Nuva Labs Bridgetower

Addressing AI's "Black Box" Problem

with Blockchain-Based Solutions for Government and Corporate Entities

Introduction

The rapid advancement of Artificial Intelligence (AI) presents transformative opportunities across various sectors, particularly in government and corporate entities such as financial services, healthcare, legal services, and government. However, the inherent "black box" nature of many AI systems poses significant challenges, hindering adoption and creating risks related to compliance, accountability, and trust.

This white paper explores the key issues associated with AI's lack of end-to-end data transparency in these industries, and introduces Bridgetower's Sovereign AI Data Lineage solution in conjunction with NUVA Labs' BlockVault as a means of addressing these problems, detail how this solution functions, outline its benefits, and conclude with a summary of these key points.

The Problem

AI's "Black Box" in Government and Corporate Entities

The core issue is that many AI systems operate opaquely, with their decision-making processes hidden from users. This lack of transparency, often referred to as the "black box" problem, creates significant obstacles for government and corporate entities, which require clear audit trails and accountability.

Entities seeking to incorporate AI into their offerings face a myriad of challenges, including:

- **Data Security and Integrity:** Government and corporate entities handle sensitive data that must be protected from tampering and unauthorized access. Traditional AI systems do not always provide robust mechanisms for verifying the integrity of the data used to train the model or the data that is used for the AI inference.

- **Lack of Transparency:** AI models, especially complex machine learning models, often produce outputs without detailing how those outputs were derived. This makes it difficult to understand why an AI system made a particular decision, hindering trust and compliance. For example, in healthcare, AI algorithms may generate a diagnosis without showing the specific features of the image that led to that conclusion.

- **Compliance and Legal Issues:** In highly regulated sectors, transparency and auditability are critical for meeting compliance standards. The lack of insight into AI decision-making processes can create legal and compliance issues downstream. For example, financial institutions must be able to verify the origin of funds for regulatory compliance. Without transparent systems, they risk penalties for non-compliance. In the legal system, AI tools are being used for investigations but trade secret laws block public scrutiny of how they work, creating a "black box" hindering AI applications within the criminal justice system.

- **Potential for Bias:** AI systems are trained on data, and if this data contains biases, the AI will perpetuate these biases, leading to unfair or discriminatory outcomes. For example, facial recognition technology has been shown to have higher error rates when identifying people of different races and skin tones. If such biased AI is used in law enforcement or healthcare, it could exacerbate existing inequities.

- **Data Security and Integrity:** Government and corporate entities handle sensitive data that must be protected from tampering and unauthorized access. Traditional AI systems do not always provide robust mechanisms for verifying the integrity of the data used to train the model or the data that is used for the AI inference.

- **Hinderance to AI Adoption:** The "black box" issue has hindered the adoption of AI in these industries because of concerns about the reliability and accountability of AI systems.

In summary, the lack of transparency in AI systems creates a significant barrier to trust and compliance for government and corporate entities. This necessitates a solution that enables clear audit trails, ensures data integrity, and promotes accountability.

The Solution

Bridgetower's Sovereign AI Data Lineage Solution Powered by BlockVault and Provenance Blockchain

Bridgetower's Sovereign AI Data Lineage solution, powered with NUVA Labs' BlockVault, offers a robust and innovative approach to address the "black box" problem by leveraging the power of blockchain technology. This integrated solution is designed to provide transparency, traceability, and auditability for AI systems, enabling their secure and compliant use in government and corporate entities. The solution integrates Generative AI, Retrieval-Augmented Generation (RAG), Confidential Computing, and advanced automation with blockchain-based proof of origin to enhance data processing, reduce errors, and enable secure, traceable data handling.

Key features of this solution include:

- **Blockchain-Based Proof of Origin:** By recording data provenance on a historically immutable blockchain, the solution ensures that all data and AI decisions are auditable and traceable. This helps to establish a clear chain of custody for sensitive data and enables the ability to verify the integrity of data at every stage.

- **Secure Data Handling:** The solution is built on top of AWS infrastructure with robust security mechanisms, such as encryption and access controls. This ensures that data is protected both in transit and at rest and that it is accessed only by authorized personnel.

- **Transparent AI Model Versioning:** The solution logs AI model versions on the blockchain, providing clear documentation of how models evolve and ensuring that AI-driven decisions are traceable to specific model versions.

- **Integration with Generative AI:** Leveraging tools like NUVA Labs' BlockVault and Amazon Bedrock, the solution enhances data processing capabilities while maintaining transparency and auditability. This ensures compliance-friendly and auditable outputs.

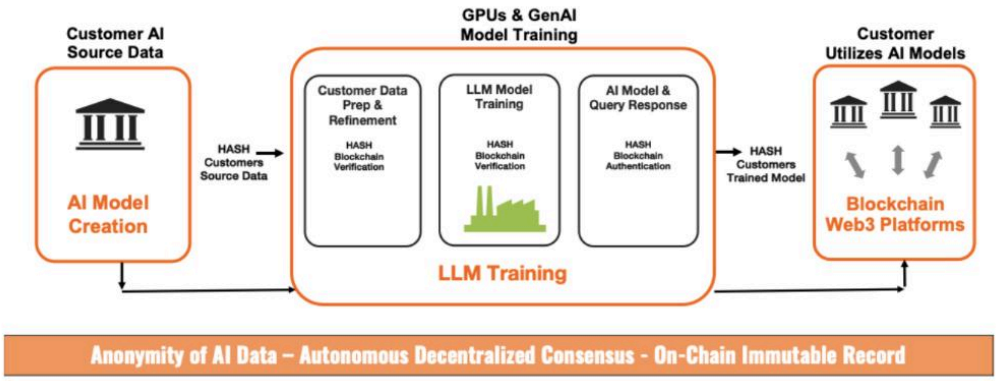
- **Retrieval-Augmented Generation (RAG):** RAG is used to allow for faster document retrieval and search by incorporating a vectorized search layer.

- **Scalability and Adaptability:** As a horizontal solution, this technology is adaptable across various document-heavy industries and is designed to scale to handle increasing data volumes and processing needs.

Deep Dive

How Bridgetower's AI Sovereign Data Solution Works

The Bridgetower solution leverages **Generative AI and blockchain-based proof of origin** to ensure AI sovereignty, robust data lineage, and auditability, delivering transformative workflow improvements across government and corporate entities such as government, healthcare, and finance. Here's how it works:



- **Data Ingestion and Transformation:** The solution is designed to handle data ingestion from on-premises systems, moving it into AWS for secure processing. Each piece of sensitive data is obfuscated and then processed across different AWS S3 stages (Raw, Stage, Processed), providing structured data for downstream tasks. This modular approach mirrors AWS's IDP strategy of preparing and structuring data for analysis, allowing government and finance organizations to process documents while protecting sensitive information.

- **AI Versioning and Real-Time Inference:** Using AWS SageMaker and Amazon Bedrock, the solution supports the fine-tuning, versioning, and inference of AI models, with each version logged immutably on the blockchain. Hash values track these model versions, providing auditability for AI-driven decisions. This ensures that any AI inference or decision within the document processing pipeline is traceable to a specific model version, aligning with IDP requirements for accountability and compliance in AI processing.

- **Data Lineage and Traceability:** By combining blockchain with AWS DynamoDB for record-keeping, the solution maintains immutable logs of data transformations, model fine-tuning steps, and AI decisions. AWS Lambda functions retrieve context and augment prompts, ensuring comprehensive traceability for all stages of document handling. This setup provides an unbroken chain of custody for each document, essential for government and corporate entities where document integrity and traceability are paramount. It aligns with IDP's emphasis on transparent and auditable document data handling.

- **Blockchain:** Leveraging Provenance Blockchain, hashed versions of data, AI models, and their transformations are ledgered on-chain. This blockchain ledger is critical for tamper-proof documentation of every transformation, processing step, and AI decision. Immutable blockchain records ensure that every document processed can be verified back to its original state, with all transformations documented securely. This meets IDP requirements by offering unparalleled transparency, which is essential for compliance and audit trails.

- **Regulatory Compliance and Data Security:** The solution includes a regulatory compliance dashboard for real-time visibility into data lineage, supported by AWS KMS for encryption in transit and at rest. AWS CloudWatch provides continuous monitoring, logging, and alerts for any anomalies in data handling or AI decision-making. This alignment with AWS IDP enables organizations to comply with data protection regulations while securely processing documents. Real-time alerts and compliance tools ensure that every document meets stringent industry standards for data security and privacy.

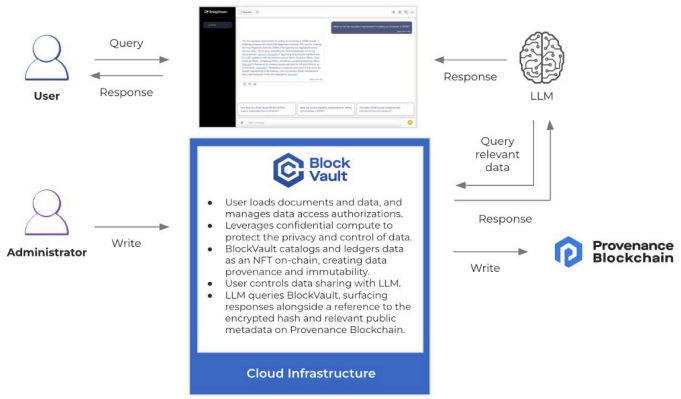
Deep Dive

Role of BlockVault and Provenance Blockchain

Powered by NUVA Labs and available only on Provenance Blockchain, BlockVault is an innovative off-chain system designed to handle large-scale data. The common link between a record ledgered on Provenance Blockchain and the BlockVault data is an unique encrypted identifier (hash-id). The hash-id is an identifier that only authorized users can read.

The NUVA Labs' BlockVault component of the solution provides a user interface for administrators, and a transparent, traceable, and auditable record of immutable truth that's ledgered and referenceable on-chain. In the case of the Sovereign AI Data Lineage solution, BlockVault lives in AWS Cloud and leverages the ProvConnect API to communicate with Provenance Blockchain, the leading layer 1 blockchain for real-world assets and AI, with over $10B in total asset value locked (TVL) and more than $40B in financial transactions supported.

The following diagram and processes outlined illustrate BlockVaults contributions to the Sovereign AI Data Lineage workflow:



- **Data Loading and Cataloging:** The system allows administrators to load documents and data that their Large Language Models (LLM) will need to reference. BlockVault catalogs all of the data loaded by the user.

- **User Authorization and Permissioning:** Administrators can set and manage user authorization and permissioning, ensuring that only authorized users can access secure data.

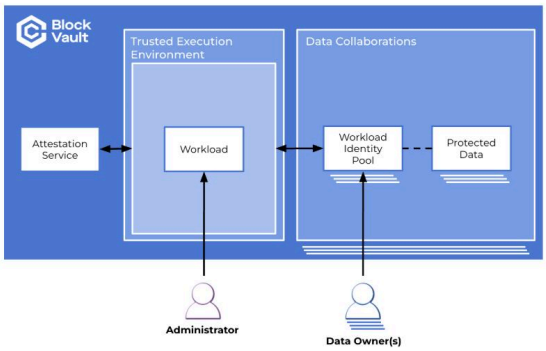
- **User-Controlled Indexing Process:** BlockVault enables a user-controlled indexing process that allows the LLM to read secure data and write responses. This model is shared with the LLM provider.

- **Data Ledgers as NFTs:** BlockVault ledgers the data as a non-fungible token (NFT) on the Provenance Blockchain, creating data provenance and immutability. When ledgered on-chain, a unique encrypted hash-id of the data is created to be used as a reference code. That code is stored in the block on the blockchain and in BlockVault, associated with a specific piece of ledgered data.

- **LLM Queries BlockVault:** When a user queries the AI, the LLM simultaneously queries BlockVault. The LLM formulates a response that includes (a) a source, (b) a reference (hash-id) to the data on-chain and in BlockVault, and (c) any relevant public metadata. The reference can only be used by permissioned users to access the secure data. The end result is a compliant and secure AI infrastructure that meets the needs of legal and compliance teams, and regulators.

The BlockVault system leverages confidential computing to mitigate a wide-range of threats and to ensure data remains encrypted and isolated, and only visible to the owners / administrators and those authorized by the owners / administrators. In other words, BlockVault is like a secure vault that keeps data protected even while you're using it, and where only authorized individuals can see what's inside, and everyone else sees a locked vault door. This is particularly important when working with individuals or companies not fully trusted, but with whom you need to share sensitive information.

Leveraging a model initially developed by the team at Google Cloud, the following diagram illustrates the confidential computing workflow within BlockVault:



Benefits of the Solution

Beyond enabling transparency, auditability, and traceability, Bridgetower's Sovereign AI Data Lineage solution has the potential to significantly improve processes and reduce overhead burdens. Let's take two highly government and corporate entities for example:

Financial Services:

- **Faster Verification:** Blockchain-based proof of origin for funding sources can expedite verification processes by 60%, reducing delays during loan applications, bank transfers, or audits. Financial institutions could verify funding sources within hours instead of days, allowing for faster customer onboarding and service delivery.

- **Reduced Compliance Costs:** By providing automated proof of funds on an immutable ledger, banks can cut compliance costs associated with traditional verification methods by up to 40%, as verified by similar blockchain use cases in finance. Collateral verification processes for non-cash reserves can also be expedited through the use of BlockVault, demonstrating the real-world backing of any blockchain-ledgered asset without disclosing private contents thereof.

- **Enhanced Fraud Detection:** The ability to trace the origin of funds back to verified sources helps mitigate fraud and money laundering risks. This can lead to a reduction in fraud-related losses and/or cost of noncompliance potentially improving financial security by over 50%, based on data from blockchain implementations in anti-money laundering programs.

Government:

- **Time Efficiency:** By automating document provenance tracking, the solution can reduce the time spent on manual verification by up to 70%, as government entities can quickly access authenticated versions of documents, avoiding redundant checks and delays.

- **Accuracy:** With real-time blockchain recording, users access the latest, verified version of documents, reducing errors due to outdated information. This will reduce misinformation by 50%, based on initial estimates in similar public-sector blockchain projects.

- **Transparency and Trust:** Proof of origin provides citizens and businesses with confidence in the information's authenticity, fostering transparency and trust. For the UAE government, this helps in presenting accurate regulatory information to international investors, improving ease of business.

Conclusion

The "black box" problem of AI poses significant challenges to government and corporate entities, hindering adoption and creating compliance risks. **Bridgetower's Sovereign AI Data Lineage solution, powered by NUVA Labs' BlockVault and built on the Provenance Blockchain, offers a comprehensive approach to address these challenges.** By leveraging blockchain technology, the solution provides **immutable data lineage, secure data handling, transparent AI model versioning, and auditability.** This ensures that data and AI systems are not only efficient, but also compliant, accountable, and trustworthy.

Key benefits of this solution include reduced manual effort, improved accuracy, faster document retrieval, lower compliance costs, enhanced fraud detection, and increased transparency and trust, allowing consumers to focus on core business functions, not on whether critical data inputs are trustworthy, biased, or otherwise unreliable. These benefits collectively enable organizations to leverage the power of AI while meeting the stringent regulatory and compliance requirements of their respective industries.

By integrating blockchain with AI, the Bridgetower and NUVA Labs solution sets a new standard for responsible and transparent AI implementation in regulated sectors. This approach promotes greater trust, accountability, and ultimately, a more efficient and reliable future for AI adoption in the most critical areas of our economy and society.