Can the Responsible Use of Generative AI be Governed by Blockchain's Decentralization and Immutability? Test Results on Financial Advisory

line 1: 1st Given Name Surname
line 2: *dept. name of organization*
(of Affiliation)
line 3: *name of organization*
(of Affiliation)
line 4: City, Country

line 1: 2nd Given Name Surname
line 2: *dept. name of organization*
(of Affiliation)
line 3: *name of organization*
(of Affiliation)
line 4: City, Country

***Abstract*— Organizations around the world have cautiously opted to restrict the use of Generative AI tools on workplace computers due to concerns related to data security and human control over AI. This research attempts to find an equilibrium between responsible usage and human professional governance of decisions and outputs generated by Explainable AI and Generative AI tools through automated auditing of AI metadata by blockchain. It presents experimental results on the effectiveness of blockchain-based governance in regulating Generative AI, in the context of providing financial advisory services for small business loans.**

Keywords—Generative AI, Responsible AI, Auditing, Blockchain

I. Introduction

Generative AI tools powered by Large Language Models (LLMs) have demonstrated advanced reasoning and articulation capabilities. The technological leap has raised two questions [1]: Can these tools completely substitute human expertise or specialized eXplainable AI (XAI) algorithms trained to provide explainable decisions for a task? How to address the data security concerns on leakage of sensitive information through user prompts? Due to these unresolved challenges, many firms have cautiously opted to restrict the use of Generative AI tools on workplace computers.

Blockchain is a decentralized technology that allows the permanent (immutable) storage of data within a peer-to-peer computer network. This paper presents the pilot study results on monitoring the usage of XAI and Generative AI by financial advisors in making and drafting small business loan decisions, respectively. The proposed framework audits the metadata of AI-generated decisions and textual content with its cryptographic hash stored in the blockchain.

II. Generative AI Governance by Blockchain

The development of successful blockchain applications requires multiple technology integrations to meet a firm's stakeholder demands for compliance with data protection laws such as the General Data Protection Regulation (GDPR) on "Data Minimization" and "Right to be Forgotten," which mandates the retention of necessary sensitive data and restricts the permanent storage of personal information on the decentralized blockchain network to ensure data confidentiality, respectively. The proposed framework utilizes blockchain's decentralization and immutability feature to detect and resist tampering of AI outcomes by malicious actors which aim to manipulate historical decisions, alter algorithmic parameters, or modify the system architecture to evade accountability tracing.

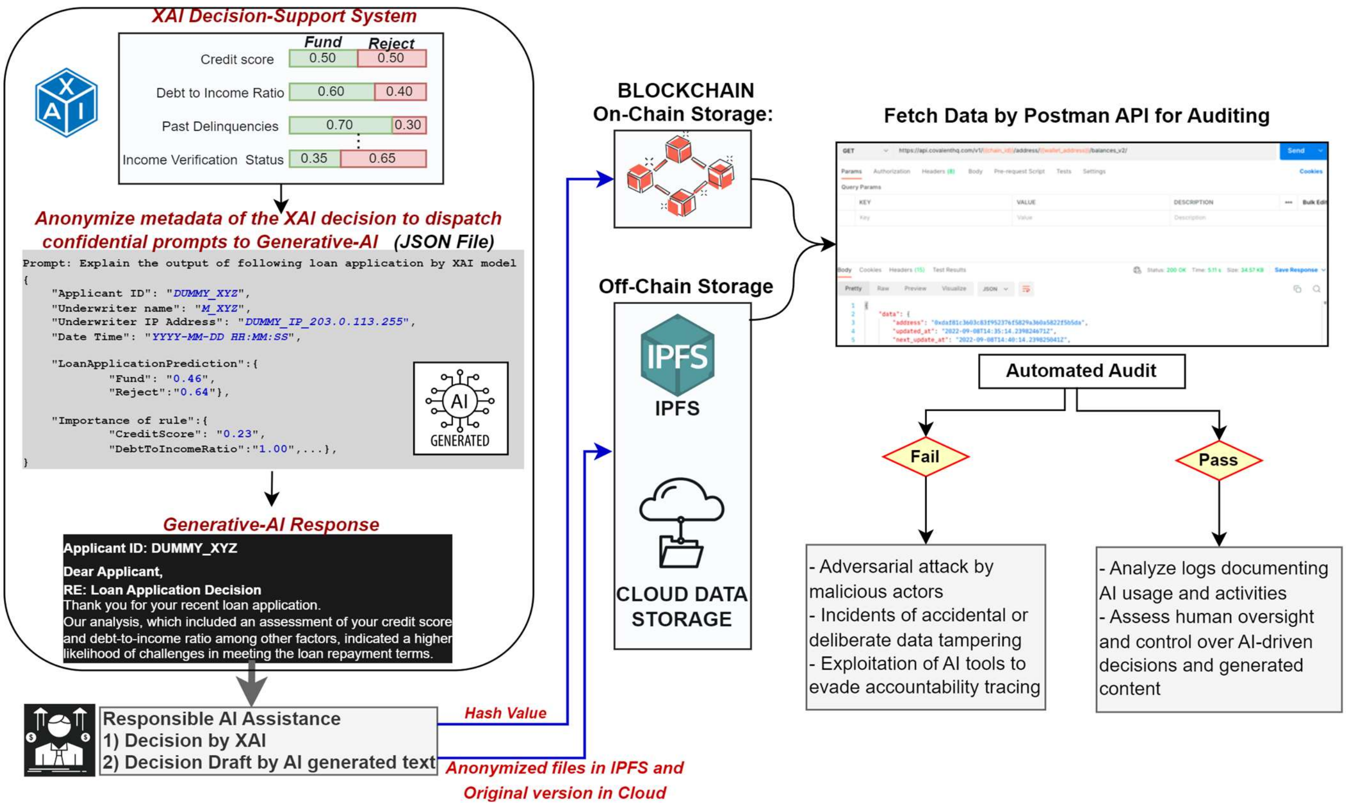
It utilizes a hybrid off-chain and on-chain data storage strategy. A smart contract (or chain code) records the hash, a unique hexadecimal string representing algorithmic decisions by XAI, and content produced by Generative AI tools on a blockchain network. Cloud stores financial documents and AI algorithmic metadata; however, it is frequently accessed by internal stakeholders such as developers and domain experts. It is vulnerable to unintentional tampering and malicious attacks. InterPlanetary File System (IPFS) is paired with Cloud and Blockchain as a scalable solution for storing large off-chain data and recovery of original untampered files if Cloud storage is compromised. It enables robust auditing for data integrity checks and accountability in AI decisions. The belief-rule-base was utilized to generate the explainable decisions to fund or reject loan applications [2]. These outputs are structured as human-readable text in JSON (JavaScript Object Notation) format, as illustrated in Figure 1. To mitigate the risk of data leakage, an anonymized text explanation is concatenated with a fixed prompt statement to send in a Generative AI API to access LLM algorithms such as Bard, gpt-text-davinci-003, and GPT-4 to get a response.

TABLE I. On-Chain and Off-Chain Data Storage

On-Chain: β (Blockchain Network)	*Separate Hash*: Human advisor's ID and the static IP address of the authorized workplace computer. *Combine Hash*: A hashed text file containing the XAI decision, with the corresponding prompt and response sent to the Generative AI.
Off-chain: α	*Cloud*: Financial documents of customers and metadata related to the XAI algorithm.
	IPFS: Anonymized text file XAI decisions and associated content from the Generative AI.
	Key Storage: Blockchain node identifiers, financial advisor and other administrative IDs, and loan application IDs.

AI's metadata contains information about past decisions, generated textual content, identification of human financial experts, and static IP addresses of workplace computers to trace the accountability back to humans for inconsistent adversarial decisions. Any alteration of the metadata and questionable decisions indicate deviations from established ethical standards. The tampering state (τ) indicates tampering if the hash value of specific data types: $x \in \{$Advisor ID, Static IP address, AI Metadata file$\}$ does not match with the immutable hash stored in the blockchain.

$$\tau = \begin{cases} 0, \; h_{\beta}^{x} \neq Cloud\_h_{\alpha}^{x} \; and \; h_{\beta}^{x} \neq IPFS\_h_{\alpha}^{x} \\ 1, \; other\ wise \end{cases} \quad (1)$$


XAI Decision-Support System
Fund
Reject
Credit score 0.50 0.50
Debt to Income Ratio 0.60 0.40
Past Delinquencies 0.70 0.30
Income Verification Status 0.35 0.65
Anonymize metadata of the XAI decision to dispatch confidential prompts to Generative-AI (JSON File)
AI GENERATED
Generative-AI Response
Responsible AI Assistance
1) Decision by XAI
2) Decision Draft by AI generated text
Hash Value
Anonymized files in IPFS and Original version in Cloud
BLOCKCHAIN On-Chain Storage:
Off-Chain Storage
IPFS
IPFS
CLOUD DATA STORAGE
Fetch Data by Postman API for Auditing
GET
Send
Params
Authorization
Headers (8)
Body
Pre-request Script
Tests
Settings
Cookies
Query Params
KEY
VALUE
DESCRIPTION
Bulk Edit
Body
Cookies
Headers (15)
Test Results
Status: 200 OK
Time: 5.11 s
Size: 34.57 KB
Save Response
Pretty
Raw
Preview
Visualize
JSON
Automated Audit
Fail
Pass

```
Prompt: Explain the output of following loan application by XAI model
{
    "Applicant ID": "DUMMY_XYZ",
    "Underwriter name": "M_XYZ",
    "Underwriter IP Address": "DUMMY_IP_203.0.113.255",
    "Date Time": "YYYY-MM-DD HH:MM:SS",

    "LoanApplicationPrediction":{
            "Fund": "0.46",
            "Reject":"0.64"},

    "Importance of rule":{
            "CreditScore": "0.23",
            "DebtToIncomeRatio":"1.00",...},
}
```

Applicant ID: DUMMY_XYZ

Dear Applicant,

RE: Loan Application Decision
Thank you for your recent loan application.
Our analysis, which included an assessment of your credit score and debt-to-income ratio among other factors, indicated a higher likelihood of challenges in meeting the loan repayment terms.

```
https://api.covalenthq.com/v1/{{chain_id}}/address/{{wallet_address}}/balances_v2/
```

```
{
    "data": {
        "address": "0xdaf81c3603c83f952376f5829a360a5822f5b5da",
        "updated_at": "2022-09-08T14:35:14.239824671Z",
        "next_update_at": "2022-09-08T14:40:14.239825041Z",
```

- Adversarial attack by malicious actors
- Incidents of accidental or deliberate data tampering
- Exploitation of AI tools to evade accountability tracing

- Analyze logs documenting AI usage and activities
- Assess human oversight and control over AI-driven decisions and generated content

Fig. 1. Automated Auditing of XAI Decisions and Generative AI Content by Blockchain

III. Test Results

The blockchain-based AI audit was evaluated by two metrics: throughput and latency on a public (Ethereum) and private (Hyperledger Fabric) blockchain network. Throughput measures the number of valid transactions per second (TPS), while latency measures the time taken for a network to broadcast a transaction via a node controlled by an organization. The demand to update information on a blockchain platform escalates with an increase in the number of users and nodes (servers). Figure 2 illustrates the decrease in throughput and increase in latency beyond 10 nodes. In this experiment, the network traffic (number of transactions) was uniformly allocated.

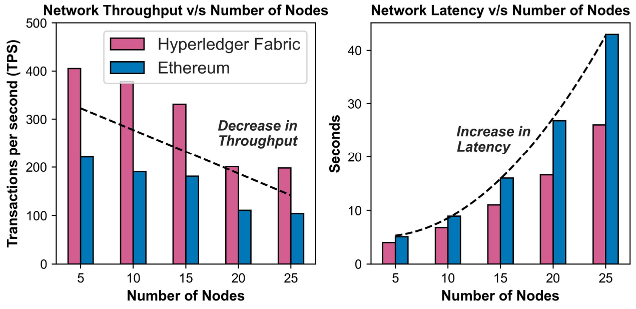


Fig. 2. Comparative Analysis of Blockchain Throughput and Latency

The efficiency of the audit process was assessed by introducing random alterations in 2% to 20% of the files stored in off-chain mediums. The goal was to ensure that the recomputed hash of off-chain storage mediums (IPFS and Cloud) matched their respective hash permanently stored in a blockchain platform (on-chain) to pass the automated audit. The completion time of the audit process grows with the increase in the number of tampered files in both Ethereum and Hyperledger Fabric, as shown in Figure 3.

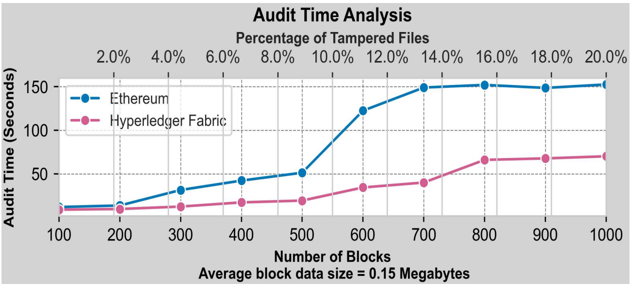


Fig. 3. Audit Trails Analysis

IV. Conclusion

The research presents the promising potential of the governance of generative AI tools and XAI decision-support systems in an organization by utilizing the immutable nature of blockchain. It can successfully manage the integrity of metadata containing the decisions by an XAI model and textual explanation by Generative AI. Automated auditing promotes the responsible use of AI technologies and reduces inconsistencies in tracking the accountability of adversarial decisions.

Acknowledgment

This work is supported by the xxx (Grant: xxx *Double-blind*).

References

[1] *Double-blind review*: Anonymous authors, Engineering Applications of Artificial Intelligence, 20XX.

[2] J.B.Yang, J. Liu, J. Wang, H.S. Sii and H.W. Wang, “Belief rule-base inference methodology using the evidential reasoning approach-RIMER,” IEEE Transactions on systems, Man, and Cybernetics-part A: Systems and Humans, vol. 36, p. 266-285, 2006.