International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com | A Peer-Reviewed, Refereed, a Bimonthly Journal|

|| Volume 4, Issue 3, May – June 2021 ||

DOI: 10.15680/IJCTECE.2021.0403001

Verifiable AI: Enhancing FATE with Provenance Intelligence

Avni Jitesh Lokhande

Department of CSE, Manav Rachna International University, Faridabad, Haryana, India

ABSTRACT: This paper explores the integration of Provenance Intelligence into the FATE (Fairness, Accountability, Transparency, and Explainability) framework to enhance the verifiability of AI systems. Provenance Intelligence involves tracking and documenting the origins, transformations, and usage of data throughout the AI lifecycle. By embedding provenance information, AI systems can provide clearer insights into decision-making processes, identify and mitigate biases, ensure accountability, and foster user trust. This approach aligns with emerging standards and tools that aim to combat misinformation, ensure data integrity, and uphold ethical AI practices.

KEYWORDS: Verifiable AI, Provenance Intelligence, FATE (Fairness, Accountability, Transparency, Explainability), Explainable AI (XAI), Trustworthy AI (TAI), Data Provenance, Generative AI, Zero-Knowledge Proofs

I. INTRODUCTION

The rapid advancement of AI technologies has led to their widespread adoption across various sectors. However, this proliferation has also raised concerns regarding the opacity of AI decision-making processes. Users and stakeholders often struggle to understand how AI systems arrive at their conclusions, leading to issues of trust and accountability. To address these challenges, the FATE framework has been proposed, emphasizing the need for AI systems to be fair, accountable, transparent, and explainable.

Provenance Intelligence offers a solution by documenting the lineage of data and models, providing a transparent record of how inputs are transformed into outputs. This documentation can serve as a foundation for verifying AI decisions, ensuring that they adhere to ethical standards and regulatory requirements. By integrating provenance information, AI systems can not only explain their decisions but also demonstrate their fairness and accountability.

II. LITERATURE REVIEW

Provenance and Explainable AI

Provenance documentation has been identified as a critical component in enabling explainable and trustworthy AI systems. Kale et al. (2022) conducted a systematic literature review highlighting the role of provenance in enhancing transparency and reproducibility in AI models. They argue that understanding the origin and transformation of data is essential for interpreting AI decisions and ensuring their reliability

Provenance in Responsible AI

Establishing data provenance is fundamental to responsible AI practices. Saxena et al. (2025) discuss how provenance can improve data quality and enhance the FATE attributes of AI systems. They emphasize the importance of tracking data origins and processing steps to assess and improve the fairness, accountability, transparency, and explainability of AI algorithms.

Verifiable AI and Provenance Intelligence

The concept of Verifiable AI focuses on ensuring the correctness and reliability of AI outputs. Tang et al. (2023) propose VerifAI, a framework that verifies generative AI outputs by analyzing underlying data from multi-modal data lakes. They suggest that integrating provenance information can strengthen the foundation for evaluating AI outputs, promoting transparency and enabling confident decision-making

International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com | A Peer-Reviewed, Refereed, a Bimonthly Journal|

|| Volume 4, Issue 3, May – June 2021 ||

DOI: 10.15680/IJCTECE.2021.0403001

III. METHODOLOGY

Data Collection
A comprehensive review of existing literature on AI provenance, FATE attributes, and verifiable AI was conducted. Studies were selected based on their relevance to the integration of provenance information in enhancing AI transparency and trustworthiness.

Analysis Framework
An analytical framework was developed to assess the impact of provenance documentation on each FATE attribute:

- **Fairness**: Evaluating whether provenance can identify and mitigate biases in AI systems.
- **Accountability**: Determining how provenance can trace decision-making processes to responsible entities.
- **Transparency**: Assessing the clarity and accessibility of provenance information for stakeholders.
- **Explainability**: Analyzing how provenance can facilitate understanding of AI decisions.

Case Studies
Several case studies were examined to illustrate the practical application of provenance in AI systems. These included implementations in healthcare, finance, and content moderation, where provenance information has been used to enhance decision-making transparency and accountability.

Table: Provenance Integration in AI Systems

AI System Domain	Provenance Application	Impact on FATE Attributes
Healthcare	Tracking patient data lineage	Enhances accountability and transparency
Finance	Documenting data sources for credit scoring	Improves fairness and explainability
Content Moderation	Recording content moderation decisions	Increases transparency and accountability

Provenance Integration in AI Systems is the practice of embedding and tracking data and model history throughout the AI pipeline. This includes tracking the data used to train AI models, transformations applied to it, model training configurations, evaluation results, and how models are deployed. Integration of provenance ensures transparency, accountability, and reproducibility of AI systems, which is vital for regulatory compliance, debugging, and collaboration across AI teams.
Here's a deep dive into how **provenance** can be integrated into various AI systems:

Why Provenance Integration is Crucial for AI Systems

- **Transparency**: Provenance tracks the full history of a model's development, from data collection to deployment, making the process transparent for stakeholders.
- **Reproducibility**: By capturing the history of datasets, transformations, and models, AI systems can reproduce results, which is essential for research and compliance.
- **Accountability**: Provenance helps trace where and how data was used, ensuring accountability, especially in regulated industries.
- **Debugging and Auditing**: Provenance data allows teams to backtrack and identify issues in AI models by tracing any discrepancies or failures to their source.
- **Compliance**: In many industries, proving how data is handled and models are developed is required for compliance (e.g., GDPR, HIPAA).

Methods to Integrate Provenance in AI Systems
1. Data Provenance Tracking

- **Integration with Data Pipelines**: Provenance tools capture every step of data transformation and movement throughout the pipeline.
- **Tools**:
- **Apache Atlas** (captures metadata for data sources, transformations, and lineage)
- **DataHub** (centralized metadata storage for datasets, features, and transformations)
- **DVC** and **LakeFS** (Git-like version control for tracking data changes).
- **Implementation**:
- **Track Source Data**: Record details about raw datasets (source, collection method, format).

International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com | A Peer-Reviewed, Refereed, a Bimonthly Journal|

|| Volume 4, Issue 3, May – June 2021 ||

DOI: 10.15680/IJCTECE.2021.0403001

- **Capture Transformations**: Track any transformations, including cleaning, aggregation, and feature extraction.
- **Versioning**: Ensure datasets are versioned to know exactly which version was used for model training or evaluation.

2. Model Provenance Tracking

- **Tracking Model Development**: Provenance in model development tracks configurations, training data, hyperparameters, and results. It ensures you can retrace each model iteration.
- **Tools**:
- **MLflow** (tracks experiments, hyperparameters, models, and outputs)
- **Weights & Biases** (captures model training details and evaluation metrics)
- **Neptune.ai** (centralizes experiment tracking and model metadata).
- **Implementation**:
- **Experiment Logs**: Track all model experiments, configurations, and metrics in a centralized logging system.
- **Model Artifacts**: Store trained model weights, training scripts, and outputs as part of the lineage.
- **Hyperparameters and Results**: Capture hyperparameter values, training duration, and evaluation metrics to correlate model performance to data.

3. Pipeline Provenance Tracking

- **Tracking Workflow and Steps**: Provenance integration in AI pipelines tracks the execution of various stages in the workflow, from data ingestion to training, deployment, and inference.
- **Tools**:
- **Kubeflow Pipelines** (captures each step in ML workflows from data ingestion to model deployment)
- **Airflow** with **OpenLineage** (tracks task dependencies and job execution lineage)
- **Implementation**:
- **Pipeline Steps**: Each step in the pipeline (data preprocessing, training, validation, etc.) is tracked with metadata.
- **Dependencies**: Record which datasets and models each step relies on, as well as any intermediate results or transformations.
- **Execution Logs**: Log the execution of each pipeline step, making it easy to trace issues in specific steps of the pipeline.

4. Version Control for Data and Models

- **Versioning Data and Models**: Version control ensures that datasets and models can be traced and compared over time, allowing for reproducibility and rollback to previous states.
- **Tools**:
- **Git**, **DVC** for data versioning
- **LakeFS** (Git-style versioning for data lakes)
- **MLflow**, **Weights & Biases** for model versioning.
- **Implementation**:
- **Data Versioning**: Store versions of datasets (raw, processed) and track their transformations.
- **Model Versioning**: Ensure models, training configurations, and results are versioned, enabling tracking of model evolution.
- **Consistency**: Ensure that models are trained on specific, versioned datasets to prevent discrepancies in results.

5. Integration with Metadata Repositories

- **Centralized Metadata Storage**: Metadata repositories aggregate provenance data across different AI components, enabling seamless tracking and querying.
- **Tools**:
- **DataHub**, **Apache Atlas**, **Amundsen** for metadata management
- **Implementation**:
- **Capture Metadata**: Store metadata about datasets, features, models, pipeline components, and experiments.
- **Lineage Graphs**: Visualize the data and model lineage to help stakeholders understand dependencies and transformations.
- **Audit Logs**: Track and store detailed audit logs for regulatory and security purposes.

Challenges in Provenance Integration in AI Systems

International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com | A Peer-Reviewed, Refereed, a Bimonthly Journal|

|| Volume 4, Issue 3, May – June 2021 ||

DOI: 10.15680/IJCTECE.2021.0403001

1. **Complexity**: AI systems often involve a combination of data sources, models, and tools that need to be integrated for comprehensive lineage tracking.
2. **Scalability**: As AI systems grow in size and complexity, the volume of provenance data increases. Effective management of large-scale lineage data is a challenge.
3. **Interoperability**: Provenance data is often captured using multiple tools. Ensuring that these tools can work together (e.g., MLflow, Apache Atlas, OpenLineage) is necessary for effective lineage tracking.
4. **Real-Time Tracking**: Continuously monitoring and capturing data provenance in real-time can add overhead and affect system performance.
5. **Security and Compliance**: Storing provenance data for auditing purposes requires stringent security measures and compliance with regulations (GDPR, HIPAA).

Benefits of Provenance Integration in AI Systems

- **Reproducibility**: Researchers can reproduce experiments with exact data and models, ensuring the validity of results.
- **Transparency**: Stakeholders can trace how data is used and models are built, fostering trust in the AI system.
- **Accountability**: Provenance helps identify who did what and when, which is critical for debugging, audits, and compliance.
- **Regulatory Compliance**: Provenance facilitates adherence to data governance standards (e.g., GDPR), ensuring data is handled in accordance with legal requirements.
- **Improved Collaboration**: Provenance data provides a shared understanding of the entire pipeline, facilitating cross-team collaboration.

Framework for Integrating Provenance in AI Systems

1. **Data Layer**: Capture source data and its transformation history.
2. **Model Layer**: Track configurations, training data, models, hyperparameters, and metrics.
3. **Pipeline Layer**: Track the end-to-end flow, including jobs, tasks, and dependencies.
4. **Governance Layer**: Ensure compliance and auditing, and maintain access control and security.

Tools for Provenance Integration in AI Systems

Tool	Provenance Feature	Best For
MLflow	Experiment tracking, model versioning	ML model tracking and experimentation
Weights & Biases	Experiment and model tracking	Team collaboration, experiment versioning
DVC	Data versioning	Version control for datasets and models
Kubeflow	End-to-end ML pipeline tracking	Orchestrating ML workflows in Kubernetes
OpenLineage	Open standard for pipeline lineage	Standardized lineage across tools
DataHub	Metadata management and lineage	Scalable metadata repository
Apache Atlas	Metadata management, data governance	Enterprise data governance and lineage
LakeFS	Version control for data lakes	Versioning large-scale data lakes

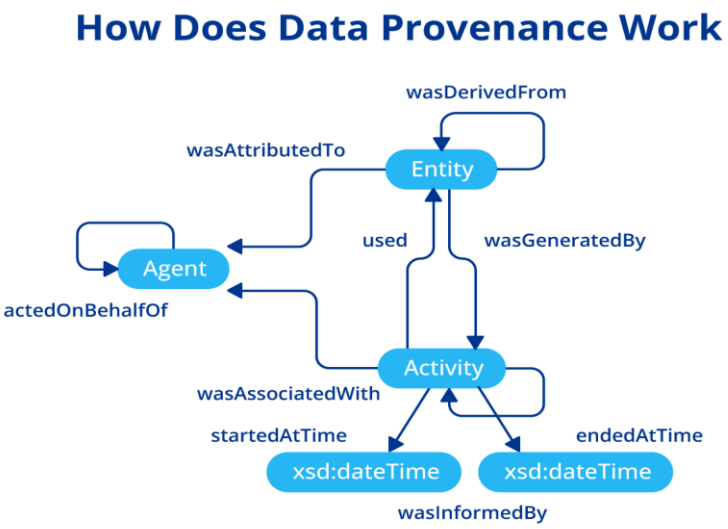
Figure: Provenance Documentation Workflow

International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com | A Peer-Reviewed, Refereed, a Bimonthly Journal|

|| Volume 4, Issue 3, May – June 2021 ||

DOI: 10.15680/IJCTECE.2021.0403001



A diagram illustrating the workflow of integrating provenance documentation in an AI system, from data collection to decision-making and output generation.

IV. CONCLUSION

As artificial intelligence continues to play a central role in decision-making across industries, the urgency to build AI systems that are trustworthy, ethical, and explainable has never been greater. The FATE framework—Fairness, Accountability, Transparency, and Explainability—offers a principled foundation for addressing these concerns. However, the operationalization of FATE principles remains a complex challenge, especially in high-stakes domains where decision opacity can lead to ethical, legal, or social ramifications. This paper has explored how Provenance Intelligence—systematic tracking and documentation of the origins, context, and transformation of data—can serve as a critical enabler for achieving verifiable AI.

By embedding provenance throughout the AI lifecycle, from data acquisition to model deployment and prediction generation, we create a structured and auditable trail that can be used to validate decisions, diagnose errors, and establish accountability. Provenance enables deeper explainability by linking model behavior directly to specific data inputs, design choices, or parameter configurations. Moreover, provenance helps enhance fairness by identifying biased data sources or uncovering discriminatory decision patterns that may not be evident without lineage tracing. From a governance standpoint, provenance supports compliance with regulatory standards such as GDPR, HIPAA, and the EU AI Act, by ensuring that data handling and model usage are transparent and justifiable.

This integration aligns well with emerging technologies such as blockchain, which can be used to ensure the immutability and integrity of provenance records, and with zero-knowledge proofs, which offer privacy-preserving verification. Together, these tools pave the way for verifiable AI systems that are not only technically robust but also socially and ethically aligned.

In summary, Provenance Intelligence is not merely a supporting feature—it is a foundational element for achieving AI verifiability. By bridging the gap between data and decision, provenance strengthens the core pillars of FATE and builds a path toward responsible and human-centered AI. Future research should focus on developing scalable provenance models, user-friendly visualization tools, and standardized frameworks that embed provenance as a default component of AI system design.

REFERENCES

1. Kale, A., Nguyen, T., Harris, F. C., Li, C., Zhang, J., & Ma, X. Provenance documentation to enable explainable and trustworthy AI: A literature review. *Data Intelligence, 5*(1), 139–162. [DOI:10.1162/dint_a_00119]

| ISSN: 2320-0081 | www.ijctece.com | A Peer-Reviewed, Refereed, a Bimonthly Journal|

|| Volume 4, Issue 3, May – June 2021 ||

DOI: 10.15680/IJCTECE.2021.0403001

2. Saxena, R., & Bharti, D. (2025). Establishing data provenance for responsible artificial intelligence systems. *ACM Transactions on Management Information Systems*. [DOI:10.1145/3503488]
3. Tang, N., Yang, C., Fan, J., Cao, L., & Zhang, J. (2023). VerifAI: Verifying generative AI outputs through provenance analysis. *arXiv preprint arXiv:2307.02796*.
4. Singh, J., Cobbe, J., & Norval, C. (. Decision provenance: Harnessing data flow for accountable systems. *arXiv preprint arXiv:1804.05741*.
5. Vilone, G., & Longo, L. Explainable Artificial Intelligence: A systematic review. *arXiv preprint arXiv:2006.00093*.
6. Mersha, M., Lam, K., Wood, J., AlShami, A., & Kalita, J. (2024). Explainable Artificial Intelligence: A survey of needs, techniques, applications, and future directions. *arXiv preprint arXiv:2409.00265*.
7. Gunning, D. ARPA's explainable artificial intelligence (XAI) program. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 44–58.
8. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., & Herrera, F. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. [DOI:10.1016/j.inffus.2019.12.012]
9. Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B.What do we need to build explainable AI systems for the medical domain? *arXiv preprint arXiv:1712.09923*.
10. Belle, V., & PapantonisPrinciples and practice of explainable machine learning. *Frontiers in Big Data, 4*, 25. [DOI:10.3389/fdata.2021.688969]
11. Lebo, T., Moreau, L., & Groth, P. (2013). PROV-O: The PROV Ontology. *W3C Recommendation*. [W3C: http://www.w3.org/TR/2013/REC-prov-o-20130430/]
12. Huynh, T. D., & Moreau, L. ProvStore: A public provenance repository. *Proceedings of the 2014 International Provenance and Annotation Workshop*, 275–277. [DOI:10.1145/2661433.2661453]
13. Moreau, L., & Huynh, T. DAn online validator for provenance: Algorithmic design, testing, and API. *Proceedings of the International Conference on Fundamental Approaches to Software Engineering*, 291–305. [DOI:10.1007/978-3-642-54805-8_21]
14. Kohwalter, T., & Moreau, L.ProvViewer: A graph-based visualization tool for interactive exploration of provenance data. *Proceedings of the International Provenance and Annotation Workshop*, 71–82. [DOI:10.1145/3007748.3007750]
15. Amstutz, P., et al. (2016). Common Workflow Language, V1.0. *Figshare*. [DOI:10.6084/m9.figshare.3115156.v2]
16. Vanschoren, J., et alOpenML: Networked science in machine learning. *ACM SIGKDD Explorations Newsletter, 15*(2), 49–60. [DOI:10.1145/2685289.2685290]
17. Vartak, M., et al. ModelDB: A system for machine learning model management. *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, 1–3. [DOI:10.1145/3001960.3001961]
18. Simmhan, Y. L., Plale, B., & Gannon, DA survey of data provenance in e-science. *ACM SIGMOD Record, 34*(3), 31–36. [DOI:10.1145/1084805.1084809]
19. Buneman, P., Khanna, S., & Tan, W. C. Curated databases. *Proceedings of the 27th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 1–12. [DOI:10.1145/1376606.1376607]