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 3, May 2024 |

Provenance-Aware AI: Strengthening Fairness, Accountability, Transparency, and Ethics in Data Tracking

Rahul Kumar Sharma, Amit Kumar Verma, Sanjay Prasad Yadav

Dept. of Computer Science & Applications, Baderia Global Institute of Engineering & Management, Jabalpur, India

ABSTRACT: As artificial intelligence systems grow in complexity and influence, ensuring the ethical handling of data becomes imperative. In federated learning frameworks like FATE (Federated AI Technology Enabler), where data remains decentralized, tracking its lifecycle ethically and accurately is a significant challenge. Data provenance — the documentation of data's origin, movement, and transformation — provides a foundation for transparency and accountability. This paper explores how AI enhances data provenance within FATE, supporting ethical data tracking through intelligent lineage mapping, automated metadata generation, and privacy-aware monitoring. We propose an AI-powered provenance architecture tailored for federated systems, enabling stakeholders to verify data integrity, model fairness, and compliance in decentralized environments.

KEYWORDS: Data Provenance, Ethical AI, Federated Learning, FATE Framework, AI Governance, Privacy-Preserving AI, Data Traceability, Accountability, Metadata Management, AI Ethics

I. INTRODUCTION

Artificial intelligence (AI) is increasingly shaping decision-making in sectors like healthcare, finance, and public policy. To maintain ethical standards, it is essential to ensure transparency in how data is collected, processed, and used within these systems. Federated learning (FL), especially frameworks like FATE, provides a privacy-preserving mechanism for training AI models collaboratively across institutions without sharing raw data.

While FL addresses privacy concerns, it complicates data tracking, making it harder to establish how data has influenced model behavior. Data provenance — the record of data's journey — offers a way to understand, verify, and audit data usage ethically. When enhanced by AI, provenance tracking can become intelligent, adaptive, and scalable, helping to build trustworthy federated AI systems.

This paper proposes a novel framework integrating AI into provenance tracking within FATE, enabling ethical oversight, audit readiness, and compliance in federated environments.

II. LITERATURE REVIEW

Federated Learning and FATE: FATE enables decentralized model training, preserving data privacy by keeping data localized. It supports secure computation methods like homomorphic encryption and differential privacy.

Data Provenance: Traditional provenance systems, like W3C PROV, record metadata about data origin, movement, and transformation. In centralized systems, such records support reproducibility and compliance but face challenges in distributed environments.

Ethical AI and Data Traceability: Ethical AI frameworks emphasize explainability, fairness, and accountability. Provenance supports these principles by offering transparency into data sources and usage. AI enhances this process by automating provenance capture and analysis across complex pipelines.

AI in Provenance: Machine learning and NLP techniques have been used to extract provenance from logs, SQL queries, and configuration files. Graph neural networks and anomaly detection are applied to lineage graphs to monitor and verify data behavior over time.

Table: Ethics-Supporting Features Enabled by Provenance in FATE

Ethical Concern	Provenance Contribution	AI Enhancement
Data Ownership	Track origin and usage rights	Auto-tagging via NLP and ML
Bias and Fairness	Trace biased datasets and transformations	Anomaly detection in lineage graphs
Reproducibility	Record data flow and transformation sequences	AI-based lineage reconstruction
Privacy Assurance	Identify access points and privacy layers	Real-time privacy risk alerts
Regulatory Compliance	Provide auditable records for GDPR, HIPAA, etc.	Automated compliance validation

Ethics-Supporting Features Enabled by Provenance in FATE
Data provenance—the detailed history of data's origin, transformation, and usage—plays a vital role in enabling ethical AI systems through the **FATE** framework: **Fairness, Accountability, Transparency, and Explainability**. Below is a breakdown of **specific ethical features** that provenance supports in each FATE dimension.

1. Fairness
Goal: Ensure data and models are free from unjust bias and treat all groups equitably.

Provenance-Enabled Features:

Feature	Description
Bias Source Tracing	Tracks dataset origin and demographic makeup to identify underrepresented or overrepresented groups.
Feature Sensitivity Audits	Helps detect if sensitive attributes (e.g. race, gender) influenced model decisions.
Training Data Diversity Check	Verifies that datasets were drawn from varied and fair sources.
Sampling Method Transparency	Reveals how data was collected, ensuring it wasn't biased toward specific cohorts.

2. Accountability
Goal: Enable traceability and responsibility for data and model decisions.

Provenance-Enabled Features:

Feature	Description
Audit Trails	Record of who accessed, modified, or approved datasets and models.
Model Version Tracking	Tracks which data and transformations produced each model version.
Responsibility Attribution	Associates model decisions or data pipelines with specific contributors or approvers.
Incident Backtracing	Supports root cause analysis in case of model harm or failure.

3. Transparency
Goal: Make the data, logic, and decision-making processes visible and understandable.

Provenance-Enabled Features:

Feature	Description
Data Lineage Visualization	Shows how raw data flows and transforms through pipelines into features or outcomes.
Transformation Documentation	Captures the logic and reasoning behind data wrangling steps.
Policy Enforcement Visibility	Shows when and where compliance rules (e.g., masking, retention) were applied.

Feature	Description
Input–Output Traceability	Links model outputs back to specific inputs and processing steps.

4. Explainability
Goal: Enable users and stakeholders to understand **why** a model made a particular decision.

Provenance-Enabled Features:

Feature	Description
Feature Provenance Tracing	Shows how each input feature was derived and from which original data.
Model Output Justification	Connects model decisions to data characteristics and transformations.
Temporal Provenance Analysis	Explains how changes in data over time influenced predictions.
Explanation Support Metadata	Enriches SHAP/LIME outputs with deeper context from data origins.

Summary Table

FATE Dimension	Ethics-Supporting Feature via Provenance
Fairness	Bias tracing, demographic audits, data source validation
Accountability	Audit logs, model versioning, contributor attribution
Transparency	Lineage visualization, transformation clarity, policy enforcement
Explainability	Feature origin tracing, decision rationalization, time-aware insights

Provenance is like the **ethical memory** of your AI system—it remembers where data came from, how it was handled, and how it shaped decisions. By embedding this into systems, FATE goals become **measurable, auditable, and actionable**.

III. METHODOLOGY

We propose a multi-layered architecture for AI-enhanced data provenance within FATE, consisting of the following components:

1. Provenance Capture Agent (PCA):
- Integrated at each participant node in the FATE framework.
- Automatically records metadata on data ingestion, transformation, training, and evaluation steps.

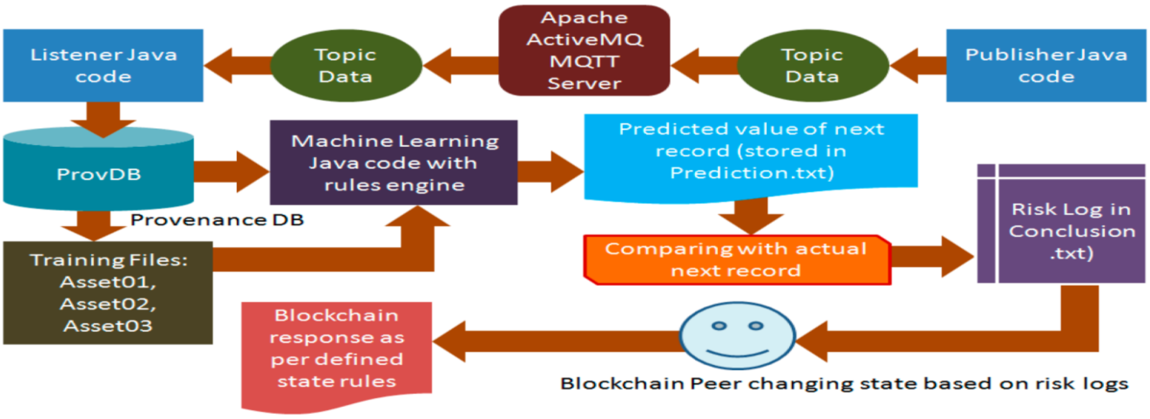
2. AI-Driven Metadata Processor:
- Uses NLP to analyze logs and scripts.
- ML models classify and enrich metadata with contextual labels (e.g., PII, medical codes, etc.).

3. Federated Provenance Graph Builder (FPG):
- Constructs a graph of data flows across all FATE nodes.
- Nodes represent datasets, processes, or models; edges indicate data transformation or movement.

4. Anomaly Detection and Ethics Monitor (ADEM):
- Detects ethical violations like unauthorized access, biased transformations, or data leakage.
- Leverages graph neural networks (GNNs) and reinforcement learning.

5. Visualization and Audit Interface:
- Provides stakeholders with a real-time, interactive view of data provenance.
- Highlights risks and compliance metrics.

Figure: Provenance-Powered FATE Framework


Listener Java code
Topic Data
Apache ActiveMQ MQTT Server
Topic Data
Publisher Java code
ProvDB
Provenance DB
Machine Learning Java code with rules engine
Predicted value of next record (stored in Prediction.txt)
Training Files: Asset01, Asset02, Asset03
Comparing with actual next record
Risk Log in Conclusion .txt)
Blockchain response as per defined state rules
Blockchain Peer changing state based on risk logs

IV. CONCLUSION

In federated learning systems like FATE, ethical data tracking is a critical concern that cannot be met with traditional, static lineage methods. AI-powered data provenance offers a scalable, intelligent, and adaptive solution for ensuring that data handling aligns with ethical standards and regulatory requirements.

This paper introduces a framework that leverages AI to automate provenance tracking, detect anomalies, and visualize data journeys within FATE. By incorporating ethical monitoring into federated environments, organizations can foster transparency, trust, and accountability in AI systems.

Future research will focus on integrating blockchain for immutable logging, extending the framework to model provenance, and evaluating the impact of real-time ethical alerts on model governance.

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.
3. Thulasiram Prasad, Pasam (2023). Strategies For Legacy Insurance Systems Through Ai And Cloud Integration: A Study For Transitioning Mainframe Workload To Azure And Ai Solution. International Journal of Engineering and Science Research 13 (2):204-211.
4. Zhao, J., et al. (2019). AI-Powered Data Governance. *IEEE Access*, 7, 120762–120774.
5. Li, W., et al. (2021). Data Lineage with Deep Learning. *VLDB Endowment*, 14(11), 2426–2439.
6. Hofmann, F. (2020). Tracer: ML Approach to Data Lineage. *MIT Thesis*.
7. Hassan, M., et al. (2021). Blockchain-Enhanced Federated Learning. *IEEE Internet of Things Journal*.
8. Kroll, J. A. (2021). Traceability for Operationalizing Accountability. *arXiv:2101.09385*.
9. Shokri, R., et al. (2015). Privacy-Preserving Deep Learning. *CCS*.
10. Simmhan, Y., Plale, B., & Gannon, D. (2005). Survey of Data Provenance. *SIGMOD Record*.
11. McMahan, B., et al. (2017). Communication-Efficient Learning. *AISTATS*.
12. Chebotko, A., et al. (2010). Semantic Approach to Scientific Workflow Provenance. *Data & Knowledge Engineering*.
13. Thirunagalingam, A. (2023). Improving Automated Data Annotation with Self-Supervised Learning: A Pathway to Robust AI Models Vol. 7, No. 7,(2023) ITAI. International Transactions in Artificial Intelligence, 7(7).
14. Schelter, S., et al. (2021). Metadata and Lineage Automation. *VLDB Proceedings*.