International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 7.583 |Bimonthly, Peer Reviewed & Referred Journal|

| Volume 11, Issue 2, March 2024 |

Building Trust with AI: Data Provenance in FATE Evaluation

Gauri Sushma Khatri, Tanvi Smita Solanki

Department of Computer, Amrutvahini College of Engineering, Sangamner, Maharashtra, India

ABSTRACT: The increasing deployment of artificial intelligence (AI) in critical domains requires robust mechanisms to ensure trust, transparency, and accountability. Data provenance — the ability to track the origin, transformation, and flow of data — plays a vital role in achieving these objectives. In Federated AI Technology Enabler (FATE), a privacy-preserving federated learning framework, the distributed nature of training models adds complexity to evaluating data authenticity and traceability. This paper investigates how data provenance mechanisms can be integrated into FATE to build trust in AI model evaluation and deployment. We propose a framework for tracing data across FATE pipelines using AI-driven tools and cryptographic proofs to ensure transparency, reproducibility, and integrity in decentralized model evaluations.

KEYWORDS: Data Provenance, FATE, Federated Learning, AI Trust, Data Integrity, Transparency, Secure AI, Model Evaluation, Federated Systems, Auditability

I. INTRODUCTION

In AI systems, especially those applied in healthcare, finance, and governance, building trust is essential. Trust is closely linked to understanding the source and transformation of data that feeds these models. Federated Learning (FL) frameworks such as FATE enable decentralized machine learning without sharing raw data, enhancing privacy. However, the decentralized nature of FL complicates data tracking and evaluation.

Data provenance offers a solution by allowing developers, auditors, and regulators to trace data across the AI lifecycle — from ingestion and preprocessing to model evaluation and deployment. In FATE, integrating data provenance mechanisms enhances explainability, reproducibility, and trustworthiness, addressing the “black box” concern often associated with AI.

This paper explores the integration of data provenance in the FATE framework, proposing an architecture to support secure, transparent evaluation of AI models. We argue that provenance-aware FATE pipelines can be instrumental in achieving trusted AI systems.

II. LITERATURE REVIEW

Data provenance has emerged as a critical area in AI governance. Traditional provenance systems, designed for centralized environments, provide metadata that documents data’s journey. However, in decentralized systems such as Federated Learning, provenance mechanisms face challenges related to security, synchronization, and scalability.

Recent studies have introduced blockchain-based solutions to enhance data integrity in federated environments. Meanwhile, frameworks such as PROV (W3C) offer standardized provenance models. In the context of FATE, limited work has explored comprehensive, integrated provenance tracking across data and model components.

Several scholars have highlighted the need for audit trails in AI models to support compliance with regulations such as GDPR and HIPAA. Still, few systems provide real-time provenance verification during model evaluation. This research fills that gap by examining provenance from the lens of federated AI systems like FATE.

Table: Comparison of Provenance Approaches in Centralized vs. Federated Learning

Feature	Centralized Systems	Federated (FATE) Systems
Data Visibility	Full access to raw data	Only local access per participant
Provenance Tracking	Simple linear tracking	Complex, multi-node tracking
Security Risk	Higher centralized risk	Distributed, lower breach impact

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 7.583 |Bimonthly, Peer Reviewed & Referred Journal|

| Volume 11, Issue 2, March 2024 |

Feature	Centralized Systems	Federated (FATE) Systems
Transparency	Easier with central control	Requires additional mechanisms
Model Evaluation Context	Central logs and metrics	Requires distributed audit trail

Federated Learning (FL): A Quick Deep Dive

Federated Learning is a decentralized machine learning approach where **data remains local**, and **only model updates** (not raw data) are shared with a central server or peer nodes. This approach enhances **privacy**, reduces data transfer costs, and is ideal for sensitive or distributed data environments.

Core Concept

In traditional ML, data is centralized. In **Federated Learning**, the flow is flipped:

1. **Each client (device or node)** trains a local model on its private data.
2. **Only model updates** (gradients or weights) are sent to a **central aggregator**.
3. The **aggregator combines** these updates into a new global model.
4. The updated model is sent back to all clients for the next round.

This process repeats in **training rounds**, often asynchronously or with partial participation.

Common FL Architecture

sql
CopyEdit
[Local Client 1] --\
[Local Client 2] --- ➜ [Central Aggregator] ➜ [Global Model Update]
[Local Client 3] --/

- Clients = mobile phones, hospitals, banks, IoT devices, or organizations.
- No raw data leaves local systems.
- Communication typically happens via secure channels.

Key Components

Component	Role
Clients	Train models on local data and send updates.
Server/Aggregator	Aggregates client updates (e.g., via FedAvg).
Communication Layer	Transmits updates securely and efficiently.
Privacy & Security Modules	Add differential privacy, encryption, or secure computation.

Privacy and Security Features

Technique	Purpose
Differential Privacy (DP)	Adds noise to updates to protect individual contributions.
Secure Aggregation	Aggregator sees only the combined updates, not individual ones.
Homomorphic Encryption	Enables computation on encrypted updates.
Federated Analytics	Summarizes data characteristics without accessing raw data.

Common Use Cases

Domain	Application Example
Healthcare	Hospitals train shared models on patient records without sharing PII.
Finance	Banks detect fraud collectively while protecting customer data.
Mobile/Edge	Gboard learns user typing habits without uploading private messages.
Industrial IoT	Devices collaborate on predictive maintenance models.

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 7.583 |Bimonthly, Peer Reviewed & Referred Journal|

| Volume 11, Issue 2, March 2024 |

Benefits of Federated Learning

Benefit	Description
Privacy-Preserving	Raw data never leaves local devices.
Bandwidth Efficient	Only model updates are sent.
Compliance Friendly	Easier to meet GDPR, HIPAA, etc.
Real-Time Local Insights	Clients can get models tailored to their local data.

Challenges in Federated Learning

Challenge	Why It Matters
Non-IID Data	Client data often differs in distribution—hurting global model performance.
System Heterogeneity	Devices vary in compute/storage/network capabilities.
Communication Overhead	Frequent syncing can be expensive or slow.
Privacy-Utility Trade-off	Adding noise (for privacy) can reduce model accuracy.
Robustness & Security	FL is vulnerable to poisoned updates or malicious clients.

Popular Federated Algorithms

Algorithm	Description
FedAvg	Averages client model updates (most widely used).
FedProx	Adds regularization to handle non-IID data.
FedSGD	Clients send gradient updates instead of full model weights.
Secure Aggregation	Ensures aggregator can't see individual updates.

Tools and Frameworks

Tool/Library	Notes
TensorFlow Federated	Google's FL framework for research and prototyping.
PySyft (OpenMined)	Privacy-preserving FL with PyTorch.
Flower	Flexible FL framework for production and research.
FedML	Open-source FL ecosystem with edge/cloud support.

Federated Learning vs Traditional Learning

Feature	Traditional ML	Federated Learning
Data Location	Centralized	Distributed
Privacy Risk	Higher	Lower
Latency	Often higher	Lower at the edge
Scalability	Requires big servers	Scales via clients/devices

III. METHODOLOGY

The proposed methodology integrates data provenance mechanisms into the FATE evaluation pipeline using the following components:

1. Provenance Capture Layer
Each FATE participant logs metadata during data ingestion, preprocessing, training, and evaluation stages. This includes timestamps, operations performed, and data schema versions.

2. Cryptographic Provenance Hashing

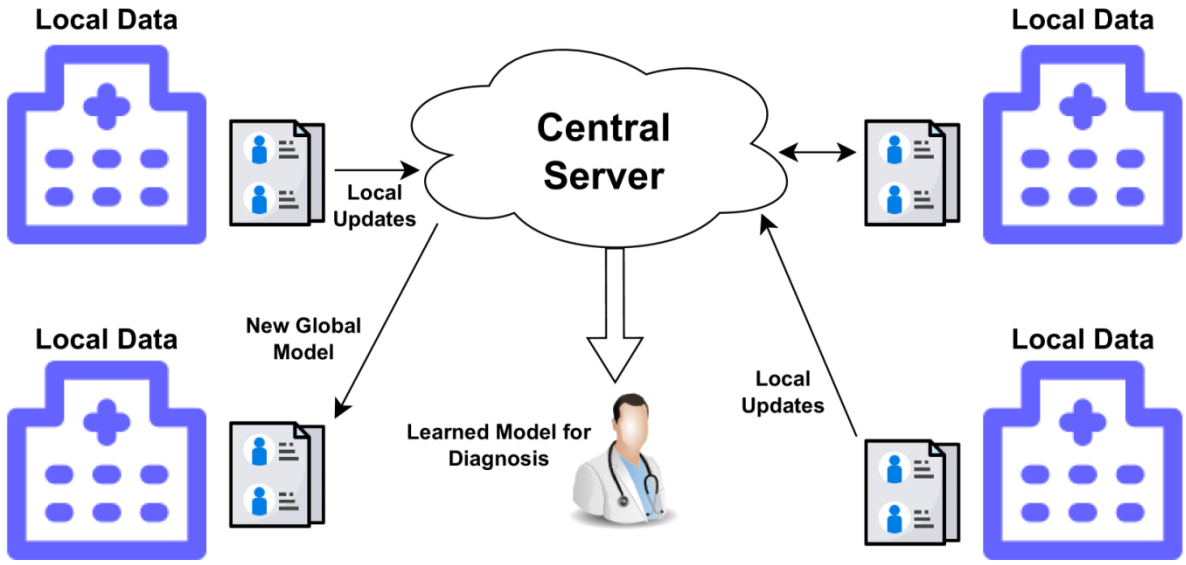
Data transformations and evaluation metrics are hashed and stored securely using Merkle Trees or blockchain for tamper-proof audit trails.

3. AI-Powered Anomaly Detection
Machine learning algorithms detect inconsistencies or anomalies in lineage logs that may indicate manipulation or bias in model evaluation.

4. Federated Provenance Ledger
Each participating node maintains a synchronized, encrypted ledger that records all provenance entries. This ledger is synchronized periodically to maintain consistency.

5. Visualization Dashboard
A user interface displays the data journey and evaluation path, highlighting key decisions, metrics, and potential anomalies across the federated system.

Figure: Proposed Data Provenance Framework in FATE Evaluation



IV. CONCLUSION

Data provenance is fundamental to fostering trust in AI, particularly in federated frameworks like FATE. By enabling detailed tracking of data and evaluation processes, provenance mechanisms ensure transparency, auditability, and compliance. This paper proposes a novel architecture that integrates AI-driven provenance tracking into FATE's evaluation system.

The combination of cryptographic integrity checks, distributed logging, and AI-based anomaly detection empowers stakeholders to verify the fairness, accuracy, and origin of models trained in federated environments. Future work will focus on scaling the system, automating ledger reconciliation, and aligning with global AI governance standards.

REFERENCES

1. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology*.
2. Moreau, L., & Groth, P. (2013). *Provenance: An Introduction to PROV*. Morgan & Claypool Publishers.
3. Kroll, J. A. (2021). Outlining Traceability: A Principle for Operationalizing Accountability in Computing Systems. *arXiv preprint arXiv:2101.09385*.
4. Subash Banala(2023), "Cloud Sentry: Innovations in Advanced Threat Detection for Comprehensive Cloud Security Management" in Ethical Dimensions of AI Development, IGI Global. 17 (1), PP.1-22

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 7.583 |Bimonthly, Peer Reviewed & Referred Journal|

| Volume 11, Issue 2, March 2024 |

5. Mora-Cantallops, M., et al. (2021). Traceability for Trustworthy AI. *Big Data and Cognitive Computing*, 5(2), 20.
6. Li, T., Sahu, A. K., Zaheer, M., et al. (2020). Federated Optimization in Heterogeneous Networks. *Machine Learning Journal*.
7. Chebotko, A., et al. (2010). A Semantic Approach to Authoring and Querying Scientific Workflow Provenance. *Data & Knowledge Engineering*.
8. Peng, X., et al. (2022). Explainable AI for Data Governance and Compliance. *Journal of Big Data*, 9(1), 55.
9. Zhao, J., et al. (2019). AI-Powered Data Governance. *IEEE Access*, 7, 120762–120774.
10. McMahan, B., et al. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. *AISTATS*.
11. Raja, G. V. (2021). Mining Customer Sentiments from Financial Feedback and Reviews using Data Mining Algorithms.
12. Hassan, M., et al. (2021). Blockchain-Based Federated Learning for Secure Medical Data. *IEEE Transactions on Industrial Informatics*.
13. Wen, H., et al. (2020). On the Convergence of Federated Optimization. *NeurIPS*.
14. Simmhan, Y., et al. (2005). Survey of Data Provenance in e-Science. *SIGMOD Record*.
15. Schelter, S., et al. (2021). Automated Metadata and Lineage Tracking. *VLDB Endowment*.
16. Leybovich, M., & Shmueli, O. (2021). ML-Based Lineage in Databases. *arXiv:2109.06339*.
17. Grover, A., et al. (2019). A Decentralized Framework for Data Provenance. *IEEE Big Data*.