IJSRA

International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

Int. J. Sci. Res. Arch.
IJSRA
International Journal of Science and Research Archive
Research Journal Archive, INDIA

(Review Article)

Federated learning for privacy-preserving, secure and scalable data intelligence in hybrid cloud systems

Emmanuel Ezeakile [1], Abdulateef Oluwakayode Disu [2], Cynthia Alabi [3], Toyosi Mustapha [4] and Moses Oluwasegun Odewale [5, *]

[1] *Department of Electrical and Information Engineering, College of Engineering, Covenant University, Ota, Ogun State, Nigeria.*
[2] *Department of Computer Science, School of Computing and Engineering Sciences, Babcock University Ilishan-Remo, Ogun, Nigeria.*
[3] *School of Geography and Natural Sciences, Northumbria University, United Kingdom.*
[4] *College of Business, Southern New Hampshire University, Manchester, New Hampshire, USA.*
[5] *College of Business, Lamar University, Beaumont,Texas, U. S. A.*

International Journal of Science and Research Archive, 2026, 18(01), 499-512

Publication history: Received on 05 December 2025; revised on 12 January 2026; accepted on 14 January 2026

Article DOI: https://doi.org/10.30574/ijsra.2026.18.1.3301

Abstract

The convergence of federated learning and hybrid cloud computing represents a transformative paradigm for privacy-preserving data intelligence. This review examines federated learning implementations in hybrid cloud environments, analyzing security mechanisms, privacy-preserving capabilities, and scalability challenges. We explore architectural frameworks and deployment strategies while analyzing security and privacy challenges from technical, organizational, and regulatory perspectives. The study highlights synergistic benefits of combining federated learning with hybrid cloud infrastructure and discusses emerging trends including homomorphic encryption, differential privacy, and blockchain integration. Through comprehensive literature analysis of publications from 2016 to 2024, key findings reveal that federated learning in hybrid clouds offers unprecedented opportunities for privacy-preserving analytics while introducing unique challenges in communication efficiency and cross-environment orchestration. Organizations can effectively leverage federated learning by implementing layered security architectures and maintaining continuous adaptation to evolving privacy regulations. This analysis provides valuable insights for practitioners and researchers navigating the intersection of federated learning and hybrid cloud computing.

Keywords: Federated Learning; Hybrid Cloud Computing; Privacy-Preserving Machine Learning; Data Intelligence; Distributed Learning; Edge Computing; Security Architectures

1. Introduction

The exponential growth of data generation across distributed environments has created unprecedented challenges for traditional centralized machine learning paradigms. Organizations worldwide generate massive volumes of sensitive data across geographically dispersed locations, including edge devices, private data centers, and public cloud infrastructure. Conventional approaches requiring data centralization for model training face significant obstacles related to privacy regulations, data sovereignty requirements, bandwidth limitations, and security concerns. Federated learning has emerged as a revolutionary distributed machine learning paradigm that enables collaborative model training without centralizing raw data, fundamentally reshaping how organizations approach data intelligence in distributed environments[1].

[*] Corresponding author: Cynthia Alabi

The global federated learning market is expanding rapidly, with strong industry forecasts indicating sustained, double-digit annual growth through 2028, driven by increasing demand for privacy-preserving machine learning and decentralized data analytics. Simultaneously, hybrid cloud adoption continues accelerating, with recent surveys indicating that 87% of enterprises have adopted hybrid cloud strategies to balance flexibility, control, and cost-efficiency[2]. The intersection of these two technological paradigms presents both compelling opportunities and complex challenges that warrant comprehensive investigation.

Federated learning, first conceptualized by Google in 2016 for mobile keyboard prediction, has evolved from a niche research area into a practical framework for privacy-preserving machine learning across diverse applications including healthcare analytics, financial fraud detection, autonomous vehicles, and smart city infrastructure. The fundamental principle of federated learning bringing computation to data rather than data to computation aligns naturally with hybrid cloud architectures that distribute workloads across public and private infrastructure based on security, compliance, and performance requirements. This alignment creates opportunities for organizations to leverage the scalability and cost-efficiency of public clouds while maintaining sensitive data within controlled private environments[3].

Despite these advantages, implementing federated learning in hybrid cloud environments introduces significant technical and organizational challenges that must be carefully addressed. Communication overhead between distributed nodes can severely impact training efficiency, particularly when spanning heterogeneous network environments with varying bandwidth and latency characteristics. Statistical heterogeneity across data silos creates convergence challenges that complicate model training, while ensuring consistent security policies across hybrid infrastructure requires sophisticated orchestration mechanisms. Privacy-preserving techniques such as differential privacy and secure aggregation introduce computational overhead that must be carefully balanced against model accuracy requirements[4]. Furthermore, regulatory compliance becomes increasingly complex when federated learning systems span multiple jurisdictions with divergent data protection frameworks, requiring organizations to navigate a complex web of legal requirements.

Recent developments in privacy-enhancing technologies offer promising solutions to these challenges. Homomorphic encryption enables computation on encrypted data without decryption, while secure multi-party computation allows collaborative computation without revealing individual inputs. Trusted execution environments provide hardware-isolated secure computation zones that protect sensitive operations from external observation. The integration of blockchain technologies provides transparent and auditable mechanisms for federated learning governance, enabling trustless collaboration across organizational boundaries. Edge computing capabilities enable localized data processing that reduces communication costs and latency while maintaining data proximity. These technological advances, combined with sophisticated orchestration frameworks adapted for distributed learning environments, are making federated learning in hybrid clouds increasingly practical for real-world deployments[5].

This paper provides a comprehensive analysis of federated learning implementations in hybrid cloud systems, addressing several critical dimensions of this emerging paradigm. We examine architectural patterns and deployment models that enable effective federated learning across hybrid infrastructure, considering both centralized and decentralized approaches. The security challenges are analyzed from multiple perspectives, including privacy leakage, model poisoning, communication security, and regulatory compliance. We evaluate privacy-preserving mechanisms including differential privacy, secure multi-party computation, homomorphic encryption, and trusted execution environments, assessing their applicability and trade-offs in hybrid cloud contexts. Scalability considerations are explored, focusing on communication efficiency, heterogeneity management, and cross-cloud orchestration. We also investigate emerging technologies and future directions that will shape the evolution of federated learning, including edge-cloud federations, blockchain integration, and quantum-resistant cryptography. Finally, we provide strategic recommendations for organizations considering federated learning adoption, covering assessment, architecture selection, security implementation, and operational excellence.

The integration of federated learning with hybrid cloud computing represents more than a technical advancement, it embodies a fundamental shift toward privacy-centric, distributed data intelligence that aligns with evolving societal expectations and regulatory requirements[6]. As regulatory frameworks worldwide increasingly emphasize data protection and user privacy, organizations must adopt approaches that enable value extraction from distributed data while maintaining compliance and security. Understanding how to effectively implement federated learning in hybrid cloud environments is essential for organizations seeking to leverage collaborative intelligence while respecting privacy boundaries and regulatory constraints. This review provides the foundation for such understanding, synthesizing current knowledge and identifying future directions in this rapidly evolving field.

2. Overview of Federated Learning and Hybrid Cloud Computing

2.1. Federated Learning: Foundations and Principles

Federated learning represents a paradigm shift in machine learning, enabling multiple parties to collaboratively train models without exchanging raw data [7]. The fundamental architecture involves distributed data sources (clients) that perform local model training on their private datasets, communicating only model updates typically gradients or weight changes to a central aggregation server. This server combines the distributed updates into a global model, which is then redistributed to clients for subsequent training rounds. This iterative process continues until the model converges to satisfactory performance.

The federated learning framework can be categorized into three primary architectures. Horizontal federated learning applies when participants share the same feature space but different sample spaces, common in scenarios where multiple hospitals collaborate on disease prediction using similar patient records. Vertical federated learning addresses situations where participants have different feature spaces for overlapping sample sets, typical in cross-industry collaborations such as banks and retailers analyzing shared customer segments. Federated transfer learning extends these concepts to scenarios with both different feature spaces and different sample distributions, enabling knowledge transfer across heterogeneous domains [8].

The mathematical foundation of federated learning centers on distributed optimization [9]. The objective is to minimize a global loss function that aggregates local losses across all participating clients. The Federated Averaging algorithm, the most widely adopted approach, computes weighted averages of local model parameters based on dataset sizes. More sophisticated aggregation mechanisms account for statistical heterogeneity, communication constraints, and adversarial scenarios where some participants may contribute corrupted updates.

2.2. Hybrid Cloud Architecture for Federated Learning

Hybrid cloud computing integrates public cloud services, private cloud infrastructure, and potentially on-premises systems into a unified, orchestrated environment[10]. This architecture enables organizations to maintain sensitive workloads in private environments while leveraging public cloud resources for less critical operations or handling variable demand. Modern hybrid cloud architectures for federated learning typically incorporate multiple layers: the infrastructure layer encompasses physical and virtualized computing resources across public and private environments; the platform layer provides unified management interfaces, identity and access management, and workload orchestration capabilities; and the application layer hosts diverse federated learning workloads distributed across environments based on security requirements, regulatory constraints, performance needs, or cost optimization objectives.

Several deployment patterns have emerged for hybrid cloud implementations of federated learning. The edge-private-public pattern maintains local training at edge devices, performs regional aggregation in private cloud infrastructure, and coordinates global model updates through public cloud services. The data residency pattern keeps sensitive data in private environments while utilizing public cloud for computation, often employing confidential computing or homomorphic encryption. The multi-cloud pattern extends hybrid architecture across multiple public cloud providers to avoid vendor lock-in and optimize cost-performance trade-offs [11].

2.3. Convergence: Federated Learning in Hybrid Cloud Environments

The integration of federated learning with hybrid cloud infrastructure creates synergistic opportunities for privacy-preserving, scalable data intelligence. Hybrid clouds provide the computational infrastructure and orchestration capabilities necessary for managing distributed federated learning workflows across heterogeneous environments[12]. Federated learning, in turn, enables organizations to extract value from data distributed across hybrid infrastructure without centralizing sensitive information.

This convergence addresses several critical requirements in modern data intelligence systems. Data sovereignty and compliance requirements often mandate that certain data categories remain within specific jurisdictions or infrastructure types. Federated learning enables collaborative analytics across these boundaries without data movement [13]. Bandwidth and latency constraints in edge-to-cloud scenarios can be mitigated through local model training with efficient update communication. Security and privacy concerns are addressed through privacy-preserving aggregation mechanisms that prevent raw data exposure while enabling model improvement.

The hybrid cloud environment provides essential capabilities for federated learning implementations [14]. Scalable compute resources in public clouds can handle intensive aggregation and orchestration tasks. Private cloud infrastructure maintains control over sensitive operations and data. Edge computing capabilities enable local training on resource-constrained devices. Container orchestration platforms such as Kubernetes facilitate deployment and management of federated learning components across heterogeneous infrastructure. Service mesh architectures provide secure, observable communication channels between distributed learning nodes.

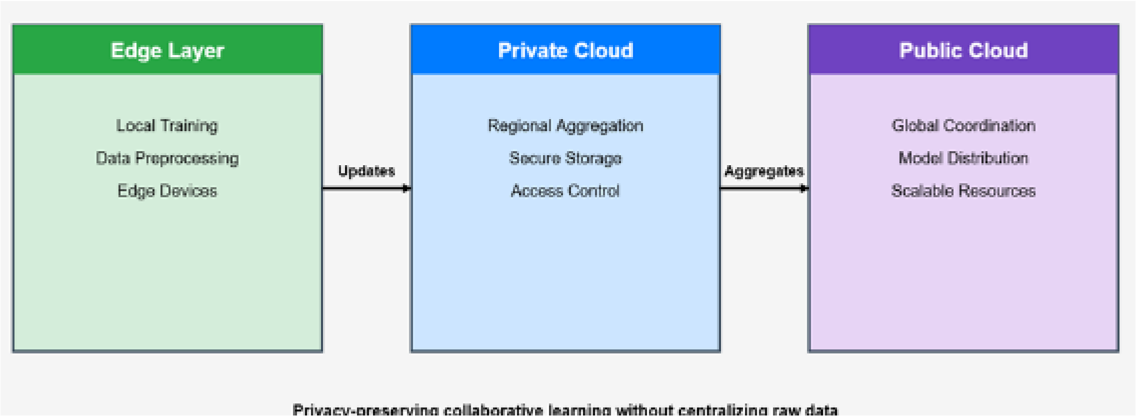


Figure 1 Federated learning Architecture in hybrid cloud

3. Federated Learning Architectures in Hybrid Cloud Systems

3.1. Centralized Federated Learning Architecture

The centralized federated learning architecture employs a single aggregation server that coordinates the training process across distributed clients. In hybrid cloud deployments, this server typically resides in a secure private cloud or trusted public cloud region with appropriate compliance certifications. Clients, which may include edge devices, on-premises servers, or public cloud instances, perform local training and communicate updates to the central server through secure channels[15].

This architecture offers simplicity in coordination and aggregation but introduces potential bottlenecks and single points of failure. The central server must handle communication with potentially thousands of clients, performing secure aggregation of their updates[16]. In hybrid cloud contexts, network topology and latency variations between public and private infrastructure can significantly impact training efficiency. Organizations often deploy content delivery network mechanisms or regional aggregation proxies to mitigate these communication challenges.

3.2. Hierarchical Federated Learning Architecture

Hierarchical federated learning introduces multiple levels of aggregation to improve scalability and efficiency in geographically distributed or multi-organizational scenarios[17]. Edge aggregators collect updates from local client clusters, performing initial aggregation before communicating with regional or global aggregators. This multi-tier approach aligns naturally with hybrid cloud architectures where edge devices connect to edge computing infrastructure, which aggregates to private cloud controllers, eventually federating to central coordination in public cloud environments.

The hierarchical model offers several advantages for hybrid deployments[18]. Communication costs are reduced through local aggregation before cross-environment transmission. Fault tolerance improves as failures in one regional aggregator do not necessarily impact the entire federation. Privacy can be enhanced through progressive aggregation that obscures individual contributions at each level. However, this architecture introduces complexity in coordination, requiring sophisticated protocols to manage hierarchical aggregation while ensuring model convergence and preventing gradient staleness.

3.3. Decentralized Federated Learning Architecture

Decentralized or peer-to-peer federated learning eliminates central coordination, with clients directly communicating and exchanging model updates with neighboring nodes [19]. This architecture offers enhanced privacy through elimination of trusted central aggregators and improved resilience against single points of failure. In hybrid cloud contexts, decentralized federations may span public cloud regions, private data centers, and edge infrastructure without requiring centralized coordination.

Blockchain technologies increasingly underpin decentralized federated learning, providing transparent, immutable records of model updates and enabling trustless collaboration across organizational boundaries. Smart contracts can automate aggregation logic, incentive mechanisms, and access control policies. However, decentralized architectures face significant challenges including complex convergence analysis, increased communication overhead from peer-to-peer exchanges, and difficulties in enforcing consistent security policies across autonomous nodes.

3.4. Cross-Silo and Cross-Device Federations

Federated learning scenarios are often categorized as cross-silo or cross-device based on participant characteristics. Cross-silo federations involve relatively few participants typically organizations or data centers with substantial computational resources and reliable connectivity[20]. Examples include hospitals collaborating on disease prediction or financial institutions jointly training fraud detection models. These federations typically operate within hybrid cloud infrastructure where each silo maintains private cloud or on-premises infrastructure while coordination occurs through secure public cloud services.

Cross-device federations involve massive numbers of resource-constrained participants such as mobile phones, IoT sensors, or edge devices. These scenarios present unique challenges including intermittent connectivity, heterogeneous hardware capabilities, and extreme communication constraints. Hybrid cloud architectures for cross-device federations typically employ edge computing infrastructure for local aggregation and preprocessing, with hierarchical aggregation through private and public cloud tiers for global model coordination[21].

4. Security Challenges in Federated Learning on Hybrid Clouds

4.1. Privacy Leakage and Inference Attacks

Despite federated learning's design principle of maintaining data privacy through local training, sophisticated inference attacks can potentially extract sensitive information from shared model updates[22]. Gradient inversion attacks reconstruct training samples from shared gradients by optimizing input data to produce similar gradient patterns. Membership inference attacks determine whether specific samples were included in training datasets by analyzing model behavior. Property inference attacks deduce aggregate properties of training data such as demographic distributions or feature correlations.

In hybrid cloud environments, these threats are amplified by heterogeneous trust boundaries. Model updates traversing public networks between private and public cloud infrastructure may be intercepted by adversaries. Malicious aggregation servers in public cloud environments could perform inference attacks on received updates. Co-tenancy in public cloud infrastructure introduces risks of side-channel attacks that could leak information about federated learning processes running on shared hardware[23].

Mitigation strategies include differential privacy mechanisms that add calibrated noise to model updates, limiting information leakage while maintaining model utility. Secure aggregation protocols based on secure multi-party computation enable servers to compute aggregate updates without accessing individual contributions. Homomorphic encryption allows computation on encrypted model updates, preventing even privileged aggregation servers from observing raw gradients. Trusted execution environments provide hardware-isolated computation zones for sensitive aggregation operations in public cloud infrastructure[24].

4.2. Model Poisoning and Byzantine Attacks

Federated learning systems are vulnerable to adversarial participants who contribute malicious model updates designed to corrupt the global model. Data poisoning attacks manipulate local training data to inject backdoors or reduce model accuracy[25]. Model poisoning attacks directly craft malicious updates without necessarily poisoning local data, often proving more effective against aggregation defenses. Byzantine attacks involve arbitrary adversarial behavior including submitting random updates or strategically designed gradients to maximize harm.

Hybrid cloud deployments may face elevated risks from compromised nodes in less-secure public cloud regions or from insider threats in private infrastructure[26]. The distributed nature of hybrid environments complicates detection and response to poisoning attacks, particularly when malicious nodes strategically time their attacks or collude across trust boundaries.

Robust aggregation algorithms provide resilience against poisoning by identifying and filtering suspicious updates[27]. Techniques include median-based aggregation, trimmed mean approaches, and machine learning-based anomaly detection that identifies updates statistically inconsistent with benign patterns. Reputation systems track participant behavior over time, reducing influence of historically malicious nodes. Zero-knowledge proofs enable participants to demonstrate computation correctness without revealing sensitive information. In hybrid clouds, blockchain-based audit trails provide transparent records of participant contributions, facilitating post-hoc analysis of potential attacks.

4.3. Communication Security and Network Attacks

The distributed communication patterns inherent to federated learning create extensive attack surfaces for network-based threats [28]. Man-in-the-middle attacks can intercept and potentially modify model updates during transmission between clients and aggregators. Distributed denial of service attacks targeting aggregation servers can disrupt training processes, while sybil attacks involve adversaries creating multiple fake identities to amplify their influence on model aggregation.

Hybrid cloud environments present particular challenges for communication security due to heterogeneous network environments spanning private networks, public internet, and cloud provider backbones. Latency and bandwidth variations complicate implementation of time-sensitive security protocols. Network segmentation across public and private infrastructure requires sophisticated key management and authentication mechanisms [29].

Mitigation approaches include authenticated encryption for all model update transmissions using protocols such as TLS 1.3 or IPsec [30]. Certificate-based mutual authentication ensures both clients and servers verify counterpart identities before communication. Intrusion detection systems specialized for federated learning traffic patterns identify anomalous communication behaviors. In hybrid clouds, virtual private networks or dedicated interconnects between public and private infrastructure provide isolated communication channels for sensitive federated learning traffic.

4.4. Resource Exhaustion and Economic Attacks

The computational and communication demands of federated learning create opportunities for resource exhaustion attacks. Free-riding attacks involve participants benefiting from the global model without contributing meaningful updates, imposing computational costs on honest participants[31]. Computation poisoning attacks submit updates requiring excessive aggregation resources, degrading overall system performance. In public cloud contexts, attackers may exploit elastic resource allocation to inflate victims' operational costs through economic denial of service attacks.

Hybrid cloud environments face unique resource management challenges balancing computational costs across public and private infrastructure while maintaining security and performance guarantees[32]. Organizations must implement sophisticated monitoring and resource allocation policies that detect and mitigate resource exhaustion attempts while accommodating legitimate heterogeneity in participant capabilities and contributions.

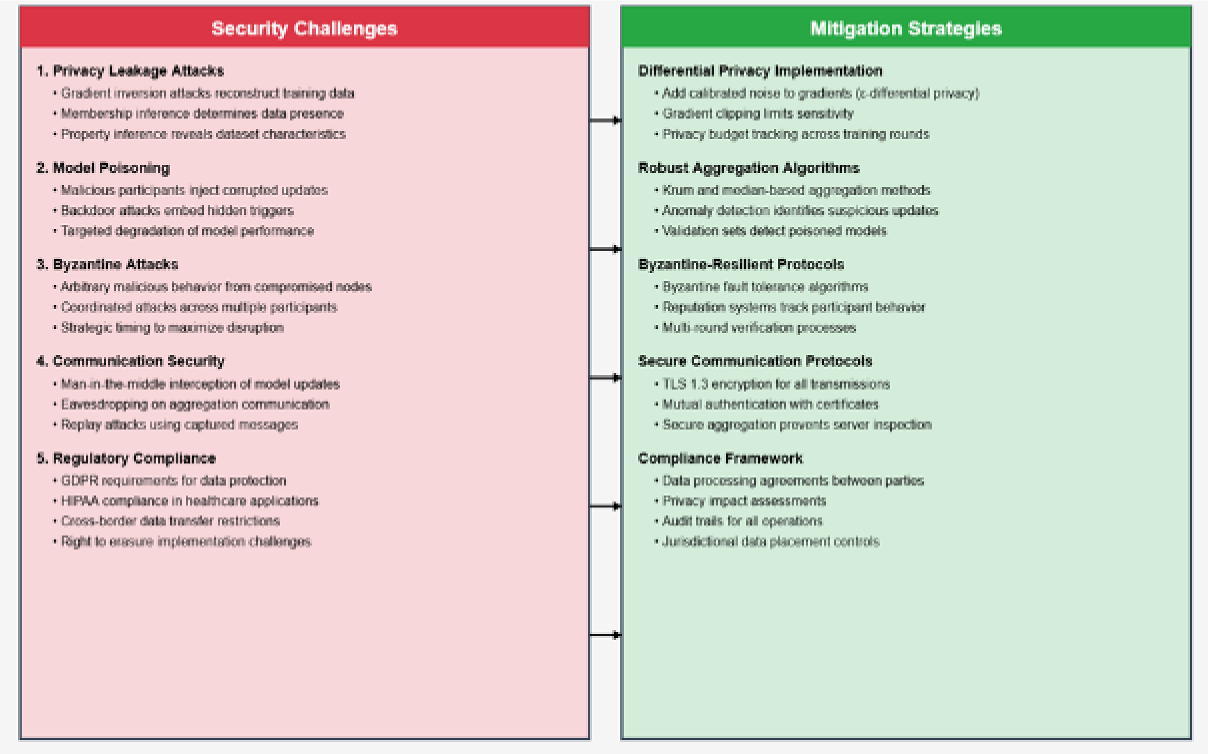
Defense mechanisms include contribution verification protocols that validate computational work without fully re-executing training. Adaptive resource allocation adjusts aggregation frequency and batch sizes based on participant capabilities and historical behavior[33]. In public cloud environments, cost monitoring and automated throttling mechanisms prevent economic attacks from generating excessive charges. Incentive mechanisms, potentially implemented through blockchain-based smart contracts, reward meaningful contributions while penalizing free-riding or malicious behavior.

4.5. Regulatory and Compliance Challenges

Federated learning in hybrid clouds must navigate complex regulatory landscapes governing data protection, privacy, and security. The European Union's General Data Protection Regulation imposes strict requirements on data processing, including federated learning systems that process personal data[34]. The California Consumer Privacy Act and similar regulations worldwide establish user rights regarding their data. Industry-specific regulations such as the Health Insurance Portability and Accountability Act for healthcare and the Payment Card Industry Data Security Standard for financial services impose additional compliance requirements.

The distributed nature of federated learning complicates compliance verification. Determining data controller and processor responsibilities becomes ambiguous when multiple parties collaboratively train models. The right to erasure under GDPR presents technical challenges for federated models where individual data contributions are aggregated into model parameters. Cross-border data flows in global federations may violate data localization requirements in various jurisdictions[35].

Hybrid cloud deployments must implement compliance-aware orchestration that ensures data and computation remain within appropriate jurisdictional and infrastructure boundaries[36]. Privacy impact assessments should evaluate federated learning systems' data protection measures. Audit mechanisms provide transparent records of data usage and model training processes. Explainability techniques help satisfy regulatory requirements for automated decision-making transparency. Legal frameworks and data processing agreements must clearly define responsibilities across federation participants and cloud providers. *Figure 2 maps the five major security challenges in federated learning to their corresponding mitigation strategies, demonstrating how privacy leakage, model poisoning, Byzantine attacks, communication vulnerabilities, and regulatory compliance issues can be addressed through layered defense mechanisms.*



Security Challenges

1. Privacy Leakage Attacks

- Gradient inversion attacks reconstruct training data
- Membership inference determines data presence
- Property inference reveals dataset characteristics

2. Model Poisoning

- Malicious participants inject corrupted updates
- Backdoor attacks embed hidden triggers
- Targeted degradation of model performance

3. Byzantine Attacks

- Arbitrary malicious behavior from compromised nodes
- Coordinated attacks across multiple participants
- Strategic timing to maximize disruption

4. Communication Security

- Man-in-the-middle interception of model updates
- Eavesdropping on aggregation communication
- Replay attacks using captured messages

5. Regulatory Compliance

- GDPR requirements for data protection
- HIPAA compliance in healthcare applications
- Cross-border data transfer restrictions
- Right to erasure implementation challenges

Mitigation Strategies

Differential Privacy Implementation

- Add calibrated noise to gradients (ε-differential privacy)
- Gradient clipping limits sensitivity
- Privacy budget tracking across training rounds

Robust Aggregation Algorithms

- Krum and median-based aggregation methods
- Anomaly detection identifies suspicious updates
- Validation sets detect poisoned models

Byzantine-Resilient Protocols

- Byzantine fault tolerance algorithms
- Reputation systems track participant behavior
- Multi-round verification processes

Secure Communication Protocols

- TLS 1.3 encryption for all transmissions
- Mutual authentication with certificates
- Secure aggregation prevents server inspection

Compliance Framework

- Data processing agreements between parties
- Privacy impact assessments
- Audit trails for all operations
- Jurisdictional data placement controls

Figure 2 Security challenges and mitigation stratergies in federated learning

5. Privacy-Preserving Mechanisms for Federated Learning

5.1. Differential Privacy

Differential privacy provides rigorous mathematical guarantees that individual data points have limited influence on model outputs, preventing inference attacks while maintaining overall utility[37]. In federated learning, differential privacy can be applied at multiple levels. Local differential privacy adds noise to individual data samples before local training, providing strong privacy guarantees but potentially significant utility loss. Distributed differential privacy adds noise to model updates before transmission to aggregators, balancing privacy and utility more effectively. Central differential privacy applies noise at aggregation servers, offering the best utility but requiring trust in aggregation infrastructure.

Implementation in hybrid clouds requires careful calibration of privacy budgets across heterogeneous participants with varying privacy requirements. Organizations maintaining sensitive data in private clouds may apply stricter privacy parameters than those using public cloud resources for less sensitive workloads. Privacy budget allocation mechanisms must account for the iterative nature of federated learning, where privacy guarantees degrade with repeated model updates[38].

Advanced techniques such as adaptive clipping and noise injection dynamically adjust privacy parameters based on update characteristics, improving utility while maintaining privacy guarantees. Privacy accounting frameworks track cumulative privacy loss across training rounds, enabling organizations to make informed decisions about training duration and model deployment[39]. In hybrid environments, differential privacy implementations must consider the trust models across public and private infrastructure, potentially applying different privacy mechanisms at each trust boundary.

5.2. Secure Multi-Party Computation

Secure multi-party computation enables multiple parties to jointly compute functions over their inputs while keeping those inputs private[40]. In federated learning, secure multi-party computation facilitates secure aggregation where aggregation servers compute the sum of model updates without observing individual contributions. This protection prevents even potentially malicious or compromised aggregation servers from performing inference attacks on individual updates.

The fundamental approach involves secret sharing, where each participant splits their model update into cryptographic shares distributed across multiple aggregation servers. These servers perform computation on shares, with the final result revealed only when sufficient shares are combined. No individual server can reconstruct original model updates from shares alone. Homomorphic secret sharing enables efficient linear operations such as averaging, which is central to most federated learning algorithms[41].

Implementation challenges in hybrid clouds include communication overhead from distributing shares across multiple aggregators and computational costs of cryptographic operations[42]. Organizations must carefully select which aggregation components reside in public versus private infrastructure, balancing security requirements against operational costs. Threshold cryptography provides resilience against compromised aggregators, requiring consensus among multiple servers before revealing aggregated results. In practice, hybrid deployments might maintain primary aggregation in trusted private infrastructure while utilizing public cloud resources for auxiliary functions such as encrypted storage or communication relay.

5.3. Homomorphic Encryption

Homomorphic encryption allows computation on encrypted data without decryption, enabling aggregation servers to process model updates while maintaining complete confidentiality [43]. Partially homomorphic encryption supports specific operations such as addition, sufficient for federated averaging algorithms. Fully homomorphic encryption enables arbitrary computation on encrypted data but incurs substantial computational overhead that currently limits practical deployment.

In federated learning, clients encrypt model updates using public keys before transmission to aggregation servers. These servers perform encrypted aggregation, computing the sum or weighted average of encrypted updates without accessing plaintext values. The aggregated result is returned to clients for decryption using their private keys, revealing only the final aggregated model while protecting individual contributions[44].

Hybrid cloud deployments can leverage homomorphic encryption to enable aggregation in less-trusted public cloud environments while maintaining privacy guarantees. The computational intensity of homomorphic operations typically requires significant cloud resources, making public cloud infrastructure attractive for cost-effective scaling. However, organizations must consider the trade-offs between enhanced privacy and substantially increased computation and communication costs. Optimization techniques such as batching operations, utilizing hardware acceleration, and hybrid approaches combining homomorphic encryption with other privacy-preserving techniques help mitigate these overheads[45].

5.4. Trusted Execution Environments

Trusted execution environments provide hardware-isolated computation zones that protect code and data from potentially malicious operating systems, hypervisors, or co-located processes[46]. Technologies such as Intel SGX, AMD SEV, and ARM TrustZone enable creation of secure enclaves where sensitive computations execute with strong confidentiality and integrity guarantees.

In federated learning, trusted execution environments can secure aggregation operations, protecting model updates from inspection by cloud providers or other privileged parties. Clients encrypt updates using keys accessible only within secure enclaves. Aggregation code executes within enclaves, processing decrypted updates with assurance that computation cannot be observed or tampered with externally [47]. This approach enables secure aggregation in public cloud infrastructure that would otherwise be considered insufficiently trusted for sensitive model updates.

Hybrid cloud deployments must navigate the heterogeneous availability of trusted execution environment capabilities across infrastructure types and cloud providers. Public cloud providers increasingly offer confidential computing services based on trusted execution environments, but capabilities, performance characteristics, and attestation mechanisms vary significantly. Organizations must evaluate trusted execution environment implementations' security properties, considering both theoretical guarantees and practical vulnerabilities discovered in specific implementations. Side-channel attacks against certain trusted execution environment implementations necessitate careful security analysis and potentially additional defensive measures[48].

6. Scalability and Performance Optimization

6.1. Communication Efficiency

Communication costs dominate federated learning performance, particularly in hybrid cloud scenarios spanning heterogeneous networks with varying bandwidth and latency characteristics[49]. The iterative nature of federated learning requires repeated rounds of model distribution and update collection, generating substantial network traffic. Cross-environment communication between edge devices, private data centers, and public clouds faces additional constraints from network segmentation, firewalls, and potentially limited interconnect capacity.

Gradient compression techniques reduce communication volume by intelligently encoding model updates. Sparsification transmits only the most significant gradient components, filtering out small updates below adaptive thresholds. Quantization reduces numerical precision of transmitted values, trading slight accuracy loss for dramatically reduced data volume. Structured updates leverage low-rank decomposition or other mathematical techniques to represent high-dimensional gradients compactly [50].

Communication-efficient protocols optimize when and how updates are transmitted. Federated averaging reduces communication frequency by performing multiple local training epochs between server communication rounds. Adaptive aggregation dynamically adjusts communication frequency based on model convergence progress and network conditions. Asynchronous protocols eliminate synchronization barriers, allowing fast participants to continue training while awaiting slower nodes[51]. In hybrid clouds, intelligent routing and regional aggregation minimize expensive cross-environment communication, performing initial consolidation at network edges before transmitting to central coordination infrastructure.

6.2. Heterogeneity Management

Federated learning across hybrid cloud environments encounters multiple dimensions of heterogeneity. Statistical heterogeneity arises when participants' data distributions differ significantly, complicating convergence and potentially leading to biased global models. System heterogeneity encompasses variations in computational capabilities, memory, and network connectivity across edge devices, on-premises servers, and cloud instances. Network heterogeneity manifests in diverse latency, bandwidth, and reliability characteristics across hybrid infrastructure[52].

Addressing statistical heterogeneity requires algorithmic innovations that account for non-IID data distributions. Personalized federated learning approaches maintain both global and local model components, enabling adaptation to participant-specific data characteristics while preserving shared knowledge [53]. Meta-learning frameworks train models capable of rapid adaptation to new data distributions with minimal local training. Clustered federated learning groups participants with similar data distributions, training specialized models for each cluster while maintaining some shared components.

System heterogeneity necessitates adaptive resource allocation and training strategies. Client selection mechanisms prioritize participants based on computational capabilities, data quantity, and connectivity characteristics [54]. Asynchronous aggregation tolerates varying completion times without blocking the training process. Model compression and knowledge distillation enable resource-constrained edge devices to participate meaningfully by training compact local models. In hybrid clouds, intelligent workload placement leverages heterogeneous infrastructure optimally, assigning compute-intensive operations to powerful cloud instances while edge devices focus on local data processing.

6.3. Cross-Cloud Orchestration

Managing federated learning workflows across hybrid infrastructure requires sophisticated orchestration that coordinates diverse components while respecting security boundaries and optimizing resource utilization [55]. Modern orchestration frameworks must handle deployment across edge devices, private data centers, and multiple public cloud providers while maintaining consistent security policies and operational monitoring.

Container orchestration platforms provide foundational capabilities for deploying and managing federated learning components. Kubernetes has emerged as the de facto standard, offering abstractions for workload scheduling, service discovery, and resource management across heterogeneous infrastructure. Federation-aware extensions enable cross-cluster orchestration, deploying components across public and private Kubernetes clusters while maintaining unified management interfaces[56].

Service mesh architectures such as Istio or Linkerd provide secure, observable communication between federated learning components. These meshes implement mutual TLS authentication, traffic encryption, and fine-grained access control policies consistently across hybrid environments. Observability features including distributed tracing and metrics collection facilitate debugging and performance optimization of complex federated workflows. Workflow orchestration tools coordinate multi-step federated learning processes including data preparation, model initialization, iterative training rounds, aggregation, and model deployment across heterogeneous infrastructure [57].

7. Future Trends and Emerging Technologies

7.1. Edge-Cloud Federated Learning

The proliferation of edge computing infrastructure creates new opportunities for federated learning. Edge servers positioned near data sources provide computational capabilities for local model training and aggregation, reducing latency and bandwidth consumption compared to direct cloud communication [58]. This edge-cloud continuum enables hierarchical federated learning architectures particularly suited for IoT and mobile scenarios generating massive distributed data volumes.

Edge-native federated learning frameworks optimize for resource-constrained environments, implementing efficient local training algorithms, compressed model representations, and opportunistic communication strategies that leverage available connectivity without strict synchronization requirements. Model splitting techniques partition neural networks between edge and cloud tiers, with computationally intensive layers executing in the cloud while privacy-sensitive early layers remain on edge devices [59]. The integration of 5G and future 6G networks will dramatically enhance edge-cloud federated learning capabilities through ultra-low latency, massive connectivity, and network slicing that provides isolated, quality-assured communication channels for federated learning traffic.

7.2. Blockchain-Enabled Federated Learning

Blockchain technology offers solutions to trust, transparency, and incentive challenges in decentralized federated learning. Immutable distributed ledgers provide auditable records of all model updates, participant contributions, and aggregation operations, enabling detection of malicious behavior and fair credit attribution [60]. Smart contracts automate aggregation logic, incentive distribution, and access control policies without centralized administration.

Tokenized incentive mechanisms reward participants for meaningful contributions, addressing free-rider problems and encouraging high-quality data and computational resource provision [61]. Reputation systems built on blockchain transaction histories enable trust establishment in permissionless federations where participants lack pre-existing relationships. However, blockchain integration introduces challenges including consensus mechanisms and transaction validation that create latency and throughput limitations. Hybrid approaches combining off-chain computation and communication with on-chain coordination, audit, and incentive mechanisms offer more practical implementations. Emerging blockchain platforms designed specifically for federated learning address these limitations through specialized consensus mechanisms optimized for machine learning workloads and efficient state channels for high-frequency off-chain updates.

7.3. Quantum-Resistant Cryptography for Federated Learning

The anticipated advent of cryptographically relevant quantum computers poses threats to current security mechanisms protecting federated learning systems[62]. Most contemporary encryption, digital signatures, and secure aggregation protocols rely on mathematical problems that quantum computers could solve efficiently, breaking their security guarantees. Transitioning to quantum-resistant cryptography is essential for long-term security of federated learning systems, particularly those handling sensitive data requiring decades of confidentiality.

The National Institute of Standards and Technology has initiated standardization of post-quantum cryptographic algorithms. Integrating these algorithms into federated learning frameworks requires careful engineering to minimize performance impacts while ensuring security against both classical and quantum adversaries. Hybrid approaches combining classical and post-quantum algorithms provide near-term security against quantum threats while maintaining compatibility with existing systems[63]. Research directions include developing quantum-resistant secure aggregation protocols and quantum key distribution for ultra-secure model update transmission.

8. Conclusion

This comprehensive review has examined the multifaceted landscape of federated learning implementations in hybrid clouds, analyzing architectural patterns, security challenges, privacy-preserving mechanisms, and scalability considerations. Key findings reveal that while federated learning introduces unique challenges in communication efficiency, heterogeneity management, and cross-environment orchestration, sophisticated solutions combining cryptographic techniques, distributed systems engineering, and machine learning innovations enable practical deployments across diverse applications. Organizations that carefully implement these mechanisms while maintaining continuous adaptation to evolving requirements can successfully leverage federated learning to achieve privacy-preserving, secure, and scalable data intelligence in hybrid cloud environments.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Lazaros K, Koumadorakis DE, Vrahatis AG, Kotsiantis S. Federated learning: Navigating the landscape of collaborative intelligence. Electronics. 2024 Nov 30;13(23):4744.

[2] Anh NH. Hybrid Cloud Migration Strategies: Balancing Flexibility, Security, and Cost in a Multi-Cloud Environment. Transactions on Machine Learning, Artificial Intelligence, and Advanced Intelligent Systems. 2024 Oct 7;14(10):14-26.

[3] Bello S. HYBRID CLOUD STRATEGIES: BALANCING CONTROL AND FLEXIBILITY IN SENSITIVE INDUSTRIES.

[4] Liu Z, Guo J, Yang W, Fan J, Lam KY, Zhao J. Privacy-preserving aggregation in federated learning: A survey. IEEE Transactions on Big Data. 2022 Jul 15.

[5] Albshaier L, Almarri S, Albuali A. Federated learning for cloud and edge security: A systematic review of challenges and AI opportunities. Electronics. 2025 Mar 3;14(5):1019.

[6] Chawki M. An effective cloud computing model enhancing privacy in cloud computing. Information Security Journal: A Global Perspective. 2024 Nov 1;33(6):635-58.

[7] Ji S, Tan Y, Saravirta T, Yang Z, Liu Y, Vasankari L, Pan S, Long G, Walid A. Emerging trends in federated learning: From model fusion to federated x learning. International Journal of Machine Learning and Cybernetics. 2024 Sep;15(9):3769-90.

[8] Gao D, Liu Y, Huang A, Ju C, Yu H, Yang Q. Privacy-preserving heterogeneous federated transfer learning. In2019 IEEE international conference on big data (Big Data) 2019 Dec 9 (pp. 2552-2559). IEEE.

[9] Konečný J, McMahan B, Ramage D. Federated optimization: Distributed optimization beyond the datacenter. arXiv preprint arXiv:1511.03575. 2015 Nov 11.

[10] Oladosu SA, Ike CC, Adepoju PA, Afolabi AI, Ige AB, Amoo OO. Advancing cloud networking security models: Conceptualizing a unified framework for hybrid cloud and on-premises integrations. Magna Scientia Advanced Research and Reviews. 2021;2(1).

[11] Hope OS. Multi-cloud strategy for enterprise applications: Cost, performance, and resilience considerations.

[12] Merseedi KJ, Zeebaree SR. The cloud architectures for distributed multi-cloud computing: a review of hybrid and federated cloud environment. The Indonesian Journal of Computer Science. 2024 Apr 1;13(2).

[13] Lazaros K, Koumadorakis DE, Vrahatis AG, Kotsiantis S. Federated learning: Navigating the landscape of collaborative intelligence. Electronics. 2024 Nov 30;13(23):4744.

[14] Merseedi KJ, Zeebaree SR. The cloud architectures for distributed multi-cloud computing: a review of hybrid and federated cloud environment. The Indonesian Journal of Computer Science. 2024 Apr 1;13(2).

[15] Fernandes DA, Soares LF, Gomes JV, Freire MM, Inácio PR. Security issues in cloud environments: a survey. International journal of information security. 2014 Apr;13(2):113-70.

[16] Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, Ramage D, Segal A, Seth K. Practical secure aggregation for privacy-preserving machine learning. Inproceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security 2017 Oct 30 (pp. 1175-1191).

[17] Pais V, Rao S, Muniyal B. Strategies for reducing the communication and computation costs in cross-silo federated learning: A comprehensive review. IEEE Access. 2025 May 26.

[18] Joblin M, Eckl B, Bock T, Schmid A, Siegmund J, Apel S. Hierarchical and hybrid organizational structures in open-source software projects: A longitudinal study. ACM Transactions on Software Engineering and Methodology. 2023 May 26;32(4):1-29.

[19] Beltrán ET, Pérez MQ, Sánchez PM, Bernal SL, Bovet G, Pérez MG, Pérez GM, Celdrán AH. Decentralized federated learning: Fundamentals, state of the art, frameworks, trends, and challenges. IEEE Communications Surveys & Tutorials. 2023 Sep 15;25(4):2983-3013.

[20] Pais V, Rao S, Muniyal B. The Critical Role of Client Selection in Cross-Silo Federated Learning: Challenges, Approaches, and Opportunities. IEEE Access. 2025 Oct 17.

[21] McCall A. EDGE-TO-CLOUD AI INTEGRATION: HYBRID ARCHITECTURES FOR REAL-TIME INFERENCE AND DATAPROCESSING IN IOTAPPLICATIONS.

[22] Rao B, Zhang J, Wu D, Zhu C, Sun X, Chen B. Privacy inference attack and defense in centralized and federated learning: A comprehensive survey. IEEE Transactions on Artificial Intelligence. 2024 Feb 8.

[23] Malomo OO. *Cybersecurity Through a Blockchain Enabled Federated Cloud Framework* (Doctoral dissertation, Howard University).

[24] Guo Z, Pan H, He A, Dai Y, Huang X, Si X, Yuen C, Zhang Y. Trusted Execution Environments for Blockchain: Towards Robust, Private, and Scalable Distributed Ledgers. IEEE Internet of Things Journal. 2025 Jul 8.

[25] Goldblum M, Tsipras D, Xie C, Chen X, Schwarzschild A, Song D, Mądry A, Li B, Goldstein T. Dataset security for machine learning: Data poisoning, backdoor attacks, and defenses. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022 Mar 25;45(2):1563-80.

[26] Albalawi A. Cloud Security and Data Protection in Hybrid Environments. InComplexities and Challenges for Securing Digital Assets and Infrastructure 2025 (pp. 175-198). IGI Global Scientific Publishing.

[27] Myakala PK, Bura C. Robust Defense Mechanisms for Agentic News Recommenders: Mitigating Data Poisoning Attacks. InCompanion Proceedings of the ACM on Web Conference 2025 2025 May 8 (pp. 1696-1704).

[28] Salim S, Moustafa N, Hassanian M, Ormod D, Slay J. Deep-federated-learning-based threat detection model for extreme satellite communications. IEEE Internet of Things Journal. 2023 Aug 3;11(3):3853-67.

[29] Baligodugula VV, Ghimire A, Amsaad F. An overview of secure network segmentation in connected IIoT environments. Computing&AI Connect. 2024 Aug 28;1(1):1-0.

[30] Hazra R, Chatterjee P, Singh Y, Podder G, Das T. Data encryption and secure communication protocols. InStrategies for E-Commerce Data Security: Cloud, Blockchain, AI, and Machine Learning 2024 (pp. 546-570). IGI Global.

[31] Lin J, Du M, Liu J. Free-riders in federated learning: Attacks and defenses. arXiv preprint arXiv:1911.12560. 2019 Nov 28.

[32] Kaul D. Optimizing resource allocation in multi-cloud environments with artificial intelligence: Balancing cost, performance, and security. JICET. 2019;4:1-25.

[33] Alsoghayer RA. Risk assessment models for resource failure in grid computing. University of Leeds; 2011.

[34] Fiero AW, Beier E. New global developments in data protection and privacy regulations: Comparative analysis of European Union, United States, and Russian legislation. Stan. J. Int'l L.. 2022;58:151.

[35] Voss WG. Cross-border data flows, the GDPR, and data governance. Wash. Int'l LJ. 2019;29:485.

[36] Bhardwaj S. Cloud Infrastructure Modernization for Regulated Industries: Balancing Innovation, Compliance, and Scalability. Journal of Computer Science and Technology Studies. 2025 Sep 23;7(9):757-67.

[37] Zhao J, Chen Y, Zhang W. Differential privacy preservation in deep learning: Challenges, opportunities and solutions. IEEE Access. 2019 Apr 9;7:48901-11.

[38] Wu X, Zhang Y, Shi M, Li P, Li R, Xiong NN. An adaptive federated learning scheme with differential privacy preserving. Future Generation Computer Systems. 2022 Feb 1;127:362-72.

[39] Lécuyer M, Spahn R, Vodrahalli K, Geambasu R, Hsu D. Privacy accounting and quality control in the sage differentially private ML platform. InProceedings of the 27th ACM Symposium on Operating Systems Principles 2019 Oct 27 (pp. 181-195).

[40] Zhao C, Zhao S, Zhao M, Chen Z, Gao CZ, Li H, Tan YA. Secure multi-party computation: theory, practice and applications. Information Sciences. 2019 Feb 1;476:357-72.

[41] Xie Q, Jiang S, Jiang L, Huang Y, Zhao Z, Khan S, Dai W, Liu Z, Wu K. Efficiency optimization techniques in privacy-preserving federated learning with homomorphic encryption: A brief survey. IEEE Internet of Things Journal. 2024 Jul 8;11(14):24569-80.

[42] Zahra WU, Amjad MT, Ahsan A, Mumtaz G. Analyzing the Limitations and Efficiency of Configuration Strategies in Hybrid Cloud Environments. Journal of Computing & Biomedical Informatics. 2024 Sep 1;7(02).

[43] Kumar M, Sethi M, Rani S, Sah DK, AlQahtani SA, Al-Rakhami MS. Secure data aggregation based on end-to-end homomorphic encryption in IoT-based wireless sensor networks. Sensors. 2023 Jul 6;23(13):6181.

[44] Chu CK, Chow SS, Tzeng WG, Zhou J, Deng RH. Key-aggregate cryptosystem for scalable data sharing in cloud storage. IEEE transactions on parallel and distributed systems. 2013 Apr 11;25(2):468-77.

[45] Oluwafemi B. Privacy-preserving computation (homomorphic encryption, MPC). Journal of Contemporary Educational Research. 2025 Sep 19;7:111-22.

[46] Lan H, Zhou Z, Zhu Q, Yan W, Hao Q, Ye X, Liu Y, Sun N. Heterogeneous Confidential Computing System for Large Language Models: A Survey. ACM Transactions on Architecture and Code Optimization. 2025.

[47] Zhao S, Zhang Q, Qin Y, Feng W, Feng D. Sectee: A software-based approach to secure enclave architecture using tee. InProceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security 2019 Nov 6 (pp. 1723-1740).

[48] Lou X, Zhang T, Jiang J, Zhang Y. A survey of microarchitectural side-channel vulnerabilities, attacks, and defenses in cryptography. ACM Computing Surveys (CSUR). 2021 Jul 13;54(6):1-37.

[49] Hosseinalipour S, Brinton CG, Aggarwal V, Dai H, Chiang M. From federated to fog learning: Distributed machine learning over heterogeneous wireless networks. IEEE Communications Magazine. 2021 Jan 1;58(12):41-7.

[50] Cichocki A, Lee N, Oseledets I, Phan AH, Zhao Q, Mandic DP. Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. Foundations and Trends® in Machine Learning. 2016 Dec 18;9(4-5):249-429.

[51] Doudou M, Djenouri D, Badache N. Survey on latency issues of asynchronous MAC protocols in delay-sensitive wireless sensor networks. IEEE Communications Surveys & Tutorials. 2012 Apr 11;15(2):528-50.

[52] Shrestha PL, Hempel M, Sharif H, Chen HH. Modeling latency and reliability of hybrid technology networking. IEEE Sensors Journal. 2013 Jul 18;13(10):3616-24.

[53] Lin X, Zhao X, Yin Z, Pan H, Jing P, Li W, Wang K, Shu Y. Knowledge-driven federated learning: A systematic literature review on approaches, challenges, and prospects: X. Lin et al. The Journal of Supercomputing. 2025 Jun 12;81(8):1016.

[54] Awaysheh FM, Alazab M, Garg S, Niyato D, Verikoukis C. Big data resource management & networks: Taxonomy, survey, and future directions. IEEE Communications Surveys & Tutorials. 2021 Jul 9;23(4):2098-130.

[55] Mary BJ. Edge-to-Cloud Federated Data Integration.

[56] Memon S. Resiliency in kubernetes federation management.

[57] Tabrizchi H, Aghasi A. Federated Cyber Intelligence: Federated Learning for Cybersecurity. Springer Nature; 2025 Apr 23.

[58] Zhao Y, Wang W, Li Y, Meixner CC, Tornatore M, Zhang J. Edge computing and networking: A survey on infrastructures and applications. IEEE Access. 2019 Jul 9;7:101213-30.

[59] Ding A, Hass A, Chan M, Sehatbakhsh N, Zonouz S. Resource-aware DNN partitioning for privacy-sensitive edge-cloud systems. InInternational Conference on Neural Information Processing 2023 Nov 15 (pp. 188-201). Singapore: Springer Nature Singapore.

[60] Zhu Q, Loke SW, Trujillo-Rasua R, Jiang F, Xiang Y. Applications of distributed ledger technologies to the internet of things: A survey. ACM computing surveys (CSUR). 2019 Nov 14;52(6):1-34.

[61] Karydas DI, Leligou HC. Federated Learning: Attacks and Defenses, Rewards, Energy Efficiency: Past, Present and Future. WSEAS Transactions on Computers. 2024;23:106-35.

[62] Nguyen DC, Uddin MR, Shaon S, Rahman R, Dobre O, Niyato D. Quantum federated learning: A comprehensive survey. arXiv preprint arXiv:2508.15998. 2025 Aug 21.

[63] Fedorov AK. Deploying hybrid quantum-secured infrastructure for applications: When quantum and post-quantum can work together. Frontiers in Quantum Science and Technology. 2023 Apr 14;2:1164428.