Scientific Research Publishing

Journal of Computer and Communications, 2026, 14(8), 26-48
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Privacy, Security, and Trust Challenges in Federated Learning: A Systematic Review and Future Research Agenda

Mitende Nicholus Nyapete, Richard Omolo, Newton Masinde

Department of Computer Science and Software Engineering, Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya
Email: mnyapete@gmail.com, comolo@gmail.com, amphinewt@gmail.com

How to cite this paper: Nyapete, M.N., Omolo, R. and Masinde, N. (2026) Privacy, Security, and Trust Challenges in Federated Learning: A Systematic Review and Future Research Agenda. *Journal of Computer and Communications*, **14**, 26-48.
https://doi.org/10.4236/jcc.2026.148002

Received: July 8, 2026
Accepted: August 15, 2026
Published: August 18, 2026

Abstract

Background: Federated Learning has emerged as a distributed machine learning paradigm that enables entities to collaboratively train artificial intelligence models without directly sharing client raw data. Federated learning enhances privacy, security, and regulatory compliance while still enabling the development of a robust, accurate, aggregated global model. Despite successes of federated learning architecture in improving the integrity, confidentiality, and availability of data between communicating entities, the model faces privacy and security challenges that hinder its widespread adoption and effectiveness in real-world applications. The systematic review aims to explore federated learning architecture that have been applied to train AI models without direct sharing of client raw data, discuss how these models have been applied in training AI to provide privacy and accountability as well as long-term elimination planning, and address the methodological strengths, limitations, and challenges in implementing them, and policy and strategic implications. **Methods**: The study presents a systematic review of federated learning architectures to investigate privacy challenges in the federated learning architecture, with a focus on the strengths, weaknesses, and practical applications of different approaches. Following PRISMA 2020 guidelines, Literature screening was done on sixteen databases (Google Scholar, Semantic Scholar, PubMed, Springer Nature, Research Gate, ScienceDirect, IEEE, Scilit, ACM digital library, Wiley online library, SciSpace, National foundation (.gov), HAL open science database, open Ukrainian citation index, open review, Iniria, and nature.com) since their inception upto march 2026. Eligible articles were screened, and data extracted from the articles, architecture type analysed, and trust challenges. **Results:** Following systematic screening, 208 studies met the inclusion criteria. The most prevalent architectures were privacy-focused architectures.

From the analysis results, the study found that limited fairness, scalability, and transparency are the major gaps in the existing literature that need to be addressed in future studies. **Conclusion**: The study concludes that future federated learning architectures should be tailored to address fairness, scalability, and transparency, which will be critical in achieving effective privacy in computer systems, thereby improving client confidence.

Keywords

Federated Learning, Attacks, Privacy, Security, Scalability, Fairness, Transparency, Trust Challenges

1. Introduction

Federated Learning has emerged as a distributed machine learning paradigm that allows entities to collaboratively train artificial intelligence models without direct sharing of client raw data [1]. This is achieved by keeping data localized on user devices. Federated learning enhance privacy, security, and regulatory compliance while still allowing development of robust and accurate aggregated global model, this decentralization of has increasingly gain popularity in critical sectors: healthcare, finance, education, Internet of Things (IoT), edge computing, environment monitoring, emergency systems, transport infrastructure management, material science, and smart systems, where data integrity, confidentiality, and availability are of major concerns [2].

Despite many successes of federated learning architecture in improving integrity, confidentiality, and availability of data between communicating entities, the model still experiences several privacy and security challenges that hinder its widespread use, adoption, and effectiveness in real-life applications. These privacy and security challenges have been found to include communication overhead due to frequent model updates, statistical heterogeneity arising from non-independent and identically distributed data, limited transparency and explainability of the learned model, fairness among clients, and vulnerability to attacks [3], additionally, scalability, trustworthiness, struggler effect, and coordination among distributed participants remains a significant drawback in the model. These trust challenges affect the performance and reliability of the model and data utility, therefore impacting the effectiveness of privacy-preserving learning. The trust challenges that have been observed in the standard federated learning architecture have led to the growing need for the development of a privacy-preserving federated learning architecture that ensures transparency and scalability. In this study, privacy was defined as protection of raw client data and model updates from unauthorized inference and or reconstruction, security was defined as resilience of the training and aggregation process against poisoning, backdoor, and inference attacks, and trust was defined as the distribution of accountability, auditability, and verifiability across participants, reducing reliance on any single party. The

study analysed articles that have carried out the implementation of federated learning architecture for privacy preservation and security to identify the existing gaps, trust challenges and innovations that are needed to improve performance of the architecture and reduce bias, improve transparency, accountability, scalability, and fairness to meet key principles of privacy in federated learning environment.

2. Methods

2.1. Study Design

A comprehensive systematic search was conducted across sixteen peer-reviewed databases: Google Scholar, Semantic Scholar, PubMed, Springer Nature, research gate, ScienceDirect, IEEE, Scilit, ACM digital library, Wiley online library, SciSpace, National foundation (.gov), HAL open science database, open Ukrainian citation index, open review, Iniria, and nature.com from inception to March 2026. The review process adhered to the PRISMA 2020 guidelines [4] for systematic reviews and meta-analyses. The research questions that were used to guide this study review were: 1) What privacy-preservation, security, and trust mechanisms for federated learning architectures have been proposed, and threats being addressed? 2) What methodological strengths, limitations, implementation challenges, and policy implications have been reported for each mechanism category?

2.2. Search Strategy

Sixteen scientific databases were searched for suitable studies using search terms "federated learning, privacy in federated learning", "security in federated learning" and "challenge", search results was restricted to title, abstract, and affiliation fields while using the corresponding search terms. The results obtained from the search for the terms and corresponding keywords were merged using 'OR', and the combined results were further combined using a Boolean operator 'AND' as shown in Table 1. The results obtained from the search were imported into Zotero software for duplicate removal and screening.

Table 1. Search concept groups and terms.

No.	Concept Group	Search Terms (combined using OR)
1	Core technology	Federated learning **OR** FL
2	Privacy	Privacy **OR** Privacy-preserving **OR** Data confidentiality **OR** Differential privacy
3	Security	Security **OR** Adversarial attack **OR** Poisoning attack **OR** Backdoor attack **OR** Gradient leakage **OR** Model inversion
4	Trust	Trust **OR** Trustworthy **OR** Blockchain **OR** Decentralized trust **OR** Auditability
5	Fairness	Fairness **OR** Bias **OR** Equitable aggregation
6	Scalability	Scalability **OR** Cross-device **OR** Resource-constrained **OR** Communication overhead
7	Transparency	Transparency **OR** Explainability **OR** Accountability

And the complete Boolean search string that was applied is "Federated learning **OR** FL" AND " Privacy **OR** Privacy-preserving **OR** Data confidentiality **OR** Differential privacy **OR** Security **OR** Adversarial attack **OR** Poisoning attack **OR** Backdoor attack **OR** Gradient leakage **OR** Model inversion **OR** Trust **OR** Trustworthy **OR** Blockchain **OR** Decentralized trust **OR** Auditability **OR** Fairness **OR** Bias **OR** Equitable aggregation **OR** Scalability **OR** Cross-device **OR** Resource-constrained **OR** Communication overhead Transparency **OR** Explainability **OR** Accountability".

Table 2 describes the inclusion and exclusion criteria that were used to filter out records that did not fit within the predetermined inclusion and exclusion criteria at this stage.

Table 2. Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
Published after January 2015	Published before January 2015
Papers proposing or evaluating a mechanism addressing privacy, security, or trust in federated learning (including fairness, scalability, or transparency as a stated design goal)	Papers not addressing privacy, security, or trust in federated learning
Papers reporting an explicit threat model or adversarial setting (such as honest-but-curious server, malicious client, external eavesdropper)	Papers with no stated threat model or adversarial assumption
Papers published in English	Studies published in languages other than English
Scientific peer-reviewed publication (Journal or Conference paper)	Systematic reviews, meta-analyses, editorials, and non-peer-reviewed preprints

2.3. Quality Evaluation

The quality of methodology of studies included was assessed, and the evaluation concentrated on major points of model development and reporting, such as the clarity of the problem and modelling objective, outcomes, transparency in model construct and modelling parameter, connection with prior federated learning architecture studies, sensitivity analysis, and trust challenges associated with federated learning architecture. There was also an evaluation of additional issues, including parameter specification and model transparency. The methodological quality of each parameter included in the study was appraised using a structured five-item rubric derived from the evaluation criteria applied throughout this review (Table 3). Each item was scored as 0-absent, 1-partially addressed, and 2-fully addressed, yielding a score of 10 per study.

Extracted data were cross-checked by a second reviewer for a random 20% sample of included studies to verify extraction accuracy; discrepancies were resolved by reference to the source text. The methodological quality of each included study was appraised using a structured five-item rubric derived from the evaluation criteria applied throughout this review (Table 4). Each item was scored 0 (absent), 1

(partially addressed), or 2 (fully addressed), yielding a total possible score of 10 per study.

Table 3. Data extraction fields.

No.	Field	Description
1	Citation/reference ID	Author (s), year, and reference number as used in this review
2	Application domain	Healthcare, finance, IoT, edge computing, general-purpose
3	Primary mechanism category	Cryptographic/Differential privacy/Blockchain-based/ Attack-and-defines/Architectural-governance
4	Trust model assumed	Honest-but-curious server, malicious adversary, or not specified.
5	Threat actor (s) considered	Server, client, external eavesdropper, colluding server-client
6	Granularity of protection	Sample-level, client-level, or update-level
7	Privacy/security mechanisms addressed,	Model poisoning, gradient leakage, model inversion, and reconstruction attack
8	Evaluation metric (s) reported	Model accuracy, privacy budget (ϵ, δ), communication cost, and latency
9	Dataset (s) used	Benchmark dataset, real-world dataset, or simulated/synthetic data
10	Reported strengths	As stated by the original authors
11	Reported limitations	As stated by the original authors
12	Quality appraisal score	Score assigned using the rubric in Table 4

Table 4. Quality appraisal rubric.

Criterion	0-Absent	1-Partial	2-Fully Addressed
Problem and modelling objective clarity	Not stated	Implied but not explicit	Explicitly stated
Outcome/results reporting	Not reported	Partially quantified	Fully quantified with defined metrics
Model and parameter transparency	Not disclosed	Partially disclosed (architecture only)	Fully disclosed (privacy budget, architecture, dataset all reported)
Threat model specification	Not stated	Vague ("adversary" undefined)	Explicit (adversary type and actor named)
Comparison with prior federated learning work	Absent	Cited without direct comparison	Explicit comparative discussion of strengths/limitations relative to prior work

2.4. Model Grouping and Evaluation Framework

In order to provide a systematic synthesis of federated learning architecture, we classified the studies that were included based on the main aim of the research. All of the eligible modelling studies were found in the process of full-text screen-

ing. Within each model, we identified and compared the essential characteristics of all the included models.

3. Results

3.1. Study Selection

A total of 395 articles were identified from the search. 30 duplicates were removed, which led to 365 articles being screened by title and abstract, with 145 articles selected for full-text evaluation. For final analysis, 73 articles [1]-[3], [5]-[74] were included in the final analysis (Figure 1). The number of full-text articles that were rejected after careful screening was 64. This comprised of studies that were not analysing privacy preservation (n = 20), reviews (n = 29), and those that did not include attacks (n = 15).

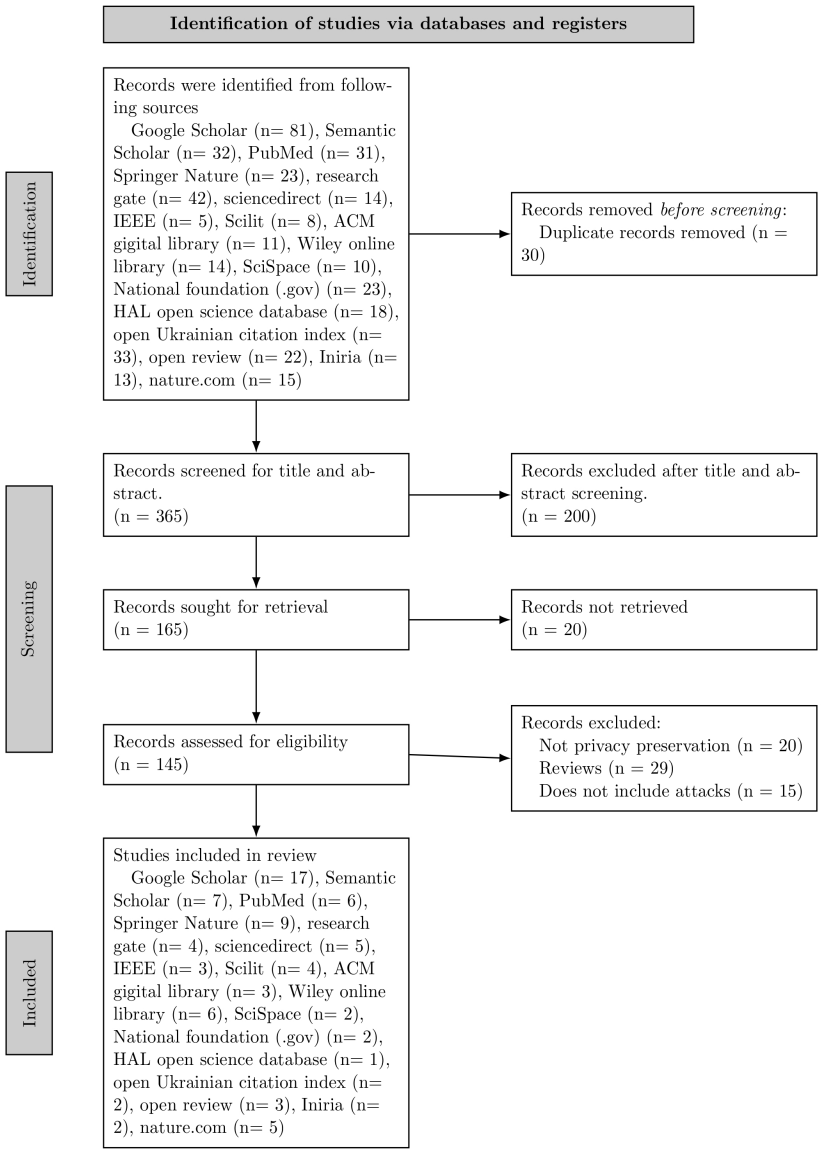

Figure 1. Summary of systematic screening of identified articles.

3.2. Classification of Privacy-Preserving Models in Federated Learning Architecture

3.2.1. Model Grouping and Evaluation Framework

To provide a systematic synthesis of the mechanisms that address privacy-preserving, security, and trust in federated learning architectures, the studies reviewed were classified according to the primary mechanism used to protect data confidentiality, resist adversarial manipulation, or distributed accountability across the federated learning lifecycle (local training, communication, and aggregation). All eligible studies identified were grouped into five broad categories based on the dominant privacy-preservation strategy employed.

1) Cryptographic privacy-preserving models which rely on secure multiparty computation, homomorphic encryption, or Zero-Knowledge Proofs (ZKPs) to mathematically guarantee that raw data or gradients are never exposed in plaintext to the aggregator or other participants.

2) Perturbation-based architectures (Differential Privacy, DP) which inject calibrated statistical noise into gradients, model updates, or aggregated outputs to bound the privacy loss of any individual participant's data.

3) Blockchain-based decentralized trust models which remove or minimize reliance on a central aggregator by distributing trust, auditability, and incentive management across a distributed ledger.

4) Attack-and-defence models which characterize the Privacy and security threat surface of federated learning (inference attacks, model inversion, poisoning, backdoors) and propose corresponding robust-aggregation or detection mechanisms.

5) Architectural and governance frameworks which address privacy holistically at the level of system design reference architectures, fairness and accountability layers, and healthcare/IoT-specific deployment patterns rather than through a single cryptographic or statistical mechanism.

Within each category, the essential characteristics of the included models were compared using the following evaluation criteria.

1) Model structure and assumptions covering trust model assumed (honest-but-curious vs malicious adversary), the threat actors considered (server, client, external eavesdropper), and the granularity of protection (sample-level, client-level, or update-level privacy).

2) Privacy mechanism address: model poisoning, gradient leakages, and inversion, and reconstruction attacks.

3) Strengths and limitations examined in terms of computational and communication overhead, scalability to cross-device settings, utility-privacy trade-off, and resistance to adaptive adversaries.

4) Comparative advantages—identifying what differentiates each model from others within or across categories.

5) Contextual use cases suggesting the settings in which each model type is most applicable, such as cross-silo healthcare consortia, cross-device mobile networks,

or IoT edge environments.

3.2.2. Cryptographic Privacy-Preserving Models

Cryptographic approaches form substantial portion of the analysed literature, which reflect the emphasis on provable, rather than merely empirical privacy guarantees. Secure multiparty computation allows multiple clients to jointly compute an aggregate model update without any single party learning the others' raw contributions. Work in this vein has been applied specifically to sensitive domains, including financial applications, where SMC has been combined with differential privacy to satisfy regulatory confidentiality requirements [5], and histopathology imaging, where cluster-based SMC has been used to protect diagnostic data shared across hospitals [6]. Extensive investigations have indicated that federated learning architecture is fundamentally multiparty computation problem, arguing that many aggregation protocols can be understood as specialized instances of general SMC theory [7]; other work has systematically catalogued the safeguards needed to secure the federated learning framework as a whole [8].

Homomorphic Encryption (HE) extends this guarantee by allowing the aggregator to perform arithmetic operations (summation of gradients) directly on encrypted ciphertexts, so that the server never has access to any client's update in plaintext. Systems such as FedML-HE have demonstrated that HE-based aggregation can be efficient enough for practical deployment [9]. At the same time, other studies have proposed multiparty homomorphic encryption schemes tailored for secure aggregation in federated learning [10] and have surveyed HE is broader contributions to privacy-preserving healthcare analytics [11]. HE has also been extended beyond conventional deep networks to spiking neural network architectures, indicating the mechanism's generalizability across model families [12]. A parallel line of work has applied simpler symmetric or asymmetric encryption models directly to federated learning update transmission to secure communication channels against eavesdropping [13].

Zero-Knowledge proofs constitute the third cryptographic pillar, which enable clients or server to prove that computation was performed honestly without revealing data or model parameters. zk-SNARK-based frameworks have been proposed to make blockchain-based FL verifiable end-to-end, ensuring that both training and aggregation steps can be audited without compromising confidentiality [14]-[16]. Related work has combined ZKPs with decentralized federated learning to eliminate the need for a trusted central verifier [17] [18], and quantum-resistant ZKP constructions have been explored for user authentication in adjacent distributed-systems contexts [19].

Comparative Advantage

Cryptographic models offer strong privacy guarantees since they do not rely on statistical assumptions about the adversary behaviour, their limitation is computational and communication overhead, which restricts their application in a resource-constrained, cross-device settings such as mobile phones or IoT sensors, and confines most reported deployments to cross-setting scenarios with a small

number of well-resourced institutional participants.

3.2.3. Differential Privacy

Differential Privacy has become the most widely adopted statistical framework for bounding privacy loss in federated learning, owing to its composability and its formal (ϵ, δ)-guarantee that is independent of the adversary's auxiliary knowledge. Foundational work establishes an algorithmic and performance trade-offs of applying differential privacy directly to the federated learning training loop, showing that carefully calibrated noise addition can preserve acceptable model utility while providing rigorous privacy bounds [20]. Subsequent studies have extended these foundations to secure, stateful aggregation protocols that combine differential privacy with cryptographic secure aggregation to reduce the amount of noise required for a given privacy budget [5], and have surveyed the wide variance in how differential privacy has been operationalized across the federated learning literature, noting substantial inconsistency in reported privacy budgets and threat models [21].

Several studies have examined differential privacy behaviour under real-world deployment constraints. Mobility and client churn have been shown to affect the achievable Privacy-utility trade-off in differentially private federated learning, since intermittent client participation complicates privacy accounting across multiple communication rounds [22]. Hybrid approaches that integrate differential privacy and lightweight cryptographic techniques have been suggested to be implemented in cross-IoT platforms in order to balance confidentiality and limited computational budget of edge devices [23]. Studies have also explored joint application of differential privacy and secure aggregation "belt and braces" strategy to defend simultaneously against curious servers and colluding clients [24]. Differential privacy has also been applied beyond the core federated learning literature to public health data sharing more broadly, illustrating its transferability as a privacy accounting framework for epidemiological and health-system data releases [25], and to location-based services, where geographic granularity introduces additional re-identification risk beyond that of conventional tabular data [26].

Comparative Advantage

Differential privacy mechanisms are largely lightweight and can be placed on top of standard FedAvg aggregation pipeline with little architectural change, making them attractive for cross-device federated learning at scale. Their central limitation is the well-documented privacy-utility trade-off: strong privacy guarantees (small ϵ) tend to degrade model accuracy, particularly for non-IID client data distributions, and most of the reviewed studies do not specify how privacy budgets should be tuned for realistic multi-round training regimes.

3.2.4. Blockchain-Based Decentralized Trust Architecture

A large and rapidly growing subset of the literature addresses privacy not through a cryptographic primitive alone, but by redesigning the federated learning architecture to remove the single point of trust represented by a central aggregation

server. Blockchain-based federated learning frameworks distribute the aggregation, auditing, and incentive functions across a distributed ledger to prevent unilateral inspection by a party, tampering or withhold client contributions. Systematic surveys have mapped this particular design for IoT deployments, and privacy guarantees across architectures [27]. Comparative studies of blockchain against centralized authentication architectures also suggest that blockchain-based designs are advantageous where IoT devices cannot rely on a persistently available, trusted third party [28].

Specific architectural contributions including committee consensus frameworks reduce computational burden of a full-network validation as it preserve decentralized trust [29]; trustworthy federated learning architectures that combine blockchain with reputation-based client selection to mitigate free-riding and malicious participation [30]; and audit-oriented designs that use blockchain to provide transparent, tamper-evident logging of encrypted training data provenance without revealing the data itself [31] [32]. Game-theoretic incentive mechanisms layered on top of blockchain have also been proposed to align individual client incentives with honest participation, addressing a privacy-adjacent concern: clients who anticipate no reputational or financial benefit have reduced incentive to protect the integrity of their contributions [33].

Blockchain has additionally been combined with the cryptographic and perturbation mechanisms discussed above to create hybrid, defence-in-depth systems: examples include zk-SNARK-verified blockchain FL [14]-[18], blockchain-secured federated learning explicitly designed to resist poisoning attacks (BPFL) [34], and frameworks integrating adaptive differential privacy with blockchain and dynamic masking specifically for Internet-of-Vehicles applications, which also incorporate explainable AI (XAI) components to support auditability [35]. Domain-specific deployments include a blockchain-entangled FL architecture for healthcare 5.0 systems [36] and blockchain-secured LSTM autoencoder models for transaction-level anomaly detection [37].

Comparative Advantage

Blockchain-based models are distinctive in addressing trust distribution rather than data confidentiality; therefore, they are valuable where the central-server threat model is of primary concern and where auditability are regulatory requirements. Their limitation include: 1) consensus latency, and 2) storage overhead of the ledger.

3.2.5. Attack-and-Defence Models

A substantial body of work approaches federated learning privacy not by proposing a protective mechanism directly, but by characterizing the attack surface that motivates such mechanisms, and by proposing corresponding defences. On the attack side, studies have demonstrated that federated learning very design, which requires clients to share gradients or model updates, creates channels for gradient leakage and model inversion attacks that can reconstruct substantial information about a client's private training data [38] [39]. Backdoor attacks, in which a mali-

cious client trains a submodel to embed a hidden trigger while evading server-side anomaly detection, have been shown to succeed even under standard FedAvg aggregation [40], and refined evasion variants of such attacks continue to be documented [41]. Data poisoning attacks that corrupt the training process by injecting mislabelled or adversarial perturbed samples have similarly been catalogued and benchmarked [42]. Attacks that specifically defeat secure aggregation protocols by exploiting model inconsistency between rounds illustrate that cryptographic protection of the aggregation channel does not, by itself, guarantee input privacy if the aggregation protocol's assumptions are violated [43] [44].

Robust-aggregation method has been proposed to detect and down weight anomalous and Byzantine client updates, thereby limiting the impact of poisoning and backdoor attacks without requiring a fully cryptographic pipeline [45] [46]. Provable defence frameworks such as FLIP explicitly target backdoor mitigation with formal guarantees rather than purely heuristic anomaly detection [47]. Comprehensive surveys have organized this attack–defence literature into structured taxonomies covering the full federated learning pipeline from data collection, through local training, to aggregation, and have highlighted persistent gaps between attacks demonstrated in controlled experimental settings and the assumptions realistic federated learning deployments can satisfy [48]-[53]. A recurring critique in this sub-literature is that many published attacks and defences rely on idealized assumptions (full visibility into the global model, a fixed, known number of malicious clients) that may not transfer to production-scale, cross-device FL systems [51].

Related, though conceptually adjacent, is a smaller set of studies that address fairness and bias as privacy-relevant concerns, since biased aggregation can itself leak information about which client populations are under- or over-represented in the global model. Fairness-aware aggregation methods based on core-stability concepts [54] and multi-gradient descent with fairness guidance (FedMDFG) [55] have been proposed alongside broader surveys of bias and fairness in machine learning that contextualize these FL-specific contributions within the wider algorithmic fairness literature [56]-[58]. Techniques for mitigating bias directly during federated aggregation have also been proposed as a complementary safeguard alongside privacy-preserving mechanisms [59].

Comparative Advantage

This category is unique in that its primary contribution is diagnostic rather than protective. It defines the threat models that the other four categories of models are designed to counter. Its main limitation, noted consistently across the surveyed literature, is a persistent gap between attack sophistication (which continues to escalate, such as refined backdoor evasion [41]) and the generalizability of defence.

3.3. Architectural and Governance Frameworks

Final category included studies that address federated learning privacy holistically

at the system architecture level, rather than through a discrete cryptographic, statistical, or ledger-based mechanism. Reference architecture articles have suggested structured taxonomies of architectural patterns for federated learning systems, incorporating privacy, security, and communication efficiency as design dimensions not as after-the-fact additions [60] [61]. Decentralized federated learning research have documented a shift from single-server aggregation to peer-to-peer and gossip-based topologies, arguing that decentralization itself constitutes a privacy-enhancing design choice by removing the aggregator as a single point of data exposure [1] [62] [63]. Scalable, fault-tolerant, and decentralized architectures have been proposed to provide and maintain privacy guarantees under node churn and partial network failures [64], and general-purpose federated learning platforms such as FED have been engineered to make privacy-preserving deployment configurations more accessible to non-specialist adopters [65].

Domain-specific governance frameworks are especially prominent in healthcare, reflecting the sector's stringent regulatory requirements. Studies have scoped the intersection of federated learning, privacy-enhancing technologies, and data protection law in medical research [66], examined FL's suitability for handling privacy-sensitive medical data more broadly [67], and reviewed clinical applications and technical architectures for healthcare federated learning specifically [68]. Neuroimaging-specific architectures have been developed to secure multi-site brain-imaging consortia [69], and privacy-preserving FL has been applied to accelerate AI adoption in internet-of-medical-things (IoMT) environments [70]. Reviews of privacy preservation for federated learning in healthcare argue that architectural and regulatory alignment over single technical mechanism is binding constraint adoption [71]. Some studies have addressed specific challenge of anonymizing healthcare data prior to its use in federated learning pipelines, treating anonymization as architectural pre-processing layer rather than a training-time mechanism [72] [73].

Comparative Advantage

The architectural and governance frameworks are the only category that clearly address organizational and regulatory constraints alongside technical privacy mechanisms. Their limitation is frequent descriptive/taxonomic rather than prescriptive, offering design guidance over deployable protocols, and their practical impact depends heavily on faithful the implementers are in translating architectural recommendations into working systems.

3.4. Strengths of Existing Architectures

Across all five set of categories, the analysis results demonstrated a rich, complementary toolkit for addressing privacy in federated learning at various points in the training pipeline. Cryptographic models offer the strongest formal guarantees against both honest but curious and malicious adversaries; Differential privacy models offer lightweight, composable privacy accounting well suited to large scale cross device deployment; blockchain based models address the structural risk

posed by a single trusted aggregator; attack and defence studies provide the empirical grounding needed to prioritize which threats matter most in practice; and architectural frameworks connect these technical mechanisms to the regulatory and organizational realities of deployment domains such as healthcare and IoT. Several studies further demonstrate that these mechanisms are not mutually exclusive: Differential privacy has been combined with secure aggregation [5] [24], and blockchain has been combined with ZKPs [14]-[17] and with adaptive differential privacy and masking [35], suggesting that defence-in-depth approaches are both feasible and increasingly common.

3.5. Limitations of Current Architectures

Despite these strengths, several limitations recur throughout the literature. Strong differential privacy guarantee reduces model accuracy, and small number of studies have provided practical guidance on selecting privacy budgets for realistic, non-IID, multi-round federated learning settings [20]. While cryptographic models offer strong guarantees, they impose computational and communication overheads which prohibits smooth cross-device and resource-constrained deployments, effectively restricting their practical use to well-resourced cross-silo consortia [9] [10]. Blockchain-based federated learning models provide consensus latency and ledger storage overhead, and, as noted above, typically require a complementary content-protection mechanism, since they secure the aggregation process rather than the substance of client updates. Attack and defence studies are frequently validated against narrow, idealized threat models and single attacks, limiting confidence that reported defences generalize to adaptive real-world adversaries [51]. Despite the architectural and governance frameworks providing valuable connections to technical and regulatory considerations, they are often descriptive rather than prescriptive, providing implementers with no concrete benchmarked configurations.

3.6. Application and Use Case

The literature demonstrated several converging applications of privacy-preserving federated learning architecture. Cryptographic and differential privacy mechanisms in healthcare have been combined with governance frameworks to support multi-institutional model training without centralizing patient data [66]-[69] [71]. Lightweight hybrid differential privacy-cryptographic schemes and blockchain-secured architectures in IoT and edge environments have been proposed to accommodate constrained device resources while still providing auditable privacy guarantees [23] [27] [35]. Secure multiparty computation and differential privacy have been jointly deployed in finance to satisfy regulatory confidentiality requirements while enabling cross-institutional fraud detection and risk modelling [5] [74]. Across all domains, attack-and-defence studies play a cross-cutting diagnostic role, informing which other mechanisms are prioritized for a given deployment's threat model.

3.7. Cross-Cutting Themes and Practical Implications

Several themes recur across the five architecture categories: "no single mechanism has achieved strong formal privacy guarantees and low computational overhead at the same time" the literature consistently frames privacy-preserving federated learning design as a trade-off space rather than a solved problem, with cryptographic models trading efficiency for guarantee strength, and differential privacy models trading accuracy for guarantee strength. "Trust-model assumptions are frequently under-specified" many studies do not clearly distinguish between honest but curious and actively malicious adversary models, which complicates comparison across proposed mechanisms and across the surveyed categories. "Collusion between the server and a subset of clients remains inadequately addressed" in a large share of the reviewed cryptographic and differential privacy literature, even though this threat model is increasingly recognized as realistic in cross-silo deployments with commercial competitors as co-participants. Auditability and non-repudiation, ability to verify, after the fact, that a training round was conducted correctly and privately, emerge as a distinct requirement that neither differential privacy nor conventional cryptographic aggregation fully satisfies on its own, motivating the growing integration of blockchain and ZKP-based verification alongside these mechanisms [14]-[18]. Fifth, "regulatory alignment", particularly in healthcare and financial domains, is treated as a binding constraint on adoption, independent of technical maturity, reinforcing the role of architectural and governance frameworks as a necessary complement to the four more technically defined categories [66] [71].

A further cross-cutting weakness, echoed across the attack-and-defence literature, is that privacy evaluation in federated learning research is rarely adversarial in the way real deployments require, most reported defences are validated against the specific attack used to motivate the paper, rather than against an adaptive adversary aware of the defence mechanism itself [51]. This mirrors a broader methodological gap in which formal privacy guarantees (ϵ-DP bounds, cryptographic security proofs) are not always matched by empirical red-team evaluation of the deployed system as a whole.

3.8. Comparative Advantages of Different Architecture Approaches

The five categories of privacy-preserving federated learning architecture reviewed here serve complementary rather than competing roles in securing the federated learning pipeline. Cryptographic architectures (SMC, HE, ZKPs) form the strongest layer of protection against a curious or malicious aggregator, and are appropriate where regulatory or contractual requirements demand provable, rather than probabilistic, privacy guarantees such as in cross-institutional healthcare or financial consortia with a small number of high-value participants [5] [9]-[11]. Differential privacy models offer a lightweight composable mechanism that is well suited to a large-scale cross-device federated learning, where cryptographic over-

head is prohibitive. However, their guarantees come at the cost of model utility, and their practical calibration remains an open challenge [20] [21]. Blockchain-based models help in addressing structurally different concern: eliminating a single trusted aggregator. They are most valuable where auditability, non-repudiation, or incentive alignment across mutually distrusting organizations is the primary requirement, rather than the confidentiality of any individual update in isolation [27] [30]. Attack-and-defence models do not, in themselves, constitute a deployable privacy mechanism. However, they determine which of the other three technical categories should be prioritized for a given threat model, and their findings increasingly motivate hybrid, defence-in-depth architectures that combine, for instance, DP with secure aggregation, or blockchain with ZKPs [5] [24] [35]. Architectural and governance frameworks provide organizational framework translating technical guarantees into regulatory compliant, auditable systems, particularly in healthcare domain where legal and technical requirements are tightly coupled [66] [68] [71].

Theretofore the is no single modelling category implemented individually can address full range of federated privacy risks. An integrative combination of cryptographic or differential privacy protection of update content, blockchain-based distribution of aggregation trust, attack-informed robust-aggregation defences, and governance frameworks aligned with domain-specific regulation offers a more complete and context-specific platform for privacy-preserving federated learning deployment than any individual mechanism reviewed in isolation.

3.9. Contextual Application of the Architectures

The applicability of each of the privacy-preserving architecture category depends on deployment context, sensitivity and regulatory status of data involved, resource constraints of participating clients, and trust relationship between participants and the aggregator.

Cryptographic models are informative in cross-platform settings with a small number of well-resourced institutional participants for instance hospitals, banks, and research consortia, where computational overhead of SMC, HE, or ZKPs is affordable and where formal, auditable privacy guarantees are a contractual or regulatory necessity [5] [9] [10].

Differential privacy models are suited to cross-device with large numbers of resource-constrained clients which include mobile phones and IoT devices, where lightweight noise injection can be layered on top of standard FedAvg-style aggregation without prohibitive overhead. However, careful privacy-budget tuning is required to preserve model utility under non-IID data [20] [22] [23].

Blockchain-based models are most applicable where no single participant is willing to act as, or be trusted as, a central aggregator, for example, in multi-organizational IoT consortia or competitive commercial settings, and where auditability of the training process is itself a requirement, independent of the confidentiality of any single update [27] [30] [36].

Attack-and-defence models are most valuable during the threat-modelling and risk-assessment phase of federated learning system design, informing which combination of the other three technical categories should be deployed for a given adversarial context, and are particularly critical in high-stakes domains where poisoning or backdoor attacks could have safety-critical consequences [40] [42] [45] [47].

Architectural and governance frameworks are most applicable in regulated domains, principally healthcare and finance, where technical privacy mechanisms must be embedded within a broader compliance and consent-management structure, and where the interoperability of multiple institutions' data governance policies is itself a design constraint [66] [68] [71].

4. Discussion

The classification aggregates the privacy, security, and trust mechanisms that has been used in federated learning architectures into five categories: 1) cryptographic models, 2) differential privacy models, 3) blockchain-based decentralized trust models, 4) attack-and-defence models, and 5) architectural and governance frameworks, each addressing distinct layer of the privacy, security, and trust challenge inherent to distributed model training. This grouping mirrors the broader observation in the federated learning literature that no single technique fully resolves the tension between data utility, communication and computation cost, and formal privacy guarantees; rather, the reviewed studies suggest that privacy in federated learning is best understood as a layered, defence-in-depth problem rather than one solvable through a single mechanism.

Cryptographic approaches that are provable guarantee are required and computational resources permit in healthcare and financial applications [5] [9]-[11]. Differential privacy dominates when scalability to large client populations is paramount, at the cost of a well-documented, only partially resolved utility trade-off [20]. Blockchain-based architectures have emerged as a response to a structurally distinct concern, the risk posed by a single trusted aggregator, and increasingly serve as a substrate on which cryptographic and differential privacy mechanisms are layered to provide end-to-end verifiable privacy [14]-[18] [35]. The attack-and-defence literature serves as the field's risk-assessment layer, continually revealing new leakage channels (gradient inversion, backdoor evasion, secure-aggregation defeating attacks) that motivate revisions to the other three categories [41] [43] [44] [51]. Architectural and governance frameworks are indispensable in translating technical mechanisms into deployable, systems compliant to regulations, especially in healthcare where legal restriction on access to patient data are significant a barrier to adoption [66] [68] [71].

Gaps observed during classification process warrant serious attention in future research. These include: the trade-off between privacy guarantee strength and computational overhead, and differentiated privacy alone cannot suffice against a determined adversary, secondly, the collusion between the aggregating server and

a subset of malicious clients is inadequately modelled across many of the reviewed literature, despite its relevance. Third, evaluation of the proposed defences against adaptive, defence-aware adversaries rather than the static, single-variant attacks remains rare, limiting confidence in real-world robustness, and lastly, the integration of privacy-preserving mechanisms with governance and regulatory frameworks, rather than treating them as an afterthought to a purely technical solution, is likely to be decisive in determining whether privacy-preserving federated learning architectures move from experimental prototypes to operational deployment at scale. The classification therefore indicates that future research should consider implementation of hybrid designs combining cryptographic and deferential privacy with blockchain-based.

5. Conclusion

This systematic review demonstrates how a federated learning architecture for privacy preservation relies on both strict methodology and practical applications. The review offers researchers and policy-makers a clear framework for assessing modelling techniques in the context of their question. A federated learning architecture should consider scalability, fairness, and transparency in its implementation to ensure robustness and enhance privacy. Federated learning architecture remains the most widely used methods for enhancing privacy and offer the benefit of handling uncertainty. While modelling has advanced, more precise work is needed to meet evolving requirements. Privacy-preservation architectures should consider addressing scalability, fairness, and transparency in future developments, thereby strengthening federated learning architecture, improving privacy and accountability.

Author Contributions

Mitende Nicholus Nyapete: Conceptualization, Formal Analysis; Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Writing-original draft, Writing review & editing; Richard Omolo: Supervision, Writing-review & editing; Newton Masinde: Supervision, Writing-review & editing.

Acknowledgements

The authors would like to thank the School of Informatics and Innovative Systems at Jaramogi Oginga Odinga University of Science and Technology for providing a conducive environment for conducting this research. Richard Omolo Newton Masinde for their constructive suggestions and comments.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Gabrielli, E., Pietro, A.D., Fenoglio, D., Pica, G. and Tolomei, G. (2026) A Survey on

Decentralized Federated Learning. arXiv:2308.04604.

[2] Li, T., Sahu, A.K., Talwalkar, A. and Smith, V. (2020) Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Processing Magazine*, **37**, 50-60. https://doi.org/10.1109/msp.2020.2975749

[3] Shahid, O., Pouriyeh, S., Parizi, R.M., Sheng, Q.Z., Srivastava, G. and Zhao, L. (2021) Communication Efficiency in Federated Learning: Achievements and Challenges. arXiv:2107.10996. http://arxiv.org/abs/2107.10996

[4] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., *et al.* (2021) The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ*, **372**, n71. https://doi.org/10.1136/bmj.n71

[5] Ball, M., Bell-Clark, J., Gascon, A., Kairouz, P., Oh, S. and Xie, Z. (2024) Secure Stateful Aggregation: A Practical Protocol with Applications in Differentially-Private Federated Learning. arXiv:2410.11368.

[6] Hosseini, S.M., Sikaroudi, M., Babaei, M. and Tizhoosh, H.R. (2022) Cluster Based Secure Multi-Party Computation in Federated Learning for Histopathology Images. In: Albarqouni, S., *et al.*, Eds., *Lecture Notes in Computer Science*, Springer, 110-118. https://doi.org/10.1007/978-3-031-18523-6_11

[7] Liu, F., Zheng, Z., Shi, Y., Tong, Y. and Zhang, Y. (2024) A Survey on Federated Learning: A Perspective from Multi-Party Computation. *Frontiers of Computer Science*, **18**, Article ID: 181336. https://doi.org/10.1007/s11704-023-3282-7

[8] Ma, C., Li, J., Ding, M., Yang, H.H., Shu, F., Quek, T.Q.S. and Poor, H.V. (2020) On Safeguarding Privacy and Security in the Framework of Federated Learning. arXiv:1909.06512. http://arxiv.org/abs/1909.06512

[9] Jin, W., Yao, Y., Han, S., Joe-Wong, C., Ravi, S., Avestimehr, S. and He, C. (2023) FedML-HE: An Efficient Homomorphic-Encryption-Based Privacy-Preserving Federated Learning System. arXiv:2303.10837. http://arxiv.org/abs/2303.10837

[10] Hosseini, E., Chen, S. and Khisti, A. (2025) Secure Aggregation in Federated Learning Using Multiparty Homomorphic Encryption. arXiv:2503.00581.

[11] Munjal, K. and Bhatia, R. (2023) A Systematic Review of Homomorphic Encryption and Its Contributions in Healthcare Industry. *Complex & Intelligent Systems*, **9**, 3759-3786. https://doi.org/10.1007/s40747-022-00756-z

[12] Nikfam, F., Casaburi, R., Marchisio, A., Martina, M. and Shafique, M. (2023) A Homomorphic Encryption Framework for Privacy-Preserving Spiking Neural Networks. *Information*, **14**, Article 537. https://doi.org/10.3390/info14100537

[13] Hussien, N., Hussien, N.M., Salman, S.A. and Aljanabi, M. (2023) Secure Federated Learning with a Homomorphic Encryption Model. *International Journal Papier Advance and Scientific Review*, **4**, 1-7. https://doi.org/10.47667/ijpasr.v4i3.235

[14] Bellachia, A.A., Bouchiha, M.A., Ghamri-Doudane, Y. and Rabah, M. (2025) VerifBFL: Leveraging ZK-SNARKs for a Verifiable Blockchained Federated Learning. *NOMS* 2025-2025 *IEEE Network Operations and Management Symposium*, Honolulu, 12-16 May 2025, 1-9. https://doi.org/10.1109/noms57970.2025.11073628

[15] Ebrahimi, E., Sober, M., Hoang, A., Ileri, C.U., Sanders, W. and Schulte, S. (2024) Blockchain-Based Federated Learning Utilizing Zero-Knowledge Proofs for Verifiable Training and Aggregation. 2024 *IEEE International Conference on Blockchain* (*Blockchain*), Copenhagen, 19-22 August 2024, 54-63. https://doi.org/10.1109/blockchain62396.2024.00017

[16] Yang, J., Zhang, W., Guo, Z. and Gao, Z. (2023) Trustdfl: A Blockchain-Based Verifiable and Trusty Decentralized Federated Learning Framework. *Electronics*, **13**, 86.

https://doi.org/10.3390/electronics13010086

[17] Ahmadi, M. and Nourmohammadi, R. (2024) ZkFDL: An Efficient and Privacy-Preserving Decentralized Federated Learning with Zero Knowledge Proof. 2024 *IEEE 3rd International Conference on AI in Cybersecurity* (*ICAIC*), Houston, 7-9 February 2024, 1-10. https://doi.org/10.1109/icaic60265.2024.10433831

[18] Byrd, D. and Polychroniadou, A. (2020) Differentially Private Secure Multi-Party Computation for Federated Learning in Financial Applications. *Proceedings of the First ACM International Conference on AI in Finance*, New York, 15-16 October 2020, 1-9. https://doi.org/10.1145/3383455.3422562

[19] García-Cid, M.I., Bodanapu, D., Gatto, A., Martelli, P., Martín, V. and Ortiz, L. (2024) Experimental Implementation of a Quantum Zero-Knowledge Proof for User Authentication. *Optics Express*, **32**, Article 15955. https://doi.org/10.1364/oe.517754

[20] Wei, K., Li, J., Ding, M., Ma, C., Yang, H.H., Farokhi, F., *et al.* (2020) Federated Learning with Differential Privacy: Algorithms and Performance Analysis. *IEEE Transactions on Information Forensics and Security*, **15**, 3454-3469. https://doi.org/10.1109/tifs.2020.2988575

[21] Fu, J., Hong, Y., Ling, X., Wang, L., Ran, X., Sun, Z., Wang, W.H., Chen, Z. and Cao, Y. (2024) Differentially Private Federated Learning: A Systematic Review. arXiv:2405.08299.

[22] Kim, E. and Lee, E. (2024) Evaluating the Impact of Mobility on Differentially Private Federated Learning. *Applied Sciences*, **14**, Article 5245. https://doi.org/10.3390/app14125245

[23] Ibrahim Khalaf, O., Ashokkumar, S.R., Algburi, S., Anupallavi, S., Selvaraj, D., Sharif, M.S., *et al.* (2024) Federated Learning with Hybrid Differential Privacy for Secure and Reliable Cross-IoT Platform Knowledge Sharing. *Security and Privacy*, **7**, e374. https://doi.org/10.1002/spy2.374

[24] Ren, X., Yang, S., Zhao, C., McCann, J. and Xu, Z. (2024) Belt and Braces: When Federated Learning Meets Differential Privacy. *Communications of the ACM*, **67**, 66-77. https://doi.org/10.1145/3650028

[25] Dyda, A., Purcell, M., Curtis, S., Field, E., Pillai, P., Ricardo, K., *et al.* (2021) Differential Privacy for Public Health Data: An Innovative Tool to Optimize Information Sharing While Protecting Data Confidentiality. *Patterns*, **2**, Article 100366. https://doi.org/10.1016/j.patter.2021.100366

[26] Wang, B., Li, H., Ren, X. and Guo, Y. (2023) An Efficient Differential Privacy-Based Method for Location Privacy Protection in Location-Based Services. *Sensors*, **23**, Article 5219. https://doi.org/10.3390/s23115219

[27] Jiang, Y., Ma, B., Wang, X., Yu, G., Yu, P., Wang, Z., *et al.* (2024) Blockchained Federated Learning for Internet of Things: A Comprehensive Survey. *ACM Computing Surveys*, **56**, 1-37. https://doi.org/10.1145/3659099

[28] Khalil, U., Malik, O.A., Uddin, M. and Chen, C.-L. (2022) A Comparative Analysis on Blockchain versus Centralized Authentication Architectures for IoT-Enabled Smart Devices in Smart Cities: A Comprehensive Review, Recent Advances, and Future Research Directions. *Sensors*, **22**, Article 5168. https://doi.org/10.3390/s22145168

[29] Li, Y., Chen, C., Liu, N., Huang, H., Zheng, Z. and Yan, Q. (2021) A Blockchain-Based Decentralized Federated Learning Framework with Committee Consensus. *IEEE Network*, **35**, 234-241. https://doi.org/10.1109/mnet.011.2000263

[30] Lo, S.K., Liu, Y., Lu, Q., Wang, C., Xu, X., Paik, H.-Y. and Zhu, L. (2021) Block-

Chain-Based Trustworthy Federated Learning Architecture. arXiv:2108.06912.

[31] Noh, S. and Rhee, K. (2024) Transparent and Accountable Training Data Sharing in Decentralized Machine Learning Systems. *Computers, Materials & Continua*, **79**, 3805-3826. https://doi.org/10.32604/cmc.2024.050949

[32] Sun, Z., Wan, J., Yin, L., Cao, Z., Luo, T. and Wang, B. (2022) A Blockchain-Based Audit Approach for Encrypted Data in Federated Learning. *Digital Communications and Networks*, **8**, 614-624. https://doi.org/10.1016/j.dcan.2022.05.006

[33] Tang, W., Liu, E., Ni, W., Qu, X., Huang, B., Li, K., *et al.* (2025) Game-Theoretic Incentive Mechanism for Blockchain-Based Federated Learning. *IEEE Transactions on Mobile Computing*, **24**, 10363-10376. https://doi.org/10.1109/tmc.2025.3567355

[34] Ren, Y., Hu, M., Yang, Z., Feng, G. and Zhang, X. (2024) BPFL: Blockchain-Based Privacy-Preserving Federated Learning against Poisoning Attack. *Information Sciences*, **665**, Article 120377. https://doi.org/10.1016/j.ins.2024.120377

[35] Narkedimilli, S., Sriram, A.V., Makam, S., Sathvik, M. and Mallellu, S.P. (2025) FAPL-DM-BC: A Secure and Scalable FL Framework with Adaptive Privacy and Dynamic Masking, Blockchain, and XAI for the IoVs. arXiv:2501.01063.

[36] Rehman, A., Abbas, S., Khan, M.A., Ghazal, T.M., Adnan, K.M. and Mosavi, A. (2022) A Secure Healthcare 5.0 System Based on Blockchain Technology Entangled with Federated Learning Technique. *Computers in Biology and Medicine*, **150**, Article 106019. https://doi.org/10.1016/j.compbiomed.2022.106019

[37] Vijay Anand, R., Magesh, G., Alagiri, I., Brahmam, M.G., Balusamy, B., Selvan, C.P., *et al.* (2025) Design of an Improved Model Using Federated Learning and LSTM Autoencoders for Secure and Transparent Blockchain Network Transactions. *Scientific Reports*, **15**, Article No. 1615. https://doi.org/10.1038/s41598-024-83564-4

[38] Liu, P., Xu, X. and Wang, W. (2022) Threats, Attacks and Defenses to Federated Learning: Issues, Taxonomy and Perspectives. *Cybersecurity*, **5**, Article No. 4. https://doi.org/10.1186/s42400-021-00105-6

[39] Sikandar, H.S., Waheed, H., Tahir, S., Malik, S.U.R. and Rafique, W. (2023) A Detailed Survey on Federated Learning Attacks and Defenses. *Electronics*, **12**, Article 260. https://doi.org/10.3390/electronics12020260

[40] Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D. and Shmatikov, V. (2020) How to Backdoor Federated Learning. *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, **108**, 2938-2948.

[41] Wang, Q., Wu, Y., Xuan, H. and Wu, H. (2024) FLARE: A Backdoor Attack to Federated Learning with Refined Evasion. *Mathematics*, **12**, Article 3751. https://doi.org/10.3390/math12233751

[42] Tolpegin, V., Truex, S., Gursoy, M.E. and Liu, L. (2020) Data Poisoning Attacks against Federated Learning Systems. In: Chen, L., Li, N., Liang, K. and Schneider, S., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 480-501. https://doi.org/10.1007/978-3-030-58951-6_24

[43] Pasquini, D., Francati, D. and Ateniese, G. (2022) Eluding Secure Aggregation in Federated Learning via Model Inconsistency. Proceedings of the 2022 *ACM SIGSAC Conference on Computer and Communications Security*, Los Angeles, 7-11 November 2022, 2429-2443. https://doi.org/10.1145/3548606.3560557

[44] Zhang, Y., Behnia, R., Yavuz, A.A., Ebrahimi, R. and Bertino, E. (2024) Uncovering Attacks and Defenses in Secure Aggregation for Federated Deep Learning. 2024 *IEEE International Conference on Data Mining Workshops* (*ICDMW*), Abu Dhabi, 9 December 2024, 650-656. https://doi.org/10.1109/icdmw65004.2024.00090

[45] Shi, J., Wan, W., Hu, S., Lu, J. and Yu Zhang, L. (2022) Challenges and Approaches for Mitigating Byzantine Attacks in Federated Learning. 2022 *IEEE International Conference on Trust, Security and Privacy in Computing and Communications* (*TrustCom*), Wuhan, 9-11 December 2022, 139-146. https://doi.org/10.1109/trustcom56396.2022.00030

[46] Wan, W., Hu, S., Lu, J., Zhang, L.Y., Jin, H. and He, Y. (2022) Shielding Federated Learning: Robust Aggregation with Adaptive Client Selection. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, Guangzhou, 29-31 August 2025, 753-760. https://doi.org/10.24963/ijcai.2022/106

[47] Zhang, K., Tao, G., Xu, Q., Cheng, S., An, S., Liu, Y., Feng, S., Shen, G., Chen, P.-Y., Ma, S. and Zhang, X. (2023) FLIP: A Provable Defense Framework for Backdoor Mitigation in Federated Learning. arXiv:2210.12873.

[48] Benmalek, M., Benrekia, M.A. and Challal, Y. (2022) Security of Federated Learning: Attacks, Defensive Mechanisms, and Challenges. *Revue d'Intelligence Artificielle*, **36**, 49-59. https://doi.org/10.18280/ria.360106

[49] Chen, Y., Gui, Y., Lin, H., Gan, W. and Wu, Y. (2022) Federated Learning Attacks and Defenses: A Survey. 2022 *IEEE International Conference on Big Data* (*Big Data*), Osaka, 17-20 December 2022, 4256-4265. https://doi.org/10.1109/bigdata55660.2022.10020431

[50] Moshawrab, M., Adda, M., Bouzouane, A., Ibrahim, H. and Raad, A. (2024) Securing Federated Learning: Approaches, Mechanisms and Opportunities. *Electronics*, **13**, Article 3675. https://doi.org/10.3390/electronics13183675

[51] Wainakh, A., Zimmer, E., Subedi, S., Keim, J., Grube, T., Karuppayah, S., *et al.* (2022) Federated Learning Attacks Revisited: A Critical Discussion of Gaps, Assumptions, and Evaluation Setups. *Sensors*, **23**, Article 31. https://doi.org/10.3390/s23010031

[52] Xie, X., Hu, C., Ren, H. and Deng, J. (2024) A Survey on Vulnerability of Federated Learning: A Learning Algorithm Perspective. *Neurocomputing*, **573**, Article 127225. https://doi.org/10.1016/j.neucom.2023.127225

[53] Zakaria, F. and Khalid, S.K. (2025) A Review of Federated Learning Attacks: Threat Models and Defence Strategies. *International Journal of Advanced Computer Science and Applications*, **16**, 544-554. https://doi.org/10.14569/ijacsa.2025.0160754

[54] Chaudhury, B.R., Li, L., Kang, M., Li, B. and Mehta, R. (2022) Fairness in Federated Learning via Core-Stability. *Advances in Neural Information Processing Systems* 35, New Orleans, 28 November-9 December 2022, 5738-5750. https://doi.org/10.52202/068431-0415

[55] Pan, Z., Wang, S., Li, C., Wang, H., Tang, X. and Zhao, J. (2023) FedMDFG: Federated Learning with Multi-Gradient Descent and Fair Guidance. *Proceedings of the AAAI Conference on Artificial Intelligence*, **37**, 9364-9371. https://doi.org/10.1609/aaai.v37i8.26122

[56] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2021) A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, **54**, 1-35. https://doi.org/10.1145/3457607

[57] Mukhtiar, N., Mahmood, A. and Sheng, Q.Z. (2025) Fairness in Federated Learning: Trends, Challenges, and Opportunities. *Advanced Intelligent Systems*, **7**, Article 2400836. https://doi.org/10.1002/aisy.202400836

[58] Taik, A., Chehbouni, K. and Farnadi, G. (2025) Fairness in Federated Learning: Fairness for Whom? *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, **8**, 2456-2469. https://doi.org/10.1609/aies.v8i3.36730

[59] Abay, A., Zhou, Y., Baracaldo, N., Rajamoni, S., Chuba, E. and Ludwig, H. (2020) Mitigating Bias in Federated Learning. arXiv:2012.02447.

[60] Lo, S.K., Lu, Q., Paik, H. and Zhu, L. (2021) FLRA: A Reference Architecture for Federated Learning Systems. In: Biffl, S., Navarro, E., Löwe, W., Sirjani, M., Mirandola, R. and Weyns, D., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 83-98. https://doi.org/10.1007/978-3-030-86044-8_6

[61] Lo, S.K., Lu, Q., Zhu, L., Paik, H., Xu, X. and Wang, C. (2021) Architectural Patterns for the Design of Federated Learning Systems. arXiv:2101.02373. http://arxiv.org/abs/2101.02373

[62] Gabrielli, E., Pica, G. and Tolomei, G. (2023) A Survey on Decentralized Federated Learning. arXiv:2308.04604. http://arxiv.org/abs/2308.04604

[63] Hallaji, E., Razavi-Far, R., Saif, M., Wang, B. and Yang, Q. (2024) Decentralized Federated Learning: A Survey on Security and Privacy. *IEEE Transactions on Big Data*, **10**, 194-213. https://doi.org/10.1109/tbdata.2024.3362191

[64] De Lacour, D., Lacoste, M., Südholt, M. and Traoré, J. (2023) Towards Scalable Resilient Federated Learning: A Fully Decentralised Approach. 2023 *IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events* (*PerCom Workshops*), Atlanta, 13-17 March 2023, 621-627. https://doi.org/10.1109/percomworkshops56833.2023.10150295

[65] Ekmefjord, M., Ait-Mlouk, A., Alawadi, S., Akesson, M., Singh, P., Spjuth, O., *et al.* (2022) Scalable Federated Machine Learning with FEDn. 2022 22*nd IEEE International Symposium on Cluster, Cloud and Internet Computing* (*CCGrid*), Taormina, 16-19 May 2022, 555-564. https://doi.org/10.1109/ccgrid54584.2022.00065

[66] Brauneck, A., Schmalhorst, L., Kazemi Majdabadi, M.M., Bakhtiari, M., Völker, U., Baumbach, J., *et al.* (2023) Federated Machine Learning, Privacy-Enhancing Technologies, and Data Protection Laws in Medical Research: Scoping Review. *Journal of Medical Internet Research*, **25**, e41588. https://doi.org/10.2196/41588

[67] Aouedi, O., Sacco, A., Piamrat, K. and Marchetto, G. (2023) Handling Privacy-Sensitive Medical Data with Federated Learning: Challenges and Future Directions. *IEEE Journal of Biomedical and Health Informatics*, **27**, 790-803. https://doi.org/10.1109/jbhi.2022.3185673

[68] Teo, Z.L., Jin, L., Liu, N., Li, S., Miao, D., Zhang, X., *et al.* (2024) Federated Machine Learning in Healthcare: A Systematic Review on Clinical Applications and Technical Architecture. *Cell Reports Medicine*, **5**, Article 101419. https://doi.org/10.1016/j.xcrm.2024.101419

[69] Stripelis, D., Gupta, U., Saleem, H., Dhinagar, N., Ghai, T., Anastasiou, C., *et al.* (2024) A Federated Learning Architecture for Secure and Private Neuroimaging Analysis. *Patterns*, **5**, Article 101031. https://doi.org/10.1016/j.patter.2024.101031

[70] Rachakonda, S., Moorthy, S., Jain, A., Bukharev, A., Bucur, A., Manni, F., *et al.* (2023) Privacy Enhancing and Scalable Federated Learning to Accelerate AI Implementation in Cross-Silo and IoMT Environments. *IEEE Journal of Biomedical and Health Informatics*, **27**, 744-755. https://doi.org/10.1109/jbhi.2022.3185418

[71] Pati, S., Kumar, S., Varma, A., Edwards, B., Lu, C., Qu, L., *et al.* (2024) Privacy Preservation for Federated Learning in Health Care. *Patterns*, **5**, Article 100974. https://doi.org/10.1016/j.patter.2024.100974

[72] Choudhury, O., Gkoulalas-Divanis, A., Salonidis, T., Sylla, I., Park, Y., Hsu, G. and Das, A. (2020) Anonymizing Data for Privacy-Preserving Federated Learning. arXiv:2002.09096. http://arxiv.org/abs/2002.09096

[73] Olatunji, I.E., Rauch, J., Katzensteiner, M. and Khosla, M. (2024) A Review of Anonymization for Healthcare Data. *Big Data*, **12**, 538-555. https://doi.org/10.1089/big.2021.0169

[74] Liu, T., Wang, Z., He, H., Shi, W., Lin, L., An, R., *et al.* (2023) Efficient and Secure Federated Learning for Financial Applications. *Applied Sciences*, **13**, Article 5877. https://doi.org/10.3390/app13105877