# Blockchain and Federated Learning Integrated Architecture for Secure Internet of Things Data Sharing

Ankit M. Bansal
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
bansalankit@colostate.edu

Neeraj Batra
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
neeraj.batra@ucf.edu

**Abstract**

The proliferation of Internet of Things devices has generated vast quantities of sensory and operational data, yet the security and privacy of sharing such data across heterogeneous networks remain pressing concerns. This paper presents a comprehensive architectural framework that integrates blockchain technology with federated learning to establish a secure, decentralized, and privacy-preserving data sharing ecosystem for IoT environments. The proposed architecture leverages blockchain as an immutable ledger for audit trails, incentive mechanisms, and decentralized identity management, while federated learning enables collaborative model training without exposing raw data. The discussion emphasizes system-level trade-offs including latency, computational overhead, scalability, and energy consumption, and examines structural design choices such as on-chain versus off-chain data storage, consensus protocol selection, and the role of smart contracts in automating governance. Infrastructure deployment considerations across edge, fog, and cloud layers are analyzed alongside sustainability metrics related to carbon footprint and hardware longevity. Robustness against adversarial attacks, fairness in reward distribution among heterogeneous participants, and policy implications for regulatory compliance are explored through cross-domain comparisons with existing centralized and hybrid approaches. The paper argues that while the integrated architecture introduces additional complexity, its capacity for transparency, accountability, and user sovereignty offers a viable pathway toward trustworthy IoT data sharing, provided that careful attention is paid to incentive alignment, network partitions, and cryptographic overhead. Forward-looking perspectives on interoperability standards, quantum-resistant cryptography, and decentralized autonomous organizations are discussed to guide future research and deployment.

**Keywords**

blockchain, federated learning, Internet of Things, data sharing, security, privacy, decentralized governance, smart contracts, edge computing.

## 1. Introduction

The Internet of Things has expanded rapidly across domains such as smart cities, industrial automation, healthcare, and environmental monitoring, leading to an exponential increase in data generation at the network edge. This data holds significant value for training machine learning models that can improve operational efficiency, predictive maintenance, and

personalized services. However, traditional approaches to IoT data sharing rely on centralized cloud servers or trusted third parties, which introduce single points of failure, data silos, and privacy vulnerabilities. Regulatory frameworks such as the General Data Protection Regulation further mandate strict controls over personal data, making centralized collection and processing increasingly untenable [1], [2]. Consequently, researchers and practitioners have turned to decentralized architectures that combine blockchain with federated learning to mitigate these challenges.

Blockchain provides a tamper-resistant distributed ledger that records transactions and data provenance without requiring a central authority. Its integration with IoT devices enables transparent audit trails, secure device identity, and automated incentive mechanisms through smart contracts [3], [4]. Federated learning, on the other hand, allows multiple parties to collaboratively train a shared machine learning model while keeping raw data locally on the devices, thereby reducing privacy risks and communication overhead [5]. The synergy between these two technologies offers a promising foundation for secure IoT data sharing, yet the architectural design involves numerous trade-offs that must be carefully balanced.

This paper proposes a holistic architectural framework that unifies blockchain and federated learning for IoT data sharing. The discussion adopts a system-level perspective, focusing on structural trade-offs, governance models, infrastructure deployment, sustainability, robustness, fairness, and policy implications. Rather than presenting a specific implementation, the paper aims to synthesize existing research and identify critical design dimensions that influence the viability and performance of such integrated systems. The remainder of the paper is organized as follows. Section 2 reviews background and related work. Section 3 presents the proposed integrated architecture. Section 4 examines system-level trade-offs and design considerations. Section 5 discusses governance, infrastructure, and deployment. Section 6 addresses sustainability and robustness. Section 7 explores fairness and policy implications. Section 8 concludes with forward-looking perspectives.

## 2. Background and Related Work

Federated learning was introduced as a means to train deep neural networks on decentralized data distributed across mobile devices, enabling privacy-preserving model updates without central aggregation of raw samples [5]. Subsequent extensions have addressed challenges such as communication efficiency, heterogeneous data distributions, and user dropout, making federated learning suitable for IoT environments where devices are resource-constrained and intermittently connected [6]. Meanwhile, blockchain technology, originally conceived for cryptocurrency, has evolved into smart contract platforms like Ethereum that support decentralized applications across various sectors [7]. The combination of blockchain and federated learning has attracted considerable attention in recent literature, with proposals ranging from audit mechanisms for gradient updates to decentralized marketplaces for data contributions [8], [9].

Early works focused on using blockchain to record model updates and rewards in federated learning, thereby providing verifiability and non-repudiation [10]. However, storing large volumes of gradient data on-chain leads to high storage costs and latency, prompting research into off-chain storage solutions and lightweight consensus protocols appropriate for IoT networks [11]. Other studies have explored the use of blockchain for device registration, revocation, and access control, integrating with federated learning to enforce data usage policies [12]. Despite these advances, most existing proposals address isolated aspects rather

than offering a comprehensive architecture that accounts for the interplay between blockchain and federated learning components under realistic IoT constraints.

The present paper builds upon this body of work by synthesizing key insights and extending the discussion to include governance, sustainability, and fairness considerations that are often overlooked in technical designs. By adopting a systems perspective, the paper aims to provide researchers and practitioners with a structured framework for evaluating and implementing blockchain-federated learning integrated architectures in diverse IoT scenarios.

## 3. Proposed Integrated Architecture

The proposed architecture comprises three primary layers: the device layer, the network layer, and the blockchain-federated learning coordination layer. At the device layer, IoT edge nodes collect sensor data and perform local model training using federated learning algorithms. Each device maintains a local copy of the global model and computes gradient updates based on its own data. Rather than transmitting raw data, devices only share encrypted model updates or compressed gradients, thereby preserving data locality [13]. The network layer consists of edge servers or fog nodes that aggregate local updates from multiple devices within a geographical region. These aggregators also validate the integrity of incoming updates using digital signatures and verify that updates are computed from authentic data via lightweight proofs.

The blockchain-federated learning coordination layer manages the global model lifecycle, incentive distribution, and identity management. A permissioned blockchain, such as Hyperledger Fabric or Quorum, is employed to maintain membership records, transaction logs, and smart contract state. Smart contracts automate the reward mechanism, distributing tokens or credits to devices based on the quality and quantity of their contributed updates. They also enforce data usage policies by checking that participating devices have not exceeded their assigned quotas or violated privacy constraints [14]. The global model is stored off-chain in a distributed file system like IPFS, with its hash recorded on the blockchain to ensure integrity and version control. Model aggregation is performed off-chain by a set of trusted aggregators, and only the final model hash and aggregated metadata are committed to the blockchain periodically.

This hybrid on-chain/off-chain design balances transparency with efficiency. Full data provenance and auditability are preserved through hashed references, while the blockchain remains lightweight enough to be maintained by resource-constrained IoT nodes. The architecture also supports hierarchical aggregation, where local models from edge devices are first combined within fog domains, and domain-level updates are further aggregated at the global level, reducing communication overhead and latency [15]. The system can accommodate both synchronous and asynchronous training rounds, with smart contracts handling timeout and drop-out scenarios.

## 4. System-Level Trade-offs and Design Considerations

The integration of blockchain and federated learning introduces several fundamental trade-offs that influence system performance and practicality. One critical trade-off involves the choice of consensus protocol. Proof-of-work consensus, while highly decentralized, imposes prohibitive energy and computational demands on IoT devices, making it unsuitable for most real-world deployments [16]. Proof-of-stake or practical Byzantine fault tolerance variants offer lower overhead but require a certain degree of trust in validators, which may conflict with the goal of full decentralization. In a permissioned setting, using a raft or Kafka-based

consensus can provide high throughput at the expense of reducing the trust model to a set of known entities. The architectural design must therefore align the consensus mechanism with the required level of decentralization, security, and performance.

Another trade-off concerns the frequency and granularity of on-chain recording. Frequent on-chain updates enhance auditability and immutability but increase latency and gas costs. Conversely, batching multiple model updates into a single on-chain transaction reduces overhead but introduces a window of vulnerability where malicious updates could be aggregated before being recorded. The design must choose appropriate batch sizes and commit intervals based on the application's criticality and the characteristics of the IoT network. Similarly, the decision of which data to store on-chain versus off-chain affects the trade-off between verifiability and scalability. Storing full gradient vectors on-chain is infeasible, but storing only cryptographic hashes requires robust mechanisms for later verification if a dispute arises.

Network topology also presents significant design choices. In a fully decentralized architecture, every device participates in consensus, leading to high communication overhead and energy consumption. A more practical approach is to adopt a hierarchical topology where edge nodes form local clusters, each with a designated aggregator that handles consensus within that cluster. Cross-cluster coordination can be achieved through a relay blockchain or a consortium model. This hierarchical structure mirrors the physical organization of IoT systems, where devices are often grouped by location or function, and can improve scalability while maintaining security guarantees [17].

## 5. Governance, Infrastructure, and Deployment

Effective governance is essential for the long-term operation of an integrated blockchain-federated learning system. Governance encompasses rules for node participation, software updates, dispute resolution, and economic policy. In a permissioned blockchain, a consortium of stakeholders can define membership criteria and manage the addition or removal of nodes. Smart contracts encode these governance rules, allowing automated enforcement while leaving room for human oversight through multi-signature mechanisms [18]. For federated learning, governance includes determining the aggregation frequency, the quality metrics for rewarding contributions, and the process for handling malicious or non-responsive devices.

Infrastructure deployment must consider the heterogeneity of IoT devices, ranging from powerful gateways to constrained sensors. Edge computing nodes can host lightweight blockchain clients or act as validators, while resource-limited devices participate only as data contributors. Cloud servers may provide additional storage and computation for global model aggregation and blockchain node operation, but introducing cloud elements reintroduces centralization risks that the architecture aims to mitigate. A hybrid edge-cloud deployment can balance these concerns by executing latency-sensitive operations at the edge and resource-intensive tasks in the cloud, with careful partitioning of trust boundaries [19].

Security and privacy infrastructure must include secure hardware elements such as trusted execution environments (TEEs) for sensitive computations like gradient aggregation and key management. TEEs can protect against adversarial aggregators that might attempt to infer private information from model updates. Additionally, differential privacy techniques can be applied to local updates before transmission, further reducing the risk of leakage. The deployment of such advanced cryptographic primitives, however, increases computational

overhead and may require specialized hardware support, which is not universally available across IoT devices.

## 6. Sustainability and Robustness

Sustainability in blockchain-federated learning architectures is multi-faceted, encompassing energy consumption, hardware longevity, and economic viability. The energy cost of maintaining a blockchain network, even with lightweight consensus, contributes to the overall carbon footprint of the IoT system. Proof-of-stake and delegated proof-of-stake consensus reduce energy requirements compared to proof-of-work, but they still involve continuous communication and validation that strain battery-powered devices [20]. Energy harvesting and low-power wide-area network technologies can alleviate some of these concerns, but system designers must carefully model the energy budget for each device, factoring in the overhead of cryptographic operations and network transmissions.

Hardware longevity is another sustainability concern. Frequent participation in consensus and model training accelerates wear on device components, particularly memory and processors. A sustainable design should incorporate mechanisms for duty cycling, allowing devices to enter low-power sleep modes when not actively contributing to a training round. Additionally, the system should support graceful degradation: devices with limited resources can opt to only participate in local training without engaging in blockchain validation, receiving proportionally lower rewards.

Robustness to adversarial behavior is critical given the open nature of many IoT ecosystems. Attacks such as gradient poisoning, model inversion, and Byzantine failures can compromise the accuracy and security of federated learning [21]. The blockchain component helps by providing transparent logs of updates, enabling forensic analysis after an attack is detected. However, real-time defense remains challenging. Robust aggregation algorithms like Krum, trimmed mean, or median-based aggregation can filter out anomalous updates, but they rely on a threshold of benign participants. The integration with blockchain allows for reputation systems that track device behavior over time, penalizing devices that consistently produce outlier updates [22]. Smart contracts can automate the slashing of collateral deposited by devices, disincentivizing malicious behavior.

Network partitions and device churn pose robustness challenges for synchronous federated learning. The architecture must handle stragglers and dropped connections without stalling the training process. Asynchronous training methods, together with blockchain-based checkpointing, can maintain progress while ensuring that stale or duplicated updates are detected and excluded. The design should also incorporate fallback mechanisms: if the blockchain becomes temporarily inaccessible, devices can continue local training and synchronize later, albeit with some risk of conflicting updates that require reconciliation.

## 7. Fairness and Policy Implications

Fairness in blockchain-federated learning systems addresses both the equitable distribution of rewards and the representational balance of the global model across heterogeneous data distributions. Devices with larger or more diverse datasets may contribute more valuable updates, but they also may have greater resource consumption. A straightforward reward scheme based solely on data volume can disadvantage small-scale participants, leading to centralization of model quality among powerful nodes. Smart contracts can implement more nuanced reward functions that account for data quality, computational cost, and participation frequency, using Shapley value approximations to fairly attribute contributions [23]. However,

computing Shapley values is computationally expensive, and approximations may introduce biases that favor certain participants.

Data heterogeneity, often referred to as non-IID data, poses fairness challenges for the global model. If certain regions or device types are underrepresented, the model may perform poorly for those groups, raising concerns about algorithmic equity. The federated learning component can incorporate personalization techniques or multi-task learning to mitigate this, but these approaches increase complexity and may conflict with the goal of a single global model. Blockchain governance mechanisms can facilitate community voting on fairness objectives, allowing stakeholders to define acceptable levels of model performance disparity [24].

Policy implications extend to regulatory compliance with data protection laws. The architecture must ensure that neither the blockchain nor the federated learning process inadvertently exposes personally identifiable information. While raw data remains on devices, model updates have been shown to leak sensitive information through gradient inversion attacks. Differential privacy, when applied with sufficient noise, can reduce leakage but comes at the cost of model accuracy. The choice of privacy budget must comply with legal standards, which may vary across jurisdictions. Additionally, the immutability of blockchain conflicts with the right to erasure under the GDPR. Solutions such as off-chain storage with revocable references or chameleon hash functions have been proposed, but they introduce trust assumptions that weaken the security guarantees of the blockchain [25]. Policy frameworks need to evolve to accommodate decentralized systems where data sovereignty is shared among multiple parties.

## 8. Conclusion

This paper has presented a comprehensive integrated architecture combining blockchain and federated learning for secure IoT data sharing, emphasizing system-level trade-offs across governance, infrastructure, sustainability, robustness, fairness, and policy. The hybrid on-chain/off-chain design enables transparent auditability while maintaining scalability and efficiency for resource-constrained IoT devices. However, the integration introduces significant complexity that demands careful alignment of consensus protocols, reward mechanisms, privacy protections, and hierarchical topologies with specific application requirements.

Looking forward, several research directions merit attention. Interoperability standards for cross-chain communication will enable federated learning across multiple blockchain networks, expanding the scope of data sharing. The advent of quantum computing poses a threat to current cryptographic primitives, necessitating the exploration of post-quantum signatures and hashes for both blockchain and federated learning. Decentralized autonomous organizations offer a governance model that could automate decision-making for reward distribution, policy updates, and dispute resolution, further reducing human intervention while increasing transparency. Finally, empirical evaluations on large-scale IoT testbeds are needed to validate the performance trade-offs discussed in this paper and to inform the design of production-ready systems. The path toward secure and trustworthy IoT data sharing is multifaceted, but the integration of blockchain and federated learning, when implemented with a holistic systems perspective, holds considerable promise for enabling decentralized intelligence that respects both privacy and accountability.

## References

1. European Parliament and Council. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L119, 1-88.

2. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., ... & Roselander, J. (2019). Towards federated learning at scale: System design. Proceedings of Machine Learning and Systems, 1, 374-388.

3. Christidis, K., & Devetsikiotis, M. (2016). Blockchains and smart contracts for the Internet of Things. IEEE Access, 4, 2292-2303.

4. Dorri, A., Kanhere, S. S., Jurdak, R., & Gauravaram, P. (2017). Blockchain for IoT security and privacy: The case study of a smart home. In 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) (pp. 618-623). IEEE.

5. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics (pp. 1273-1282). PMLR.

6. Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492.

7. Buterin, V. (2014). A next-generation smart contract and decentralized application platform. White Paper.

8. Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2020). Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. IEEE Transactions on Industrial Informatics, 16(6), 4177-4186.

9. Li, Z., Kang, J., Yu, R., Ye, D., Deng, Q., & Zhang, Y. (2020). Consortium blockchain for secure energy trading in industrial Internet of Things. IEEE Transactions on Industrial Informatics, 14(8), 3690-3700.

10. Kim, H., Park, J., Bennis, M., & Kim, S. L. (2019). Blockchained on-device federated learning. IEEE Communications Letters, 24(6), 1279-1283.

11. Nguyen, D. C., Pathirana, P. N., Ding, M., & Seneviratne, A. (2021). Integration of blockchain and cloud of things: Architecture, applications, and open research issues. IEEE Internet of Things Journal, 8(1), 146-161.

12. Shafagh, H., Burkhalter, L., Hithnawi, A., & Duquennoy, S. (2017). Towards blockchain-based auditable storage and sharing of IoT data. In Proceedings of the 2017 on Cloud Computing Security Workshop (pp. 45-50).

13. Wang, S., Tuor, T., Salonidis, T., Leung, K. K., Makaya, C., He, T., & Chan, K. (2019). Adaptive federated learning in resource constrained edge computing systems. IEEE Journal on Selected Areas in Communications, 37(6), 1205-1221.

14. Novo, O. (2018). Blockchain meets IoT: An architecture for scalable access management in IoT. IEEE Internet of Things Journal, 5(2), 1184-1195.

15. Liu, L., Zhang, J., Song, S. H., & Letaief, K. B. (2020). Client-edge-cloud hierarchical federated learning. In ICC 2020-2020 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.

16. Zhang, Y., Wang, K., Zhou, Y., & Duan, L. (2020). A blockchain-based federated learning framework for secure and efficient IoT data sharing. IEEE Internet of Things Journal, 7(7), 6400-6411.

17. Kang, J., Yu, R., Huang, X., Qin, M., & Zhang, Y. (2019). Enabling localized peer-to-peer electricity trading among plug-in hybrid electric vehicles using consortium blockchains. IEEE Transactions on Industrial Informatics, 13(6), 3154-3164.

18. Beck, R., Czepluch, J. S., Lollike, N., & Malone, S. (2016). Blockchain–the gateway to trust-free cryptographic transactions. In Twenty-Fourth European Conference on Information Systems (ECIS) (pp. 1-14).

19. Satyanarayanan, M. (2017). The emergence of edge computing. Computer, 50(1), 30-39.

20. Sedlmeir, J., Buhl, H. U., Fridgen, G., & Keller, R. (2020). The energy consumption of blockchain technology: Beyond myth. Business & Information Systems Engineering, 62(6), 599-608.

21. Blanchard, P., El Mhamdi, E. M., Guerraoui, R., & Stainer, J. (2017). Machine learning with adversaries: Byzantine tolerant gradient descent. In Advances in Neural Information Processing Systems (pp. 119-129).

22. Cao, X., Liu, J., & Shen, X. (2020). Reputation-aware federated learning for secure and efficient data sharing in vehicular networks. IEEE Transactions on Vehicular Technology, 69(12), 15238-15252.

23. Ghorbani, A., & Zou, J. (2019). Data Shapley: Equitable valuation of data for machine learning. In International Conference on Machine Learning (pp. 2242-2251).

24. Hynes, N., Vohra, V., & Raj, B. (2018). A democratic approach to data sharing. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (pp. 267-280).

25. Ateniese, G., Magri, B., Venturi, D., & Andrade, E. (2017). Redactable blockchain–or–rewriting history in Bitcoin and friends. In 2017 IEEE European Symposium on Security and Privacy (EuroS&P) (pp. 111-126).