

Research Article



Decentralized Data Trust System for Privacy-Preserving Federated Learning

Mohd Salique Khan,^{1,*} Samiya Khan,² Sania Shaikh,² and Luminita Moraru³

¹ Department of Computer Engineering, M.H. Saboo Siddik College of Engineering, Mumbai, Maharashtra, 400008, India

² Department of Computer Science Engineering (Data Science), Kalsekar Technical Campus, Navi Mumbai, Maharashtra, 410206, India

³ Faculty of Sciences and Environment, the Modelling & Simulation Laboratory, Dunarea de Jos University of Galati, 47 Domneasca Str., Galati, 800008 Romania

*Corresponding Author

Mohd Salique Khan

✉ msaliquekhan@gmail.com

Received: 25 July 2026

Revised: 04 September 2026

Accepted: 17 September 2026

Published Online: 22 September 2026

Citation

M. S. Khan, S. Khan, S. Shaikh, L. Moraru, Decentralized data trust system for privacy-preserving federated learning, *Journal of Cybersecurity and Intelligent Systems*, 2026, 1(1), 26504, <https://doi.org/10.64189/cis.26504>.

Open Access

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/), which permits the non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as appropriate credit to the original author(s) and the source is given by providing a link to the Creative Commons License and changes need to be indicated if there are any. The images or other third-party material in this article are included in the article's Creative Commons License, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons License and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this License, visit: <https://creativecommons.org/licenses/by-nc/4.0/>

© The Author(s) 2026

Abstract

The increasing use of Artificial Intelligence (AI) in sensitive and regulated domains has raised concerns regarding data privacy, trust, governance, accountability, and fair value distribution. Although federated learning (FL) enables collaborative model training without transferring raw data, conventional FL does not inherently provide decentralized auditability, trust management, or contribution-aware incentives. In this paper, a decentralized data trust system for privacy-preserving federated learning based on a four-layer architecture integrating federated training, privacy protection, blockchain-based governance, and contribution-aware incentives is proposed. The framework uses secure aggregation and differential privacy for privacy protection, blockchain and smart contracts for verifiable governance and auditability, and an approximate Shapley value mechanism for contribution-based reward allocation. The proposed system connects participant activity, trust evaluation, contribution assessment, and reward distribution within a unified workflow. A comparative analysis with conventional FL, blockchain-enabled FL, and data marketplace approaches demonstrates the broader functional coverage of the proposed architecture across privacy, auditability, trust, incentives, and regulatory traceability. As this work is conceptual and architectural, quantitative prototype validation remains part of future work to evaluate model utility, communication overhead, scalability, and incentive-computation cost.

Keywords: Federated learning; Blockchain; Data privacy; Trust management; Incentive mechanism; Shapley value; Healthcare.

1. Introduction

The rapid expansion of artificial intelligence across sensitive and high-impact domains has intensified concerns related to data privacy, trust, and governance. Modern AI systems depend on large-scale datasets, and more than 90% of the global data have been generated in recent years, much of which are managed through centralized infrastructures. These centralized models have become major sources of data breaches and misuse, with billions of personal records compromised globally and the financial impact of major breaches reaching millions of dollars per incident. Regulatory frameworks such as data protection laws now impose penalties of up to 4% of global annual revenue, highlighting the urgent need for stronger control and accountability in AI data usage. Federated learning is a distributed machine learning approach that enables model training without centralizing data. Each participant trains the model locally on its own device or server, and only the resulting model updates, rather than the raw data, are transmitted for aggregation.^[1] While this substantially mitigates privacy risk, it does not eliminate it entirely. Previous work has demonstrated that model updates and gradients can still leak information about the underlying data under certain conditions, indicating that FL alone does not fully resolve the privacy problem.^[2]

A blockchain is a decentralized ledger that records events without the need for a central authority to vouch them. Pairing it with FL gives an audit trail of every training round, who participated, what was submitted, and when, without a trusted third party sitting in the middle.

Data valuation and incentive mechanisms involve measuring how much each participant's data actually contributed to a trained model's performance. The most common approach researchers use is the Shapley value, a concept from cooperative game theory that splits credit on the basis of marginal contribution.^[3] Computing it properly means retraining the model many times, which is expensive; thus, recent work focuses on making it less expensive to approximate.

Regulatory compliance means that an AI system must stay compliant, not just once, as models are retrained and data sources are added.^[4-6] Frameworks such as the GDPR and the EU AI Act are pushing exactly this kind of ongoing traceability and governance requirement.^[4,5]

This research is further grounded in the principles of privacy-preserving AI, distributed AI systems, and secure model collaboration, which are techniques that collectively allow multiple parties to jointly train models without exposing raw data or individual updates.^[1,2] Underlying all of this is the broader goal of trustworthy AI, defined by regulatory bodies as AI that is lawful, ethical, and technically robust.^[4] Central to the proposed framework is the principle of data ownership, ensuring

that each participant retains full legal and operational control over its own data even while contributing to a shared model, an idea this research formalizes through the proposed decentralized data trust system.

Without integrated trust mechanisms that tie these pieces together, federated training, blockchain-based governance, fair incentive computation, and continuous compliance, uncontrolled AI data ecosystems risk increasing privacy violations, regulatory non-compliance, and loss of public trust. These risks are what this research is aimed at addressing.

Key challenges in AI data ecosystems:

- Privacy exposure: Centralized data storage increases the scale and impact of data breaches.
- Limited transparency: Data owners lack visibility and control over downstream data usage.
- Regulatory pressure: Continuous compliance and auditability are difficult to enforce in current AI systems. Unfair value distribution: Data contributors rarely receive compensation proportional to their contribution.
- Centralization risks: Single points of control create systemic security and governance vulnerabilities.

These five factors, spanning privacy, transparency, regulation, fairness, and governance, are summarized in Fig. 1.

2. Literature survey

2.1 Literature review

Recent research has increasingly combined federated learning (FL) with blockchain, privacy-preserving techniques, trust management, and incentive mechanisms.^[7-13] Studies have shown that blockchain can improve decentralization, auditability, and participant accountability while introducing challenges related to scalability, communication overhead, privacy, and computational cost.^[12,13]

Privacy preservation remains an important concern because model updates may still reveal sensitive information. Differential privacy, secure aggregation, homomorphic encryption, and secure multiparty computation have therefore been explored to increase FL privacy.^[14] However, these techniques alone do not provide decentralized governance, contribution tracking, or fair value distribution.

Contribution valuation and incentive mechanisms have also received increasing attention.^[15-23] Shapley value-based methods estimate participant contributions, but exact computation becomes expensive as the number of participants increases. Recent approaches, such as π FL, combine Shapley valuation with secure computation and smart contracts for privacy-preserving incentive distribution.^[24] Similarly, blockchain-enabled incentive-based FL has been investigated in applications such as the Internet of Vehicles, demonstrating the potential for

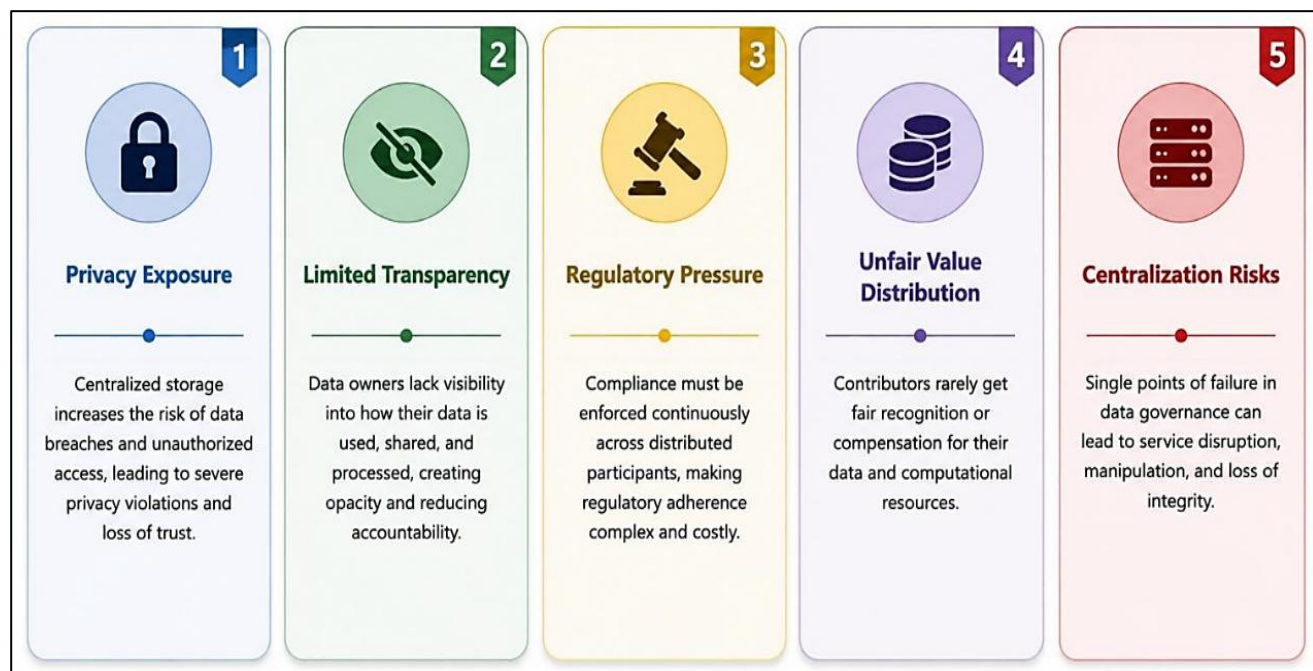


Fig. 1: Key challenges in existing AI data ecosystems.

verifiable coordination while highlighting performance and scalability challenges.^[25] More recent work, including COZO, further addresses storage efficiency, fairness, and contribution valuation in blockchain-enhanced FL.^[26]

Despite these advances, existing approaches generally address privacy, blockchain coordination, trust, or incentives as separate components. A clear gap remains in the integration of privacy preservation, decentralized governance, participant trust, contribution valuation, incentive distribution, and regulatory traceability within a unified FL workflow.

Therefore, this research proposes a decentralized data trust system that connects these functions through a four-layer architecture comprising a data layer, a federated training layer, a trust and governance layer, and an incentive layer. The framework explicitly links participant activity, trust evaluation, contribution assessment, and reward allocation within a common architecture.

2.2 Current flaws identified in existing research

On the basis of prior studies and empirical findings, the following critical flaws are consistently observed:

- **Loss of data control:** Data owners lose oversight once data or model updates are shared.
- **Limited auditability:** Inability to trace how data contributes to model updates over time.
- **Centralization risks:** Reliance on centralized coordinators creates single points of failure.
- **Privacy leakage risks:** Model updates may still leak sensitive information under certain conditions.
- **Unfair value distribution:** Absence of standardized mechanisms to evaluate data contribution.

- **Regulatory fragility:** Difficulty in enforcing continuous compliance across AI lifecycle updates.

These flaws indicate that existing AI systems not only are vulnerable in the present but also pose long-term systemic risks as AI adoption scales further across critical domains. These flaws and their associated risks are summarized in Table 1.

Table 1: Key problems identified in existing AI data ecosystems and their reported impact

Problem	Cause	Risk/Impact (Reported)
Data Breaches	Centralized AI Databases	Billions of records leaked globally → identity theft, fraud
Privacy Misuse	No usage tracking	60%+ orgs can't track secondary data use
Financial Loss	Large-scale breaches	₹80–100 crore+ per major breach (avg)
Trust Loss	Black-box AI decisions	Users stop adopting AI systems
Systemic Failure	Single control authority	One breach affects millions of users

2.3 Problem identification: Current gaps and challenges

Artificial intelligence systems increasingly rely on large-scale and sensitive data to operate effectively, especially in domains such as healthcare, finance, and public services. However, despite rapid AI adoption, data management and governance practices remain largely centralized. With more than 90% of global data generated in recent years, centralized control models have amplified risks related to privacy breaches, misuse, and loss of trust. Existing privacy-preserving and decentralized technologies address individual concerns

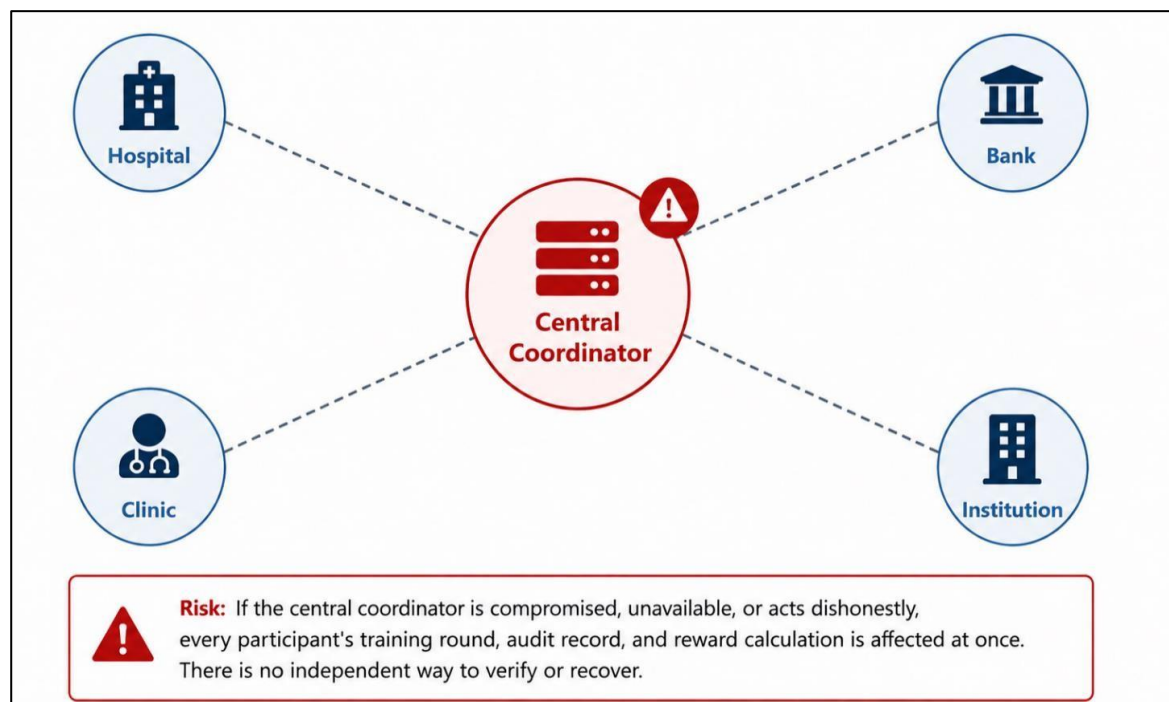


Fig. 2: Single point of failure in centralized coordination.

but fail to provide an integrated framework for trust, accountability, and fairness in collaborative AI ecosystems. As a result, organizations face rising regulatory pressure, financial losses, and reduced willingness among data owners to participate in AI-driven collaboration.

2.4 Critical trust failures in current AI data ecosystems

- Lack of transparent data governance: Once data are used for AI development, data owners lose visibility and control throughout their lifecycle. Studies report that 60%+ organizations cannot track secondary data usage, weakening long-term trust.
- Privacy and regulatory vulnerabilities: Centralized AI infrastructures remain prime targets for breaches, with billions of personal records compromised globally. Continuous regulatory compliance during frequent model updates is difficult to enforce, increasing legal and operational risk.
- Persistent centralization in federated systems
- Despite federated learning, many deployments still rely on centralized coordinators, creating single points of failure and concentrated governance power.
- Residual privacy leakage risks: Research shows that shared model updates and gradients can still leak sensitive information under certain conditions, indicating that local data retention does not guarantee full privacy.
- Security and robustness threats federated learning (FL) systems are vulnerable to model poisoning and backdoor attacks, where malicious participants can

influence global models without immediate detection.^[11]

- Unequal value distribution: The absence of a standardized data contribution measurement results in economic imbalance, where contributors receive little or no reward despite generating AI value.

If these challenges remain unresolved, AI ecosystems risk escalating privacy violations and regulatory failures, economic inequality, and declining public trust, posing serious barriers to the sustainable and ethical growth of artificial intelligence. This risk of centralized failure is illustrated in Fig. 2.

3. Existing solution landscape and limitations

3.1 Existing solutions and industry approaches

- Current direction in industry: The AI industry is gradually moving away from centralized data collection because of privacy risks and regulatory pressure. The focus is on reducing raw data movement while enabling collaborative learning across multiple data owners.
- Federated learning (FL): Federated learning (FL) allows collaborative model training without transferring raw data to a central server. Models are trained locally, and only updates are shared, reducing data exposure while supporting cross-organization collaboration.
- Adoption by major technology firms: Large technology companies have adopted federated and privacy-preserving AI in real-world systems. Google uses federated learning for on-device personalization, Apple applies it to privacy-focused services such as

Siri, Microsoft focuses on confidential computing, and IBM emphasizes AI governance and auditability.

- Government regulations and policy initiatives: Governments have introduced data protection and AI governance laws such as the GDPR and the EU AI Act, emphasizing privacy, transparency, and accountability. Although these laws do not mandate decentralized trust markets, they encourage privacy-preserving and auditable AI systems.
- Cryptographic privacy mechanisms: Secure aggregation protects individual model updates by exposing only aggregated results, while differential privacy limits individual data influence through noise addition. These techniques improve privacy but do not address governance, ownership, or fair compensation.
- Blockchain-enabled federated learning: Blockchain-based federated learning aims to decentralize the coordination and verification of training updates. Although it improves auditability, it increases communication overhead and security complexity.
- Data marketplaces and compute-to-data models: Decentralized data marketplaces promote compute-to-data approaches where computation moves to data sources. However, linking data access to verifiable AI training outcomes remains a challenge.

3.2. Limitations of existing solutions:

- Fragmented solution landscape: Existing approaches address privacy, security, governance, or incentives independently, resulting in siloed systems that lack end-to-end trust enforcement.
- Limited integration with ai workflows: Blockchain-based solutions are mostly restricted to logging and audit trails and are not deeply integrated with AI training and model update pipelines.
- Absence of fair incentive mechanisms: There is no standardized or scalable method to measure data

contribution, making fair and transparent reward distribution impractical.

- Persistent centralization risks: Many federated learning deployments still rely on centralized coordinators, creating single points of failure and concentrated governance power.
- Weak continuous compliance: Current systems struggle to enforce dynamic regulatory compliance across frequent model updates and cross-border collaborations.

3.3 Comparative analysis of existing approaches and the proposed system

Comparison with recent approaches: Recent blockchain-enabled FL studies have increasingly addressed privacy, decentralized coordination, trust, and incentive mechanisms, but these capabilities are often evaluated separately or with application-specific designs.^[12-14,24-26] The proposed system differs in that it integrates privacy protection, blockchain-based governance, trust scoring, contribution valuation, and incentive distribution within a single FL workflow, as summarized in Table 2. Therefore, the comparison highlights the proposed framework's system-level integration rather than claiming superior numerical performance over existing approaches.

4. Proposed solution: decentralized data trust system for federated learning (FL)

The limitations of existing privacy-preserving and decentralized AI approaches highlight the need for a unified, trust-centric framework for data collaboration. To address these challenges, this research proposes a decentralized data trust system for federated learning (FL) that integrates privacy, decentralized governance, trust management, and value exchange within a common architecture.

- Core concept: Decentralized data trust system: A

Table 2: Comparative analysis of conventional federated learning, blockchain-logged federated learning, data marketplace approaches, and the proposed decentralized data trust system across raw data exposure, auditability, contribution-based rewards, privacy leakage resistance, regulatory traceability, trust scoring, coordination overhead, and long-term participation incentives. FL, federated learning; DP, differential privacy.

Evaluation Parameter	Plain Federated Learning	Blockchain-Logged FL	Data Marketplaces	Proposed System
CRRaw Data Exposure	None	None	Depends on platform policy	None
Auditability	Low	High	Medium	High
Contribution-Based Reward	Not Supported	Limited	Manual/Platform-Dependent	Automated (Shapley-Based)
Privacy Leakage Resistance	Medium	Medium	Medium	High (Differential Privacy+ Secure Aggregation)
Regulatory Traceability	Low	Medium	Low	High
Trust Scoring	Not Available	Limited	Reputation-Based	Explicit Trust Model
Coordination Overhead	Low	High	Medium	Medium
Long-Term Participation Incentive	Low	Low	Medium	High

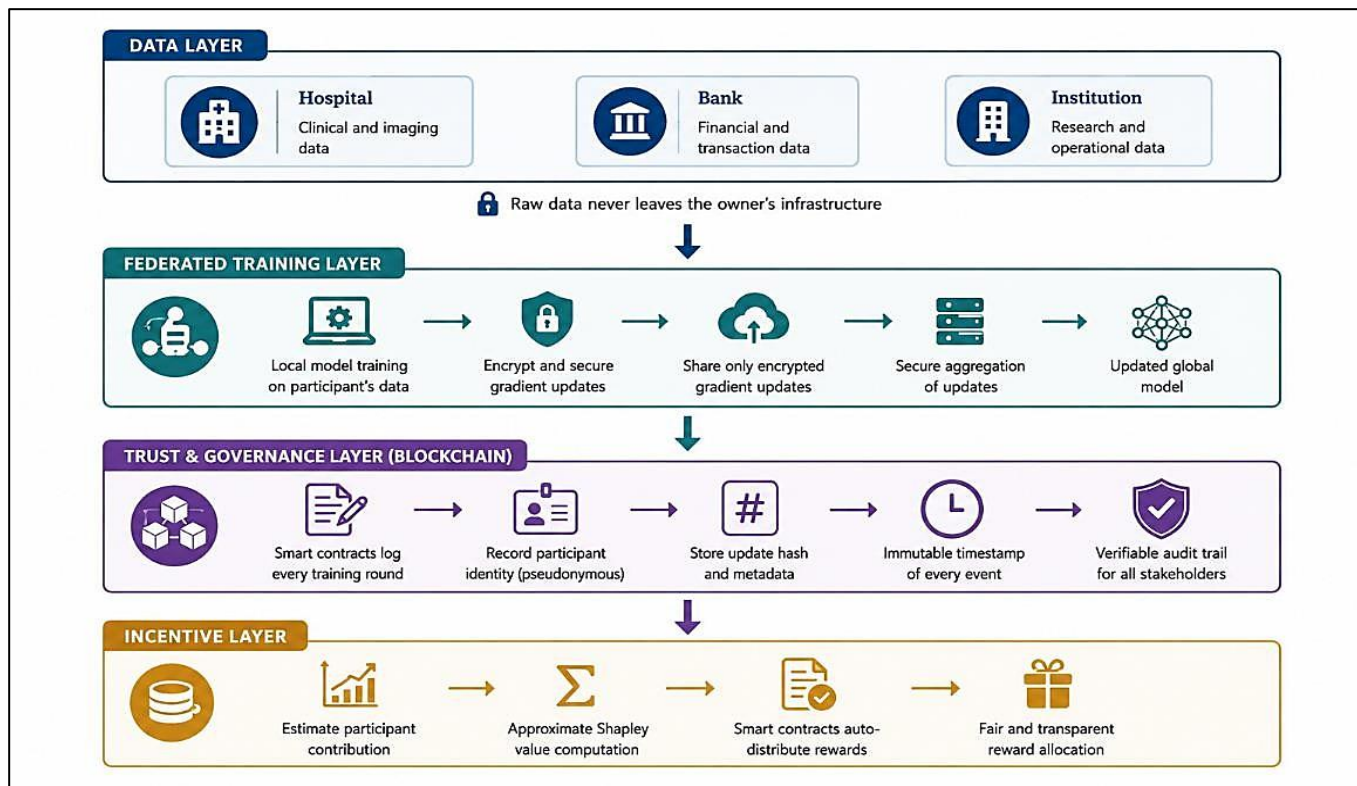


Fig. 3: System architecture of the proposed decentralized data trust system.

decentralized data trust system enables data owners to retain control over their data while AI training is performed locally. Decentralized mechanisms support transparency, accountability, and fair value exchange throughout the AI lifecycle.

- **Role of federated learning (FL):** Federated learning enables collaborative model training without transferring raw data. Participants train models locally and share model updates, supporting privacy-preserving collaboration across organizations.
- **Role of blockchain in trust enforcement:** Blockchain provides a decentralized governance layer through immutable audit trails and smart contracts. It records participation, update-related information, and governance events without requiring centralized control.
- **Fair incentives and value exchange:** The framework links participant rewards to estimated contribution through automated mechanisms, supporting transparent and contribution-aware incentive distribution.

The proposed system consists of four integrated layers. The data layer keeps raw data within the infrastructure of its owner. The federated training layer performs local model training and exchanges model updates without exposing the underlying data. The trust and governance layer uses blockchain and smart contracts to record participation, update hashes, and timestamps, providing a verifiable audit trail. Finally, the incentive layer estimates participant contribution using

an approximate Shapley value and distributes rewards through smart contracts. This architecture connects privacy protection, governance, trust evaluation, contribution assessment, and incentives within a unified FL workflow. The complete four-layer architecture is illustrated in Fig. 3.

4.1 Discussion of the proposed algorithm

Algorithm 1. Proposed federated training procedure with trust and incentive layer input:

Local datasets $D_1, D_2, \dots, D_n$; initial global model M_0

Output:

Updated global model M ; trust scores $\{T_i\}$; participant rewards $\{R_i\}$

- 1: Initialize the global model M_0 .
- 2: while the convergence criterion is not satisfied do
- 3: Broadcast the current global model M to all participants.
- 4: for each participant P_i do
- 5: Train the local model using dataset D_i .
- 6: Apply Differential Privacy:
 $W_i' = W_i + N(0, \sigma^2)$
- 7: Encrypt the protected model update W_i' .
- 8: Compute Hash(W_i').
- 9: Submit the encrypted model update and its hash to the blockchain network.
- 10: end for
- 11: Verify the received model updates using their corresponding blockchain hashes.
- 12: Perform Secure Aggregation to generate the updated global model M .

```

13:   for each participant  $P_i$  do
14:     Compute the contribution score  $C_i$  using
the approximate Shapley value.
15:     Compute the trust score:  $T_i = \alpha C_i + \beta S_i + \gamma H_i$ 
16:     Compute the participant reward:  $R_i = C_i \times R_{Total}$ 
17:   end for
18: Execute the smart contract to distribute rewards.
19: Record model updates, rewards, and consent
information on the blockchain.
20: end while
21: Return the updated global model  $M$ , trust scores  $\{T_i\}$ ,
and rewards  $\{R_i\}$ .

```

Algorithm 1 describes the complete training workflow of the proposed decentralized data trust system, connecting federated learning with privacy protection, blockchain-based governance, trust evaluation, and contribution-aware incentives. The procedure begins with the initialization of the global model M_0 , after which the participating organizations perform local training using their own datasets. The raw data remain within the participants' infrastructure, while only the required model information is prepared for further processing.

Before model updates are shared, privacy-preserving mechanisms are applied to reduce the risk of information leakage. Secure aggregation prevents individual updates from being directly exposed during aggregation, whereas differential privacy can introduce controlled noise to limit the possibility of inferring sensitive information from shared updates. This allows collaborative learning while maintaining the principle that raw participant data do not leave its original infrastructure. The blockchain-based Trust and Governance Layer is then used to record relevant training-round information, including participant identity, update-related hashes, and timestamps. These records provide an immutable and verifiable audit trail. Consequently, participation and training events can be independently verified without storing the underlying sensitive training data on the blockchain. Following each training round, participant behavior and contribution information are used to update the trust scores. A participant who consistently provides valid and useful updates can maintain a higher level of trust, whereas abnormal or low-quality behavior can negatively affect its level of trust assessment. This mechanism provides a continuous trust-management process rather than relying only on initial participant authentication.

The contribution evaluation stage estimates the marginal contribution of each participant using an approximate Shapley value mechanism. The use of an approximation is important because exact Shapley-value computation requires the evaluation of a large number of participant combinations and can therefore become

computationally expensive as the federation grows. The estimated contribution scores are subsequently used by the incentive layer to determine proportional rewards.

Finally, the updated global model, trust scores, and participant rewards are produced as the outputs of the procedure. Smart-contract-based execution can automate the recording of relevant events and the distribution of rewards according to predefined policies. Thus, Algorithm 1 connects the technical training process with governance and economic mechanisms, which is a central feature of the proposed framework. The algorithm also introduces practical trade-offs. Privacy mechanisms may affect model utility depending on the level of protection applied, blockchain operations introduce communication and storage overhead, and contribution valuation adds computational cost. These factors should be quantitatively evaluated during prototype implementation to determine the practical scalability and efficiency of the proposed system.

4.2 Mathematical modeling

The proposed decentralized data trust system for privacy-preserving federated learning is mathematically represented using a set of equations describing global model aggregation, participant contribution evaluation, incentive distribution, privacy preservation, trust computation, and overall system optimization. These formulations provide a conceptual representation of the proposed framework.

4.2.1 Global model aggregation

Assume that there are N participants in the federated learning network. Each participant trains the model locally using its own dataset. The global model is obtained by aggregating the local model parameters according to the size of each participant's dataset.

$$W^{(t+1)} = \sum_{i=1}^N \left(\frac{n_i}{\sum_{k=1}^N n_k} \right) W_i^{(t)} \quad (1)$$

where,

$W_i^{(t)}$ = Local model update from participant

n_i = Number of local training samples of participants

N = Total number of participants

$W^{(t+1)}$ = Updated global model at round $t + 1$

This weighted aggregation gives greater importance to participants contributing larger datasets.

4.2.2 Data contribution score

The contribution of each participant is measured on the basis of its impact on the improvement in global model accuracy.

$$\phi_i = \sum_{S \subseteq \mathcal{N} \setminus \{i\}} \frac{|S|! (N - |S| - 1)!}{N!} [v(S \cup \{i\}) - v(S)] \quad (2)$$

subject to

$$\hat{\phi}_i = \frac{1}{K} \sum_{k=1}^K [v(S_k^i \cup \{i\}) - v(S_k^i)] \quad (3)$$

where,

ϕ_i = Shapley contribution value of participant i

$\mathcal{N}$ = Set of all participants

S = Coalition of participants excluding participant i

$|S|$ = Number of participants in coalition S

$v(S)$ = Utility obtained from coalition S

$v(S \cup \{i\}) - v(S)$ = Marginal contribution of participant i

N = Total number of participants

The approximate Shapley value $\hat{\phi}_i$ is used as the basis for determining the participant contribution score C_i .

4.2.3 Reward distribution

The incentive mechanism distributes rewards according to each participant's contribution score.

$$R_i = C_i \times R_{\text{Total}} \quad (4)$$

where,

R_i = Reward allocated to participant i

C_i = Participant contribution score derived from approximate Shapley evaluation

R_{Total} = Total reward pool

This is intended to support fair and transparent reward distribution proportional to each participant's estimated contribution.

4.2.4 Privacy protection

To reduce the possibility of reconstructing sensitive information from model updates, differential privacy is applied before local parameters are transmitted.^[27]

$$W_i' = W_i + \mathcal{N}(0, \sigma^2) \quad (5)$$

where,

W_i' = Privacy-protected model update

W_i = Original local model update

$\mathcal{N}(0, \sigma^2)$ = Gaussian noise with mean 0 and variance σ^2

σ = Noise parameter

Increasing the value of σ improves privacy protection but may slightly reduce model accuracy.

4.2.5 Trust score computation

Each participant is assigned a trust score on the basis of its contribution, security behavior, and historical reputation.

$$T_i = \alpha C_i + \beta S_i + \gamma H_i \quad (6)$$

subject to

$$\alpha + \beta + \gamma = 1 \quad (7)$$

where,

T_i = Trust score of participant i

C_i = Contribution score

S_i = Security score

H_i = Historical reputation score

α, β , and γ = Weighting coefficients

Participants with higher trust scores are given greater preference during future rounds of collaboration.

4.2.6 Overall objective function

The objective of the proposed framework is to maximize model performance and participant trust while minimizing privacy risk.

$$\max(\lambda_1 A + \lambda_2 T - \lambda_3 P) \quad (8)$$

subject to

$$\lambda_1 + \lambda_2 + \lambda_3 = 1 \quad (9)$$

where,

A = Global model accuracy

T = Overall trust score

P = Privacy risk

$\lambda_1, \lambda_2, \lambda_3$ = Weight coefficients balancing accuracy, trust, and privacy

The optimization objective is to achieve an effective balance between learning performance, trustworthiness, and privacy preservation in decentralized federated learning environments.

4.3 Computational complexity and scalability analysis

Compared with conventional federated learning, the proposed framework introduces additional computational and communication operations. The overall computational cost is associated mainly with local model training, privacy-preserving processing, secure aggregation, blockchain transaction handling, trust score computation, and contribution valuation. Local model training remains the dominant computation at participating institutions and depends primarily on the model architecture, dataset size, and number of local training epochs.

The blockchain layer introduces additional overhead for recording participant-related metadata, update hashes, timestamps, and reward transactions. However, sensitive model updates and raw training data are not stored directly on the blockchain. This design limits blockchain storage requirements to verification and governance-related information rather than complete model or dataset contents.

Contribution valuation represents another important scalability consideration. Exact Shapley-value computation can require evaluating a large number of participant coalitions and therefore becomes computationally expensive as the number of participants increases. The proposed framework consequently uses an approximate Shapley value mechanism to reduce this computational burden while retaining contribution-

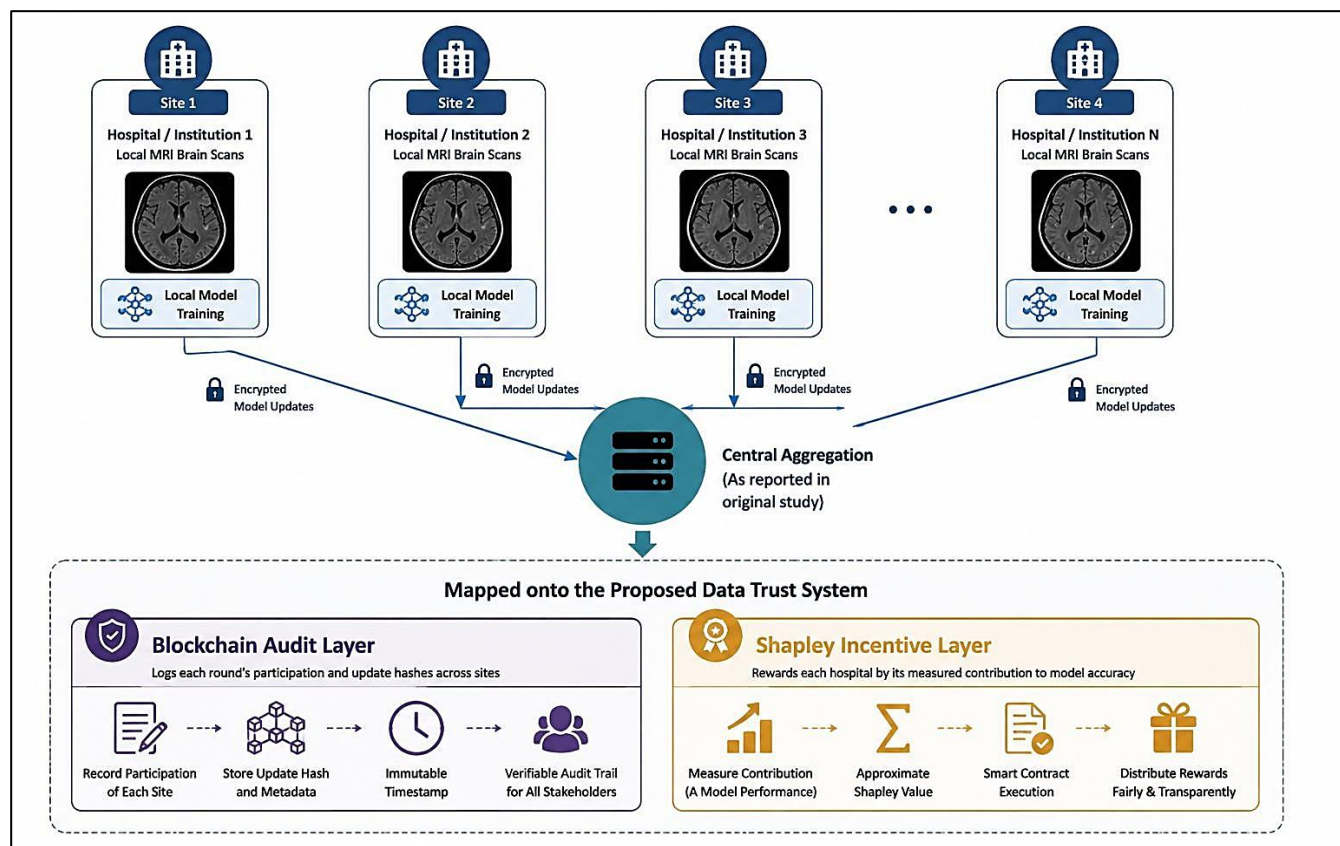


Fig. 4: Case study flow-multi-institution brain tumor segmentation.

aware reward allocation.

The communication overhead is also greater than that of conventional federated learning because the proposed workflow requires additional interactions associated with blockchain verification, privacy-preserving operations, and incentive management. Nevertheless, the decentralized architecture distributes computation across participating institutions rather than relying on a single centralized coordinator. The framework is therefore designed to support scalability through distributed processing, lightweight on-chain metadata, and approximate contribution evaluation.

In practical deployment, scalability should be evaluated by varying the number of participating clients, federation rounds, model update sizes, and levels of data heterogeneity. Important measurements include training time per round, communication volume, blockchain transaction latency, blockchain storage growth, contribution evaluation time, and total system execution time. These measurements provide quantitative evidence for determining the scalability of the proposed framework under realistic federated learning workloads.

5. Case study: Multi-institutional brain tumor segmentation using federated learning

The corresponding workflow is depicted in Fig. 4.

Case study: Multi-institutional brain tumor segmentation (medical imaging)

Sheller *et al.* reported a feasibility study in which federated learning was applied for multi-institutional brain tumor segmentation, enabling deep learning collaboration without the need to share patient data.^[28] Their setup still uses iterative aggregation “at a centralized server,” showing that privacy-preserving training can work while coordination often remains centralized in practice. Follow-up analysis in related federated tumor segmentation research indicates that differences in data distributions across sites can significantly impact federated model performance, illustrating a key limitation for cross-silo real deployments.

Sheller *et al.* reported that federated models trained across four institutions achieved segmentation performance within approximately 1–4% of a centrally trained baseline, despite the added coordination overhead, demonstrating the feasibility of federated learning while preserving data locality.^[19,28]

How this maps to “data trust system” (example flow):

In the proposed decentralized data trust framework, each hospital can define participation terms, including training permissions, model purpose, and privacy constraints, while federated learning enables local training with only model updates and relevant metrics being exchanged. Secure aggregation and differential privacy can be applied to reduce leakage from shared updates, whereas blockchain-based coordination can

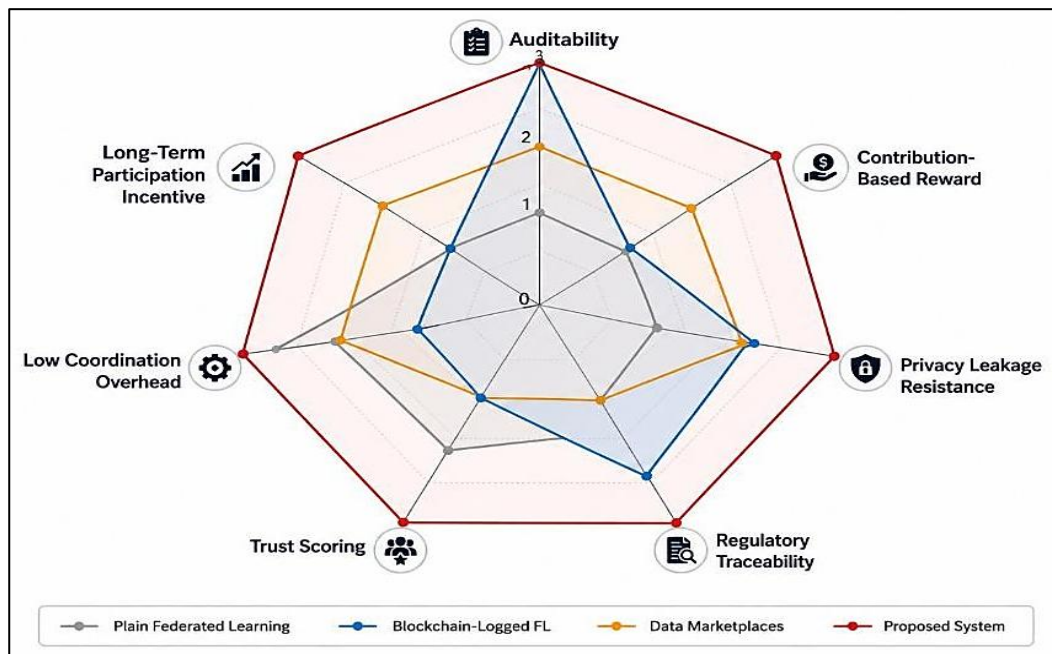


Fig. 5: Comparative analysis of the proposed system and existing approaches.

provide verifiable records of contributions and training-round participation. Incentives can be linked to measurable participant contributions using Shapley value-based valuation. The case study also highlights the practical need to address non-IID data and mitigate adversarial or low-quality contributions in decentralized healthcare environments.

6. Results and discussion

6.1 Market relevance and adoption of trustworthy AI

As artificial intelligence adoption accelerates across industries, trustworthy AI has emerged as a critical market requirement rather than an optional feature. Organizations increasingly demand AI systems that are transparent, reliable, and accountable, particularly in regulated and safety-critical domains. While global investment in AI continues to grow rapidly, a lack of trust remains a major barrier to large-scale deployment. Studies indicate that poorly calibrated trust in AI systems can lead to financial losses, legal liabilities, safety failures, and reduced user acceptance. Consequently, trustworthiness is becoming a key differentiator for enterprise and public-sector AI solutions.

6.2 Applications of trustworthy AI across domains

- Increased financial and operational losses
- Higher regulatory and legal exposure
- Safety risks in high-stakes environments
- Slower adoption of AI-driven systems

These risks highlight the growing market demand for AI systems that embed trust and accountability by design.

6.3 Applications of trustworthy AI across domains

- **Healthcare:** AI systems support diagnosis, medical

imaging, and treatment planning. Over-trusting AI can result in incorrect clinical decisions, while under-trusting reduces efficiency. Trustworthy AI ensures that AI functions as a decision-support system, maintaining clinician oversight.

- **Autonomous vehicles:** Self-driving systems require precise human-AI trust calibration. Misaligned trust has already contributed to real-world accidents. Trust-aware AI mechanisms improve safety and public acceptance.
- **Banking and finance:** AI is widely used for credit scoring, fraud detection, and risk analysis. Trustworthy AI helps professionals understand when automated decisions require manual review, reducing financial and compliance risks.
- **Public services and governance:** AI is being increasingly applied in welfare distribution, policing, and public decision-making. Trustworthy AI is essential for ensuring fairness, accountability, transparency, and ethical use.

6.4 Comparative evaluation and discussion

The comparative evaluation presented in Fig. 5 examines the proposed system against plain federated learning, blockchain-logged federated learning, and data marketplaces across auditability, contribution-based reward, privacy leakage resistance, regulatory traceability, trust scoring, coordination overhead, and long-term participation incentives. The comparison is based on the functional capabilities defined in the proposed architecture and existing solution landscape rather than on experimentally measured benchmark values. Plain-federated learning provides the advantage of keeping raw data local and maintaining relatively low

coordination overhead; however, it does not inherently provide immutable auditability, contribution-based rewards, or explicit trust scoring. Blockchain-logged FL improves auditability by recording training-related events but introduces additional coordination requirements and does not by itself establish a complete mechanism for contribution-based reward allocation. Data marketplace approaches support data valuation and economic exchange, but governance, privacy enforcement, and regulatory traceability may remain dependent on the underlying platform.

The proposed system provides broader functional coverage by integrating privacy preservation, decentralized governance, trust management, contribution valuation, and incentive distribution within a common workflow.

Differential privacy and secure aggregation are used to strengthen protection against privacy leakage, while the blockchain layer provides verifiable records of participation and model-update metadata. The approximate Shapley value mechanism further connects participant contribution with reward allocation, providing a basis for transparent and contribution-aware incentives. These characteristics correspond to the higher capability levels shown for the proposed system in the comparative evaluation.

The comparison also highlights important trade-offs. Compared with conventional federated learning, the additional blockchain, privacy, trust, and contribution-valuation mechanisms can increase computational and communication requirements. In particular, Shapley value computation may become costly as the number of participants increases, which motivates the use of an approximate valuation approach. Similarly, privacy-enhancing mechanisms may introduce a trade-off between privacy protection and model utility. Therefore, the proposed framework should be interpreted as a system-level improvement in governance, accountability, privacy, and fairness rather than as a claim of superior numerical model accuracy.

Overall, the comparative analysis supports the central objective of the proposed framework: integrating functions that are commonly treated separately in existing approaches. The framework connects participant activity, privacy protection, trust evaluation, contribution assessment, auditability, and reward allocation within a unified architecture. However, because the present study is conceptual and architectural, quantitative validation on standard federated learning datasets is needed to measure model accuracy, communication overhead, scalability, and incentive-computation cost in future work.

6.5 Future outlook for decentralized data trust systems and federated data

- **Learning enterprise-scale adoption:** Healthcare, banking, insurance, autonomous systems, and smart cities will increasingly adopt decentralized data trust systems to enable secure, privacy-preserving AI collaboration across organizations.
- **Blockchain-driven AI governance:** Blockchain will evolve from passive logging to active governance, enforcing data usage policies, auditability, contribution tracking, and regulatory compliance through smart contracts.
- **Privacy-enhancing AI techniques:** Advanced federated learning, secure aggregation, and differential privacy mechanisms will mature to further reduce data leakage while maintaining high model performance.
- **Fair data economies:** Data contribution valuation models will enable transparent and incentive-driven data sharing, encourage long-term participation and reduce data monopolization.

6.6 Prototype validation and experimental evaluation plan

The proposed framework is currently presented as a conceptual and architectural system; therefore, the present study does not claim experimental validation or measured performance improvements. To validate the proposed design in future work, a prototype implementation will be developed using standard federated learning datasets and multiple simulated participant nodes. The planned prototype will implement the four proposed layers, including local federated training, privacy protection, blockchain-based governance, and contribution-aware incentive computation.

The experimental evaluation will measure the model utility, privacy-utility trade-offs, communication overhead, blockchain transaction and storage overhead, computational cost of contribution evaluation, trust score behavior, and reward distribution. Standard federated learning performance measures such as accuracy, precision, recall, F1 score, and convergence behavior are used where appropriate for the selected application dataset.

The proposed framework is compared with conventional federated learning, blockchain-enabled federated learning, and other relevant decentralized approaches under comparable experimental conditions. Different numbers of participants and data distributions are considered to examine scalability and the effect of non-IID data. The resulting measurements provide empirical evidence for determining whether the proposed integration of privacy, trust, governance, and incentives provides practical benefits while maintaining acceptable computational and communication costs.

6.7 Quantitative evaluation methodology

The quantitative evaluation will be conducted using standard federated learning datasets and multiple simulated clients under comparable experimental conditions. The proposed framework is evaluated against conventional federated learning and blockchain-enabled federated learning baselines. Model performance will be assessed using accuracy, precision, recall, F1 score, and convergence behavior, while system-level performance will be evaluated using communication overhead, computational cost, blockchain transaction and storage overhead, and contribution evaluation time.

To examine scalability, experiments will be conducted with different numbers of participating clients and varying levels of non-IID data distribution. The effects of privacy-preserving mechanisms will also be evaluated by analyzing the trade-off between model utility and privacy protection. These measurements provide a quantitative basis for assessing the practical benefits and overheads introduced by the proposed decentralized trust, privacy, and incentive mechanisms.

7. Conclusion

This research set out to address a gap that current federated learning systems leave open: privacy alone is not enough without the trust, accountability, and fair incentives built into the system by design. Federated learning solves the data-sharing problem, but on its own, it does not provide mechanisms to verify training activities, maintain accountability, or reward participants according to their contribution. The proposed decentralized data trust system addresses this gap by combining three key components that existing approaches often separate. Federated learning keeps raw data with its owner. A blockchain layer provides an immutable record of training activities, supporting transparency and auditability. A Shapley value-based incentive mechanism links participant rewards to their estimated contribution to model performance, supporting transparent and contribution-aware reward allocation. The literature review and comparative analysis in this paper show that existing solutions generally address privacy, blockchain-based auditability, data valuation, or incentives as separate components. The proposed system is designed to integrate these functions with trust management and regulatory traceability within a unified federated learning workflow. This integration provides a system-level approach to privacy, accountability, fairness, and governance rather than treating these requirements independently. Notably, this work represents a conceptual and architectural contribution rather than an empirically benchmarked system. Future work should therefore focus on prototype implementation and evaluation using standard federated datasets, particularly in domains such as healthcare and

finance. Such evaluation determines whether the proposed integration of privacy, trust, fairness, and regulatory traceability provides practical benefits while maintaining acceptable computational and communication overhead.

Acknowledgment

The authors would like to thank M.H. Saboo Siddik College of Engineering and Kalsekar Technical Campus for providing the necessary academic support and facilities to carry out this research.

CRedit Author Contribution Statement

Mohd Salique Khan: Conceptualization, Methodology, Supervision, Project Administration, Writing – Original draft. **Samiya Khan:** Investigation, Formal Analysis, Methodology, Validation, Writing – Review & editing. **Sania Shaikh:** Investigation, Formal Analysis, Visualization, Validation, Writing – Review & editing. **Luminita Moraru:** Methodology, Formal Analysis, Writing – Review & editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

No data were generated or analysed during the current study. Therefore, data sharing is not applicable to this article.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Artificial Intelligence (AI) Use Disclosure

The authors confirm that no artificial intelligence (AI)-assisted technologies were used in the writing of the manuscript, and no images were generated or manipulated using AI. AI-based tools were used solely for language editing to improve grammar, clarity, and readability, in accordance with journal policy. The authors take full responsibility for the accuracy, originality, and integrity of the work.

Supporting Information

Not applicable.

References:

- [1] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. Arcas, Communication-Efficient Learning of Deep Networks from Decentralized Data, Proceedings of the 20th International

- Conference on Artificial Intelligence and Statistics, PMLR, 2017, **54**, 1273-1282.
- [2] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. QS Quek, H. V. Poor, Federated learning with differential privacy: Algorithms and performance analysis, *IEEE transactions on information forensics and security*, 2020, **15**, 3454-3469, doi: 10.1109/TIFS.2020.2988575.
 - [3] L. S. Shapley, A Value for n-Person Games, in Contributions to the Theory of Games, Princeton University Press, 1953, 307-317.
 - [4] European Union, Artificial intelligence act. European Parliament: European Parliamentary Research Service, 2024, 1689, <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>, Accessed 29 June 2026.
 - [5] European Parliament and Council of the European Union, Regulation (EU) 2023/2854 (Data Act), Official Journal of the European Union, 2023, <https://eur-lex.europa.eu/eli/reg/2023/2854/oj>, Accessed 29 June 2026.
 - [6] J. Kelly, S. A. Zafar, L. Heidemann, J. -V. Zacchi, D. Espinoza, N. Mata, Navigating the EU AI Act: A Methodological Approach to Compliance for Safety-critical Products, IEEE Conference on Artificial Intelligence (CAI), 2024, 979-984, doi: 10.1109/CAI59869.2024.00179.
 - [7] S. K. M. Sameera, S. Nicolazzo, M. Arazzi, A. Nocera, R. R. K. A., V. P., M. Conti, Privacy-Preserving in Blockchain-Based Federated Learning Systems, Computer Communications, 2024, **222**, 38-67, doi: 10.1016/j.comcom.2024.04.024.
 - [8] K. Li, A Blockchain-Integrated Federated Learning Approach for Secure Data Sharing and Privacy Protection in Multi-Device Communication, *Applied Artificial Intelligence*, 2025, **39**, doi: 10.1080/08839514.2024.2442770.
 - [9] X. H. Yang, T. C. Li, A Blockchain-Based Federated Learning Framework for Secure Aggregation and Fair Incentives, Connection Science, 2024, **36**, 2316018, doi: 10.1080/09540091.2024.2316018.
 - [10] Y. Li, C. Chen, N. Liu, H. Huang, Z. Zheng, Q. Yan, A Blockchain-Based Decentralized Federated Learning Framework with Committee Consensus, *IEEE Network*, 2021, **35**, 234-241, doi: 10.1109/MNET.011.2000263.
 - [11] A. Yazdinejad, A. Dehghantanha, H. Karimipour, G. Srivastava, R. M. Parizi, A Robust Privacy-Preserving Federated Learning Model Against Model Poisoning Attacks, *IEEE Transactions on Information Forensics and Security*, 2024, **19**, 6693-6708, doi: 10.1109/TIFS.2024.3420126.
 - [12] J. Liu, C. Chen, Y. Li, L. Sun, Y. Song, J. Zhou, B. Jing, D. Dou, Enhancing Trust and Privacy in Distributed Networks: A Comprehensive Survey on Blockchain-Based Federated Learning, *Knowledge and Information Systems*, 2024, **66**, 4377-4403, doi: 10.1007/s10115-024-02117-3.
 - [13] Y. Tang, Y. Zhang, T. Niu, Z. Li, Z. Zhang, H. Chen, L. Zhang, A Survey on Blockchain-Based Federated Learning: Categorization, Application and Analysis, *Computer Modeling in Engineering & Sciences (CMES)*, 2024, **139**, 2451, doi: 10.32604/cmes.2024.030084.
 - [14] Y. Lu, X. Huang, K. Zhang, S. Maharjan, Y. Zhang, Blockchain and Federated Learning for 5G Beyond, *IEEE Network*, 2021, **35**, 1, 219-225, doi: 10.1109/MNET.011.1900598.
 - [15] P. Guo, Y. Yang, W. Guo, Y. Shen, A Fair Contribution Measurement Method for Federated Learning, *Sensors*, 2024, **24**, 4967, doi: 10.3390/s24154967.
 - [16] H. Zhu, Z. Li, D. Zhong, C. Li, Y. Yuan, Shapley-value-based Contribution Evaluation in Federated Learning: A Survey, 2023 IEEE 3rd International Conference on Digital Twins and Parallel Intelligence (DTPI), 2023, 1-5, doi: 10.1109/DTPI59677.2023.10365410.
 - [17] R. H. L. Sim, X. Xu, B. K. H. Low, Data Valuation in Machine Learning: Ingredients, Strategies, and Open Challenges, in Proc. IJCAI, 2022, 5607-5614.
 - [18] N. Tastan, S. Fares, T. Aremu, S. Horvath, K. Nandakumar, Redefining contributions: Shapley-driven federated learning, arXiv preprint, 2024, doi: 10.48550/arXiv.2406.00569.
 - [19] S. Zheng, Y. Cao, M. Yoshikawa, Secure Shapley Value for Cross-Silo Federated Learning, Proc. VLDB Endowment, 2023, **16**, 1657-1670, doi: 10.14778/3587136.3587141.
 - [20] J. T. Wang, R. Jia, Data Banzhaf: A Robust Data Valuation Framework for Machine Learning, Proceedings of the 26th International Conference on Artificial Intelligence and Statistics, 2023, 6388-6421.
 - [21] Z. Liu, Y. Chen, H. Yu, Y. Liu, L. Cui, GTG-Shapley: Efficient and Accurate Participant Contribution Evaluation in Federated Learning, *ACM Transactions on intelligent Systems and Technology*, 2022, **13**, 21, doi: 10.1145/3501811.
 - [22] Q. Sun, X. Li, J. Zhang, L. Xiong, W. Liu, J. Liu, Z. Qin, K. Ren, ShapleyFL: Robust Federated

- Learning Based on Shapley Value, In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23), Association for Computing Machinery, 2023, 2096–2108, doi: 10.1145/3580305.3599500.
- [23] Q. P. Nguyen, B. K. H. Low, P. Jaillet, Trade-off Between Payoff and Model Rewards in Shapley-Fair Collaborative Machine Learning, in Advances in Neural Information Processing Systems (NeurIPS), 2022, **35**, 30542–30553, doi: 10.52202/068431-2214.
- [24] K. Chen, J. Zhang, X. Liu, Z. Feng, X. Yang, π FL: Private, atomic, incentive mechanism for federated learning based on blockchain, *Blockchain: Research and Applications*, 2025, **6**, 100271, doi: 10.1016/j.bcr.2024.100271.
- [25] F. Zhao, B. Yang, Z. Su, C. Li, Y. Ding, A Blockchain-Enabled Privacy-Preserving and Incentive Mechanism-Driven Federated Learning Scheme for IoV, 2025, *Computer Networks*, **264**, 111262, doi: 10.1016/j.comnet.2025.111262.
- [26] N. Liu, Y. Zhou, Y. Qin, Q. Feng, H. Qi, COZO: A Secure and Efficient Blockchain-Enhanced Federated Learning Paradigm with Optimized Storage and Equitable Contribution Valuation, *Future Generation Computer Systems*, 2026, **179**, 108362, doi: 10.1016/j.future.2025.108362.
- [27] C. Dwork, Differential Privacy, in Proc. ICALP, Springer, 2006, doi: 10.1007/11787006_1.
- [28] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen, S. Bakas, Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations Without Sharing Patient Data, *Scientific Reports*, 2020, **10**, 12598, doi: 10.1038/s41598-020-69250-1.

Publisher Note: The views, statements, and data in all publications solely belong to the authors and contributors. GR Scholastic is not responsible for any injury resulting from the ideas, methods, or products mentioned. GR Scholastic remains neutral with respect to jurisdictional claims in published maps and institutional affiliations.