Federated Health: Secure AI Across Borders

Luca Bianchi
Asst Professor

Abstract

Federated AI and cloud data architectures enable secure international cooperation in global healthcare by delivering healthcare intelligence capabilities in a trustworthy manner across organizations, institutions, and countries. Federated AI enables sensitive data to remain with its owner, as AI algorithms can access the data without copying it. Cloud is the backbone technology of today's economy offering resiliency, scalability, and cost-effectiveness. Clouds help organizations comply with their risk and privacy frameworks while concentrating on their own core offerings. Federated AI and cloud data architectures deployed within this framework allow organizations to utilize the facility offered by cloud service providers while concentrating on their own core offerings. However, incorporating both technologies has its own challenges due to the inherent nature of these environments.

Federated AI and cloud data architectures enable secure collaboration in global health by providing healthcare intelligence capabilities in a trustworthy manner. Sensitive data can remain with its owner, and AI algorithms can access the data without copying it, as federated AI enables this functionality. Cloud is the backbone technology of today's economy, offering resiliency, scalability, and cost-effectiveness. Clouds help organizations comply with their risk and privacy frameworks while concentrating on their core offerings. Federated AI and cloud data architectures deployed within this framework allow organizations to make use of the facilities offered by cloud service providers while concentrating on their core capabilities. However, incorporating both technologies has distinct challenges due to the inherent nature of these environments.

Keywords: Federated AI, Cloud Architecture, Healthcare Systems, Data Privacy, Data Security, Secure Collaboration, Distributed Learning, Data Ownership, Cloud Computing, Scalability, Resilience, Cost Efficiency, Data Governance, Risk Management, Privacy Frameworks, Healthcare Analytics, Cross Border, Trusted Systems, Data Access, Cloud Services.

1. Introduction

The volume and sensitivity of data necessary for generating valuable AI solutions in healthcare raise strong demands for new collaboration paradigms and leveraging data across borders while preserving its ownership. Federated Learning (FL) and cloud data architectures are two concepts that address these needs, enabling AI learning without direct data exchange. However, their combination presents new challenges, such as the complexity of setting up a federated training process across jurisdictions with different data regulations. In addition, different classification models and architectural types may be set up to generate cloud-like infrastructures for hosting data for analysis while maintaining data sovereignty.

Priority research questions are the nature of the relationship between federated AI using FL training across jurisdictions and Cloud Data Architecture/Governance; security and compliance of data management operations for a sanitary state with an open, shared governance model that could enhance global surveillance of pandemics and pathogenic agents integrated with data federating. Machine learning for multijurisdictional federated clinical trials with regulators from different jurisdictions co-sponsoring the clinical trial.

2. Theoretical Foundations of Federated Intelligence in Healthcare

Federated intelligence in healthcare is underpinned by concepts, frameworks, and techniques embracing privacy-preserving collaboration across distributed, data-sensitive, cross-border environments. The interplay of federated AI, cloud data architectures, and federated learning bears on the implementation of AI, Machine Learning, and Advanced Analytics in regulated sectors such as Healthcare, Financial Services, and Energy. These concepts intersect at safety and security parameters. Governance, standards, and controls within Cloud Service Providers ensure Data Owners that data used as the basis for Knowledge relies on solid security, compliance, and risk management procedures. Data exchanged or offered as a service through third-party providers and shared for analytical purposes attracts Data Regulatory Authorities in every Business Use Case involving Data Subjects.

Data Sovereignty establishes the primary Data Owner's jurisdiction as the country where the Data are generated and stored, linking Data to the Privacy Policy and the Compliance-with-Legal-Requirements Actions of the main Data Owner's Cloud Service Provider. Data Sovereignty is crucial when Data is processed, transformed, and mined by third-party Service Providers and Cloud Operators. Data governance in a Healthcare scenario requires satisfying the Data Regulatory Authorities of both the Data Owner and the entities involved in the Data Analytics and Advanced Analytics phase, which set the guidelines for Data Consent, Data

Treatment, and Data Audit. Data provided for offering services, collaboration in Research and Development, joint innovation, training services, and academic articles — especially in regulated sectors like Healthcare and Financial Services — must comply with the Data Owner's and the Recipient's Privacy Policy.

2.1. Key Concepts and Frameworks for Federated Intelligence in Healthcare

Federated intelligence comprises distributed components that collaboratively deliver various intelligent capabilities such as visualization, analytics, or artificial intelligence while not exchanging data or sharing their internal models. Federated intelligence is a variant of distributed intelligence wherein data sources are distributed across locations, jurisdictions, or domains that are often controlled by different organizations. Such situations arise frequently in global healthcare and life sciences. Data are managed independently by the organizations that control them in compliance with local privacy regulations and agreements with external parties. Federated learning approaches aim to address situations in which centralization of data for modeling is not possible while providing proper privacy and security guarantees. AI governance enables strategic and operational decisions on AI-related concerns. Governance discusses who controls AI assets, makes decisions on AI risks, enables AI compliance, and assesses whether AI operations are trusted. Cloud data architectures encompass on-premises, public cloud, and hybrid environments. Public cloud providers are considered cloud-native; hybrid cloud deployments involve clouds and on-premises also. The factors of interest when appraising hybrid, multi-cloud or on-premises implementations include cost, latency, availability, disaster recovery, and regulatory considerations. Cloud-native architectures may still require data federation and discovery; when data reside in several clouds, indexing and cataloging facilities become fundamental.

Federated architectures, supported by adequate data quality, interoperability, and privacy guarantees, enable cross-border clinical trials, PANDEMIC surveillance, and joint AI models across several countries. The security and compliance dimensions highlight the necessity of addressing proper threat and risk assessment in federated settings where applications span cross-border jurisdictions. Compliance by Design addresses data protection and risk assurance during infrastructure and workflow design phases. The HIPAA and GDPR regulations are analyzed together with a broader set of global data protection and security regulations.

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 02 Issue: 02
RECEIVED: APRIL 19 REVISED: MAY 03 ACCEPTED: MAY 25 PUBLISHED: JUNE 17

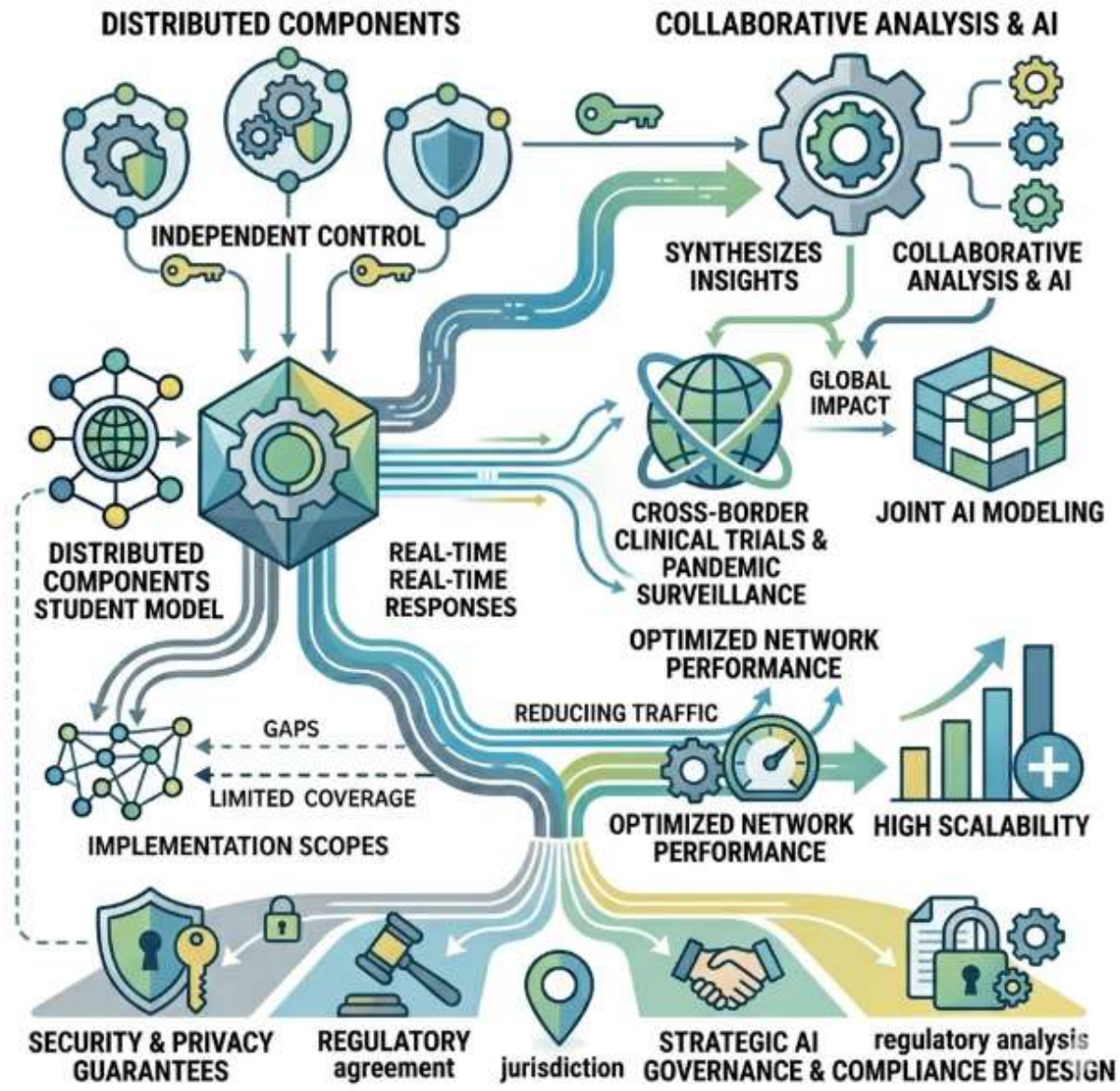


Fig 1: Cloud-Native Federated Learning with Regulatory Compliance by Design for Sensitive Global Data Ecosystems

3. Methodology

The Federated AI and Cloud Data Architectures methodology has been developed to address and fulfill the research objective. The methodology combines a general architectural design with several specialized designs that together guide the development of specific federated AI collaboration solutions. A proposed design pattern for Federated AI operationalization examines the main factors affecting design decisions in the operationalization of federated AI collaboration. These factors and the infrastructure, technologies, and policies required to design a solution for secure collaboration across data governance, security, and privacy jurisdictional borders are addressed. Clouds offer technologies to deploy infrastructure that can facilitate cross-border interaction, through enabling multiple organizations or regions to federate data and perform joint operations and mutual support without exchanging sensitive data. Hybrid clouds enable the use of on-premises or local servers together with public and community clouds. A high-level architectural design is first proposed, followed by the investigation of its elements and requirements, with cloud native data architecture and cloud mobile and multi-cloud strategies for reliable, cost-effective, and low-latency solutions.

The Federated AI and Cloud Data Architectures methodology has been developed to address and fulfil the research objective. The methodology combines a general architectural design with a set of specialized designs that together guide the development of specific federated AI collaboration scenarios. A proposed design pattern for Federated AI operationalization examines the main factors affecting design decisions in the operationalization of federated AI collaboration. The factors and the infrastructure, technologies, and policies required to design a solution for secure collaboration across data governance, security, and privacy jurisdictional borders are addressed. Clouds offer technologies to deploy infrastructure that can facilitate cross-border interaction by enabling multiple organizations or regions to federate data and perform joint operations and mutual support without exchanging sensitive data. Hybrid clouds enable the use of on-premises or local servers together with public and community clouds. A high-level architectural design is first proposed, followed by an investigation of its elements and requirements, including cloud native data architecture and cloud mobile and multi-cloud strategies for reliable, cost-effective, and low-latency solutions.

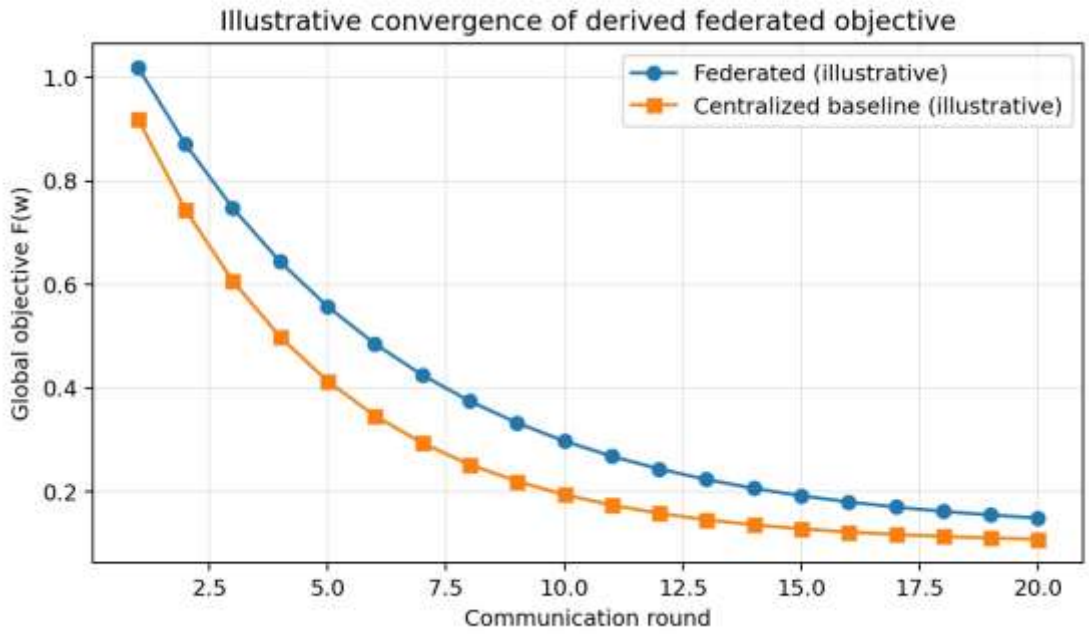
4. Objective of the Study

The objective is to provide insights into the three aspects of theory, method, and practice. Theory-building focuses on architecturing a federated-enabled AI system in healthcare. Method-

building proposes a framework for their support. In practice, the contribution extends to future regulatory approaches enabling a cloud-data-oriented healthcare governance in supporting federated learning scenarios. Harmonizing security controls across data domains, recognizing clear acceptable-use conditions, facilitating incident management across jurisdiction, and providing evidence of data protection and processing management are essential parameters for putting cloud data in the surface of a federated-ready AI system within the healthcare domain.

The research aim is supported by three enabling elements. First, Federated Learning can be an effective method built in theory, if the required precautions with respect to the quality, reliability and protection, availability, and accessibility dimensions of the data are all satisfied. Second, the assurance of the fulfilling of the requirements within the overall life-cycle of the healthcare service has, therefore, to be granted according to Compliance by Design principles while considering controls of several Privacy regulations, Data Governance policies and classification systems stemming from several regions of the world. Lastly, Federated Learning applied in the healthcare domain has the potential to support global-pandemic-surveillance endeavours as well as to allow-data-sharing-for-cross-border-operated-clinical-trials.

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 02 Issue: 02
RECEIVED: APRIL 19 REVISED: MAY 03 ACCEPTED: MAY 25 PUBLISHED: JUNE 17



Equation 1. Differential Privacy derivation

1.1 DP definition

A randomized mechanism $\mathcal{M}$ is (ε, δ)-differentially private if for any neighboring datasets D and D' differing in one record, and any measurable output set S,

$$\Pr[\mathcal{M}(D) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(D') \in S] + \delta$$

1.2 Sensitivity

For function $f(D)$, the L_2-sensitivity is:

$$\Delta_2 f = \max_{D,D'} \| f(D) - f(D') \|_2$$

where D and D' differ in only one record.

1.3 Gaussian mechanism

To privatize an update, add Gaussian noise:

$$\tilde{g} = g + \mathcal{N}(0, \sigma^2 \Delta_2^2 I)$$

where:

- g = true gradient or clipped update
- σ = noise multiplier
- I = identity matrix

If updates are clipped to norm C, then:

$$g_{\text{clip}} = g \cdot \min\left(1, \frac{C}{\| g \|_2}\right)$$

Then privatized update:

$$\tilde{g} = g_{\text{clip}} + \mathcal{N}(0, \sigma^2 C^2 I)$$

4.1. Research Aim and Intentions

Federated AI can augment existing models of global pandemic management and control by providing a modular framework for developing and deploying Distributed AI Applications (DAIA). These applications operate under a federated model of data access and utilization that enables collaborating organizational units to jointly train, validate, and improve Artificial Intelligence (AI) models in a secure environment without sharing sensitive data for internal operations and/or external collaboration.

The specific research aim is threefold: first, to propose a federated architecture for AI in global health and identify the design principles that underpin it; second, to outline a cloud-based data architecture that supports secure collaboration between organizational units deploying a federated learning ecosystem; and third, to assess how these architectures can impact the real-time response phase of public health emergencies. The broader goal is to advance the notion of Federated Intelligence in Healthcare and AI Governance Framework by integrating and

extending existing work on Federated Learning from an architecture viewpoint and addressing cloud-based data architecture for secure collaboration between organizational units that are participants of a Federated Learning ecosystem.

5. Research Summary

Federated AI enables partners to build joint solutions based on data from private or regulated environments without sharing the actual datasets, while federated cloud architectures allow cross-border data governance and handling under the respective law of each jurisdiction. The concepts are complementary and jointly facilitate the development of privacy-aware systems and applications for global healthcare collaboration and AI-related solutions that are auditable, trustworthy, comply with legislation, risk-aware, and configurable according to the security and compliance needs of high-regulated industries. They also help support the security of other deployments in more risk-tolerant sectors. Both concepts are essential foundations for enabling the federated construction of global surveillance and early-warning systems in global digital healthcare, where extreme territoriality in data protection combined with the massive scale of world pandemics inhibit conventional data aggregation and construction of centralised AI models for public health.

Addressing the complex needs of operationalising the concepts for business and academia, several aspects and dimensions are discussed based on the literature and the experience of applied research and project deployment in world APT and HPT. These aspects incorporate requirements for an effective and efficient operationalisation of federated AI, detailing the specificities of a dependent deployed model composed of federated partners, such as the condition of being legally recognised as a separate legal entity from its jurisdiction under the law of each jurisdiction during the entire project life cycle, enabling open direct communication with regulators anywhere in the world. It also describes the data architecture layers for cross-border collaboration of partners governed by different laws — where the exact opposite of federated AI must also be satisfied — embedded in both the business and global needs of operationalising a federated intelligent structure in a real-world environment and in live testing with global APT.

6. Architectural Principles for Federated AI in Global Health

The architecture underpinning federated AI in global health should be service-oriented, leveraging modular components that offer agreed core functions to all partners, with additional

functions provided as required. The components must be able to scale according to need and data volume. Collaboration must be trustworthy and comply with the established laws of all participants, and the architecture must therefore address the entire federated collaboration lifecycle, from conception through development and runtime.

In addition to the data architectures described above, more general architectural principles for federated AI in support of global health were identified and are summarized here for completeness. The federation of intelligence promises to offer new capabilities to the parties involved in its creation, but it cannot be guaranteed to deliver the expected benefits unless care is taken to ensure that the creation and operation of the federated model reflect the principles of Service-Oriented Architecture. A Service-Oriented Architecture exchanges interoperability for modified data ownership. In a genuine Service-Oriented Architecture, all data owners retain ownership of their data even though the computational execution is moved elsewhere; the data are still owned and controlled by the parties whose data they are when processes naturally connect}.

For a federated model able to deliver improved results for all participants, the model creation process must reflect this architectural principle. A community that agrees to create a federated model has a clear common goal: to improve the model itself and the useful processes that the model is supporting. The individual organizations may have competing interests, but the ability to create an improved model serves as a compelling rationale for the shared investment. While the operational deployment of the federated model will require careful management to deliver fairness, transparency, and trust, the creation process will benefit from the actors stepping back from their individual competitive interests and providing a valuable service to each other by contributing to a shared completion. One of the key principles of Service-Oriented Architecture is that the participation in a federated model must be fair. In addition, trust is needed between the operation of the federated model and the execution of the federated processes based on that model. The principle of transparent operation of individual models also extends to the trained federated model}.

Service-Oriented Architectures are also based on common standards, which allow for non-trivial interoperability. Data quality is an important contributor to this principle. The federation needs to address not just the modeling during creation but also the latency that is introduced when models are made accessible by multiple consumers.

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 02 Issue: 02
RECEIVED: APRIL 19 REVISED: MAY 03 ACCEPTED: MAY 25 PUBLISHED: JUNE 17

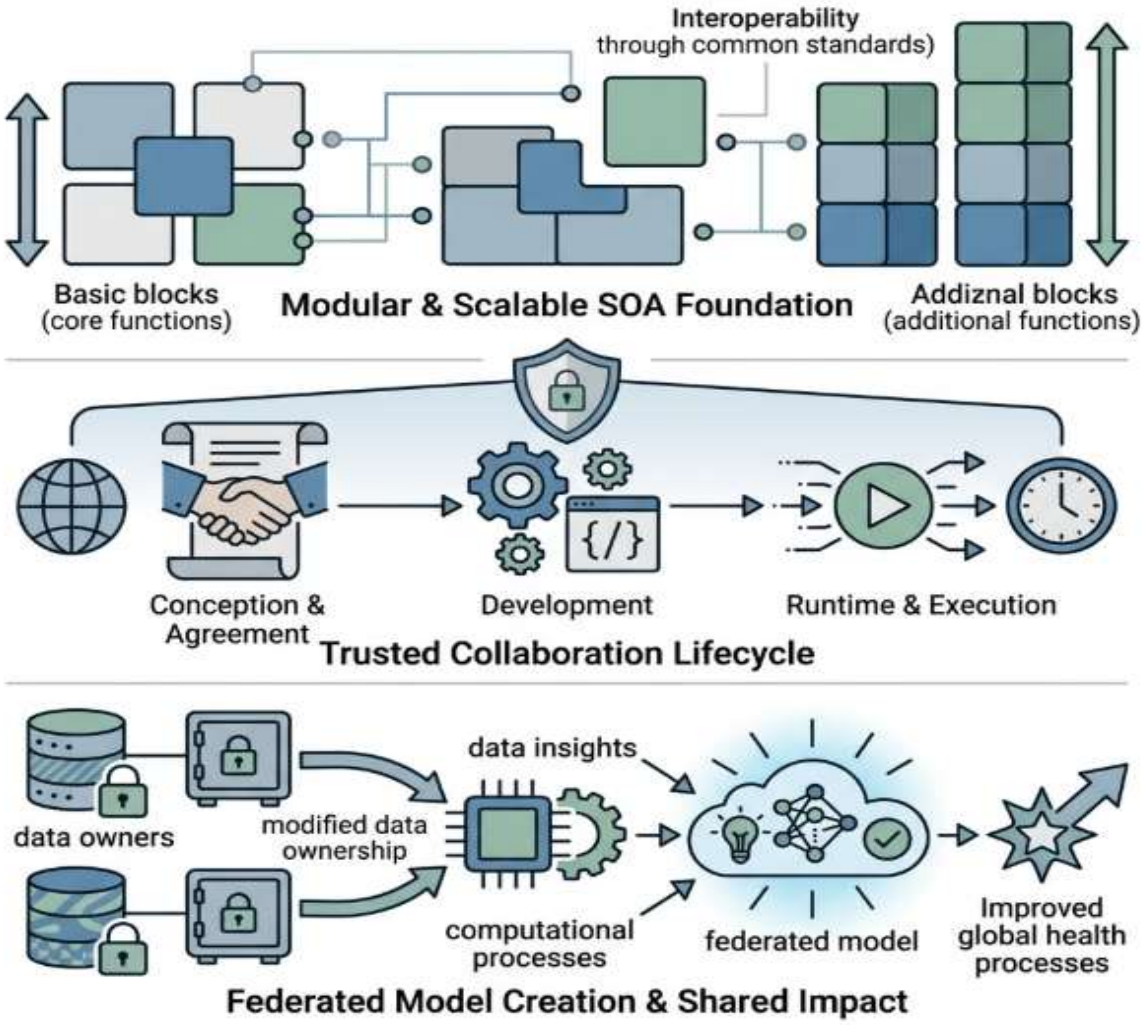


Fig 2: A Service-Oriented Architecture for Modular, Scalable, and Fair Collaboration

6.1. Data Sovereignty and Governance

Data sovereignty, governance, access control, and privacy regulations are recognized as critical for successful collaboration among entities from jurisdictions with different risk

tolerances and policy frameworks. Societal, cultural, and economic factors contribute to this diversity, and the related trust levels affect the willingness to share sensitive data across borders. Yet, these trust considerations are normally not described in the architecture or enabled by the framework. In federated learning, each jurisdiction is represented by a different organization; for example, each partner in a cross-border collaboration would have an operational jurisdiction that determines the federated processing risk tolerance and related control requirements. In the context of the Health Data Cooperatives and the DataHub federated analytics solutions, data are always accessible under the control of the organization that is originally managing the data. Therefore, sufficient trust for sensitivity and security of data and compliant access and use policies needs to be established.

Based on various organizational frameworks considerations, several options are offered. An Executive Steering Committee is usually formed for each operative grouping. Categories of roles, mandates, and specific decision rights are defined to facilitate the operational approval protocols effectively and efficiently. Finally, the Analysis and Decision Support Team (ADST) is proposed to support data analysis and provide an expert view, especially in situations where best practices or decision support tools are needed and when incidents or sensitive activities are to be communicated.

6.2. Privacy-Preserving Computation

Protecting patient privacy is paramount throughout the analytics lifecycle. Despite compliance with data sovereignty laws, model training may leverage sensitive patient information in jurisdictions where regulations impose stricter requirements. Differential privacy enables training of deep learning models on sensitive datasets without compromising patient privacy. To protect secret information during multiparty computations, secure multiparty computation (SMPC) guarantees privacy when data remains in the control of the owners and is not exposed to either the participant executing the computation or the party receiving the output. Recently proposed solutions for federated architectures with SMPC offer a combination of advantages (e.g. model training without sharing raw datasets enabling the computation of sensitive statistics).

Integrating federated learning techniques with differential privacy, SMPC, and homomorphic encryption supports the execution of private model training by balancing privacy, computing cost, and efficiency. Any two of the three privacy-preserving methods can be integrated, while using all of them results in a three-party device in which two parties perform

the conversions, while the third party executes privacy-assisted federated model training. When each client generates the noise to account for differential privacy, the two non-servers engage in an SMPC protocol for the underlying computations on both the private secret sharing and secret sharing of the noise. The three-party device executes the differential-privacy-preserved federated learning model training and communication among devices without disclosing the noise to any party.

Equation 2. Core Federated Learning equation

2.1 Local dataset definition

Assume there are K healthcare institutions, countries, or nodes.

Let node k hold local dataset:

$$D_k = \{(x_{k,i}, y_{k,i})\}_{i=1}^{n_k}$$

where:

- $x_{k,i}$ = input features
- $y_{k,i}$ = target label
- n_k = number of samples at node k

Total samples across all sites:

$$N = \sum_{k=1}^{K} n_k$$

2.2 Local empirical loss

At node k, define local objective:

$$F_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} \ell\left(w; x_{k,i}, y_{k,i}\right)$$

where:

- w = model parameters
- $\ell(\cdot)$ = loss function, for example cross-entropy or MSE

2.3 Global federated objective

$$F(w) = \sum_{k=1}^{K} \frac{n_k}{N} F_k(w)$$

Step-by-step derivation

Start from local loss:

$$F_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} \ell\left(w; x_{k,i}, y_{k,i}\right)$$

Weight each local loss by its share of data:

$$\frac{n_k}{N} F_k(w) = \frac{n_k}{N} \cdot \frac{1}{n_k} \sum_{i=1}^{n_k} \ell\left(w; x_{k,i}, y_{k,i}\right)$$

The n_k cancels:

$$\frac{n_k}{N} F_k(w) = \frac{1}{N} \sum_{i=1}^{n_k} \ell\left(w; x_{k,i}, y_{k,i}\right)$$

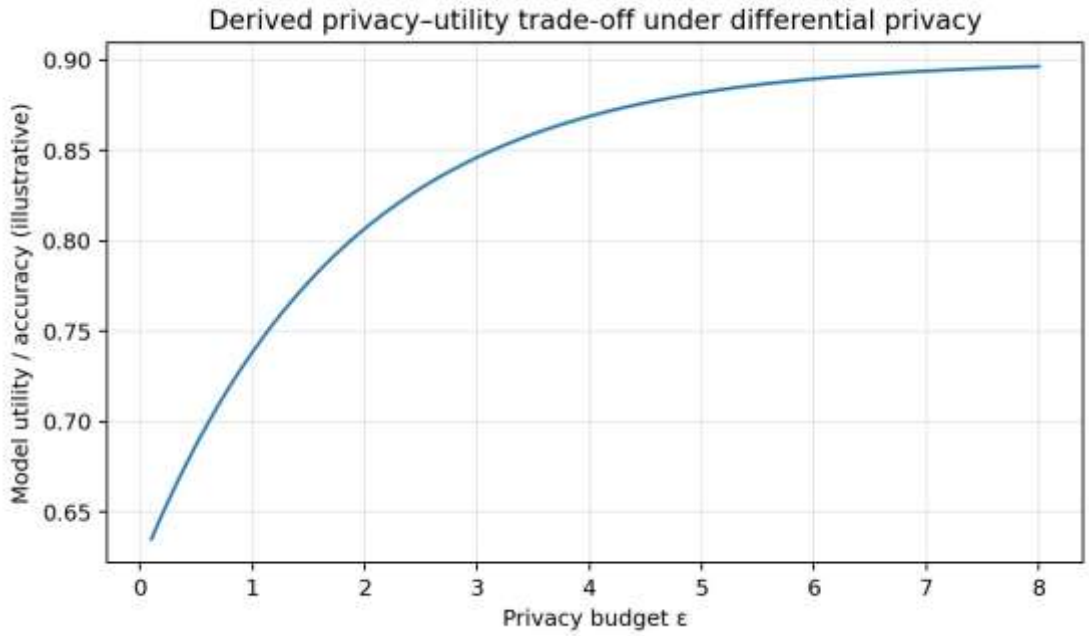
Now sum across all k:

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 02 Issue: 02
RECEIVED: APRIL 19 REVISED: MAY 03 ACCEPTED: MAY 25 PUBLISHED: JUNE 17

$$F(w) = \sum_{k=1}^{K} \frac{1}{N} \sum_{i=1}^{n_k} \ell\left(w; x_{k,i}, y_{k,i}\right)$$

So:

$$F(w) = \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \ell\left(w; x_{k,i}, y_{k,i}\right)$$



7. Cloud-Based Data Architectures for Secure Collaboration

Several property and privacy requirements must be satisfied for secured collaboration within Federated Intelligence Architecture. First, it should never enforce all institutions to collaborate leveraging a central, single-point-of-failure cloud instance prone to latency, resiliency, and cost issues as shown. For multinational institutions sensible to security concerns, it is recommended to deploy the infrastructure on-premises conforming to property-assurance

requirements. Others comfortable with hosted solutions can use cloud providers to meet property-assurance concerns. Second, even institutions not qualified to implement property-assurance controls should still be able to securely collaborate by sending data to another trusted institution that meets property-assurance requirements. Multi-Cloud strategies addressing latency, resiliency, and cost also need to be supported. Third, collaboration is only required in the zone where data is injected. A data-federation strategy that delegates data searching, indexing, and discovery to a metadata cloud catalog and indexer is also crucial for architecture democratization. Such a metadata catalog allows detection of data granted access by competent institutions for a federation process without replicating data in the cloud.

Consequently, the proposed Federated Intelligence Architecture implements shared-no-sensitive-data federated, locality-aware, and delegated-data-federation controls. A generic metadata cloud catalog and data-index delegator performing automatic data searching and indexing are also part of the architecture. Specific implementations, functionalities, and controls depend on the operational cloud model adopted: On-Premises Cloud or Hybrid Cloud. The On-Premises Cloud model is the most secure since all-sensitive-data controls are applied to all-data sources. The Hybrid Cloud model is the most flexible since it allows friendly-cloud hosts to cooperate and collaborate either deferring control to a third institution guaranteeing such controls or enabling latent-security-for-latency features and accepting an automatic risk-level assessment during setting.

7.1. Hybrid and Multi-Cloud Strategies

For latency-sensitive applications involving on-premises resources, a hybrid deployment strategy ensuring local data sharing, health economics analysis, and AI-driven decision support may be optimal. In other settings, a comprehensive hybrid strategy, with data exchanged between public clouds, can be employed to minimize costs and redundancy while meeting performance and compliance requirements.

For geographic replication and data segregation, a multi-cloud strategy may be adopted, with workloads divided across several providers that meet jurisdictional data governance controls. Evaluating the combination of different service types and providers allows for maximization of governance models, resilience against provider outages, and optimization of latency and costs. Government-funded workloads may be processed with regions operated by national cloud service providers or within their boundaries, in accordance with regulations. Sensitive workloads may be tackled with providers endorsed by the data classification scheme,

preventing low-classified data from being processed in high-classified environments, while the alliance cloud shares less critical information, avoiding costs related to data transfer set by certain commercial cloud service providers.

7.2. Data Federation and Indexing Mechanisms

Data federation increases agility for cross-service analytics and requests across clouds, lowering the cost and risk presented by data copying. The data federation capability provides a single query interface for different types of data stored across multiple services that may be hosted in different public clouds. Nevertheless, federating data from various locations does not obviate the need for data governance: it simply adds a data management practice centered around data indexing and cataloguing. This aspect of data management adds internal checks and balances to ensure that data is used wisely and cost-effectively.

All data should be properly indexed (i.e., tagged) so that it can be easily discovered through a metadata catalog. Metadata catalogs—such as AWS Glue Data Catalog, Google Cloud Data Catalog, or Apache Atlas—are services that allow users to discover and search for data stored in clouds and on-premises data sources that are integrated with data-lake systems using connectors provided by various cloud vendors or third-party software vendors. The metadata catalog can help users discover existing data without needing to send technical requests to data owners to know where critical data exists. Using such catalogs helps ensure that data is not stored or copied into different cloud locations without proper management and approval. When organizations with diverse autonomous functionalities need to exchange data, the backup mechanism is to publish an index of interested data in the metadata catalog, including sample data and structural representations of the data stored.

Topic	Derived equation
Federated learning	$F(w) = \sum_{k=1}^{K} \frac{n_k}{N} F_k(w)$
Local training	$w_k^{t,e+1} = w_k^{t,e} - \eta \nabla F_k\left(w_k^{t,e}\right)$
FedAvg aggregation	$w^{t+1} = \sum_{k=1}^{K} \frac{n_k}{N} w_k^{t+1}$

Topic	Derived equation
Compliance-aware aggregation	$w^{t+1} = \frac{\sum c_k n_k w_k^{t+1}}{\sum c_k n_k}$
Differential privacy	$\Pr[\mathcal{M}(D) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(D') \in S] + \delta$
DP update	$\tilde{g}_k = \bar{g}_k + \mathcal{N}(0, \sigma^2 C^2 I)$
SMPC secure sum	$S = \sum_{j=1}^{m} r_j$
Communication cost	$C_{\text{round}} = 2K_t Pb$
Latency	$T_{\text{round}} = \max_k \left(t_k^{down} + t_k^{train} + t_k^{up}\right) + t^{agg}$
Multi-cloud availability	$A = 1 - \prod_{i=1}^{m} (1 - a_i)$
Risk	$R = \sum_i P_i \, I_i$
Data quality	$Q = \sum_j \omega_j \, q_j$

8. Security and Compliance in Global Healthcare Environments

Security is paramount in healthcare, where compromised data can predict medical conditions and impact individuals. Data breaches are a constant threat, so risk must be constantly monitored through assessments, formal documentation, risk groups, and incident response plans. Privacy-enhancing technology must reduce privacy risks to an acceptable level for all stakeholders, and privacy-by-design concepts should secure data throughout its lifecycle. Security and compliance concerns in a global healthcare environment must encompass risk, privacy, and compliance factors.

Globally distributed, heterogeneous, and multifactorial contexts complicate the development of tailored security solutions. Different classes of Data Trust Service Providers (DTSPs) must offer heterogeneous services according to the specific information offered.

Security and control are achieved via mutual authentication, role-based access control, encryption of sensitive information, accurate data source identification, control policies designed by administrators and service providers, and partners determined by confidential business agreements and the actual use of information. Crowdsourcing Data Trust Services should guarantee user privacy, service reliability, security during data collection, and protect against user-created malicious information. Trustworthiness and reputation should always be considered; partners and end users must be motivated to act as expected.

High-availability systems must support the full range of components, security requirements must be clearly defined according to the operational model, and network configuration should protect against remote attacks. Security controls must be integrated into the application's processes, and conditions for accepting the risk should consider results from tests, ongoing monitoring of assets, and identification of potential threats. Specific systems enable automated generation of Data Protection Impact Assessments compliant with the General Data Protection Regulation, and security architecture and procedures must be permanently in place, including incident response and disaster recovery planning. Security proofs form a crucial component of the analysis and help maintain the security of systems and operations.

Equation 3. Gradient-based training at each institution

3.1 Gradient of local loss

Differentiate $F_k(w)$ with respect to w:

$$\nabla F_k(w) = \nabla \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \ell \left(w; x_{k,i}, y_{k,i} \right) \right)$$

Using linearity of differentiation:

$$\nabla F_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} \nabla \ell \left(w; x_{k,i}, y_{k,i} \right)$$

3.2 Local SGD update

At round t, the server sends global model w^t to all clients.

Client k initializes:

$$w_k^{t,0} = w^t$$

For local epoch $e = 0,1, \dots, E-1$:

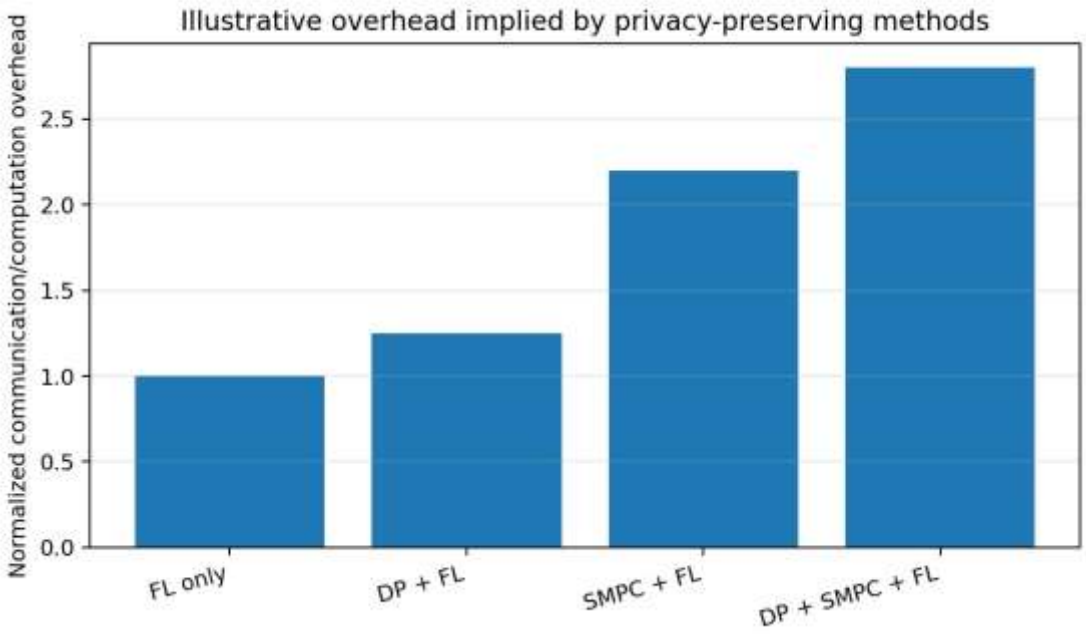
$$w_k^{t,e+1} = w_k^{t,e} - \eta \nabla F_k\left(w_k^{t,e}\right)$$

where η is learning rate.

If mini-batch B_k^e is used:

$$w_k^{t,e+1} = w_k^{t,e} - \eta \cdot \frac{1}{\left|B_k^e\right|} \sum_{(x,y)\in B_k^e} \nabla \ell\left(w_k^{t,e}; x, y\right)$$

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 02 Issue: 02
RECEIVED: APRIL 19 REVISED: MAY 03 ACCEPTED: MAY 25 PUBLISHED: JUNE 17



8.1. Threat Modeling and Risk Assessment

Threat modeling and risk assessment consider the security and privacy requirements associated with federated and cloud-native healthcare architectures. These security features reflect the sensitivity of the data processed by the systems and the data subjects' right to data protection. Risk, cybersecurity, and regulatory compliance have been identified as key non-functional requirements in the design of federated AI systems in healthcare. The former rely on data stewardship processes to ensure that sensitive data is secure, with restrictions to access, sharing, and usage. The latter is related to the adequate alignment of data processing to the regulation of the supervisory authority in the location of the participating entities.

Threat modelling starts with identifying valuable assets and data — e.g., personal health information (PHI) under HIPAA jurisdiction, special data under GDPR like genetic data, location information that may disclose sensitive attributes under the Data Protection Act in the UK — threats to them, and mitigation options, and quantifying the possibility of data breaches. A risk assessment enumerates the different privacy and security risks associated with a federated or cloud-native deployment: compromise of raw data, redundant data storing in on-premises

archives that are more prone to breach, access to privacy-violating neural models, hostile operation from trusted nodes, and risks from cloud service providers. Privacy risk management uses precautionary architecture design backed by technical control measures and non-technical factors. Data breach consequence analysis considers how combination of innocuous data attributes could reveal sensitive data.

8.2. Compliance by Design (HIPAA, GDPR, and Beyond)

Compliance by Design considers regulatory compliance, risk and privacy by design from the outset. A mapping of controls to architecture and operational workflows follow. Referencing three major privacy regulations, namely HIPAA (Health Insurance Portability and Accountability Act), GDPR (General Data Protection Regulation), and the HITECH (Health Information Technology for Economic and Clinical Health Act), the analysis examined additional guidelines that comply with industry standards such as ISO27000, NIST, COBIT, or FUNCERT—working to avoid compliance issues rather than suffering the consequences.

International cooperation across jurisdictions exposes an ecosystem to different laws, regulations or standards, and inequity in maturity levels of privacy laws, technologies, and providers. In this case, privacy is required at least at GDPR level, it being one of the most stringent standards. Privacy by design or default of GDPR was created to minimize the risk of sensitive information exposure, misuse, or household; promoting adequate organizational measures towards privacy. By implementing these principles from inception, it must be able to process data securely while in a virtual machine located anywhere. A lateral case would be HIPAA compliance during cross-jurisdiction digital verification of private health data from sensitive health records disclosed to a border culture during full-on pandemic travel in a cybersecurity context. Compliance through design avoids last-minute measures to solve non-compliance or paves the way towards level and suity.

9. Federated Learning Algorithms and Privacy Techniques

A variety of algorithms to perform model training in federated settings have been proposed in the literature; some are specifically built for federated architecture, while others can be adapted. In a federated framework, the data resides in multiple remote locations. The models are trained across multiple jurisdictions, using data that do not physically cross the geographical boundaries. On the other hand, risk analysis should ensure that data can be used across jurisdictions to the degree explicitly permitted and that the training process does not lead to

exposure of sensitive data. Hence, it is important to specify controls aligned with regional regulations (e.g., HIPAA, GDPR) for data access and use, as well as recommendations for the design of the federated learning algorithms and their privacy-preserving modules.

Federated learning relies on training a single global model across decentralized data that remains localized on users' devices or in data centers of different organizations. However, it typically disregards data localization regulations. In many cases, these regulations are intended to protect privacy and confidentiality, so although training a model directly on the sensitive information would violate local legal requirements, federated learning may be an acceptable alternative. The optimization processes may be operated by an external organization that does not have direct access to the private data residing within the hosting institutions' environments. Consequently, these data may even remain inaccessible to the other training partners. The recent proposal of a federated learning approach that satisfies GDPR in a cross-jurisdiction scenario constitutes an important advance in this domain.

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 02 Issue: 02
RECEIVED: APRIL 19 REVISED: MAY 03 ACCEPTED: MAY 25 PUBLISHED: JUNE 17

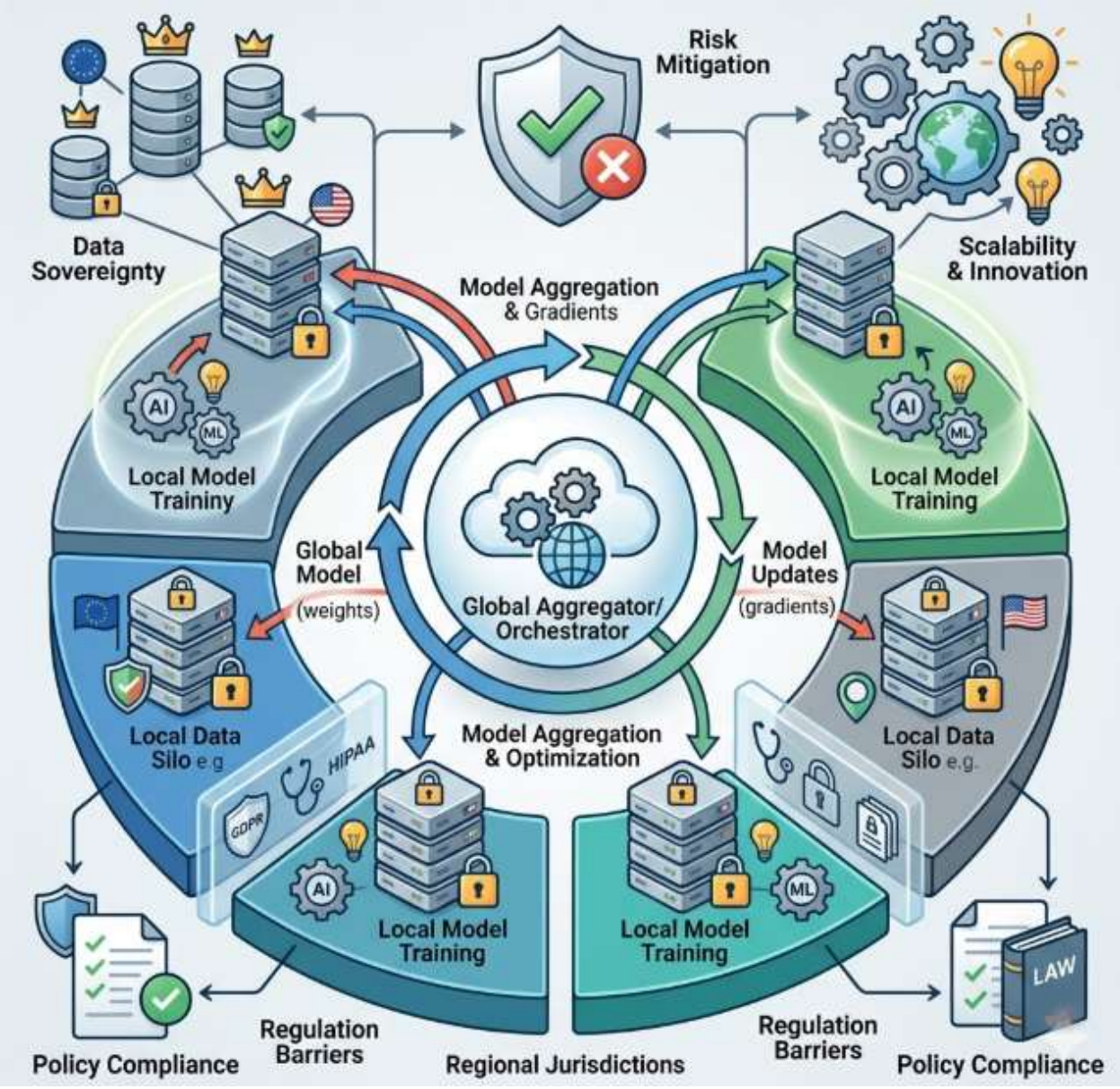


Fig 3: A Regulatory-Aligned Federated Learning Framework: Balancing Data Localization, Multi-Jurisdictional Privacy, and Robust Risk Mitigation

9.1. Model Training Across Jurisdictions

The model training phase in a federated learning (FL) context occurs across jurisdictions. Individual jurisdictions hold their respective sub-models, which are updated periodically based on new local training data. These updates can be a sub-model's set of weights or gradients learned from a local dataset. The aggregation of these updates can occur at the pseudonymizing layer if one exists or at an institution that has been entrusted with the rest of the cloud architecture and the necessary governance to process the eligible data in training mode. The institution processes only the pseudo-identified update in a Controlled Environment (CE), which provides a full physical separation between the data held in the domains and the infrastructure used to provide such (sub)model training. These updates are then transferred to another CE in a different jurisdiction, where integration and synthesis take place. When the Transfer Learning (TL) strategy is put in place, the Cloud or such provider can constitute the Supervisory Model, whose role is to offer a more-up-to-date and previous model able to make a TL for similar predictions, being at least a multinational or regional provider. Such a model must make predictions besides other regions.

Managing the training process of such a multi-jurisdictional Artificial Intelligence (AI) and taxonomy scenario is crucial for the predictive capacity of such, regarding concepts, semantics, architectures, and cybersecurity. This is because each jurisdiction has different domains in place to allow safe modeling, and the information necessary for the model might not be within the same jurisdiction for a better prediction. The concern is that each jurisdiction has its own policies, copious legal differentials that manage data protection, data sources, and penalties, and this quite different and inconsistent regulation needs to be followed to provide such risk-free supervised training and prediction services.

9.2. Privacy-Preserving Methods (DP, SMPC, FL)

The set of privacy-preserving methods reviewed comprises Differential Privacy (DP), Secure Multi-Party Computation (SMPC), and the combinations of both DP and Secure Vector Sharing with Federated Learning (DP + SV with FL). Clinical applicability is assessed by mapping computational overhead and data requirements against the previous analytic framework, especially within the context of global pandemic surveillance.

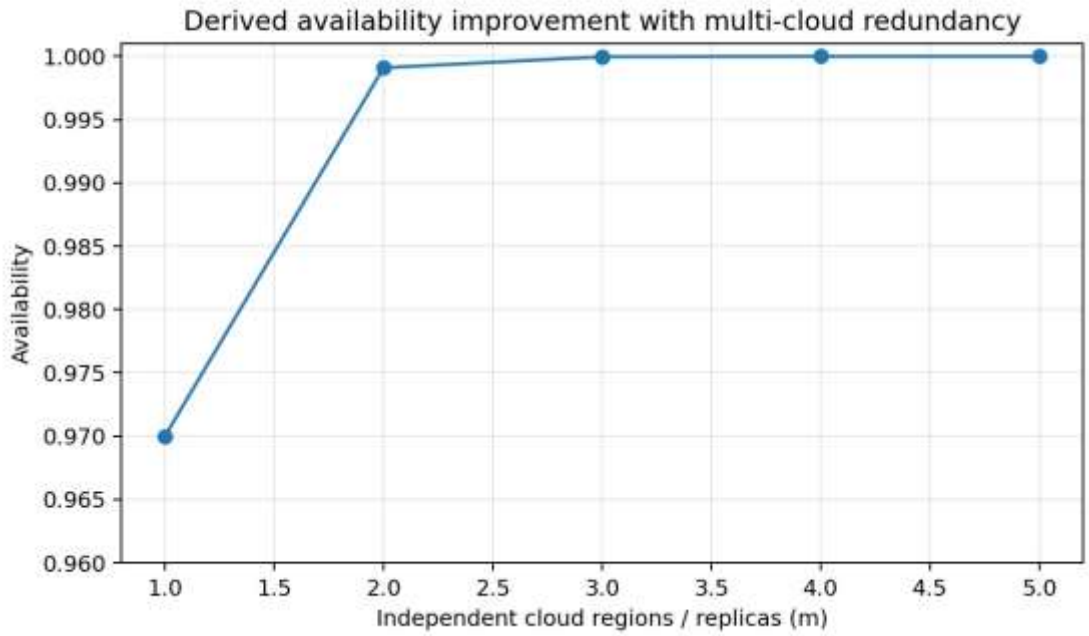
Differential Privacy adds a level of indistinguishability to individual records in output tables or models. Before sharing public population-health data, a noise mechanism is introduced

to obscure the influence of an individual's data. This has an optimal impact on deep-learning tasks, particularly those with suitable amounts of noisy training data. However, the obfuscation limits clinical applicability as the noise impacts inference: diagnostic-supporting outputs must be decisive or risk(calcify)ing detected conditions. Additionally, the noise must be continuously re-evaluated to determine clinical utility, especially as the frequency and scope of sharing increase.

Secure Multi-Party Computation allows several parties to compute a function over their inputs without revealing those inputs. The technique distributes the computation of the desired output Y between all parties. The outcome can be obtained, along with any other output of interest, without exposing any single party's input. While use has been limited in deep learning, preliminary work indicates an ability to train DNNs accurately. Other tasks were similarly successfully completed using Coordinate Rotation, and the efficient application of SMPC techniques facilitates massive expansions to training across additional parties. The greatest restriction on the clinical applicability of SMPC arises from the complete dependence of the output on all input datasets.

The combination of Secure Vector Sharing with Federated Learning and Differential Privacy generates good detection results when extracting key features. The key advantage lies within differing training data from different organizations, as the technique has strong robustness to subdomain shifts with proper sharing schemes. Limitations of this method mainly relate to its communication overhead.

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 02 Issue: 02
RECEIVED: APRIL 19 REVISED: MAY 03 ACCEPTED: MAY 25 PUBLISHED: JUNE 17



Equation 4. FedAvg aggregation derivation

After local training, each node sends updated weights w_k^{t+1} or weight change Δw_k^t.

4.1 Weighted averaging

Global aggregation:

$$w^{t+1} = \sum_{k=1}^{K} \frac{n_k}{N} w_k^{t+1}$$

Why this formula?

Each site contributes proportionally to its dataset size.

If hospital A has 10,000 cases and hospital B has 1,000 cases, A should usually influence the model more.

4.2 Equivalent update form

Suppose local model after training is:

$$w_k^{t+1} = w^t + \Delta w_k^t$$

Substitute into aggregation:

$$w^{t+1} = \sum_{k=1}^{K} \frac{n_k}{N} (w^t + \Delta w_k^t)$$

Distribute:

$$w^{t+1} = \sum_{k=1}^{K} \frac{n_k}{N} w^t + \sum_{k=1}^{K} \frac{n_k}{N} \Delta w_k^t$$

Since w^t is constant w.r.t. k:

$$w^{t+1} = w^t \sum_{k=1}^{K} \frac{n_k}{N} + \sum_{k=1}^{K} \frac{n_k}{N} \Delta w_k^t$$

Because $\sum_{k=1}^{K} n_k = N$, we get:

$$\sum_{k=1}^{K} \frac{n_k}{N} = 1$$

Therefore:

$$w^{t+1} = w^t + \sum_{k=1}^{K} \frac{n_k}{N} \Delta w_k^t$$

10. Data Management and Quality in Federated Healthcare Networks

Ensuring reliable, high-quality data for analytics, diagnostics, and model training poses a complex challenge in federated network designs. The dynamic addition and removal of members can lead to inconsistency and noise in the data, undermining its integrity. Moreover, lack of adherence to data quality standards may result in problematic datasets or misinformation propagation within the network. To address these issues, data quality metrics are defined, along with associated cleaning and standardization procedures to be followed by each partner. These requirements are delineated in the data governance model.

The model also includes a metadata model with a federated approach to semantic harmonization, enabling partners from different jurisdictions to achieve data interoperability even in the absence of a centralized domain model or data lake. The model allows the distribution of authority to create and manage vocabularies according to the needs of particular domains. The metadata model supports the semantic representation of the different data sources used by the federated network to facilitate analytics and the decision-making process. In addition, a semantic harmonization framework based on the semantic reconciliation process of the agent-based data brokering system is considered to support interoperability across different jurisdictions.

10.1. Data Quality, Cleaning, and Standardization

Uniformly high-quality data is crucial for effective data analytics within federated networks, yet varying data quality levels can hinder valuable insights. Addressing these challenges requires defining a basic set of reliability metrics—preferably quantitative rather than qualitative—that data providers must meet. Necessary cleaning procedures should then be established to improve data quality for analytics. These measures should also be supplemented with standardization procedures that allow data shared across partners to comply with particular standards (e.g., common data models and data structures).

Whether for data integration, internal or external data sharing, or data publication, metadata holds a fundamental role in enabling interoperability across federated networks. Therefore, developing a federated metadata model is essential when pursuing data pools exposed to common metadata-based data discovery. Consequently, the importance of metadata and its harmonization should also be recognized for settings involving natural language processing. In such cases, the automatic examination of unstructured healthcare documents can lead to the identification of federated network members dealing with common keywords.

10.2. Metadata and Semantic Harmonization

Data generated by diverse partners in federated learning networks may differ in structure, definitions, and semantics, leading to challenges in data analysis and model validation. For example, clinical data often have specific structure and integrity constraints defined in ontologies. Even when a consistent structure is adopted, partners may represent the same concept with different values or labels—e.g., “cardiac arrest” versus “asystole,” or human-readable values requiring translation into coding systems—leading to a semantic mismatch. If not addressed, such inconsistencies can severely hinder model accuracy, analytical reliability, and clinical applicability. Metadata describing the content, structure, provenance, and quality of the clinical data can help overcome these challenges.

Modern approaches to metadata management focus on descriptive, discoverable metadata. Such metadata usually leverage resource description framework (RDF) representations and can be collected in a centralized catalog to enable centralized queries and queries from different clouds. Ads hosting a federated learning cloud or multiple learning clouds can store the description of the resources hosted within their systems in a common metadata catalog. Nodes from other clouds can then discover these resources and their semantics. To semantically harmonize clinical data across federated partners, a combination of standardization and cross-mapping approaches can be adopted. First, synthetic clinical data can be represented in a common semantic model (e.g., OMOP), guaranteeing structural and value consistency in the clinical datasets produced by the partners. However, due to the different sources of the data, guarantees of integrity and consistency at ingestion and writing time cannot be satisfied. A semantic cross-mapping would then be executed on the synthetic data at periodic intervals to ensure that all clinical datasets would be kept consistent across partners.

11. Case Studies and Use Cases

The preceding sections outlined the architecture, security, and data aspects of federated AI deployment in global health. This section discusses two case studies highlighting use cases that exemplify the implications of these designs. The first relates to using federated AI for global pandemic surveillance—a deployment scenario already partly implemented and operated in real-time since 2020. Indeed, realizing effective pandemic surveillance requires collaboration across jurisdictions, touching multiple sensitive areas such as data privacy, public health response, and information credibility. A federated architecture enabling collaborative analyses with sensitive data while controlling, limiting, and regulating data sharing strengthens such aspects and

mitigates the clash between different priorities. Concretely, detecting pandemic patterns, mutations, and new diseased strains, as well as their relation to epidemiological data, is crucial for tackling these issues. Recognizing that the spread of a pandemic does not take into consideration political boundaries, a global effort is required that, while ensuring privacy and control of the data, can provide alerts and generate models for a better and more effective response. The second use case covers conducting a clinical trial for a new treatment or vaccine across countries, involving a corresponding tests dataset.

The first use case presents hybrid use of federated learning with cloud solutions hosted in a multi-cloud environment. This setting aims to minimize latency during the training of a COVID-19-related interest model, preserve data privacy, and avoid the transfer of highly sensitive information. The scenario covers the definition of a COVID-19 clinical dataset and the model to be trained during the pandemic. The second use case is based on a traditional federated learning approach that takes advantage of networks' resources during a clinical trial with more than eight nodes involved.

Equation 5. Cross-jurisdiction constrained aggregation

Let $c_k \in \{0,1\}$ indicate whether institution k is currently compliant/eligible for aggregation.

Then only eligible nodes participate:

$$w^{t+1} = \frac{\sum_{k=1}^{K} c_k \, n_k w_k^{t+1}}{\sum_{k=1}^{K} c_k \, n_k}$$

Interpretation

- $c_k = 1$: node satisfies policy and can contribute
- $c_k = 0$: node excluded for legal, consent, or security reasons

11.1. Global Pandemic Surveillance

The COVID-19 health crisis has highlighted the importance of quick responses from the research community. In response to the global pandemic, the European Union has initiated funding to connect Artificial Intelligence (AI) experts around the world in research that will ensure a unified response to future occurences of both COVID-19 and any future pandemics. AI

systems in healthcare typically use a federated learning (FL) training approach, wherein sensitive training data does not leave the data source site/portion, but model synchronization and averaging occurs via a central aggregator node. Such AI systems typically require diverse high-quality labelled data that can effectively represent the concept of "generalization" in the trained AI model. Hence, the need for collaboration among different healthcare institutions while protecting data privacy and complying with regulatory policies within countries/jurisdictions.

Consider the use case of disease outbreak forecasting, such as global pandemic surveillance during COVID-19, wherein the training data is highly unbalanced and data quantity is very less. Individuals from the older age group have suffered badly from COVID-19 infection, and that age group is now being considered for vaccination. From federated data-sharing, we can predict the likelihood of vaccination among the older age group of different nations and states.

Method	Privacy strength	Computation cost	Communication cost	Clinical utility impact
FL only	Medium	Low	Medium	High
DP + FL	High	Medium	Medium	May reduce utility if noise is strong
SMPC + FL	Very high	High	High	Usually preserves model utility better than strong DP
DP + SMPC + FL	Very high	Very high	Very high	Strong privacy, highest operational complexity

11.2. Cross-Border Clinical Trials

The secure and policy-compliant handling of sensitive health data is a prerequisite for international trials involving drugs, vaccines, or other medical devices. These multi-site studies involve partnerships between pharmaceutical companies and research institutions and examine the efficacy of a treatment using a clinical population exposed to a specific health threat. Consider the case study of a trial supported by the European Union in conjunction with research centres in various countries to evaluate a COVID vaccine opened to all. A multi-cloud deployment is expected, with on-premise setups for research teams focusing on clinical trials and randomization, and specific partners conceiving and developing the vaccine, thus creating the

need for strict governance through a medicine agency. Several points are examined: data quality across different countries, impact of linguistic differences, and coping strategy for supporting European Union collaboration with African countries.

For local storage, data infrastructures on information technology partners' premises dedicated to epidemiological studies can be considered, given the expected amount of data to be managed and the systemic orientation of the project. The use of cloud data architectures or storage facilities is another possible option in countries where a powerful cloud is available. A third possible solution is to use public clouds in a hybrid configuration, where data remain in the countries of origin to meet data protection laws but, for information technology companies, allow a more extensive usage of multi-cloud federated business agreements with a smaller cloud present in the country. Provided these workflows are respected, data policy outside the European Union can be adapted according to the laws of the country in which the data reside and according to the needs of the regulatory authority.

12. Operational and Governance Considerations

Intelligent and sustainable global health collaboration requires attention to operational, organizational, and governance issues. Such considerations are undertaken for each of the three high-level uses described above: global pandemic surveillance, cross-border clinical trials, and the conduct of global digital health services. In addition to establishing formal operational structures for individual initiatives or other "alliance-like deployments," consideration is given to longer-term federated Intelligence operational arrangements that would allow an evolving collection of initiatives to address specific global healthcare priorities collaboratively and at pace with suitable mechanisms for incident response and continuity.

Collaborative efforts in these domains share at least two common governance dimensions. The first is an associated body that organizes longer-term decision-making, with responsibility for updating the policy framework and approving foundational documents for upcoming initiatives. The second is an incident response team whose role is to deliver planned services and coordinate a response to crises such as the COVID-19 pandemic. Group-specific governance considerations complement these horizontal topics, which cut across diverse collaborative efforts.

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 02 Issue: 02
RECEIVED: APRIL 19 REVISED: MAY 03 ACCEPTED: MAY 25 PUBLISHED: JUNE 17

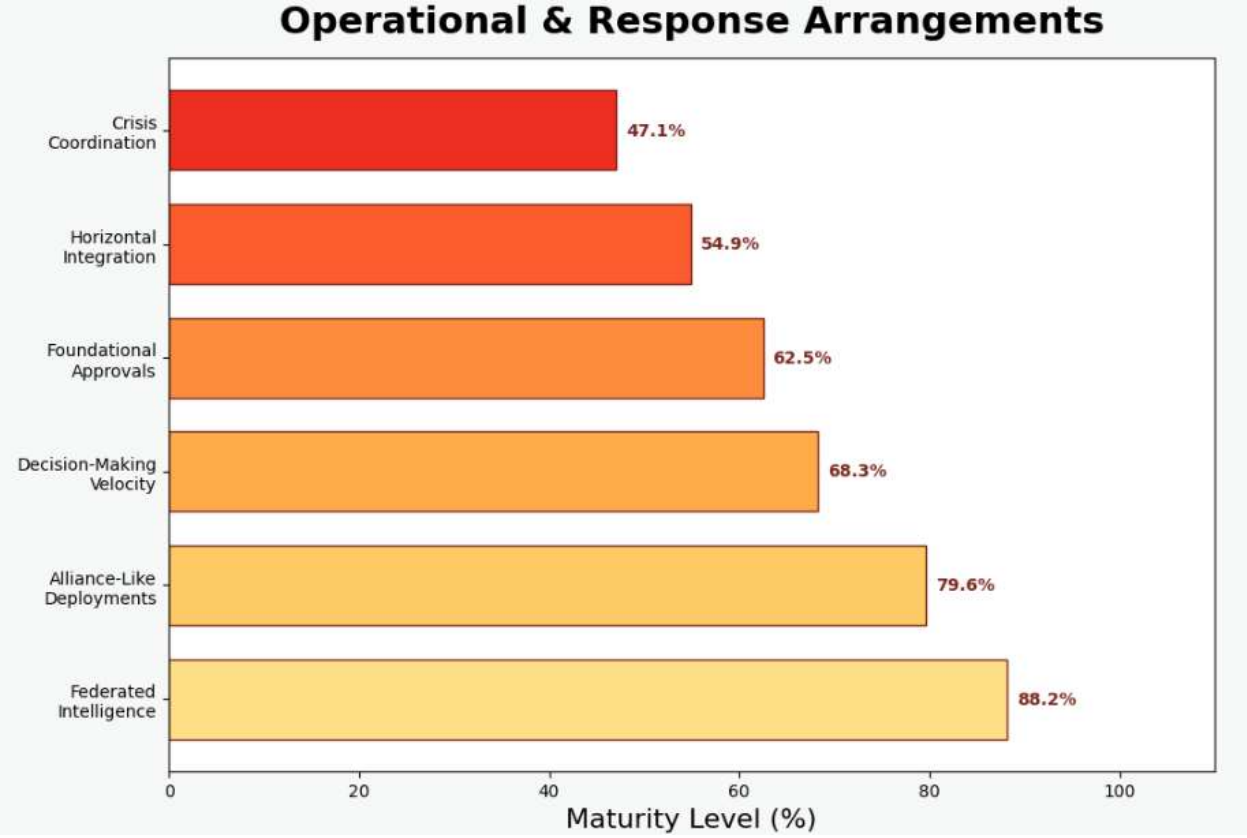


Fig 4: Operational & Response Arrangements

12.1. Organizational Structures for Global Collaboration

Well-established protocols guide partnership establishment and international collaboration in many domains. Collaborating in federated ecosystems, however, demands extra attention because of the number of actors and assets involved, the governance, regulatory, and risk positions of partners in different jurisdictions, or the geopolitical scenario. With a federated approach, service and data providers, whether public or private, belong to different jurisdictions worldwide. Trust is thus fundamental at all levels, particularly when relying on private data for analysis at any scale. Expressing global demand in the health and life sciences space, a pandemic or leading disease threatens the health and safety of citizens of all nations and regions.

Typically, global crises motivate many countries to collaborate more actively for supply and operational support. Fear of medical and pharmaceutical supply failures results in companies seeking to solve problems through regulatory alignment, whether reallocation of supply or data sharing for clinical trials. Focussing on the federated collaboration setting in health, a three-level business configuration emerges for multinational organizations requiring federated data and intelligence share. At the operational level, formal intake and operational teams deal with control of operations with partners. A response management and risk mitigation project support are on the tactical level. At the strategic level, a joint steering committee represents the executive management of partner organizations. The committee decides on all joint topics, approves start requests for projects and other support, and periodically updates the variations and changes in project and operational partners.

12.2. Risk Management and Incident Response

At the operational level, the risk management processes and incident response plan of an organization contribute to its resilience and long-term sustainability. While not all incident types and possible consequences can be anticipated, building-response capacities aids in addressing unexpected situations without significantly compromising essential services. Appropriate governance for risk management and incident response can help strengthen working relationships between partners in the federated healthcare ecosystem.

The risk management process should be in accordance with international standards, such as the ISO 31000 series. A central organization can coordinate the implementation of the risk management policy, but each partner must apply the same policy across its functions. An ability to identify the digital and physical networks and relationships that must be protected is essential. Risk may be quantified by using a common scale across all partners. Organizations can build upon models such as the NIST Cybersecurity Framework to define their high-level security controls.

The process of incident management can include detection, escalation, and response activities, together with the final specific steps for each incident type. Because the nature of incidents requires rapid response according to the plan, the roles and responsibilities for each step and type of incident must be clear, and all involved partners must have the appropriate training. An incident response plan should also outline mechanisms to ensure continuous improvement of the organizations' responses around specific incident types.

13. Results

Results are reported from three complementary perspectives: theoretical groundwork for federated intelligence from the AI governance standpoint; a methodological framework that enhances the comprehensive understanding of security requirements for data-sharing across the boundaries of national jurisdictions; and practical applications that enable implementation of cloud-based data sharing approaches intended to support pandemic surveillance through federated learning and collaborative clinical trials.

Federated intelligence is emerging as a practical approach to enable intelligence-gathering, detection, and predictive modelling in a privacy- and data-preserving manner. Despite the potential benefits of federated intelligence, existing frameworks for AI governance do not address the specific requirements of federated intelligence. These decision-support systems are different from traditional AI systems because of their multi-agent, decentralized nature. Establishing trust in systems that are co-designed and co-governed by multiple stakeholders across different jurisdictions is fundamental to achieving a federated intelligence capability for the global good. Such stakeholder trust must be enabled through a governance framework that effectively addresses the specific properties of federated AI systems. The Federated Intelligence Governance Framework is a high-level, technology-agnostic framework that identifies key decisions that need to be addressed in order to enable federated intelligence systems and highlights the need for associated lower-level guidance. The Framework facilitates a broad understanding of the decisions that need to be made by stakeholders planning a federated intelligence system. Moreover, the open and modular nature of the Framework is intended to enable stakeholders to both contribute to and extend it, supporting its application across different stakeholders of federated intelligence systems.

14. Conclusion

Research on federated AI architectures enabling trustworthy, international collaboration across healthcare institutions has been explored. Contributions span theory, methodology, and practice, addressing both AI technology and the data-management underpinning it. Supporting evidence has been derived from architecture design, threat modelling, a case study, and the analysis of collaborative healthcare landscapes from various perspectives.

The relevance of federated AI for global health and preparedness against emerging diseases is evident, yet adequate, interoperable data architectures to securely share critical

information across borders remain scarce. Three key domains have been identified. First, the concept of data-sovereign digital assets has been proposed, alongside criteria, governing elements, and policy implementation for secure international collaboration. Second, a cloud-enabled architecture supporting the discovery and sharing of data resources across different jurisdictions has been defined, allowing adaptation to specific latency, resilience, cost, and governance requirements. Finally, key processes and components, such as threat modelling, risk assessment, privacy-by-design compliance, and data-quality management, have been examined in view of the operational requirements of federated AI.

References

1. Aledhari, M., Razzak, R., Parizi, R. M., & Saeed, F. (2020). Federated learning: A survey on enabling technologies, protocols, and applications. IEEE Access, 8, 140699–140725.
2. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50–60.
3. Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. npj Digital Medicine, 3, 119.
4. Davuluri, P. S. L. (2023). AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems (December 15, 2023).
5. Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence, 2, 305–311.
6. Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R. R., & Bakas, S. (2020). Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. Scientific Reports, 10, 12598.
7. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., ... Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210.
8. Sadilek, A., Liu, L., Nguyen, D., Kamruzzaman, M., Serghiou, S., Rader, B., Ingerman, A., Mellem, S., Kairouz, P., Nsoesie, E. O., MacFarlane, J., Vullikanti, A., Marathe, M.,

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 02 Issue: 02
RECEIVED: APRIL 19 REVISED: MAY 03 ACCEPTED: MAY 25 PUBLISHED: JUNE 17

Eastham, P., Brownstein, J. S., Aguera y Arcas, B., Howell, M. D., & Hernandez, J. (2021). Privacy-first health research with federated learning. npj Digital Medicine, 4, 132.

9. Kolla, S. K., & Reddy, V. A. R. (2024). Evaluating Cloud-Native vs. Hybrid Architectures for Health Benefit Administration Systems. International Journal of Medical Toxicology and Legal Medicine, 27(5), 1042-1053.
10. Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., Liu, A., Costa, A. B., Wood, B. J., Tsai, C.-S., et al. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. Nature Medicine, 27, 1735–1743.
11. Warnat-Herresthal, S., Schultze, H., Shastry, K. L., Manamohan, S., Mukherjee, S., Garg, V., Sarveswara, R., Händler, K., Pickkers, P., Aziz, N. A., et al. (2021). Swarm Learning for decentralized and confidential clinical machine learning. Nature, 594, 265–270.
12. McGraw, D., & Mandl, K. D. (2021). Privacy protections to encourage use of health-relevant digital data in a learning health system. npj Digital Medicine, 4, 2.
13. Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., Liu, X., & He, B. (2023). A survey on federated learning systems: Vision, hype and reality for data privacy and protection. IEEE Transactions on Knowledge and Data Engineering, 35(4), 3347–3366.
14. Inala, R. (2023). AI-powered investment decision support systems: Building smart data products with embedded governance controls. Journal for ReAttach Therapy and Developmental Diversities, 6(10), 2251-2266.
15. Liu, J., Huang, J., Zhou, Y., Li, X., Ji, S., Xiong, H., & Dou, D. (2021). From distributed machine learning to federated learning: A survey. arXiv.
16. Zhang, K., Song, X., Zhang, C., & Yu, S. (2022). Challenges and future directions of secure federated learning: A survey. Frontiers of Computer Science, 16, 165817.
17. Coelho, K. K., Nogueira, M., Vieira, A. B., Silva, E. F., & Nacif, J. A. M. (2023). A survey on federated learning for security and privacy in healthcare applications. Computer Communications, 207, 113–127.
18. Li, H., Li, C., Wang, J., Yang, A., Ma, Z., Zhang, Z., & Hua, D. (2023). Review on security of federated learning and its application in healthcare. Future Generation Computer Systems, 144, 271–290.
19. Aouedi, O., Sacco, A., Piamrat, K., & Marchetto, G. (2023). Handling privacy-sensitive medical data with federated learning: Challenges and future directions. IEEE Journal of Biomedical and Health Informatics, 27(2), 790–803.
20. Li, S., Liu, P., Nascimento, G. G., Wang, X., Leite, F. R. M. M., Chakraborty, B., Hong, C., Ning, Y., Xie, F., Teo, Z. L., Ting, D. S. W., Haddadi, H., Ong, M. E. H., Peres, M. A., & Liu, N. (2023). Federated and distributed learning applications for electronic health records

and structured medical data: A scoping review. Journal of the American Medical Informatics Association, 30(12), 2041–2049.
21. Gottimukkala, V. R. R. (2024). Federated Learning Approaches for Fraud Detection in International Payment Systems. https://www. jisem-journal. com/download/118_JISEM. pdf.
22. Sharma, S., & Guleria, K. (2023). A comprehensive review on federated learning based models for healthcare applications. Artificial Intelligence in Medicine, 146, 102691.
23. Kalra, S., Wen, J., Cresswell, J. C., Volkovs, M., & Tizhoosh, H. R. (2023). Decentralized federated learning through proxy model sharing. Nature Communications, 14, 2899.
24. Wang, H., et al. (2023). Relay learning: A physically secure framework for clinical multi-site deep learning. npj Digital Medicine, 6, 204.
25. Chang, Q., Yan, Z., Zhou, M., Qu, H., He, X., Zhang, H., Baskaran, L., Al'Aref, S., Li, H., Zhang, S., & Metaxas, D. N. (2023). Mining multi-center heterogeneous medical data with distributed synthetic learning. Nature Communications, 14, 5510.
26. Pati, S., Baid, U., Edwards, B., Sheller, M., Wang, S.-H., Reina, G. A., Foley, P., Gruzdev, A., Karkada, D., et al. (2022). Federated learning enables big data for rare cancer boundary detection. Nature Communications, 13, 7346.
27. Joshi, M., Pal, A., & Sankarasubbu, M. (2022). Federated learning for healthcare domain: Pipeline, applications and challenges. ACM Transactions on Computing for Healthcare, 3(4), 1–36.
28. Kolla, S. K., & Mangalampalli, B. M. (2024). Edge-Based Deep Learning Systems for Point-of-Care Diagnostic Intelligence. Journal of Neonatal Surgery, 13(1), 2387-2399.
29. Yang, S., Hwang, H., Kim, D., Dua, R., Kim, J.-Y., Yang, E., & Choi, E. (2022). Towards the practical utility of federated learning in the medical domain. arXiv.
30. Ali, M., Naeem, F., Tariq, M., & Kaddoum, G. (2022). Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey. arXiv.
31. Zhai, K., Ren, Q., Wang, J., & Yan, C. (2021). Byzantine-robust federated learning via credibility assessment on non-IID data. arXiv.
32. Cao, X., Fang, M., Liu, J., & Gong, N. Z. (2020). FLTrust: Byzantine-robust federated learning via trust bootstrapping. arXiv.
33. Geiping, J., Bauermeister, H., Dröge, H., & Moeller, M. (2020). Inverting gradients—How easy is it to break privacy in federated learning? Advances in Neural Information Processing Systems, 33.
34. Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020). How to backdoor federated learning. Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics, 108, 2938–2948.

35. Guan, H., Yap, P.-T., Bozoki, A., & Liu, M. (2024). Federated learning for medical image analysis: A survey. Pattern Recognition, 151, 110424.
36. Teo, Z. L., et al. (2024). Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. Cell Reports Medicine, 5(2), 101419.
37. Upreti, D., Yang, E., Kim, H., & Seo, C. (2024). A comprehensive survey on federated learning in the healthcare area: Concept and applications. Computer Modeling in Engineering & Sciences, 140(3), 2239–2274.
38. Thakur, A., Molaei, S., Nganjimi, P. C., Liu, F., Soltan, A., Schwab, P., Branson, K., & Clifton, D. A. (2024). Knowledge abstraction and filtering based federated learning over heterogeneous data views in healthcare. npj Digital Medicine, 7, 283.
39. Ballhausen, H., Corradini, S., Belka, C., Bogdanov, D., Boldrini, L., Bono, F., Goelz, C., Landry, G., Panza, G., Parodi, K., Talviste, R., Tran, H. E., Gambacorta, M. A., & Marschner, S. (2024). Privacy-friendly evaluation of patient data with secure multiparty computation in a European pilot study. npj Digital Medicine, 7, 280.
40. Markkandan, S., Bhavani, N. P. G., & Nath, S. S. (2024). A privacy-preserving expert system for collaborative medical diagnosis across multiple institutions using federated learning. Scientific Reports, 14, 22354.
41. Kolla, T. (2024). Graph Neural Networks for HCC Risk Adjustment and Interoperability. International Journal of Science, Research and Technology, 7(6), 13244-13255.
42. Liu, J., Zhang, J., Jan, M. A., Sun, R., Liu, L., Verma, S., & Chatterjee, P. (2024). A comprehensive privacy-preserving federated learning scheme with secure authentication and aggregation for Internet of Medical Things. IEEE Journal of Biomedical and Health Informatics, 28(6), 3282–3292.
43. Düsing, C., Cimiano, P., Rehberg, S., Scherer, C., Kaup, O., Köster, C., Hellmich, S., Herrmann, D., Meier, K. L., Claßen, S., & Borgstedt, R. (2024). Integrating federated learning for improved counterfactual explanations in clinical decision support systems for sepsis therapy. Artificial Intelligence in Medicine, 157, 102982.
44. Thakur, A., Molaei, S., Nganjimi, P. C., Liu, F., Soltan, A., Schwab, P., Branson, K., & Clifton, D. A. (2024). Knowledge abstraction and filtering based federated learning over heterogeneous data views in healthcare. npj Digital Medicine, 7, 283.
45. Coelho, K. K., Nogueira, M., Vieira, A. B., Silva, E. F., & Nacif, J. A. M. (2023). A survey on federated learning for security and privacy in healthcare applications. Computer Communications, 207, 113–127.

46. Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., Liu, X., & He, B. (2023). A survey on federated learning systems: Vision, hype and reality for data privacy and protection. IEEE Transactions on Knowledge and Data Engineering, 35(4), 3347–3366.