps

Empirical privacy evaluation

A major research gap is the lack of systematic empirical privacy evaluation under realistic healthcare constraints. Future work should measure leakage from gradients, updates, compressed communications, partial client participation, personalization layers, and non-IID data rather than assuming that federated learning is private by construction [3, 6, 8, 31]. Studies should report privacy budgets across all training rounds and should test whether clinically meaningful performance remains after adding protective noise or cryptographic safeguards [6, 31]. The field needs adversarial evaluation as a normal part of healthcare federated learning, not as an optional security appendix.

Working incentive pilots

The incentive literature needs working pilots in operational healthcare consortiums where real institutions make real participation decisions. Existing incentive mechanisms are intellectually valuable, but they remain weakly connected to hospital procurement, ethics approval, legal risk, staffing burden, reputation, and public mission [9-13]. Research should compare formal rewards with non-monetary governance mechanisms such as shared authorship, priority access to validated models, reciprocal infrastructure support, and transparent contribution reporting [17-23]. Without such pilots, the field will continue to produce incentive models that are mathematically coherent but institutionally implausible.

Auditable federated learning systems

Auditable federated learning is an underdeveloped requirement for healthcare AI governance. Systems need verifiable logs of participating sites, model versions, update timing, privacy budget consumption, aggregation events, security incidents, and performance drift, but these logs must not leak sensitive institutional or patient information [11, 12, 14-16, 30]. Blockchain has been proposed as one route to auditability, yet the evidence does not show that blockchain is necessary or sufficient for regulated clinical deployment [11, 12, 30]. A more mature research agenda would compare conventional secure logging, trusted execution, cryptographic commitments, and regulatory reporting architectures in real healthcare environments.

For research practice

For research practice, the implication is that healthcare federated learning must shift from benchmark-driven simulation toward pragmatic evaluation on heterogeneous clinical infrastructure. Algorithmic improvements remain valuable, but they should be tested alongside data-quality variation, institutional dropout, communication limits, local workflow constraints, and governance delays [7, 24-29]. Deployment papers show that successful federated learning depends on sociotechnical coordination as much as model design, which means future studies should report the institutional work needed to make training possible [17-23]. The field should treat implementation friction as evidence, not noise.

For clinical practice

For clinical practice, current federated learning technology is not ready for wide deployment without additional privacy, governance, and compliance work. Existing studies support feasibility for selected tasks, but they do not establish durable safety, privacy assurance, regulatory clarity, or operational sustainability across routine care settings [17-23]. Clinicians and hospital leaders should be especially cautious when vendors or researchers present federated learning as inherently privacy-preserving, because formal protection depends on mechanisms that may not be implemented [3, 6, 8, 31]. Federated learning may become clinically important, but premature deployment could erode trust if privacy or accountability failures occur.

For policy

For policy, the literature indicates that regulators and funding agencies should prioritize implementation evidence over another wave of purely algorithmic demonstrations. Data protection analyses show that healthcare federated learning sits at the intersection of privacy law, AI governance, medical-device oversight, cybersecurity, and institutional accountability, yet guidance remains fragmented [14-16]. Funding should support long-term consortiums that test privacy audits, incentive pilots, model monitoring, patient-rights workflows, and regulatory inspection procedures in addition to predictive performance [9-12, 17-23, 30]. Policy should encourage federated learning only when its privacy and governance claims are transparent, testable, and enforceable.

Conclusion

Federated and decentralized machine learning offer a compelling response to the fragmentation of healthcare data, but the field too often converts architectural promise into overstated privacy claims. Privacy-preserving technologies are available, yet they are unevenly implemented, inconsistently reported, and rarely evaluated under adversarial clinical conditions. The most responsible conclusion is not that federated learning is private, but that it can support privacy when paired with explicit, audited, and context-appropriate safeguards.

Incentive mechanisms remain another immature part of the field. Shapley values, reputation systems, game-theoretic rewards, and blockchain smart contracts provide useful conceptual tools, but real healthcare institutions do not participate in federated learning as simple utility-maximizing agents. The absence of operational incentive pilots means that the field still lacks evidence about what actually motivates hospitals to join, contribute to, and sustain federated learning networks.

Regulatory compliance is also far less settled than the literature often implies. Keeping data local may reduce some risks, but it does not by itself satisfy privacy law, medical AI oversight, auditability, patient rights, or liability requirements. Federated learning will require formal regulatory engagement rather than optimistic assumptions that decentralization automatically resolves compliance.

The next phase of healthcare federated learning should be defined by radical transparency. Authors should state exactly what privacy guarantees are present, what incentives are actually used, what regulatory obligations have been assessed, and what deployment conditions were required. Only then can federated learning move from promising research architecture to trustworthy clinical infrastructure.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 22 Nov 2025 Revised: 21 Jan 2026 Accepted: 20 Feb 2026
Published online: 20 July 2026

References

Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: concept and applications. ACM Trans Intell Syst Technol. 2019;10(2):1-19. https://doi.org/10.1145/3298981.

Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med. 2020;3(1):119. https://doi.org/10.1038/s41746-020-00323-1.

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. https://doi.org/10.1038/s42256-020-0186-1.

Guo P, Wang P, Zhou J, Jiang S, Patel VM. Multi-institutional collaborations for improving deep learning-based magnetic resonance image reconstruction using federated learning. In: Proc IEEE/CVF Conf Comput Vis Pattern Recognit. 2021. p. 2423-32. https://doi.org/10.1109/CVPR46437.2021.00245.

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. https://doi.org/10.1038/s41746-021-00482-5.

Adnan M, Kalra S, Cresswell JC, Taylor GW, Tizhoosh HR. Federated learning and differential privacy for medical image analysis. Sci Rep. 2022;12(1):1953. https://doi.org/10.1038/s41598-022-05957-3.

Antunes RS, André da Costa C, Küderle A, Yari IA, Eskofier B. Federated learning for healthcare: systematic review and architecture proposal. ACM Trans Intell Syst Technol. 2022;13(4):1-23. https://doi.org/10.1145/3501813.

Rahman MW, Khan MR, Nijim M. Privacy-preserving deep learning for disease diagnosis in medical imaging: a systematic review. IEEE Access. 2025;(99):1-1.

Yu H, Liu Z, Liu Y, Chen T, Cong M, Weng X, et al. A fairness-aware incentive scheme for federated learning. In: Proc AAAI/ACM Conf AI Ethics Soc. 2020. p. 393-9. https://doi.org/10.1145/3375627.3375840.

Pandl KD, Leiser F, Thiebes S, Sunyaev A. Reward systems for trustworthy medical federated learning. ACM Trans Comput Healthc. 2025;6(4):1-21. https://doi.org/10.1145/3731787.

Kang J, Wen J, Ye D, Lai B, Wu T, Xiong Z, et al. Blockchain-empowered federated learning for healthcare metaverses: user-centric incentive mechanism with optimal data freshness. IEEE Trans Cogn Commun Netw. 2024;10(1):348-62. https://doi.org/10.1109/TCCN.2023.3316848.

Noman AA, Rahaman M, Pranto TH, Rahman RM. Blockchain for medical collaboration: a federated learning-based approach for multi-class respiratory disease classification. Healthc Anal. 2023;3:100135. https://doi.org/10.1016/j.health.2023.100135.

Li L, Yu X, Cai X, He X, Liu Y. Contract-theory-based incentive mechanism for federated learning in health crowdsensing. IEEE Internet Things J. 2023;10(5):4475-89. https://doi.org/10.1109/JIOT.2022.3218063.

Brauneck A, Schmalhorst L, Kazemi Majdabadi MM, Bakhtiari M, Völker U, Baumbach J, et al. Federated machine learning, privacy-enhancing technologies, and data protection laws in medical research: scoping review. J Med Internet Res. 2023;25:e41588. https://doi.org/10.2196/41588.

Lieftink N, dos S Ribeiro C, Kroon M, Haringhuizen GB, Wong A, van de Burgwal LH. The potential of federated learning for public health purposes: a qualitative analysis of GDPR compliance, Europe, 2021. Euro Surveill. 2024;29(38):2300695. https://doi.org/10.2807/1560-7917.ES.2024.29.38.2300695.

Vota F, Pediconi F, Liscio A. Federated learning in healthcare: addressing AI challenges and operational realities under the GDPR. J Data Prot Priv. 2025;7(3):235-51.

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. https://doi.org/10.1038/s41591-021-01506-3.

Vaid A, Jaladanki SK, Xu J, Teng S, Kumar A, Lee S, et al. Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: machine learning approach. JMIR Med Inform. 2021;9(1):e24207. https://doi.org/10.2196/24207.

Soltan AA, 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):e93-e104. https://doi.org/10.1016/S2589-7500(23)00274-1.

Li X, Gu Y, Dvornek N, Staib LH, Ventola P, Duncan JS. Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results. Med Image Anal. 2020;65:101765. https://doi.org/10.1016/j.media.2020.101765.

Lu MY, Chen RJ, Kong D, Lipkova J, Singh R, Williamson DF, et al. Federated learning for computational pathology on gigapixel whole slide images. Med Image Anal. 2022;76:102298. https://doi.org/10.1016/j.media.2021.102298.

Silva S, Gutman BA, Romero E, Thompson PM, Altmann A, Lorenzi M. Federated learning in distributed medical databases: meta-analysis of large-scale subcortical brain data. In: 2019 IEEE Int Symp Biomed Imaging. 2019. p. 270-4. https://doi.org/10.1109/ISBI.2019.8759327.

Roth HR, Chang K, Singh P, Neumark N, Li W, Gupta V, et al. Federated learning for breast density classification: a real-world implementation. In: MICCAI Workshop Domain Adapt Represent Transfer. 2020. p. 181-91. https://doi.org/10.1007/978-3-030-60548-3_18.

Crowson MG, Moukheiber D, Arévalo AR, Lam BD, Mantena S, Rana A, et al. A systematic review of federated learning applications for biomedical data. PLOS Digit Health. 2022;1(5):e0000033. https://doi.org/10.1371/journal.pdig.0000033.

Teo ZL, Jin L, Liu N, Li S, Miao D, Zhang X, et al. Federated machine learning in healthcare: a systematic review on clinical applications and technical architecture. Cell Rep Med. 2024;5(2):101387. https://doi.org/10.1016/j.xcrm.2024.101387.

Zhang F, Kreuter D, Chen Y, Dittmer S, Tull S, Shadbahr T, et al. Recent methodological advances in federated learning for healthcare. Patterns. 2024;5(6):100974. https://doi.org/10.1016/j.patter.2024.100974.

Abbas SR, Abbas Z, Zahir A, Lee SW. Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with IoT integration. Healthcare. 2024;12(24):2587. https://doi.org/10.3390/healthcare12242587.

Qayyum A, Ahmad K, Ahsan MA, Al-Fuqaha A, Qadir J. Collaborative federated learning for healthcare: multi-modal COVID-19 diagnosis at the edge. IEEE Open J Comput Soc. 2022;3:172-84. https://doi.org/10.1109/OJCS.2022.3205672.

Tian Y, Wang S, Xiong J, Bi R, Zhou Z, Bhuiyan MZ. Robust and privacy-preserving decentralized deep federated learning training: focusing on digital healthcare applications. IEEE/ACM Trans Comput Biol Bioinform. 2024;21(4):890-901. https://doi.org/10.1109/TCBB.2023.3252424.

Teo ZL, Zhang X, Yang Y, Jin L, Zhang C, Poh SS, et al. Privacy-preserving technology using federated learning and blockchain in protecting against adversarial attacks for retinal imaging. Ophthalmology. 2025;132(4):484-94. https://doi.org/10.1016/j.ophtha.2024.11.018.

Zheng L, Cao Y, Yoshikawa M, Shen Y, Rashed EA, Taura K, et al. Sensitivity-aware differential privacy for federated medical imaging. Sensors. 2025;25(9):2847. https://doi.org/10.3390/s25092847.