NextGen Scientific Publication
American International Journal of Computer Science and Technology

Volume 7 Issue 6, Pg. No. 104- 118, AIJCST-V7I6P111, 2025
https://doi.org/10.63282/3117-5481/AIJCST-V7I6P111

Original Article

Federated AI and Big Data Architectures for Global Healthcare Collaboration

***Shashikala Valiki**
Independent Researcher, USA.

Abstract:

Advances in Big Data architectures and AI techniques based on Federated Learning improve the opportunities for collaboration in healthcare for use cases such as pandemic surveillance and precision medicine. Big Data infrastructures have recently evolved from classical data warehouses to include data lakes and data hubs. Fleets of data lakes and data hubs form a cloud of trust, enabling federated computing across organization boundaries. These concepts provide the foundations for a Big Data architecture for global healthcare based on semantically compliant integration of data hubs and lakes using open data exchange and description standards created by the international community. Recent developments in open-source software libraries for Federated Learning support data protection and data sovereignty with minimal data exposure to other organizations. Proposed solutions combine the benefits of Federated Learning with the elastic, self-service, and pay-as-you-go aspects of the public cloud for AI in healthcare.Future developments target global health, with specific applications to pandemic surveillance and response, and to precision medicine and genomics. The evolution of architectures for AI in health, based on Federated Learning and Federated Reinforcement Learning, is guided by the cross-functional collaboration model FR-FA-CFCM. The model formalizes the rules, roles, and mechanisms for co-creating software solutions among different stakeholder roles across organization boundaries, on-premise and in the cloud.

Keywords:

Federated Learning in Healthcare, Federated Reinforcement Learning, Big Data Architectures for Health, Cloud of Trust Infrastructure, Healthcare Data Lakes and Data Hubs, Semantic Data Integration Standards, Open Data Exchange Protocols, Privacy-Preserving AI Systems, Data Sovereignty Frameworks, Cross-Organizational Federated Computing, Pandemic Surveillance Analytics, Precision Medicine Platforms, Genomic Data Collaboration, Public Cloud AI for Healthcare, Elastic Self-Service Data Infrastructure, Open-Source Federated AI Libraries, Cross-Functional Collaboration Models, FR-FA-CFCM Governance Framework, Distributed Healthcare Intelligence Systems, Global Health Data Ecosystems.

Article History:

Received: 04.10.2025

Revised: 09.11.2025

Accepted: 22.11.2025

Published: 04.12.2025

1. Introduction

Artificial intelligence and big data improve healthcare, but developments are hampered by the inability to share sensitive data. Although federated learning for AI training can help, such architectures must also be integrated with big data management, storage, processing, and distribution. Global pandemic surveillance and response and precision medicine with genomics require collaboration between federated AI and big data systems. The combination enables global hospitals, pharmaceutical companies, universities, research institutes, and laboratories to work together to develop AI models while meeting strict requirements for security, privacy, data protection, and discrimination. Although innovations in AI and Big Data are advancing the healthcare

sector, there are still countless opportunities and needs where these technologies can improve healthcare processes. For example, the COVID-19 pandemic is a vivid reminder that influenza-like pandemics may still arise from animal viruses, and a proper preparatory research plan is essential for future.

AI can assist in quickly collecting early diagnosis information. AI systems can connect online healthcare services and detect infection with COVID-19 based on cough and breath. AI can summarize medical knowledge at a level comprehensible to the public, constantly update vaccination information, and give suggestions on preventing infection by wearing masks. During a pandemic, progress in designing vaccines and medicines can be accelerated through collaboration with pharmaceutical enterprises, enabling the public to obtain vaccines as early as possible.

1.1. Overview of the Global Healthcare Data Landscape

By 2025, society can expect to see the first results of substantial investments in global healthcare data architecture. The need for federated AI and big data technologies is now widely recognized by the health community, and a growing number of projects with global, regional, and national focus are laying the groundwork for a coherent, collaborative landscape. Progress is being made in building frameworks, principles, and standards to ensure security, compliance, and trust in the technologies.

Figure 1. Navigating the Global Healthcare Data Frontier: Federated Learning Frameworks for Privacy-Preserving Precision Medicine and Public-Good Research

During the pandemic, the relationships and technologies that made timely global data sharing possible were tested and strengthened. In precision medicine, the high-dimensional nature of the discovery space, combined with international data protection regulations and local restrictions, make federated-learning methods and approaches—such as data hubs and synthetic-data generation essential. Data-lake models, common in the private sector, are now emerging for public-good datasets. At the data-user level, however, the privacy and ethical concerns of individuals remain paramount. Widespread adoption of federated techniques is still some way off, but the basic research has matured, and successful pilot implementations in sensitive domains have proven the value and feasibility of deploying federated-learning techniques in real-world scenarios. Major efforts are now seeking to define an integrated set of principles, protocols, and frameworks that guide the responsible use of artificial-intelligence and federated-learning technologies while enabling community-driven formal review and approval processes to deliver representative and usable datasets for public-good research.

2. The Landscape of Global Healthcare Data

The healthcare community and other domains generate a wealth of health-related data worldwide. Hospitals record patient information and medical imaging. Pharmaceutical companies and laboratories create clinical trial results and genomics. Companies issue data from wearables and smart devices. The rapidly growing Internet of Things brings health signals and

parameters straight from homes. National health organizations and international institutions generate health authority documents. The volume is great. But the value of this information and knowledge is underused. The full potential of these assets is not being realized because more than 80% of them are generated in silos, recording rich histories and compelling patterns but seldom shared or combined with the data from other sources. Consequently, knowledge gaps remain wide open. For instance, the COVID-19 pandemic, with its surge in cases, mutations, variants, and millions of patients scattered around the world, had several consequences that could and should have been anticipated with precision by the community learning from data. Yet when the next pandemic comes, those gaps are likely to persist, still challenging pandemic surveillance and significantly increasing risks and uncertainties.

Equation 1) Step-by-step derivation of core Federated Learning equations (FedAvg)

1. Notation

- Client/site index: $k \in \{1, \dots, K\}$
- Local dataset at site k: D_k, size $n_k = |D_k|$
- Total samples: $N = \sum_{k=1}^{K} n_k$
- Model parameters: $\mathbf{w} \in \mathbb{R}^d$

2. Global objective from local objectives

Start from classical empirical risk minimization:

- **Local loss** at client k:

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

- **Weighted global loss** (each site weighted by its data volume):

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

Derivation (why this is correct):

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

So it equals the loss you would compute if all data were pooled—without actually pooling it.

3. Gradient of the global objective

Differentiate:

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

4. One FL communication round with local SGD

Local step (SGD) at site k: for local steps $s = 0, \dots, E-1$

$$\mathbf{w}_{k,s+1}^{(t)} = \mathbf{w}_{k,s}^{(t)} - \eta \, \nabla \hat{F}_k\left(\mathbf{w}_{k,s}^{(t)}\right)$$

- η learning rate
- $\nabla \hat{F}_k$ is a minibatch gradient estimator of ∇F_k

Initialize:

$$\mathbf{w}_{k,0}^{(t)} = \mathbf{w}^{(t)}$$

After E local steps, client k returns:

$$\mathbf{w}_k^{(t)} = \mathbf{w}_{k,E}^{(t)}$$

5. FedAvg aggregation (the key equation)

Coordinator forms the next global model as the **weighted average**:

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

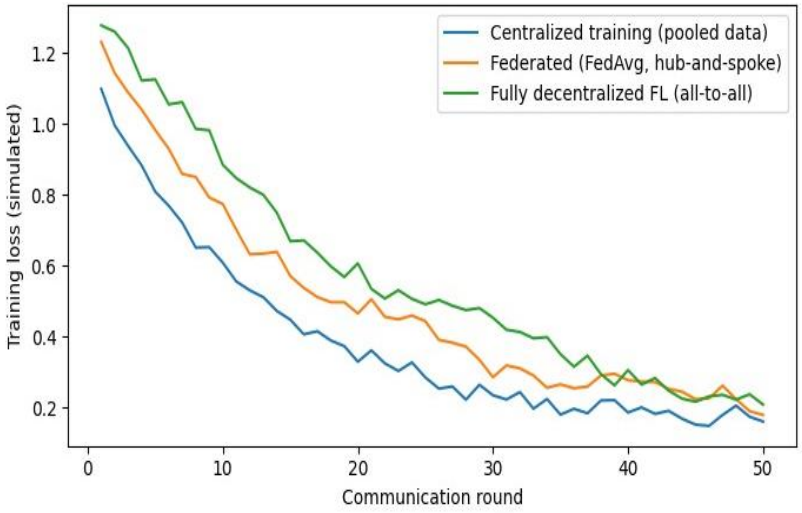

Figure 2. Illustrative Convergence behavior

2.1. Current Challenges and Opportunities in Global Healthcare Data

As previously outlined, data refers to factual information (usually numeric) collected and used for reference, analysis, or calculation, but its use is not isolated to informatics. Data is collected by individuals and groups registering it or sending it to repositories. Jumping ahead, the state of the Global Data Landscape points to a global "Data Capital" of 59 ZB in 2020, which will grow to nearly 180 ZB by 2025 and ultimately to 463 ZB by 2035 (Verizon, 2020). The last data set delineates different data types, showing that many types are expected to grow more than 30% every five years. Using the Data Capital metaphor, it is possible to state that the healthcare industry represents only 1% of these data but expects to grow three times by 2025. Public health organizations, regulatory agencies, and research organizations are large consumers of Global Data Capital, aiming to protect and maintain the health of individuals and populations. These organizations use data to study the pandemic's crisis management, biological markers of infections and other diseases, and ways to mitigate the infection spread. An initial reflection over the Global Data Capital reveals that the origin of most of this data type is the pharmaceutical industry and trading companies for Information and Technology (IT), and that they were originally consuming the "public health data" generated by a few healthcare sectors worldwide. As the speed of the pandemic shocked the entire world, it is still possible to see several societies and industries "helping" the cause. However, these types of "help" may also disappear, and data sharing may drastically decrease. Therefore, several questions and answers arise from the current disruption in the healthcare sector at the Global Data Capital level.

Table 1. Comparison of Federated Learning Architectures in Healthcare

Architecture / pattern	What is shared	Why used in healthcare	Key risks/limits
Centralized FL	Model updates ↔ aggregator	Simpler governance & ops	Single point of failure; trust in aggregator
Decentralized ring topology	Predictions to non-custodians; updates among subset nodes	Handles asymmetric participation	Potential leakage via predictions; coordination
Fully decentralized FL (FD-FL)	Peer-to-peer updates/params	Stronger decentralization, less single point	High comms + heterogeneity makes hard

3. Federated Learning in Health: Concepts and Architectures

As a subset of distributed machine learning, Federated Learning (FL) enables multiple parties to jointly train a machine-learning model without wholesale data sharing, allowing for greater privacy, confidentiality and possibly less data governance friction. Building upon a standard Federated Learning framework, two hybrid FL architectures have been established that support the integration of more advanced security methods (i.e. Secure Multi-Party Computation) as well as Privacy-Preserving Input Perturbation techniques (PPIP). Each architecture has been developed to enable distinct federated data-science use cases. By making dark data operational in global efforts such as pandemic surveillance and response and precision medicine, FL promises to deliver impactful advances across the whole span of human health.

Traditional ML and data-sharing approaches do not adequately meet the distinct privacy and regulatory requirements around global healthcare data collaboration. FL provides an alternative: COVID-19, genomics, drug discovery, and other extended

healthcare scenarios face the same privacy, governance, knowledge-gap tension as security-critical sectors such as defence, finance, and telecommunications, and for similar reasons are embracing federation as a route to data wealth creation without the associated privacy liabilities. Existing use cases highlight the importance of extending the concept of RL to FL in order to realise these outcomes.

Equation 2) Derivations for the "hybrid / decentralized" topologies

1. Ring topology update (decentralized subset nodes)

Let each node k average with neighbors $j \in \mathcal{N}(k)$. A standard formulation uses a **mixing matrix** P (row-stochastic, $P_{kj} \geq 0$, $\sum_j P_{kj} = 1$):

- Local compute step:

$$\widetilde{\mathbf{w}}_k^{(t+1)} = \mathbf{w}_k^{(t)} - \eta \nabla F_k\left(\mathbf{w}_k^{(t)}\right)$$

- Ring mixing step:

$$\mathbf{w}_k^{(t+1)} = \sum_{j \in \mathcal{N}(k) \cup \{k\}} P_{kj}\ \widetilde{\mathbf{w}}_j^{(t+1)}$$

Interpretation: you replace the "central aggregator" with neighbor averaging—consistent with ring communication.

2. Fully decentralized all-to-all (FD-FL)

Same form, but $\mathcal{N}(k)$ is large (potentially all peers). The equation is identical; only the topology changes, which increases overhead and coordination burden (the notes operational challenges for FD-FL).

3. Multi-model hybrid FL (two groups: custodians + non-custodians)

A clean mathematical way:

- Custodian group C, non-custodians U
- Custodian $c \in C$ trains model w_c using FedAvg within C (or locally), then shares w_c outward.
- A non-custodian $u \in U$ uses an ensemble prediction:

$$\hat{y}_u(x) = \sum_{c \in C} \alpha_{u,c}\ f(x; \mathbf{w}_c) \quad \text{with} \quad \alpha_{u,c} \geq 0,\ \sum_{c \in C} \alpha_{u,c} = 1$$

3.1. Federated Learning Paradigms

Federated Learning (FL) systems can be broadly categorized based on the levels of data ownership and participation. The simplest scenario occurs when all the participating institutions share a common federated model (centralized FL). In contrast, a ring topology is defined when only a subset of institutes acts as nodes in a FL process, i.e., data-custodians of their respective data repositories, while the remaining stakeholders share the model predictions derived from the nodes (decentralized FL). True privacy-preserving FL architectures, where institutes keep data within their environments without sharing model parameters or predictions, such as all-to-all communication topologies in Fully Decentralized FL (FD-FL), are less commonly implemented in operational settings due to the challenges in harmonizing model training across dissimilar institutions engaged in decentralized health projects.

The multi-model federated approach elegantly merges centralized- and decentralized-architecture concepts to help define secure and manageable communication topologies in interaction scenarios characterized by asymmetric data distribution and unequal data privacy constraints among partners. It explicitly defines two groups of disease-diagnosing partners; the model-custodians are responsible for training the models and share only model parameters with the non-custodian sites in a prediction-accessible ring topology.

3.2. Privacy-Preserving Techniques

Different privacy-enhancing technologies can be deployed to protect the local data remaining at the data provider site, the distributed model weights during training, or the final model resulting from training. With the former, the raw data, which is often patient records, is either disclosed only to authorized users or encrypted with cryprographic schemata. The final model training returns therefore the best model, without directly providing sensitive data. With the second, by using secure multi-party computation (MPC) or homomorphic encryption on the shared model weights during updates, the data provider network joins its contribution (the training results) without disclosing its sensitive information. Finally with the last, differential privacy assures that the contribution of any individual dataprovider cannot be inferred from the distributed model. Federated Learning is a promising approach as data is kept where it is generated. Regulation policies adopted in some regions impose restrictions on data exportation outside the region or impose not to keep certain classes of data for long. The integration with respective regulations need to be considered building health applications with a federated approach.

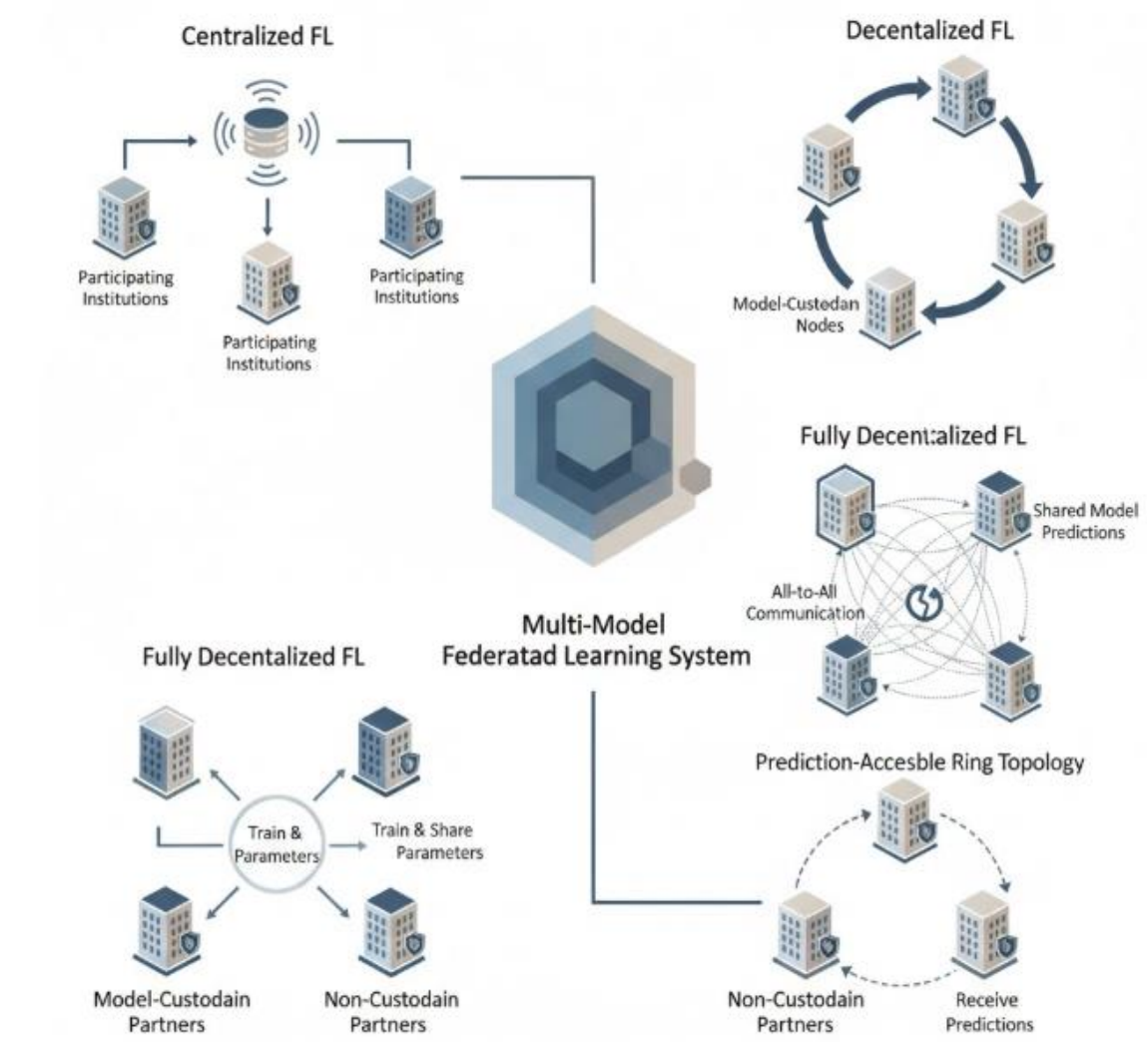

Figure 3. Hybrid Federated Learning Architectures: A Multi-Model Approach to Asymmetric Data Privacy in Decentralized Health Systems

4. Big Data Architectures for Healthcare

Considerable progress has been achieved over the past decade in implementing Big Data capabilities in Health, and there are currently numerous on-going and completed projects showing innovative use cases for these systems. However, such implementations still only cover a small number of the use cases for the Global Health sector. Critical challenges and strategic recommendations may thus be defined for Big Data systems in Health:

- Building Data Lakes and Data Hubs in Health: Organizations such as the European Data Protection Board (EDPB) underline the necessity of many Data Lakes or Data Hubs in Health to group dedicated healthcare data related to a specific context. Currently, many data feeds are being established to support pandemic surveillance, whether for COVID or any future emerging disease, but the construction of additional federated Data Lakes or Data Hubs is urgently needed to implement any of the other use cases defined in the health zone of the FD-DataSpace such as disease control, pathogenic evolution, zoonosis detection, health preparedness or human security.
- Enabling Open Data Standards for Health Data: The implementation and availability of open data standards in healthcare is essential to promote connections and interoperability of the many systems related to Health. Not only the open healthcare sediments previously mentioned, also existing reference architectures and information mapping frameworks need to be extended and equipped with open data standards to simplify and accelerate the creation of data feeds from various health sources into the dedicated Data Lakes or Data Hubs in Health. Equally important is to ensure the existence of trusted open data spaces in health.
- Guaranteeing Compliance, Security, Trust and Privacy: In addition to the above-enabling factors, it is obvious that compliance to privacy requirements, cyber-security and security of the entire system need to be ensured. It is equally

important to foster public and private trust that such requirements are indeed guaranteed and that, in turn, public and private entities benefit from establishing data feeds into the systems.

Equation 3) Step-by-step derivations of privacy-preserving techniques

1. Secure aggregation (MPC-style) for FedAvg

Goal: aggregator learns only the **sum** $\sum_k \Delta_k$ (updates), not each Δ_k.

Let client update be $\Delta_k = \mathbf{w}_k^{(t)} - \mathbf{w}^{(t)}$.

Masking derivation (classic secure-aggregation idea):

- Each pair of clients k, j agrees on random mask vector $\mathbf{r}_{k,j}$ such that $\mathbf{r}_{k,j} = -\mathbf{r}_{j,k}$.
- Client k sends masked update:

$$\tilde{\Delta}_k = \Delta_k + \sum_{j \neq k} \mathbf{r}_{k,j}$$

- Aggregator sums:

$$\sum_{k=1}^{K} \tilde{\Delta}_k = \sum_{k=1}^{K} \Delta_k + \sum_{k=1}^{K} \sum_{j \neq k} \mathbf{r}_{k,j}$$

But masks cancel pairwise because $\mathbf{r}_{k,j} = -\mathbf{r}_{j,k}$, so:

$$\sum_{k} \sum_{j \neq k} \mathbf{r}_{k,j} = 0 \Rightarrow \sum_{k=1}^{K} \tilde{\Delta}_k = \sum_{k=1}^{K} \Delta_k$$

Aggregator gets the needed total update, **not individual** updates.

2. Homomorphic encryption aggregation (conceptual math)

With additive homomorphic encryption $Enc(\cdot)$:

- Client encrypts update: $Enc(\Delta_k)$
- Aggregator computes encrypted sum:

$$Enc\left(\sum_k \Delta_k\right) = \prod_k Enc(\Delta_k)$$

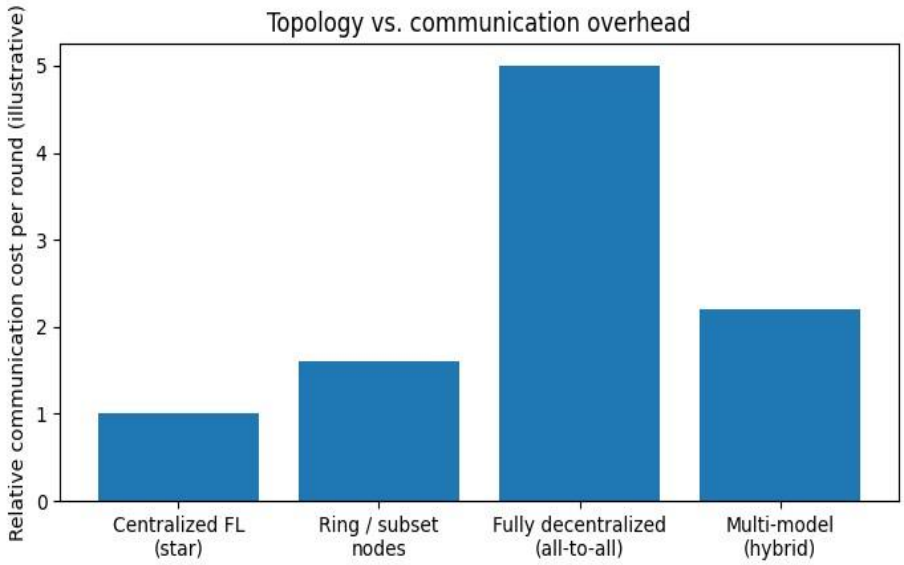

Figure 4. Comparison of Communication Overhead Across Federated Learning Topologies

4.1. Data Lakes and Data Hubs

While federated learning enables the distributed execution of training algorithms without transferring raw data, Big Data architectures such as data lakes and data hubs facilitate the sharing of data in compliance with privacy regulations and without breaching trust. Data lakes enable a large quantity of raw data, acquired from multiple sources, to be stored at low cost for potential future use. Data hubs combine semantically aligned data supplied by multiple organizations for cooperative AI analytics. In the healthcare context, regulatory frameworks discourage the transfer of personal patient data across jurisdictions. Data lakes thus collect information for future global health crises, while organizations employ data hubs to provide, rather than

acquire, data from other jurisdictions. A cross-border pandemic should be addressed through compliance with the WHO's International Health Regulations (2005) and the establishment of an operational data hub for one or more infectious disease hubs of the European Union's Digital Europe Programme.

Federated AI architectures are iteratively enhanced through federated learning updates. The Federal Risk and Authorization Management Program grants interim security authorizations to cloud services for use by U.S. federal agencies. The System and Organization Controls reporting framework assists service organizations in building trust. Achieving compliance with these standards strengthens security, heightens the perception of compliance, and augments end-user trust in Health Data Lakes. Integrating these trust-enabling elements allows the operationalization of transparent procedures designed to grant end-users confidence in the proper management of sensitive data in Health Data Lakes and across cooperative Federated Artificial Intelligence services.

4.2. Data Standards and Interoperability

Many factors impede interoperability on a global level. First and foremost, the semantic and syntactic mismatches between schemas hampers machine understanding. The compensation needs to be addressed manually with expensive annotation or mapping. Furthermore, the structure is not uniform. The original XML encoded data sources do not follow the structure necessary for a classical data hub approach. In case of access through copy-and-paste the mapping is still needed. Jurisdictional issues also arise from global operation of the virus data hub. Embedding copy-and-paste approaches makes the jurisdictional issues more flexible as the data are copied to a national data hub in the normal jurisdiction of the user.

Nonetheless, establishing common syntax and semantic schemas is the primary means of enabling interoperability. By encoding data in the Fair Data Principles and through the Genetic Data Poster drafted by the UN Office of the Secretary-General's Special Envoy for the 2030 Agenda for Sustainable Development, standard compliant data can be provided efficiently. Global networks involving a data hub within each corner of the world, to be operated by national authorities, is another way to enhance the compliance with regulations and laws.

5. Integration of Federated AI with Big Data Systems

The integration of AI with Big Data and Data Sciences is to support intelligent decision-making processes and provide public services. AI is employed to analyze the large amounts of unstructured data generated in Big Data systems, allowing predictive modeling and supporting decision-making processes. A contrasting architectural pattern utilizes decentralized intelligent systems together with global Big Data infrastructures. This approach allows collaborative public-services intelligent system development at a global scale. The Federated AI paradigm is supported by Decentralized Data Mining Foundations and allows supporting the security, compliance, and trust necessities of sensitive data analysis.

These patterns facilitate multiple Federated AI Models in parallel, with no restrictions on collaboration with other models developed worldwide. Global healthcare collaboration in 2025 is expected to support pandemic surveillance, prediction, and control with the connection of autonomous intelligent systems strategically distributed across countries. In a complementary perspective, AI models based on AI4EU foundations enable the collaborative development of predictive healthcare-related models. Federated AI Models assist in predicting students' academic performance through the cooperation of educational institutions and associated organizations. In this collaboration model, local IQ data and academic failure rates are considered sensitive.

Table 2. Comparative Analysis of Privacy-Preserving Techniques

Privacy technique	Protects	Typical cost	Trade-off
Access control / encryption at rest	Raw data locally	Low–medium	Ops complexity vs usability
Secure Multi-Party Computation (MPC) / secure aggregation	Updates during aggregation	Medium–high	Stronger confidentiality vs latency
Homomorphic encryption (HE) on updates	Updates during compute/aggregation	High	Strong security vs compute
Differential Privacy (DP)	Individual contributions to model	Low–medium	Privacy vs accuracy

5.1. Architectural Patterns for Collaboration

An architectural integration of federated AI and global healthcare data hubs addresses the need for privacy, security, and cross-organizational collaboration. It allows the construction of cross-organizational AI models while complying with data privacy regulations, and data providers controlled the sharing in a hub-and-spoke model. Beyond the healthcare domain, the proposed architecture pattern can be applied in other areas where different entities strive to collaboratively build AI models without disclosing sensitive and private data, such as finance and marketing. Increasing regulatory pressures mean that federated AI architectures supporting the privacy and protection of sensitive mission-critical data remain pertinent. Structured around a data hub, these designs eliminate the need to create point-to-point connections, allowing third parties — within as well as outside the project — to access the data without creating extra work for the data providers. Furthermore, such architectures are able to cope with data residing in different cloud infrastructures.

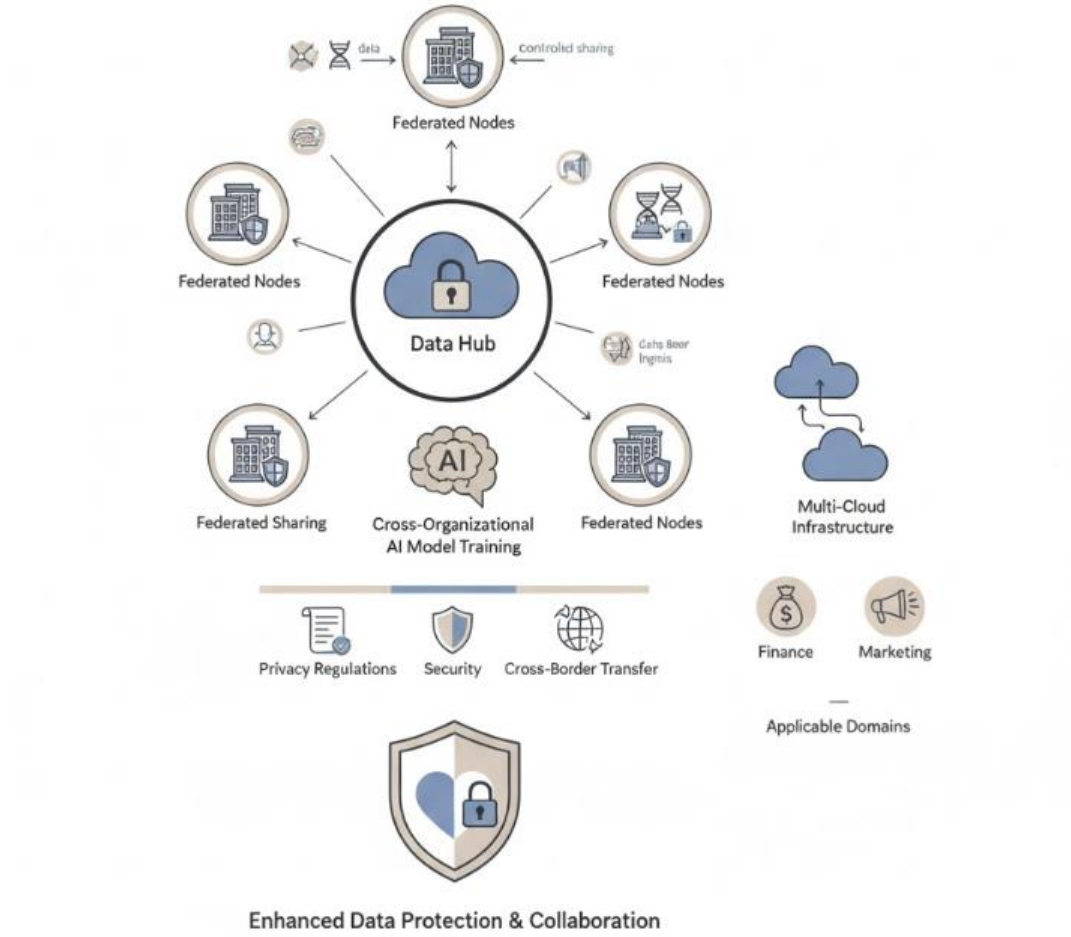

Figure 3. The Hub-and-Spoke Federated AI Model: Scaling Cross-Organizational Collaboration through Privacy-Preserving Data Hubs and Multi-Cloud Interoperability

5.2. Security, Compliance, and Trust

Global federated health systems must satisfy data security and regulatory compliance obligations for each data source yet be sufficiently flexible to support collaborative machine-learning workloads and different security and compliance models, including systems that do not require data-sharing or sharing sensitive information and models based on synthetic data. A recent proposal offers a multi-tier, multi-domain architecture pattern that addresses these requirements for the case of banking, insurance, and credit companies. Additional requirements stem from the need to build public trust in AI-based systems in general and in sensitive areas like global health in particular. A framework for responsible AI describes security and privacy aspects together with ethical, innovation, and sustainability considerations.

It encourages the adoption of AI governance and risk management frameworks and tools based on principles of privacy and data governance, fairness, reliability and safety, inclusiveness, transparency and explainability, and accountability. In global health, these principles may need to be extended to also encompass data-sharing readiness and trustworthiness indicators. Processes to incorporate privacy and ethical-by-design solutions and ensure AI solutions meet expectations and standards may need to be included as well.

6. Use Cases in Global Health

A few specific scenarios are then formulated to illustrate the integration of federated AI systems with big data architectures, establishing a direction for these technologies to support global healthcare collaboration by 2025. Federated AI in health is among the priority topics identified by the Global Coalition for AI in Health and Life Sciences, with pandemic surveillance and response as an initial opportunity. A case of accelerating pathogenic identification and epidemiological investigation exemplifies the collaboration between numerous hospitals and institutions responsible for health monitoring and assessment during a pandemic. International patients at participating clinical test institutions provide deep sequencing data, and genomic surveillance laboratories responsible for benign and malignant modification Welsh data of SARS-CoV-2. Global vaccine manufacturers provide clinical feasibility of vaccines against Omicron and its derivative strains. By organically combining all of these data and resources, a federated AI research and application system can be constructed for biomedicine, biology, chemistry, and synthetic engineering. The AI researcher Graham Hacohen provides synchronized threshold data, and a relatively independent transparent liable Technology-based analysis solves the urgent need for pathogens in a short time.

Another major area is pandemics in the fields of precision medicine and precision drug research and development. The Medical Data Interoperability Alliance is a social group dedicated to developing an interoperability specification for sharing clinical and omic data, and with the ultimate goal of enabling large global multi-dimensional integration, sharing, and collaborative modeling of medical and antioxidant clinical data, thus significantly improving the speed and accuracy of precision drug research and development. With the support of leading drug manufacturers and technical service providers, the alliance has released the OMOP model, which contains medical data standards governing patients and omic analysis; which is still limited to individual drug manufacturers. However, through allied cooperation with the Medical Data Interoperability Alliance and introduction of analytical models and comparative drugs from leading drug manufacturers, a federated AI pharmacokinetics research platform based on the OMOP model and joint clinical analysis and comparison configuration is expected to be deployed and implemented quickly.

Equation 4) Big Data side: formalizing "data lakes vs data hubs" integration

- Data lakes = store large raw data (low-cost storage, future use)
- Data hubs = semantically aligned data across orgs for cooperative analytics

A minimal mathematical abstraction:

1. Lake (raw ingestion) as a union of sources

Let raw sources be $S_1, \dots, S_m$. A lake stores:

$$\mathcal{L} = \bigcup_{j=1}^{m} S_j$$

No strict schema is required.

2. Hub (semantic alignment) as a mapping into a common schema

Let Ω be a shared ontology/schema (e.g., common coding systems). Each org k has mapping g_k:

$$g_k : \text{Schema}_k \to \Omega$$

Hub dataset:

$$\mathcal{H} = \bigcup_{k=1}^{K} g_k(D_k)$$

6.1. Pandemic Surveillance and Response

Severe illnesses, pandemics, and health-related emergencies arise almost any time across multiple countries due to various reasons. Recent pandemic-like situations, such as COVID-19, have disrupted education and business around the globe. So, surveillance is a critical requirement for healthcare stakeholders. The anticipatory surveillance is difficult if the related big data are dispersed in place or size, and health-related data of one or few countries cannot provide adequate justification in case of a severe illness in any small zone area that has never been seen earlier. Surveillance and resolution are critical requirements for pandemic-like massive health-related challenges.

AI applications supported by big data, conventional methods and models enable for adequate performance prediction, health disease classification, severity and risk prediction, logistical decision-making, forecasting, business impact analysis, and so on for various small and medium-sized categories of illness across a dispersed global region. However, these methods and models within the health domain have limitations, assumptions, validation, and justification issues. Even though comparable quality of significance can be achieved while integrating a few countries with sufficient data from a few years or even months, it always cannot be considered to attain global impact prediction. Hence, federated-designed pattern-based collaboration between

various hospitals or countries using their large-size dispersed data or condition and pattern capability is necessary for several healthcare cases.

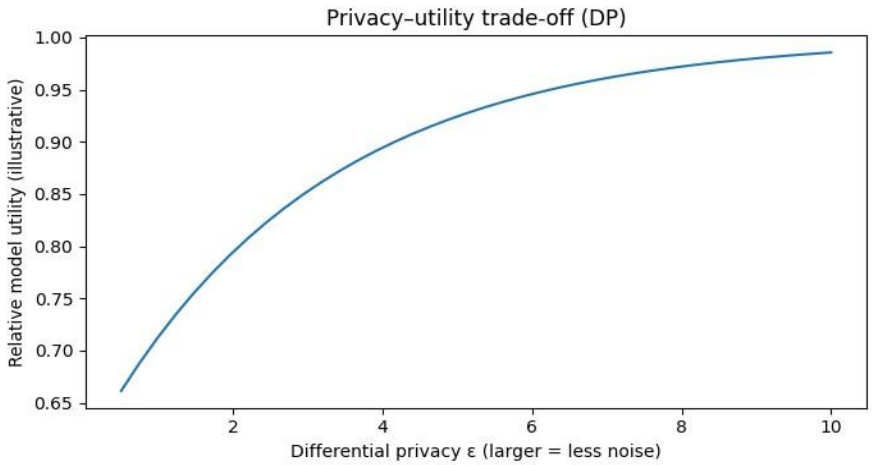

Figure 4. Privacy–Utility Trade-off under Differential Privacy

6.2. Precision Medicine and Genomics

Recent years have seen remarkable strides in molecular biology technologies such as genomics, transcriptomics, metabolomics, proteomics, and microbiome research yet breakthroughs in new drugs, effective treatment of major diseases, and a decline in overall mortality remain rare. There is a growing realization that many diseases cannot be cured or even effectively treated without better understanding of individual patient biology and genetics, which i.e. suggests that they require a more individual and HUMAN-based medicine approach—the trend known as precision medicine. Genomic studies provide fundamental resources for interpreting population variation in genes, phenotypes, and diseases in human beings, animals, and plants. All humans are 99.9% genetically identical, yet the small differences that make up the remaining 0.1% of the genome sequence constitute the substantial basis for human individuality and susceptibility to disease. Even small differences in genome sequences can have an enormous impact on human health. Genetic factors are implicated in many physiological and biochemical processes, including responses to pharmaceutical treatments makes them a key factor in personalized medicine. Genomic studies of a range of diseases have identified disease-associated variants in populations of European ancestry, but the majority of the current genome-wide association studies (GWAS) have been largely limited in their discovery and characterization of population-specific disease variants.

The exponential increase of biomedical data, including genomic sequences, transcriptome expression data, genome-wide association studies (GWAS), clinical data, and drug response and side effects data at an unprecedented scale across millions of individuals is helping to develop this new precision medicine area. However, most of the genomic data from different institutions, laboratories, and countries lie in silos, which makes it incompletely exploited and analyzed or not well analyzed at all. None of the new artificial intelligence methods for health data, especially large language models, has been effectively applied to interoperation and semantical modeling of these billion-petal data/multi-modal tasks. Furthermore, although scientific results have made humans aware that small differences in genome sequences can have huge effects on humans, the existing publicly accessible disease genomics information resources, such as the 1000 Genomes Project, GWAS Catalog, and dbGaP databases, are still insufficient for effective population genetic studies, especially for studies of non-European populations.

7. Conclusion

In conjunction with the increasing use of massively parallel computing clusters capable of AI-inferencing million-dollar AI models, federated learning is becoming an increasingly common AI training paradigm designed primarily to address the challenges of AI training in highly sensitive environments where data privacy cannot be assured. Federated-learning training frameworks promise to revolutionise national and global AI-as-a-service by enabling organisations and countries to securely share only the intelligence they can make freely available, while remaining compliant with local privacy regulations.

Federated-learning paradigms and federated vs non-federated computing frameworks have assessment-gridding support structures readily available; however, privacy-preserving mechanisms remain an emergent research area defined mainly through use-case-specific development initiatives. Federated-learning-as-a-service enables the integration of massive AI models

into healthcare data-lake and data-hub architectures, allowing the harmonisation of data hub-intelligence development without the need for the dataset owners to share their data with the intelligence developers. Security and compliance services empowered through responsibility frameworks such as the Non-Disclosure Covenant help to establish a trust architecture that enables the dynamic addition of security, compliance, and trust services to healthcare-data-hub architectures. Concomitant provisioning of global-system usage intelligence enables federated-learning-as-a-service to support the identification of optimum-specified and -commissioned model-development solutions. Any demonstrated success encourages subsequent investment as the enabling-cloud-services geopolitical milieu permits.

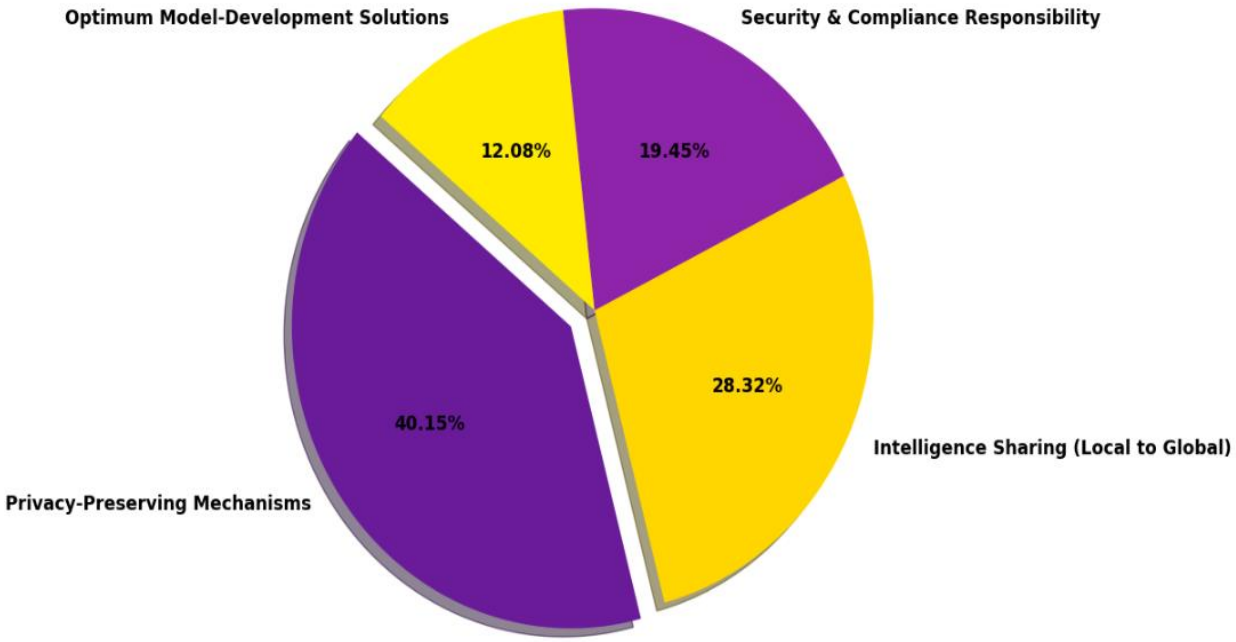

Figure 5. Strategic Weighting: Federated Healthcare Paradigm

7.1. Final Insights and Future Directions

The envisioned architecture provides an integrated approach to federated AI, big data and global health. It supports major use cases for federated AI in health, including pandemic surveillance and response, and precision medicine with particular focus on genomic data. The combination of these use cases is especially timely, as presently they are approaching critical mass, making collaboration increasingly plausible and fruitful. The concept outlines a potential architecture for collaboration between federated AI in health and big data systems. As information security, privacy and trust are paramount in health, the architecture particularly emphasizes safeguarding these factors in three ways. Compliance with legal and regulatory requirements underpinning data sharing is addressed through the use of data hubs for restricted data that demand higher levels of protection. Transparency and auditability are strengthened by operation of data lakes as a form of data hub, with details of stage 1 processes published and stage 3 custom analysis protected by segregation. Finally, a federation-wide policy framework defines the operation of the federated AI ecosystem, including participant eligibility, approval of local contributing data for each model training round, and handling of divulged fraud. These measures aim to foster a secure, compliant, privacy-preserving and trusted federated AI ecosystem that supports crucial public health imperatives and broader AI use and user acceptance across society.

The intersection of federated AI in health with big data systems is fleshed out in several dimensions, exploring architectural collaboration patterns, security and privacy requirements, and the potential role of compliance and trust. The examination draws on the operation of the European Health Data Space – a robust compliance-oriented ecosystem for data lakes and data hubs housing health data and digital services that capitalise on sharing – and the landscapes of global public health and core federated AI in health user communities. These new AI methods harness tailored data for critical health challenges without the ethical, privacy and security risks associated with conventional data sharing. Achieving success hinges on an overarching environment that addresses the primary concerns of information security, privacy and ethics in an explicit manner, and so strengthens compliance, trust and acceptance.

References

[1] NIST. (2023). Artificial intelligence risk management framework (AI RMF 1.0). National Institute of Standards and Technology.

[2] Kolla, S. K. (2021). Designing Scalable Healthcare Data Pipelines for Multi-Hospital Networks. World Journal of Clinical Medicine Research, 1(1), 1–14. Retrieved from https://www.scipublications.com/journal/index.php/wjcmr/article/view/1376

[3] Armbrust, M., Zaharia, M., Xin, R. S., et al. (2015). Apache Spark: A unified engine for big data processing. Communications of the ACM, 59(11), 56–65.

[4] Varri, D. B. S. (2024). Adaptive and Autonomous Security Frameworks Using Generative AI for Cloud Ecosystems. Available at SSRN 5774785.

[5] Batini, C., & Scannapieco, M. (2016). Data and information quality: Dimensions, principles and techniques. Springer.

[6] Babaiah, C., Dobriyal, N., Shamila, M., Aitha, A. R., Patel, S. P., & Upodhyay, D. (2025, December). Intelligent Fault Detection and Recovery in Wireless Sensor Networks Using AI. In 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG) (pp. 1-6). IEEE.

[7] Benjamens, S., Dhunnoo, P., & Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices. NPJ Digital Medicine, 3, 118.

[8] Davuluri, P. N. Integrating Artificial Intelligence into Event-Driven Financial Crime Compliance Platforms.

[9] Bertsekas, D. P. (2012). Dynamic programming and optimal control (Vol. 1). Athena Scientific.

[10] Vajpayee, A., Khan, S., Gottimukkala, V. R. R., Sharma, D., & Seshasai, S. J. (2025). Digital Financial Literacy 4.0: Consumer Readiness for AI-Driven Fintech and Blockchain Ecosystems. International Insurance Law Review, 33(S5), 963-973.

[11] Brundage, M., Avin, S., Clark, J., et al. (2018). The malicious use of artificial intelligence. arXiv.

[12] Nigam, N., Sireesha, B., Ediga, P., Segireddy, A. R., & Bokde, S. (2025, December). Comparative Evaluation of Cloud Security Algorithms Using Multiple Classifiers with an Optimized Intrusion Detection System. In 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG) (pp. 1-6). IEEE.

[13] Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19, 171–209.

[14] Pareyani, S., Goswami, S., Geetha, Y., Dimri, S. K., Niharika, D. S., & Amistapuram, K. (2025, December). Smart Resource Allocation in Wireless Sensor Networks Through AI Techniques. In 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG) (pp. 1-6). IEEE.

[15] Vijaya Rama Raju Gottimukkala. (2025). Agentic AI for Next-Generation Cross-Border Payments: Contextual Learning in Transaction Routing. Journal of Informatics Education and Research, 5(4). Retrieved from https://jier.org/index.php/journal/article/view/3794

[16] Varri, D. B. S. V. (2025). Human-AI collaboration in healthcare security.

[17] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–407.

[18] Nagubandi, A. R. (2025). Cryptocurrency Market Spillovers: Risk Contagion Across Global Financial Systems.

[19] European Parliament and Council of the European Union. (2016). General Data Protection Regulation (GDPR). Official Journal of the European Union.

[20] Yandamuri, U. S. AI-Driven Decision Support Systems for Operational Optimization in Hospitality Technology.

[21] Gentry, C. (2009). A fully homomorphic encryption scheme. Stanford University.

[22] Guntupalli, R. (2025). Federated Deep Learning for Predictive Healthcare: A Privacy-Preserving AI Framework on Cloud-Native Infrastructure. Vascular and Endovascular Review, 8(16s), 200-210.

[23] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

[24] Dutta, P., Mondal, A., Vadisetty, R., Polamarasetti, A., Guntupalli, R., & Rongali, S. K. (2025). A novel deep learning rule-based spike neural network (SNN) classification approach for diagnosis of intracranial tumors. International Journal of Information Technology, 17(9), 5705-5712.

[25] [25] He, J., Baxter, S., Xu, J., et al. (2019). The practical implementation of artificial intelligence technologies in medicine. Nature Medicine, 25, 30–36.

[26] [26] Enterprise-Scale Gen AI Orchestration Using Small LMs and LLM Agents for Intelligent ITSM and HRSD Automation in Enterprise Ecosystems. (2025). MSW Management Journal, 35(2), 1889-1897.

[27] Holzinger, A. (2016). Interactive machine learning for health informatics. Springer.

[28] FinOps Strategies for AI-Enabled Real-Time Compliance Platforms in Cloud Native Environments. (2025). MSW Management Journal, 35(2), 2080-2088.

[29] IBM. (2023). Data fabric architecture overview. IBM Redbooks.

[30] Yandamuri, U. S. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706.

[31] Sasi Kumar Kolla. (2023). Big Data–Driven Machine Learning Frameworks for Clinical Risk Prediction. International Journal of Medical Toxicology and Legal Medicine, 26(3 and 4), 44–59. Retrieved from https://ijmtlm.org/index.php/journal/article/view/1456

[32] Kelly, C. J., Karthikesalingam, A., Suleyman, M., et al. (2019). Key challenges for delivering clinical impact with AI. BMC Medicine, 17, 195.

[33] Kumar, K. M., Parasar, A., Walia, A., Inala, R., & Thulasimani, T. (2025, August). Enhancing Risk Management Strategies in Financial Institutions Using CNN and Support Vector Regression. In 2025 5th Asian Conference on Innovation in Technology (ASIANCON) (pp. 1-6). IEEE.

[34] Koller, D., & Friedman, N. (2009). Probabilistic graphical models. MIT Press.

[35] Rao, A. N., Garapati, R. S., Suganya, R. T., Kaliappan, A., & Kamaleshwar, T. (2025, August). Smart Solar Harvesting and Power Management in IoT Nodes Through Deep Learning Models. In 2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS) (pp. 1-6). IEEE.
[36] Liu, F., et al. (2025). Foundational architecture for AI agents in healthcare. Cell Reports Medicine, 6(10), 102374.
[37] Paleti, S., Baliyan, M., Aitha, A. R., Reddy, B. A., Bhadauria, G. S., & Sing, S. A. (2025, August). Graph—LSTM Hybrid Model for Improving Fraud Detection Accuracy in E-Commerce Financial Services. In 2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS) (pp. 1-6). IEEE.
[38] Moreau, L., & Groth, P. (2013). Provenance: An introduction to PROV. Morgan & Claypool.
[39] Nagabhyru, K. C., Rani, M., Reddy, D. S., & Krishnaraj, V. (2025, August). Machine Learning-Driven Fault Detection in Electric Vehicles via Hybrid Reinforcement Learning Model. In 2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS) (pp. 1-6). IEEE.
[40] Obermeyer, Z., & Emanuel, E. (2016). Predicting the future—Big data and clinical medicine. NEJM, 375, 1216–1219.
[41] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning. ACM Transactions on Intelligent Systems and Technology, 10(2), 1–19.
[42] Pearl, J. (2009). Causality (2nd ed.). Cambridge University Press.
[43] Srikanth, T., Segireddy, A. R., & Elavarasi, S. A. (2025, October). STaSFormer-SGAD: Semantic Triplet-Aware Spatial Flow-Guided Spatio-Temporal Graph for Anomaly Detection in Surveillance Videos. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 1-7). IEEE.
[44] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. NEJM, 380, 1347–1358.
[45] Kolla, S. K. (2021). Architectural Frameworks for Large-Scale Electronic Health Record Data Platforms. Current Research in Public Health, 1(1), 1–19. Retrieved from https://www.scipublications.com/journal/index.php/crph/article/view/1372
[46] Nagabhyru, K. C. (2025). Beyond Automation: The 2025 Role of Agentic AI in Autonomous Data Engineering and Adaptive Enterprise Systems.
[47] Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
[48] Lebcir, I., Mageswari, S. U., Bhosale, Y. H., Nagubandi, A. R., & Mahabooba, M. M. Agile Strategic Management in the Age of Disruption: Leveraging AI and Data Analytics for Competitive Advantage.
[49] Satyanarayanan, M. (2017). The emergence of edge computing. Computer, 50(1), 30–39.
[50] Velangani Divya Vardhan Kumar Bandi. (2024). Intelligent Data Platforms For Personalized Retail Analytics At Scale. Metallurgical and Materials Engineering, 30(4), 1011–1027. Retrieved from https://metall-mater-eng.com/index.php/home/article/view/1011-1027
[51] Sheller, M. J., Reina, G. A., Edwards, B., et al. (2020). Multi-institutional deep learning without sharing patient data. Brainlesion Workshop.
[52] Garapati, R. S., & Daram, D. S. B. (2025). AI-Enabled Predictive Maintenance Framework For Connected Vehicles Using Cloud-Based Web Interfaces. Available at SSRN 5524261.
[53] Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of AI. JAMA, 320(21), 2199–2200.
[54] Rongali, S. K. (2025, August). Deep Learning for Cybersecurity in Healthcare: A Mulesoft-Enabled Approach. In 2025 International Conference on Artificial Intelligence and Machine Vision (AIMV) (pp. 1-6). IEEE.
[55] Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning (2nd ed.). MIT Press.
[56] Siva Hemanth Kolla. (2023). Deep Learning–Driven Retrieval-Augmented Generation for Enterprise ITSM Automation: A Governance-Aligned Large Language Model Architecture . Journal of Computational Analysis and Applications (JoCAAA), 31(4), 2489–2502. Retrieved from https://www.eudoxuspress.com/index.php/pub/article/view/4774
[57] Tsamados, A., Aggarwal, N., Cowls, J., et al. (2022). The ethics of algorithms. AI & Society, 37, 215–230.
[58] Davuluri, P. S. L. N. . (2024). AI-Driven Data Governance Frameworks for Automated Regulatory Reporting and Audit Readiness. Metallurgical and Materials Engineering, 30(4), 996–1010. Retrieved from https://metall-mater-eng.com/index.php/home/article/view/1936
[59] Wooldridge, M. (2009). An introduction to multiagent systems (2nd ed.). Wiley.
[60] Garapati, R. S. (2025). An Intelligent IoT Security System: Cloud-Native Architecture with Real-Time AI Threat Detection and Web Visualization. Journal homepage: https://jmsronline. com, 2(06).
[61] Zhang, A., Xing, L., Zou, J., & Wu, J. C. (2022). Shifting ML for healthcare to deployment. Nature Biomedical Engineering, 6, 1330–1345.
[62] GUNTUPALLI, R. (2025). EXPLAINABLE AI IN CLINICAL DECISION SUPPORT: INTERPRETABLE NEURAL MODELS FOR TRUSTWORTHY HEALTHCARE AUTOMATIONEXPLAINABLE AI IN CLINICAL DECISION SUPPORT: INTERPRETABLE NEURAL MODELS FOR TRUSTWORTHY HEALTHCARE AUTOMATION. TPM–Testing, Psychometrics, Methodology in Applied Psychology, 32(S9 (2025): Posted 15 December), 462-471.
[63] Benford, S., et al. (2009). Emergent multi-agent architectures. Autonomous Agents and Multi-Agent Systems, 18, 15–45.
[64] Inala, R. (2025). A Unified Framework for Agentic AI and Data Products: Enhancing Cloud, Big Data, and Machine Learning in Supply Chain, Insurance, Retail, and Manufacturing. EKSPLORIUM-BULETIN PUSAT TEKNOLOGI BAHAN GALIAN NUKLIR, 46(1), 1614-1628.
[65] Ferber, J. (1999). Multi-agent systems: An introduction. Addison-Wesley.
[66] Bandi, V. D. V. K. (2023). Production-Grade Machine Learning Pipelines For Healthcare Predictive Analytics. South Eastern European Journal of Public Health, 189–205. Retrieved from https://www.seejph.com/index.php/seejph/article/view/7057
[67] Kephart, J. O., & Chess, D. M. (2003). The vision of autonomic computing. Computer, 36(1), 41–50.
[68] Aitha, A. R., & Jyothi Babu, D. A. (2025). Agentic AI-Powered Claims Intelligence: A Deep Learning Framework for Automating Workers Compensation Claim Processing Using Generative AI. Available at SSRN 5505223.
[69] Huhns, M. N., & Singh, M. P. (1998). Readings in agents. Morgan Kaufmann.

[70] Amistapuram, K. (2025). GENERATIVE AI FOR CLAIMS EXCEPTIONS AND INVESTIGATIONS: ENHANCING RESOLUTION EFFICIENCY IN COMPLEX INSURANCE PROCESSES. Available at SSRN 5785482.
[71] Erl, T. (2016). Microservices design patterns. Prentice Hall.
[72] Gottimukkala, V. R. R. (2025). Generative AI for Exceptions and Investigations: Streamlining Resolution Across Global Payment Systems. Journal of International Commercial Law and Technology, 6(1), 969-972.
[73] Fowler, M. (2018). Refactoring (2nd ed.). Addison-Wesley.
[74] Segireddy, A. R. (2025). GENERATIVE AI FOR SECURE RELEASE ENGINEERING IN GLOBAL PAYMENT NETWORK. Lex Localis: Journal of Local Self-Government, 23.
[75] Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). Design patterns. Addison-Wesley.
[76] Amistapuram, K. (2025). Agentic AI for Next-Generation Insurance Platforms: Autonomous Decision-Making in Claims and Policy Servicing. Journal of Marketing & Social Research, 2, 88-103.
[77] Rieke, N., Hancox, J., Li, W., et al. (2020). Federated learning for digital health. NPJ Digital Medicine, 3, 119.
[78] Zaharia, M., et al. (2010). Spark: Cluster computing with working sets. HotCloud.
[79] Rongali, S. K., & Varri, D. B. S. (2025). AI in health care threat detection. World Journal of Advanced Research and Reviews, 25(3), 1784-1789.
[80] Lakshman, A., & Malik, P. (2010). Cassandra. ACM SIGOPS Operating Systems Review, 44(2), 35–40.
[81] Nagubandi, A. R. (2025). PIONEERING SELF-ADAPTIVE AI ORCHESTRATION ENGINES FOR REAL-TIME END-TO-END MULTI-COUNTERPARTY DERIVATIVES, COLLATERAL, AND ACCOUNTING AUTOMATION: INTELLIGENCE-DRIVEN WORKFLOW COORDINATION AT ENTERPRISE SCALE. Lex Localis, 23(S6), 8598-8610.
[82] Stonebraker, M., & Çetintemel, U. (2005). One size fits all? ICDE Proceedings, 2–11.
[83] Yandamuri, U. S. (2022). Big Data Pipelines for Cross-Domain Decision Support: A Cloud-Centric Approach. International Journal of Scientific Research and Modern Technology, 227.
[84] Moreira, M. W. L., et al. (2018). IoT-based smart healthcare systems. Sensors, 18(4), 1155.
[85] Guntupalli, R. (2025). Multi-Cloud vs. Hybrid Cloud Security: Key Challenges and Best Practices. Hybrid Cloud Security: Key Challenges and Best Practices (November 21, 2025).
[86] Mell, P., & Grance, T. (2011). The NIST definition of cloud computing. NIST.
[87] Pamisetty, A., Paleti, S., Adusupalli, B., Singireddy, J., Inala, R., & Nagabhyru, K. C. (2025, September). Explainable AI Systems for Credit Scoring and Loan Risk Assessment in Digital Banking Platforms. In 2025 IEEE 13th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS) (pp. 1478-1483). IEEE.
[88] World Health Organization. (2021). Ethics and governance of artificial intelligence for health. WHO Press.
[89] [Kolla, S. H. (2024). RETRIEVAL-AUGMENTED GENERATION WITH SMALL LLMS FOR KNOWLEDGE-DRIVEN DECISION AUTOMATION IN ENTERPRISE SERVICE PLATFORMS. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 15(3), 476–486. https://doi.org/10.61841/turcomat.v15i3.15497
[90] Moreau, L., et al. (2015). The W3C PROV family of specifications. Future Generation Computer Systems, 29(7), 161–165.
[91] Rongali, S. K. (2025, August). AI-Powered Threat Detection in Healthcare Data. In 2025 International Conference on Artificial Intelligence and Machine Vision (AIMV) (pp. 1-7). IEEE.
[92] Jennings, N. R., & Wooldridge, M. (1998). Applications of intelligent agents. Springer.
[93] Van Roy, P. (2009). Self-management in distributed systems. IEEE Computer, 42(12), 40–47.
[94] Vardhan Kumar Bandi, V. D. (2024). Automated Feature Engineering Systems in Large-Scale Healthcare Data Environments. Journal of Neonatal Surgery, 13(1), 2127–2141. Retrieved from https://www.jneonatalsurg.com/index.php/jns/article/view/10004
[95] Kairouz, P., McMahan, H. B., Avent, B., et al. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210.
[96] Nagabhyru, K. C., & Babu, A. J. Human In The Loop Generative AI: Redefining Collaborative Data Engineering For High Stakes Industries.
[97] McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 1273–1282.