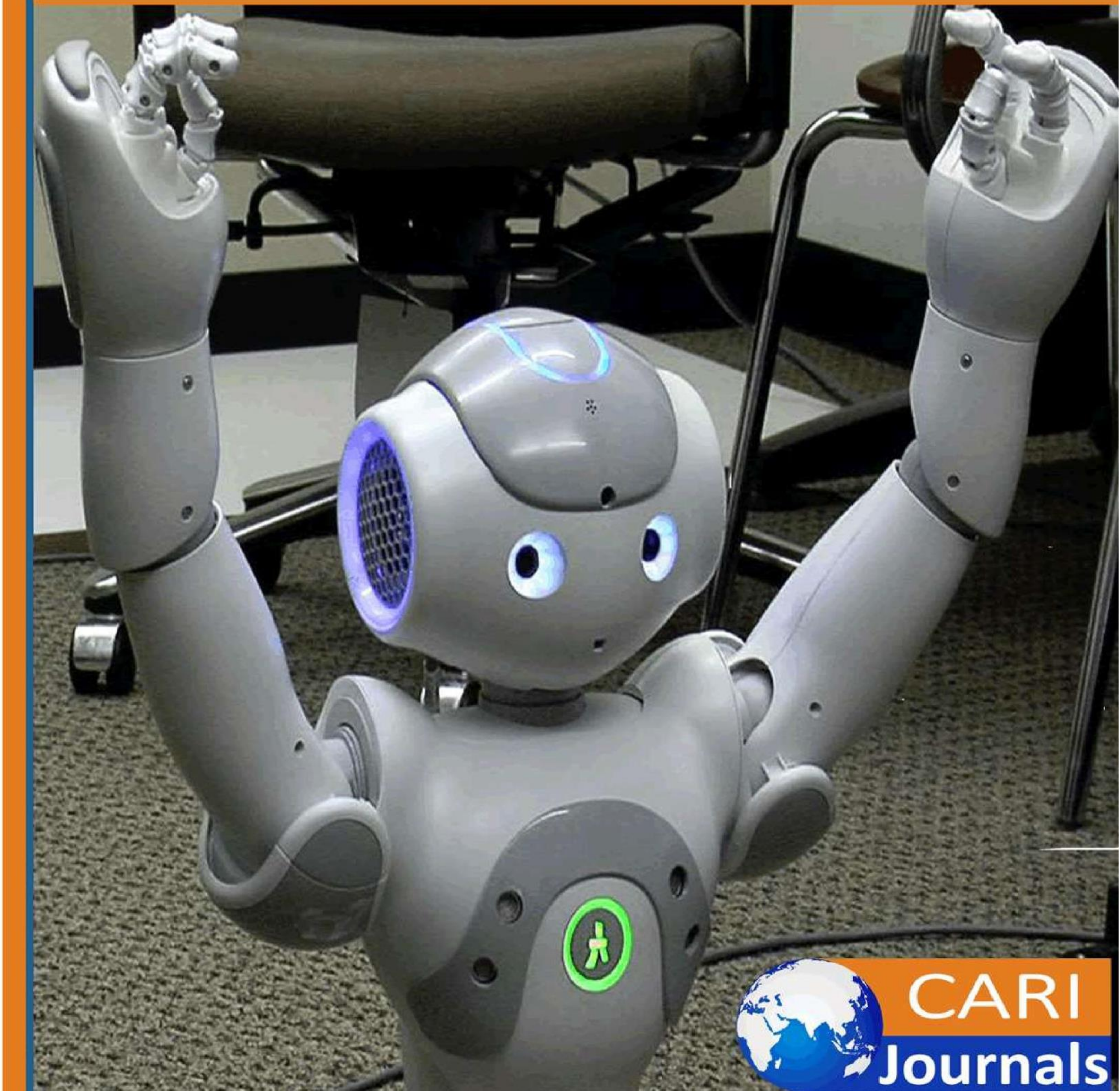
International Journal of
Computing and
Engineering
(IJCE)

Federated Machine Learning Across Hybrid Clouds: Balancing Security and Privacy



CARI
Journals

Federated Machine Learning Across Hybrid Clouds: Balancing Security and Privacy

Sri Ramya Deevi

Booz Allen Hamilton

https://orcid.org/0009-0004-6454-977X

Accepted: 1st Aug, 2023, Received in Revised Form: 10th Aug, 2023, Published: 22nd Aug, 2023

Abstract

Purpose: The purpose of this study is to examine the application of Federated Machine Learning (FML) within hybrid cloud environments, where public and private infrastructures coexist. While FML inherently enhances privacy by keeping data localized, its deployment in hybrid clouds introduces complex challenges regarding data security, compliance, and trust. This article aims to identify the critical trade-offs between model accuracy, computational efficiency, and privacy preservation, while proposing a framework to address these issues.

Methodology: Evaluating adversarial attacks and data leakage risks specific to distributed and hybrid cloud contexts. Investigating privacy-preserving techniques such as differential privacy, secure multiparty computation, and trusted execution environments, with a focus on their scalability and performance in hybrid deployments. Designing a security-aware framework that balances trust management, policy enforcement, and data protection across hybrid cloud infrastructures. Conducting scenario-based analyses to demonstrate how organizations can implement federated learning within hybrid clouds while meeting compliance and data sovereignty requirements.

Findings: The findings reveal that federated learning in hybrid clouds can provide significant benefits in terms of privacy and regulatory compliance. Organizations must balance performance metrics, such as model accuracy and training efficiency, with stringent security requirements. Differential privacy and secure multiparty computation offer strong protection but may degrade efficiency, while trusted execution environments present a middle ground with practical benefits for hybrid scenarios.

Unique Contribution to Theory, Policy and Practice: The proposed security-aware framework supports adaptable and resilient implementations, helping organizations enforce policies, manage trust relationships, and safeguard sensitive data. Effective adoption requires aligning technical safeguards with regulatory mandates, ensuring that privacy-preserving strategies remain adaptable across evolving multi-cloud ecosystems.

Keywords: *Federated Machine Learning (FML), Hybrid Cloud, Differential Privacy, Secure Aggregation, Data Sovereignty, Distributed Learning*

1. Introduction

The exponential growth of data from distributed sources such as edge devices, mobile applications, and institutional silos has spurred interest in Federated Machine Learning (FML) as a privacy-preserving alternative to centralized machine learning. FML allows model training across decentralized data sources without transferring raw data to a central server, thereby mitigating privacy risks and addressing data sovereignty concerns [1]. This decentralized approach has shown promising applications in healthcare, finance, and IoT, where sensitive data is often subject to regulatory constraints such as HIPAA and GDPR [2].

Simultaneously, hybrid cloud architectures integrating public and private cloud infrastructures have emerged as the dominant model for scalable, flexible, and cost-effective enterprise computing. The convergence of FML and hybrid cloud environments introduces new layers of complexity in terms of security, trust boundaries, and privacy enforcement. While hybrid clouds offer dynamic resource provisioning, they also expand the attack surface, complicating secure orchestration of federated tasks across heterogeneous infrastructures [3].

The integration of FML in hybrid clouds raises key challenges: maintaining model accuracy while applying privacy-preserving techniques, ensuring data protection across trust boundaries, and aligning with compliance requirements. Several technologies, including differential privacy, secure multiparty computation (SMPC), and trusted execution environments (TEEs), have been proposed to mitigate these issues [4][5]. Yet, a holistic framework to securely deploy FML across hybrid cloud ecosystems remains underexplored. This article aims to bridge this gap by analyzing existing approaches, identifying trade-offs, and proposing a privacy-aware architecture that supports federated learning in hybrid cloud settings while balancing performance, security, and compliance.

2. Background and Related Work

Federated Machine Learning (FML) was introduced to address privacy concerns associated with centralized data aggregation. Instead of transmitting raw data to a central server, local devices or nodes train models independently and only share encrypted model updates with a coordinating server. This approach was popularized by Google’s application of federated learning to mobile keyboard prediction. Since then, FML has expanded to domains requiring strict data confidentiality, including healthcare, finance, and autonomous systems.

Initial research focused on improving communication efficiency and scalability. McMahan et al. introduced Federated Averaging (FedAvg), a core algorithm that aggregates local model updates to form a global model, reducing communication overhead while maintaining accuracy [6]. This method does not inherently guarantee privacy or robustness against adversarial attacks. Later works proposed enhancements such as compression, sparsification, and adaptive learning rates to support large-scale deployments [7].

Hybrid cloud computing, characterized by the integration of private and public cloud environments, offers flexibility and resource scalability. While it supports diverse deployment scenarios, hybrid clouds introduce challenges around data residency, cross-cloud communication, and policy enforcement [8]. When FML is deployed over such environments, concerns about secure orchestration, identity federation, and trust management are amplified. Effectiveness and computational overhead vary significantly depending on the deployment context. Several recent studies have attempted to integrate these methods within hybrid cloud frameworks, but comprehensive, scalable solutions are still lacking [9].

This review of foundational work sets the stage for analyzing the unique challenges of FML in hybrid cloud environments and motivates the need for secure, privacy-aware federated architectures.

3. Federated Learning in Hybrid Cloud Environments

Federated Learning (FL) deployed within hybrid cloud environments presents a compelling approach for organizations seeking to harness distributed data sources while adhering to privacy, compliance, and scalability requirements. Hybrid clouds comprising both public and private cloud infrastructures offer flexibility by enabling sensitive data to remain on-premises or within private clouds, while leveraging public cloud resources for computationally intensive tasks such as global model aggregation or hyperparameter tuning [10].

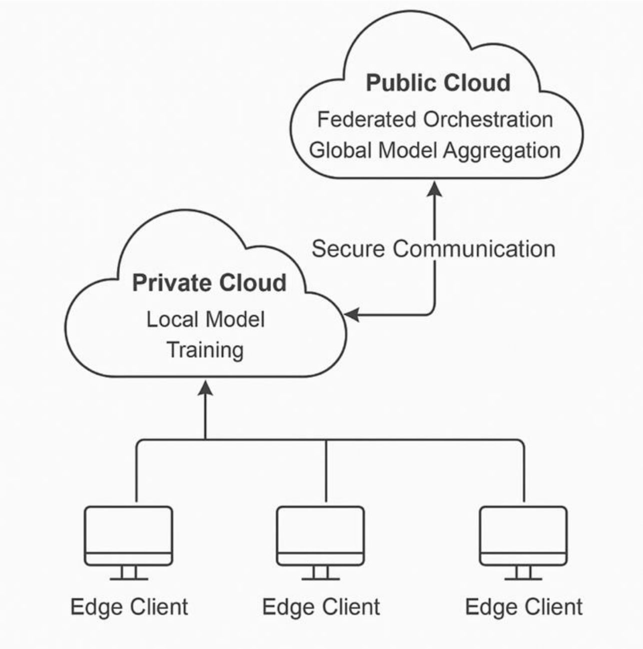


Figure 1. Federated Learning in Hybrid Cloud Environments

This architecture is particularly beneficial in regulated sectors like healthcare and finance, where data residency laws restrict data movement across jurisdictions. In such cases, FL enables model training at edge nodes hospitals or bank branches with model updates coordinated via a public cloud-based aggregator [11]. This setup introduces complexity related to infrastructure heterogeneity, trust boundaries, and orchestration of federated processes across cloud layers.

A typical FL system in a hybrid cloud includes three tiers: edge clients or data silos often in private clouds, a federated orchestration layer (hosted in the public cloud), and a secure communication protocol linking them. The hybrid cloud model supports elastic resource provisioning and high availability but also increases the risk of side-channel attacks, configuration errors, and latency issues during synchronization of model updates [12].

Security risks in such environments are exacerbated by the distributed nature of both data and compute resources. Therefore, hybrid FL deployments often require advanced privacy-preserving techniques like differential privacy, secure aggregation, and hardware-based enclaves, integrated with identity and access management systems across cloud platforms [13]. Deploying FL across hybrid clouds necessitates efficient coordination mechanisms, policy-aware workload distribution, and compatibility with diverse APIs and compliance frameworks. These challenges underscore the need for robust architectural designs that balance performance with security and privacy guarantees in hybrid federated systems [14].

4. Security and Privacy Challenges

Federated Machine Learning (FML) deployed across hybrid cloud environments introduces complex security and privacy challenges due to the interplay of distributed data sources, cross-cloud communication, and untrusted participants. Although FML is designed to enhance privacy by keeping data localized, the process of sharing model updates, gradients, and metadata can still reveal sensitive information [15]. One of the primary threats in federated settings is model inversion attacks, where adversaries reconstruct original data by analyzing gradients or updates shared during training [16]. In hybrid clouds, where aggregation often occurs in a public cloud environment, the risk is heightened if the aggregator is compromised or colludes with malicious actors.

Another concern is membership inference attacks, where an attacker determines whether a particular data point was part of the training set. This can pose serious compliance issues under regulations such as the General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA) [17]. Hybrid cloud setups expand the attack surface by introducing multiple administrative domains, which may differ in security policies, configurations, and trust assumptions. Misconfigured cloud services, lack of end-to-end encryption, and insufficient identity management mechanisms can lead to data leakage or unauthorized access [18]. Poisoning attacks where adversaries manipulate local model updates to corrupt the global model are particularly challenging in hybrid federated environments due to limited visibility across cloud boundaries and heterogeneity of nodes [19].

Securing FML in such scenarios requires robust countermeasures, including secure aggregation, differential privacy, trusted execution environments (TEEs), and policy-aware access control frameworks [20]. These solutions often involve trade-offs in terms of computational overhead, model accuracy, and communication latency, which must be carefully balanced improvement and the strategic use of Agile coaches helped it achieve significant improvements in innovation, time to market, and customer satisfaction [18].

5. Privacy-Enhancing Technologies

To address the inherent risks in Federated Machine Learning (FML) across hybrid cloud environments, a variety of privacy-enhancing technologies (PETs) have been developed. These techniques aim to minimize information leakage during model training and communication, thereby strengthening confidentiality and compliance with regulatory frameworks.

One foundational approach is differential privacy (DP), which adds calibrated noise to model updates or outputs to obscure the contribution of individual data points. This ensures that the inclusion or exclusion of a single data record has a limited impact on the model's behavior, providing formal privacy guarantees [21]. While effective, DP introduces a trade-off between privacy and model accuracy, which can be challenging to manage in hybrid cloud scenarios involving heterogeneous data distributions and compute resources.

Secure Multiparty Computation (SMPC) is another promising method, allowing multiple parties to jointly compute a function over their inputs without revealing them to one another. In the context of FML, SMPC can be used to perform secure aggregation of model updates, thereby preventing any single server a public cloud aggregator from learning sensitive information [22]. Homomorphic encryption (HE) allows computations to be performed directly on encrypted data. Though it offers strong privacy guarantees, its application in real-time federated learning remains limited due to high computational overhead and latency, especially in hybrid environments where network performance can vary [23].

Trusted Execution Environments (TEEs), such as Intel SGX, provide hardware-based isolated regions for secure execution of code and data. TEEs are particularly valuable in hybrid clouds, where the public cloud component may not be fully trusted. TEEs can host the aggregation logic, ensuring integrity and confidentiality even in hostile environments [24]. Blockchain and distributed ledger technologies have been explored to provide tamper-evident logging and decentralized coordination in federated systems. These can improve transparency and auditability but often suffer from scalability and energy consumption issues [25]. These PETs offer building blocks for secure and privacy-preserving federated learning in hybrid clouds. Selecting and configuring them appropriately requires careful consideration of performance, trust assumptions, and the sensitivity of the data involved.

6. Trade-offs and Performance Considerations

While Federated Machine Learning (FML) across hybrid cloud environments offers significant privacy and scalability benefits, it also presents trade-offs in model accuracy, system performance, communication overhead, and regulatory compliance. These considerations are critical when designing secure FML systems that operate effectively in hybrid infrastructures. One of the most pressing challenges is the accuracy privacy trade-off. Techniques such as differential privacy and secure aggregation introduce noise or computation constraints that can degrade model performance. Adding noise to model updates for privacy protection often results in reduced prediction accuracy, particularly in non-IID (non-independent and identically distributed) data scenarios common across edge and cloud nodes [26].

Another important factor is communication efficiency. FML systems must transmit model updates between edge devices, private clouds, and public cloud aggregators. Secure protocols like homomorphic encryption or SMPC significantly increase communication costs due to large ciphertext sizes or complex multi-party protocols [27]. This is further exacerbated in hybrid cloud environments, where latency and bandwidth may vary across networks and providers. Computational overhead is also a concern, particularly with resource-intensive privacy-preserving technologies. Techniques like homomorphic encryption and trusted execution environments (TEEs) demand additional compute cycles and memory, potentially slowing down the learning process, especially at edge devices with limited resources [28].

Scalability and orchestration complexity increase as FML deployments expand across diverse hybrid cloud nodes. Maintaining synchronization, secure key exchange, and federated coordination becomes more difficult and costly with increasing numbers of participants [29]. Compliance versus agility remains a fundamental tension. Organizations must balance rapid model iteration and deployment with strict regulatory requirements such as GDPR, HIPAA, or CCPA. Ensuring that models are auditable, explainable, and trained with privacy guarantees is often in tension with the agile, performance-oriented design goals of AI systems [30].

Addressing these trade-offs requires adaptive architectures that allow fine-grained control over security parameters, context-aware optimization of communication protocols, and dynamic resource allocation strategies across the hybrid cloud continuum.

7. Proposed Framework for Secure FML in Hybrid Clouds

To effectively balance privacy, security, and performance in Federated Machine Learning (FML) within hybrid cloud environments, I propose a modular and adaptive framework that integrates policy-aware orchestration, secure aggregation mechanisms, and dynamic trust management across cloud tiers. This framework is designed to support large-scale deployments involving heterogeneous devices, edge nodes, private data centers, and public cloud services.

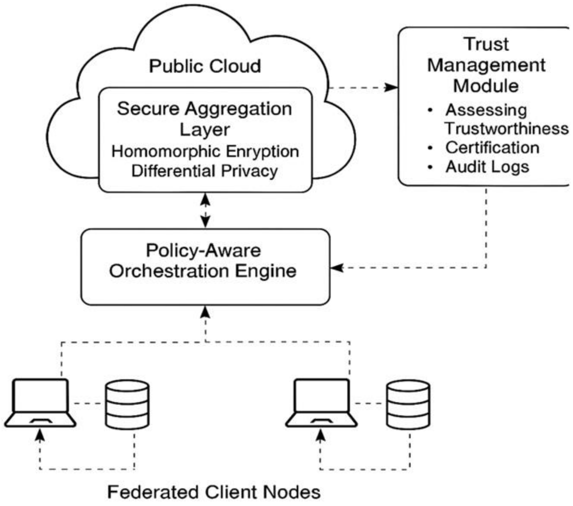


Figure 2. Framework for Secure FML in Hybrid Clouds

The proposed architecture consists of four key components

Federated Client Nodes: typically residing on edge devices or within private clouds, responsible for local model training on sensitive data

Secure Aggregation Layer: hosted on the public cloud or within a Trusted Execution Environment (TEE), which performs privacy-preserving aggregation of model updates

Policy-Aware Orchestration Engine: which enforces data residency, access control, and compliance policies

Trust Management Module: which assesses and certifies participant trustworthiness and maintains audit logs for compliance validation.

To ensure data confidentiality and integrity, the framework utilizes homomorphic encryption or Secure Multiparty Computation (SMPC) during update transmission and aggregation phases [31]. Differential privacy is optionally applied at the client side to introduce statistical noise and reduce the risk of individual re-identification [32].

The orchestration engine dynamically assigns training and aggregation roles based on resource availability, network conditions, and trust scores. This policy-driven distribution optimizes for performance while adhering to regulatory constraints such as HIPAA or GDPR [33]. TEEs are employed to ensure secure computation on untrusted public cloud nodes, protecting both the code and data involved in aggregation [34]. A blockchain-based audit layer can optionally be integrated to ensure tamper-proof logging of federated rounds, update provenance, and compliance reports, which is particularly useful in regulated sectors like healthcare and finance [35].

The modular design of this framework allows customization and scalability. High-trust environments may favor lightweight encryption, while low-trust, cross-border deployments can opt for stronger privacy guarantees with more computational overhead. This adaptability ensures that the framework is suitable across a wide range of federated learning use cases in hybrid cloud infrastructures.

8. Future Research Directions

As Federated Machine Learning (FML) continues to evolve and integrate with hybrid cloud architectures, several open challenges remain that warrant further research and innovation. These directions focus on enhancing scalability, adaptability, and trustworthiness of FML systems, particularly in complex, multi-cloud environments.

Zero-Trust Architecture for Federated Learning: Incorporating zero-trust principles into FML frameworks is a promising avenue for securing communications and enforcing least-privilege access. Future systems must authenticate and verify every entity whether a cloud component, edge device, or orchestration node before trust is established. This calls for research into lightweight, context-aware authentication mechanisms suitable for federated deployments.

Energy-Efficient Privacy Techniques: Privacy-enhancing technologies such as homomorphic encryption and secure multiparty computation are often computationally expensive. Developing lightweight, energy-efficient variants of these methods tailored for resource-constrained edge devices remains an important research challenge.

Federated Model Lifecycle Management: Automating the end-to-end lifecycle of federated models including version control, rollback, and compliance auditing will be essential in regulated industries. Integration with CI/CD pipelines and data lineage tracking are important areas for further exploration.

Cross-Cloud Interoperability Standards: Interoperability between public and private clouds is critical for hybrid FML systems. Future work should support the development of standardized APIs, secure data interchange formats, and federation protocols to enable seamless model sharing and orchestration across heterogeneous cloud providers.

Blockchain Integration for Auditable FML: Blockchain technology can be leveraged to create immutable audit trails of model updates, participant behavior, and policy enforcement. Research into scalable, low latency distributed ledger designs compatible with hybrid cloud environments will be crucial for building trustworthy federated systems.

By addressing these directions, future research can significantly enhance the resilience, transparency, and applicability of federated learning in hybrid cloud environments, fostering broader adoption in critical sectors such as healthcare, finance, and public infrastructure.

9. Conclusion

Federated Machine Learning (FML) offers a transformative approach to privacy-preserving AI by enabling decentralized model training without exposing sensitive data. When integrated into hybrid cloud environments, FML presents unique advantages in terms of scalability, compliance, and resource optimization. It also introduces complex challenges related to security, data privacy, communication efficiency, and orchestration across heterogeneous infrastructures. This article has explored the key architectural components, security threats, and privacy-enhancing technologies relevant to FML in hybrid clouds. I presented a modular framework that incorporates secure aggregation, policy-aware orchestration, and trust management to balance performance and privacy needs. I also highlighted trade-offs inherent in current solutions and outlined future research directions to address evolving technical and regulatory demands.

As FML continues to mature, it is essential to develop adaptable, secure, and interoperable systems that meet the stringent requirements of distributed, data-sensitive applications. The proposed framework serves as a foundation for building such systems, promoting trustworthy AI in domains where both innovation and compliance are paramount. By advancing secure FML architectures, organizations can unlock the full potential of collaborative machine learning while safeguarding user data and organizational integrity in hybrid cloud environments.

References

[1] Konečný, J., McMahan, H. B., et al., “Federated Learning: Strategies for Improving Communication Efficiency,” arXiv preprint arXiv:1610.05492, 2016.

[2] Rieke, N., et al., “The future of digital health with federated learning,” NPJ Digital Medicine, vol. 3, no. 1, pp. 1–7, 2020.

[3] Zhang, Y., et al., “Securing Data in Hybrid Clouds: Challenges and Solutions,” IEEE Cloud Computing, vol. 3, no. 1, pp. 58–66, Jan./Feb. 2016.

[4] Dwork, C., Roth, A., “The Algorithmic Foundations of Differential Privacy,” Found. Trends Theor. Comput. Sci., vol. 9, no. 3–4, pp. 211–407, 2014.

[5] Mohassel, P., Zhang, Y., “SecureML: A system for scalable privacy-preserving machine learning,” in IEEE Symposium on Security and Privacy (SP), 2017, pp. 19–38.

[6] McMahan, H. B., Moore, E., Ramage, D., et al., “Communication-efficient learning of deep networks from decentralized data,” in Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS), 2017, pp. 1273–1282.

[7] Sattler, F., Wiedemann, S., Müller, K.-R., et al., “Robust and communication-efficient federated learning from non-iid data,” IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 9, pp. 3400–3413, Sept. 2020.

[8] Gai, K., Qiu, M., Zhao, H., “Security and privacy issues: A survey on federated learning,” Future Generation Computer Systems, vol. 105, pp. 719–727, Apr. 2020.

[9] Qu, Y., Wu, D., Xu, J., et al., “Blockchain-based federated learning with hybrid cloud architecture,” IEEE Internet of Things Journal, vol. 9, no. 7, pp. 5072–5084, Apr. 2022.

[10] Zhang, Q., Yang, L. T., Chen, Z., et al., "A survey on deep learning for big data," Information Fusion, vol. 42, pp. 146–157, Jul. 2018.

[11] Sheller, M. J., Edwards, B., Reina, G. A., et al., "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," Scientific Reports, vol. 10, no. 1, p. 12598, 2020.

[12] Shamsabadi, A. S., Bertran, A., Zolfaghari, P., et al., "Distributed machine learning in edge computing: A review," IEEE Internet of Things Journal, vol. 10, no. 3, pp. 1572–1589, Feb. 2023.

[13] Bonawitz, K., Eichner, H., Grieskamp, W., et al., "Towards federated learning at scale: System design," in Proc. 2nd SysML Conf., Palo Alto, CA, USA, 2019.

[14] Yang, Q., Liu, Y., Chen, T., et al., "Federated machine learning: Concept and applications," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 10, no. 2, pp. 1–19, Jan. 2019.

[15] Melis, L., Song, C., De Cristofaro, E., et al., "Exploiting Unintended Feature Leakage in Collaborative Learning," in IEEE Symposium on Security and Privacy (SP), 2019, pp. 691–706.

[16] Fredrikson, M., Jha, S., Ristenpart, T., "Model inversion attacks that exploit confidence information and basic countermeasures," in Proc. 22nd ACM SIGSAC Conf. on Computer and Communications Security (CCS), 2015, pp. 1322–1333.

[17] Shokri, R., Stronati, M., Song, C., et al., "Membership inference attacks against machine learning models," in IEEE Symposium on Security and Privacy (SP), 2017, pp. 3–18.

[18] Zhang, Y., Chen, X., Wang, J., et al., "Security and Privacy in Smart City Applications: Challenges and Solutions," IEEE Communications Magazine, vol. 58, no. 3, pp. 20–26, Mar. 2020.

[19] Bhagoji, A. N., Chakraborty, S., Mittal, P., et al., "Analyzing federated learning through an adversarial lens," in Proc. 36th Int. Conf. Machine Learning (ICML), 2019, pp. 634–643.

[20] Bonawitz, K., Ivanov, V., Kreuter, B., et al., "Practical Secure Aggregation for Privacy-Preserving Machine Learning," in Proc. ACM Conf. Computer and Communications Security (CCS), 2017, pp. 1175–1191.

[21] Abadi, M., Chu, A., Goodfellow, I., et al., "Deep learning with differential privacy," in Proc. ACM SIGSAC Conf. on Computer and Communications Security (CCS), 2016, pp. 308–318.

[22] Bonawitz, K., Ivanov, V., Kreuter, B., et al., "Practical secure aggregation for privacy-preserving machine learning," in Proc. ACM Conf. Computer and Communications Security (CCS), 2017, pp. 1175–1191.

[23] Acar, A., Aksu, H., Uluagac, A. S., et al., "A survey on homomorphic encryption schemes: Theory and implementation," ACM Computing Surveys (CSUR), vol. 51, no. 4, pp. 1–35, Jul. 2018.

[24] Costan, V., Devadas, S., "Intel SGX explained," IACR Cryptology ePrint Archive, vol. 2016, p. 86, 2016.

[25] Kim, H.-M., Laskowski, M., "Toward an ontology-driven blockchain design for supply-chain provenance," Intelligent Systems in Accounting, Finance and Management, vol. 25, no. 1, pp. 18–27, Jan. 2018.

[26] Geyer, R. C., Klein, T., Nabi, M., "Differentially private federated learning: A client level perspective," arXiv preprint arXiv:1712.07557, 2017.

[27] Mohassel, P., Zhang, Y., "SecureML: A system for scalable privacy-preserving machine learning," in IEEE Symposium on Security and Privacy (SP), 2017, pp. 19–38.

[28] Tramer, F., Zhang, F., Juels, A., et al., "Stealing machine learning models via prediction APIs," in 25th USENIX Security Symposium, 2016, pp. 601–618.

[29] Kairouz, P., McMahan, H. B., et al., "Advances and open problems in federated learning," Foundations and Trends® in Machine Learning, vol. 14, no. 1–2, pp. 1–210, 2021.

[30] Veale, M., Binns, R., "Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data," Big Data & Society, vol. 4, no. 2, pp. 1–17, 2017.

[31] Li, T., Sahu, A. K., Talwalkar, A., et al., "Federated learning: Challenges, methods, and future directions," IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50–60, May 2020.

[32] Dwork, C., Roth, A., "The Algorithmic Foundations of Differential Privacy," Foundations and Trends® in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014.

[33] Kairouz, P., McMahan, H. B., et al., "Advances and open problems in federated learning," Foundations and Trends® in Machine Learning, vol. 14, no. 1–2, pp. 1–210, 2021.

[34] Costan, V., Devadas, S., "Intel SGX Explained," IACR Cryptology ePrint Archive, vol. 2016, p. 86, 2016.

[35] Zhang, R., Xue, R., Liu, L., "Security and privacy on blockchain," ACM Computing Surveys (CSUR), vol. 52, no. 3, pp. 1–34, 2019.