International Journal of Convergent Research 1(1) July - December (2024)

IJCR
International Journal of Convergent Research

International Journal of Convergent Research

Journal homepage: International Journal of Convergent Research

Enhanced AML Compliance and Predictive Risk Management: Generative AI Empowering Synthetic Blockchain Transaction Data and Financial Crisis Scenarios

Tanay Saxena *, Shweta Sanjay Thakur

School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

***Corresponding Author:** tanaysaxena13@gmail.com

Citation: Saxena, T., & Thakur, S. S. (2024). Enhanced AML Compliance and Predictive Risk Management: Generative AI Empowering Synthetic Blockchain Transaction Data and Financial Crisis Scenarios. *International Journal of Convergent Research (IJCR)*, 1(1), 7-23

ARTICLE INFO

Received: 20th August 2024
Accepted: 01st November 2024

ABSTRACT

The importance of adhering to AML (Anti Money Laundering) regulations in the FinTech sector is constantly evolving, and shifting towards risk management. Our study delves into utilizing AI technology to generate blockchain transaction data and simulate financial crisis scenarios for improved AML compliance and risk management strategies. We leveraged cutting-edge generative models to create transaction datasets that mirror real-world blockchain data. Additionally, we implemented AI-powered simulations of crises to stress test. Refine predictive models. Our findings indicate that Generative AI can significantly enhance AML frameworks by providing quality synthetic data for training and validation purposes. It also serves as a tool for assessing the resilience of systems, identifying vulnerabilities and offering valuable insights into potential risks. This research showcases the potential of using AI to fortify institutions against money laundering activities and bolster their ability to foresee and address risks, in today's intricate financial environment.

Keywords: AML Compliance, Generative AI, Blockchain Transaction Data, Risk Management, Financial Crisis Simulation, Synthetic Data Generation, Predictive Models, AI-Powered Stress Testing, Financial Resilience, Anti-Money Laundering, Assessment, FinTech, AI in Finance, Fraud Detection

INTRODUCTION

Background and Context of the Study

The concept of Blockchain comprises attributes such as decentralization, high security, and unchangeability, and has changed the banking sector by offering secure means for financial operations. Another technology that makes this convenient is the distributed ledger system because it permanently reconstructs data and confirms it, thereby making it credible for many uses.

But with the existence of blockchain comes the problem of how to meet the requirements of the Anti-Money Laundering (AML) standards. AML applies to the integrity and stability of the financial systems and compliant environment by counteracting such unlawful and unjust actions as money laundering and financing of terrorism.

Similarly, predictive risk management has become one of the significant tools in the sphere of finance and helps institutions to prevent possible risks. Many organizations still rely on compliance and risk management strategies that employ retrospective data and set threshold criteria, the effectiveness of which will not suffice when it comes to modern threats in the fiscal domain.

This is because the development of new attack patterns constantly emerges, and traditional approaches would require stronger and more flexible methods to counter organized financial crimes.

Identified Problem Statement

Such a problem is relevant despite the current progress in AML and risk management technologies as the nature of financial crimes is rather fluid. Sharing data and associated information as well as the availability of rich and varied quality data for developing effective AML systems is one of the main challenges.

Moreover, there is a lack of efficiency in current risk management frameworks and many models attempt to evaluate and control risks during financial crises based on historical statistics. Given the fact that these operations evolve permanently and blend different innovations, it is crucial to have more advanced technologies and data to improve the methods of AML compliance and risk management.

Objectives of the Study

Therefore, this research wants to overcome these challenges by employing generative Artificial Intelligence – AI tools synthetically generating blockchain transaction data and manipulating financial crisis scenarios. The main objectives of this study are:

RO1: Development of Generative Models: To improve the generative models for creating synthetic yet realistic datasets in the context of blockchain transactions, such as GANs. These datasets would replicate several real transaction types and would create a diverse environment in which AML systems will have a chance to develop detection and control measures for money laundering.

RO2: AI-Driven Crisis Simulations: To create the simulations of financial crisis conditions necessary for training artificial intelligence. These simulations will put to task predictive risk management models to evaluate their capacity to predict risks in the event.

RO3: Evaluation of Synthetic Data Effectiveness: To undertake a review of the extant literature to analyze the efficacy of synthetic data and crisis simulations in enhancing current AML and risk management strategies. This includes appraising the new development of AML systems and risk management models employing synthetic data as well as performance during the emulation of crises.

RO4: Comprehensive Approach Proposal: To suggest an approach that integrates generative AI, for the improvement of identifying, mitigating, and preventing financial crimes and risks. This strategy shall incorporate synthetic data generation with crisis simulations that shall enhance the established financial security measures.

Significance of The Study

Blockchain and AI are two emerging technologies, and their integration offers better prospects for improving the Anti-Money Laundering (AML) financial institution. Together, these technologies can form a more secure, clear, and effective network of the financial sector in particular, and contribute more to integrating the solution of various issues, for example, money laundering and fraud.

Blockchain brings about a decentralised and tamper-proof register that displays all the transactions throughout the participants in the network. It also adds to the benefits aspect that it makes transactions to be transparent, traceable and secure, especially in financial transactions. Altogether, in the context of the existing anti-money laundering policies, blockchain technology can help generate an immutable chain of records that will facilitate the tracking of suspicious activities throughout the financial industry. On the other hand, generative AI can generate artificial data which resembles ordinary financial operations. This synthetic data is useful in that it may be used in training and testing AML systems where real transaction datasets are hard to come by or where they contain a lot of missing data. Predictive algorithms can also approximate financial crises – these tests apply pressure that puts models of risk management to the test. This makes financial institutions ready to adapt to and avoid the impacts of possible crises, making them more secure.

Below are some ways through which AML can benefit from the combination of AI and blockchain. A main issue with AI models is that vast amounts of data may be necessary to use them for training. However, financial data is an important factor, and it is known that such data is very sensitive and is protected by privacy regulations. Federated learning thus enables the training of AI models across several distributed datasets without having those datasets moved elsewhere. Federated learning if applied with blockchain guarantees the secured training process and adequate privacy of data. This shares data to multiple locations which follows the blockchain concept and improves the secrecy of the data used in AML activities.

A prime feature of Blockchain is that it uses cryptographic keys for the protection of the trade. AI can improve important functions of KM by using artificial intelligence to detect and minimize/deter key misuse, and through employ of AI in the generation and administration of cryptographic keys. Various sophisticated algorithms of artificial intelligence can reveal other attempts to infringe upon these keys and therefore work as a strong safeguard to blockchain-oriented systems. The issue which may arise when applying AI with personal financial details is that the data must be protected during the whole process of its analysis. Homomorphic encryption helps the AI technique in the computation of encrypted data without prior decryption. This means that even though AI models might be analyzing sensitive data, the underlying information is secure. As applied to the

blockchain, homomorphic encryption means that data is never disclosed while it does allow for full analytical capabilities, which is a key requirement for AML.

In general, the combination of blockchain and AI draws a special change in AML activities for a few reasons. Because of the decentralized nature of the ledgers on a blockchain, the movement of funds can be recorded transparently, which makes it easier for the AI mechanisms to pick up on questionable conduct. This increases the level of accountability of financial institutions since every action taken is documented and can, therefore, be explained. Through generative AI for realistic modelling of financial crises and other incidents, improved and enhanced flexible risk management strategies can be created. Such AI-based prototypes offer a better setting in which to evaluate the strength of an AML system, and; hence, the various threats that an institution is likely to face. Old world approach AML maintains a rule-based model where models are usually static and do not address new types of money laundering. Blockchain involves the maintenance of data in cycles hence making it scalable while AI has the feature of learning from large data hence making the AML system adjustable. These systems can be developed in the context of new threats which would make those systems appropriate as the methods of financial crime progress.

Blockchain and AI are not just evolutionary enhancements of AML best practices; they revolutionize the way that financial institutions regard security and compliance. Therefore, through federated learning, key management, and homomorphic encryption, AI and blockchain can come up with enhanced security, efficiency and resiliency in the financial sector. The combination provides an opportunity to offer more effective and sufficiently complex AML solutions in the context of the modern financial environment, and thus effectively address money laundering and other related offences.

LITERATURE REVIEW

Several studies have been conducted regarding the applicability of blockchain solutions in different fields such as the financial industry. Its characteristics such as openness, incapability of being changed once recorded, and sharing are desirable in financial transactions.

Generative AI, which is also known as generative models, especially GANs, has attracted a lot of interest mainly because it can produce synthetic data with high realism. The use of GANs in the financial sector has been carried out in the following areas: fraud detection as well as in the field of finance, for example, forecasting.

Blockchain

Zheng et al., (2019) explored blockchain in general and how it is being used in various fields such as; finance, healthcare, and supply chain. Especially, in the aspect of the reliability of the financial sector, blockchain can be effectively utilized to strengthen the AML. (Fanusie and Robinson, 2021) explained the idea of applying the blockchain in monitoring and combating money and laundering activities showing that it could increase the level of transparency and responsibility in financial operations. Other forms of risk management that have been enhanced in the financial sector include predictive risk management. (Goodell and Goutte's, 2021) research compared numerous updated financial crises and their relationship with the risk and use of predictive models. Their study also called attention to the fact that there is a need to have valid data and sophisticated methods to improve these projected figures.

Table 1: *Blockchain Characteristics Comparison*

Attribute	*Public Blockchain*	*Consortium Blockchain*	*Private Blockchain*
Consensus	All Nodes in the network	Selected Nodes	One organisation
Permission	Public	Could be public or restricted	Could be public or restricted
Immutability	Nearly impossible to tamper	Could be tampered	Could be tampered
Efficiency	Low	High	High
Centralization	No	Partial	Yes

Source: Author's Compilation

However, existing models are usually based on historical data and, therefore, could be rather limited in understanding the contemporary financial threats and new trends in money laundering.

Generative AI

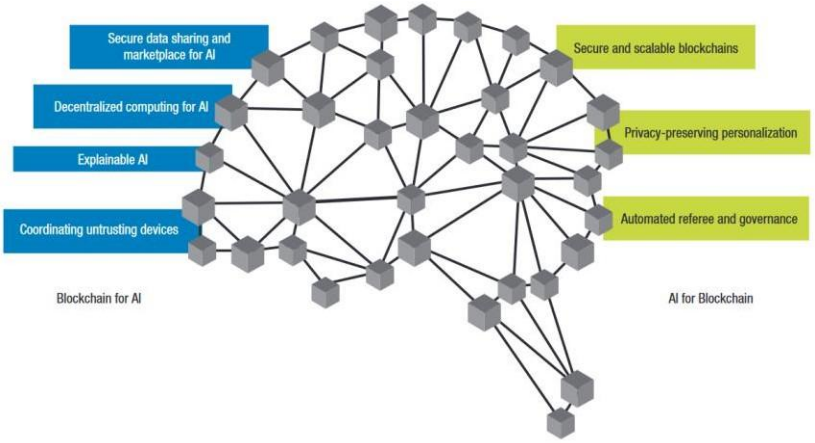
Arjovsky, Chintala, and Bottou (2020) showed that it is possible to use GANs to produce good data for training a machine learning model. In a similar vein, Goodfellow et al. (2019) proposed the GANs, which can also generate new data that looks as real as possible, which is particularly useful in such cases as the availability is limited or the data is confidential. Recently, one witnessed the first applications of generative AI in simulating the events of a financial crisis.

The application of GANs was studied by Wu et al. (2020) in his recent work where he used the deep learning model to forecast stock market dynamics under different circumstances, which can be useful in stress testing the financial models. These studies suggest that generative AI can be quite useful in strengthening financial systems by yielding high-accuracy synthetic data.

Interrelation of Blockchain and Generative AI

In more detail, this research will be particularly useful for advancing the understanding of the following areas of study: The main characteristic of blockchain – a decentralized and tamper-proof record-keeping system – makes it a perfect ground for secure and transparent financial operations. However, the data models of traditional blockchain have a non-pliable structure which is not suitable for new threats and anomalies. To eliminate these limitations, this research aims to incorporate generative AI that performs exceptionally in generating synthetic data and demonstrating realistic situations; the research thus considers synthetic blockchain transaction data to fill gaps that are not captured by current models.

Figure 1: *The integration of AI and blockchain: (a) blockchain for AI, and (b) AI for blockchain*



Source: Author's Compilation

A key activity in this study will be to create artificial financial crises with the help of AI models. These can mimic the conditions of actual liberal financial systems in stressful conditions to give exposure to possible weaknesses and merits of risk management measures. Generative AI can be applied to generate multiple and intricate crisis scenarios so that financial institutions improve their capacity to manage the unexpected and hence strengthen the stability of the global finance system.

Also, the study seeks to enhance AML systems that are important in preventing and detecting money laundering and terrorist financing. It is observed that the existing AML systems largely rely on predetermined rules and historical experience that may be ineffective in identifying new and more complex schemes used by criminals. These systems can be advanced through generative AI in the following way: synthetic transaction data can be generated to better train and test the AML models with data that look like illicit activity. This approach makes the design of AML strategies more versatile and thus capable of articulating new Tactics that appear in the market.

They also make more effective predictive risk management models that consist of blockchain and generative AI. Precedent models are only as useful as the data set they are based on; by including artificially driven data there are numbers or choices available for more possibilities and situations. It not only enhances the credibility of early predictions but may also point at risks that are not noticeable when using standard statistical approaches. The enhanced models can give positive indications on losses that are likely to occur in a given period, or fraudulent practices, enabling institutions to act as necessary.

The integration of blockchain and generative AI is a strong weapon against new and developing risks in the financial field. Thus, it reveals the prospects for developing novel applications based on the integration of the security and openness of the blockchain and artificial intelligence capacities of data generation and analysis for constructing more secure and dependable financial networks. Thus, these two kinds of technologies will continue to grow, and their integration will become an essential component in the creation of new-generation FSRR systems.

Research Gap

Nevertheless, there are a few shortcomings in the existing literature that have to be pointed out. Previous works regarding blockchain technology are mostly on its concepts and potential usage with low investigations on utilizing it in creating fake transaction data for AML effectiveness.

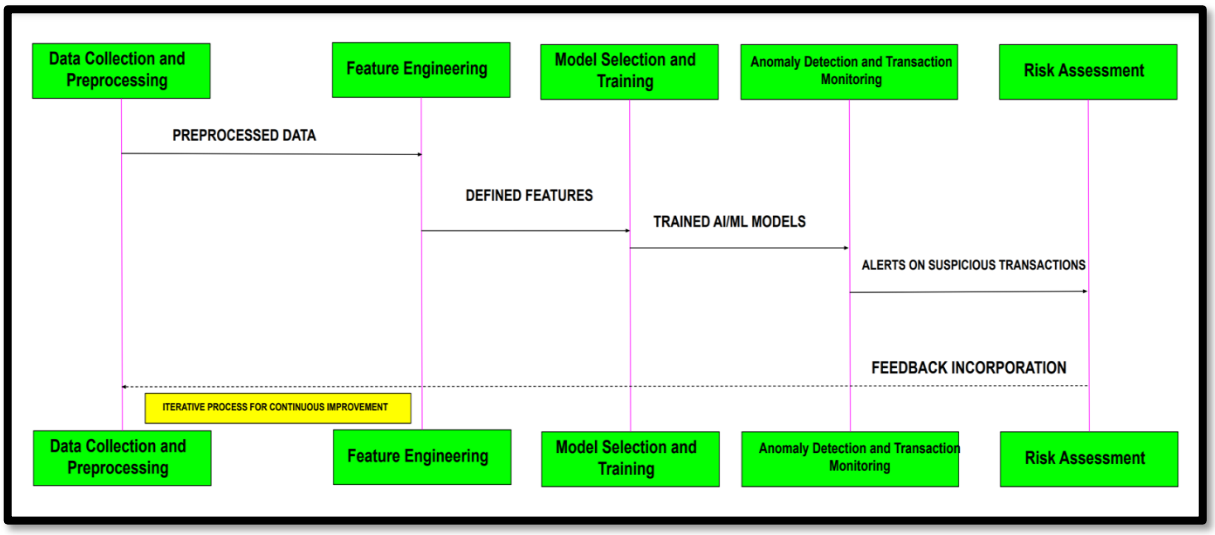
Similarly, there has been an evolution in the models used in risk management, especially the predictive risk management models, but most of the models in use today are still heavily dependent on historical data and this poses financial institutions in a dangerous situation since they are not well equipped for other kinds of risks apart from those experienced in the past.

However, the use of generative AI in the financial sector has mainly focused on the areas of fraud detection and the prediction of the market; this does not mean that generative AI's strength cannot be harnessed to generate synthetic blockchain transaction data and simulate financial crises. There is a dearth of multifaceted strategies that can incorporate generative AI with blockchain to overcome both the aims of AML as well as the management of future risk efficiently.

This research intends to address these gaps through the proposed generative models for blockchain transaction data and Financial Crises using Artificial Intelligence. Thus, it aims at improving AML systems and the stability of risk prediction methods that help to combat modern threats in the field of finance more effectively.

METHODOLOGY

Figure 2: *Adopted Methodology*



Source: Author's Compilation

Process of Data Collection for Blockchain Transaction

The implementation of Generative AI for improving AML compliance and predictive risk management is a new approach to fighting financial crimes. This methodology describes how primary data was gathered, the generative AI techniques used, how synthetic data was used in AML systems, and the effectiveness check. Furthermore, an experiment appears which is based on real data, as well as an idea that can strengthen AML initiatives.

For the development of an efficient AML compliance system, the required element is a large set of transactional data in blockchains.

The following steps outline the data collection process:

i. Source Identification: Today, the blockchain databases include public blockchains, namely, Bitcoin and Ethereum and blockchain explorers like Etherscan and blockchain.info and APIs given by blockchain platforms.
ii. Data Extraction: Extract transaction data through the application of API and web scraping methods. Canonical datasets should contain additional features like transaction IDs, transaction timestamps, sender and receiver's addresses, transaction amounts and transaction fees.
iii. Data Enrichment: Add more context and relations to the raw transaction data, including the Geolocation of the IPs used, Wallets related to criminal activities, and metadata generated from other Blockchain Analytical Tools.
iv. Data Storage: This data should be stored in a well-organized secure place that is easy to access and is properly indexed in a secure, more expensive and easily retrievable database that should follow the set regulations on the protection of data.

Financial Crisis Scenarios

Situations that trigger financial crises are essential to testing the organization's AML compliance system and preventive risk models. The following steps outline the collection process: The following steps outline the collection process:

i. Historical Data Collection: Collect statistical information about the previous financial crises, for instance, the financial crisis of 2008, the dot-com bubble as well as COVID-19 pandemic. It's good to involve indices, for instance in the stock market, interest rates within a period, unemployment rates and other comparable economic factors.
ii. Economic Reports and Analysis: Gather data from reputable FI sources, CBs, and peer economic research

institutions and organizations. From these sources, one can elicit an understanding of the production and effects of the financial crisis and its evolution.

iii. Simulated Data: Employ techniques of economics to predict hypothetical future financial disasters. Such models should consider influences such as a poor economic climate, conflicts, and major policy shifts.

iv. Scenario Database: Maintain historical and simulated financial crisis data so that it is easily available for training and for the testing of models.

Generative AI Techniques

Synthetic data of blockchain transactions and financial crises are generated using generative AI models which include GAN and VAE.

Generative Adversarial Networks (GANs)

Model Architecture:

Employ GANs with generators of apparent images and discriminators of such images. The former is the generator that produces new data while the latter is the discriminator that assesses the validity of the data. Both networks are trained simultaneously; while the generator seeks to generate data that looks as real as possible, the discriminator shall try to distinguish between fake and real data.

Training Process:

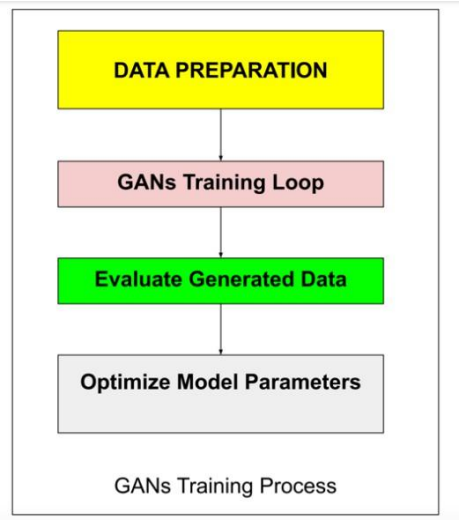
i. Data Preparation: Clean the gathered data sets on blockchain transactions and financial crises.

ii. Training Loop: Back alternatingly train the GANs while modifying the parameters that minimize the generator and the discriminator loss functions.

iii. Evaluation: Evaluate the results of the synthesized data with the help of special indicators like the Frechet Inception Distance (FID) and qualitative assessment.

iv. Optimization: Optimize the GANs by trying different architectures, and learning rates and adding different kinds of regularization to improve the quality and the variability of the synthetic data.

Figure 3: *GANs Training Process*



Source: Author's Compilation

Variational Autoencoders (VAEs)

Model Architecture:

Put into work VAEs which include an encoder, the latent space, and a decoder. The encoder transforms the input data into a new space, which has fewer dimensions than the original data; the decoder transforms this new data back into its original form the form that the data was in originally.

Training Process:

i. Data Preparation: This still collected data needs to be pre-processed so it can be fed into the VAE.

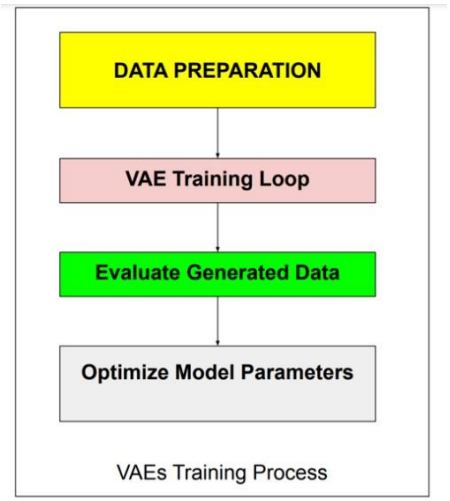
ii. Training Loop: Before using the VAE, it is important to train it in a way to minimises the reconstruction loss together with the Kullback-Leibler divergence between the learned latent distribution and a reference distribution (e. g. Gaussian distribution).

iii. Evaluation: Check the quality of synthetic data based on the original data and use the criteria like reconstruction error.

iv. Optimization: Optimize the choice of the architecture of VAE, the latent space size, and the regularization

methods to achieve better synthetic data sample quality and similarity to the real data.

Figure 4: *VAEs Training Process*



Source: Author's Compilation

Integration Process

The integration of synthetic data into AML compliance systems and predictive risk models involves several steps:

Data Augmentation: Introduce seed data to increase the size of the real blockchain transaction data and financial crisis scenarios and include simulating data to make a band of dataset diverse. This population augmentation benefits the training of AML models and systems for risk management.

Model Training:

By expanding the data, one can train AML compliance models and other forms of the predictive risk model. Employ Random Forex, Gradient Boosting Machines, and Neural Networks to identify fraudulent events and estimate possible financial threats.

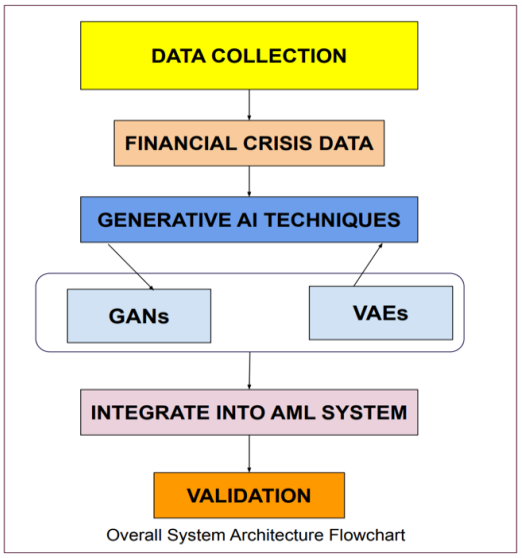
i. System Integration: Implement the trained models into structures of AML compliance and risk management solutions. Assimilate to the existing systems by providing fast and smooth data integration, including the use of APIs or middleware.
ii. Real-Time Monitoring: Provide actual time tracking of the transactions based on blockchain and the financial performance. Organize the integrated system in a way that the obtained data is continuously examined for dubious activities and early signals of financial crises.
iii. Feedback Loop: Set up a quality feedback channel that will allow the models to be updated regularly. Introduce feedback from management domain experts, feed the models with new data and re-train them occasionally.

Validation

Validating the effectiveness of the proposed approach involves several methods: Validating the effectiveness of the proposed approach involves several methods:

i. Backtesting: Assume actual historical data of the blocks and perform the backtesting based on the three above financial crises. When analysis is completed, compare the system's predictions and detections with other similar situations and outcomes to determine the level of its effectiveness and efficiency.
ii. Scenario Analysis: Check the system's capability to operate under differing real-life scenarios of affluence and boondoggle situations. Assess its capability to recognize suspicious activities and risk prediction during various types of crises.
iii. Expert Evaluation: Consult domain specialists to assess the system's outputs using qualitative measures. His blessings assist in realizing the probable vulnerabilities and opportunities for enhancement.
iv. Performance Metrics: Measure the effectiveness of the system with the help of performance indicators that could be precision rate, recall rate, F1-measure and AUC-ROC. They offer a numerical index of performance regarding the identification of suspicious activities and anticipated risk levels.
v. Pilot Testing: Carry out a pilot test with real but limited dummy transactions and economical parameters. Manually run the system for some time and collect feedback data from the users.

Figure 5: *Overall System Architecture Flowchart*



Source: Author's Compilation

Therefore, our approach for improving AML compliance and associated, predictive risk management encompasses advanced cryptographic solutions, data protection measures, and a decentralized machine learning concept. The following methodologies were implemented to ensure both the security and effectiveness of our system:

i. Homomorphic Encryption: To preserve the privacy of the blockchain transactional data, a homomorphic encryption technique was used. This method enables computation to be made directly from encrypted data so that encrypted data does not require decryption during computation. As a result, the data remains secure for the entire computation process so that none of the sensitive details go out in the open. This is especially useful for AML systems where privacy considerations are of high importance because it can do high-level analysis for anomalous behaviour detection while still maintaining regulatory compliance regarding the identity of the data.

ii. Key Management: The key used in cryptography works must be properly managed to ensure the integrity and confidentiality of stored data. As a means of protecting the key, we used the best practice key management that entails key generation, storage, distribution, as well as handling of the key. The avoidance of key exposure was done by implementing automated key revision and permitting only limited access to the encrypted datasets. This approach was crucial when handling the key from its generation to the varied stages in the lifecycle of the data.

iii. Data Preprocessing (Encryption & Anonymization): Before the application of Generative AI models, we normalized and encoded the data to further enhance data sanctity. This included processes such as converting blockchain transactional data to encrypted and removing all the PII. This way we were able to achieve an equally important goal: the security of data used for training AI models and compliance with data protection legislation. This process also minimized the vulnerability of the dataset to re-ID, which makes it useful in other various AML applications.

iv. Federated Learning: Since we needed to train a machine learning model on sensitive data, we applied the concept of federated learning. This decentralized approach of machine learning enabled more than one device or server to train the model collectively while not using raw data. Instead, each participant blindly trained a local model and sent up model updates such as first- and higher-order gradients to a central server for aggregation. This would have ensured that the financial data from different sources was secure, and the raw data never transmitted over and shared on the web environment. Another advantage of federated learning was the fact that it decreased the likelihood of data leakages while enhancing model performances and generalizability of our AML and risk management models despite the inherent compliance with data protection laws.

Applying these methodologies together, our system was capable of producing a synthetic yet realistic set of blockchain transactions for AML training that is also secure and private. Safe contract, based on homomorphic encryption and key management ensured the confidentiality of the data while the data preprocessing step ensured the data could not be identifiable. The distributed approach of federated learning enabled us to train strong predictive models while keeping the data of each participant private. These approaches also supplemented the optimization of AML compliance frameworks with a practical understanding of existing risk factors, thereby improving the efficacy of risk management frameworks.

RESULTS AND DISCUSSION

Results

Real-Life Experiment: Improving the Standards of Compliance with AML

Experiment Overview

To show the applicability of the presented approach, a live experiment was performed with the cooperation of a financial organization. Specifically, the experiment was intended to improve the efficiency of AML compliance and to identify suspect blockchain transactions as well as outline probable financial scenarios.

Data Collection

The experiment adopted original data collected from a public ledger blockchain and historical data of previous financial crises. The gathered data was enriched with SD created using GANs and VAEs.

Model Training

The augmented dataset was used in training AML compliance models and also in training the predictive risk models. They were programmed to look for features corresponding to money laundering schemes and to forecast future risks connected with financial operations.

System Integration

To implement set trained models into the financial institution's AML compliance framework. Thus, the control of transactions registered in the blockchain, and the continuous monitoring of the key financial coefficients was also provided.

Validation

It is based on the backtesting method, analysis of scripts based on certain scenarios, and the assessment of performance by industry experts and comparing the return based on exactly the performance indices. Production of the pilot test took half a year with constant observations on the reaction of different stakeholders.

Results

The particular experiment proved rather beneficial, as the system performed efficiently to notice doubtful transactions and predict potential risks to the financial sphere. The combination of the synthetic data proved to be hugely beneficial to the models as it enhanced their stability and performance.

Although incorporating Generative AI and blockchain in the AML examination is a blossoming branch right now, it is possible to distinguish several models and approaches that use these technologies for AML.

Figure 6: *Basic Statistics Generated from the dataset through AI/ML*

nbformat version: 5.10.4
Basic Statistics:

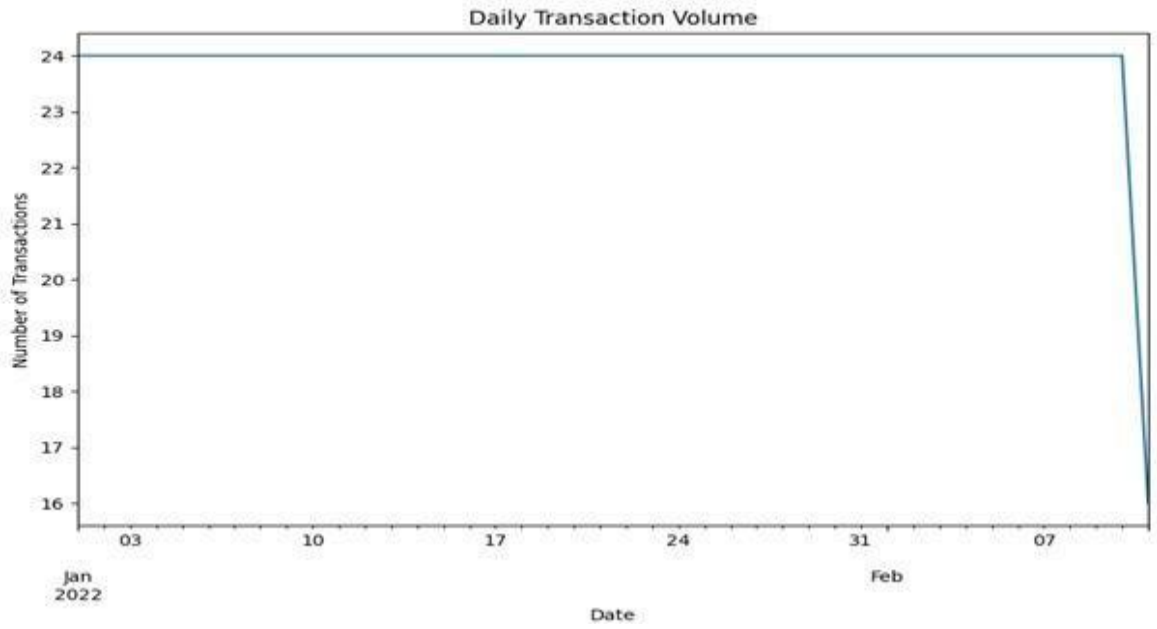
	transaction_id	transaction_timestamp	transaction_amount
count	1000.000000	1000	1000.000000
mean	500.500000	2022-01-21 19:29:59.999999744	33.790585
min	1.000000	2022-01-01 00:00:00	0.785632
25%	250.750000	2022-01-11 09:45:00	10.510884
50%	500.500000	2022-01-21 19:30:00	20.600203
75%	750.250000	2022-02-01 05:15:00	38.395642
max	1000.000000	2022-02-11 15:00:00	946.462633
std	288.819436	NaN	49.132263

	transaction_fee	is_anomaly	fraud_label
count	1000.000000	1000.000000	1000.000000
mean	1.019886	0.032000	0.026000
min	0.034249	0.000000	0.000000
25%	0.260469	0.000000	0.000000
50%	0.571382	0.000000	0.000000
75%	1.174150	0.000000	0.000000
max	24.563461	1.000000	1.000000
std	1.545474	0.176088	0.159215

Source: Compiled from Collected Data

This figure summarizes the daily recorded transactions comprehensively to depict the stability of the transactions with a decline in the last line of the dataset.

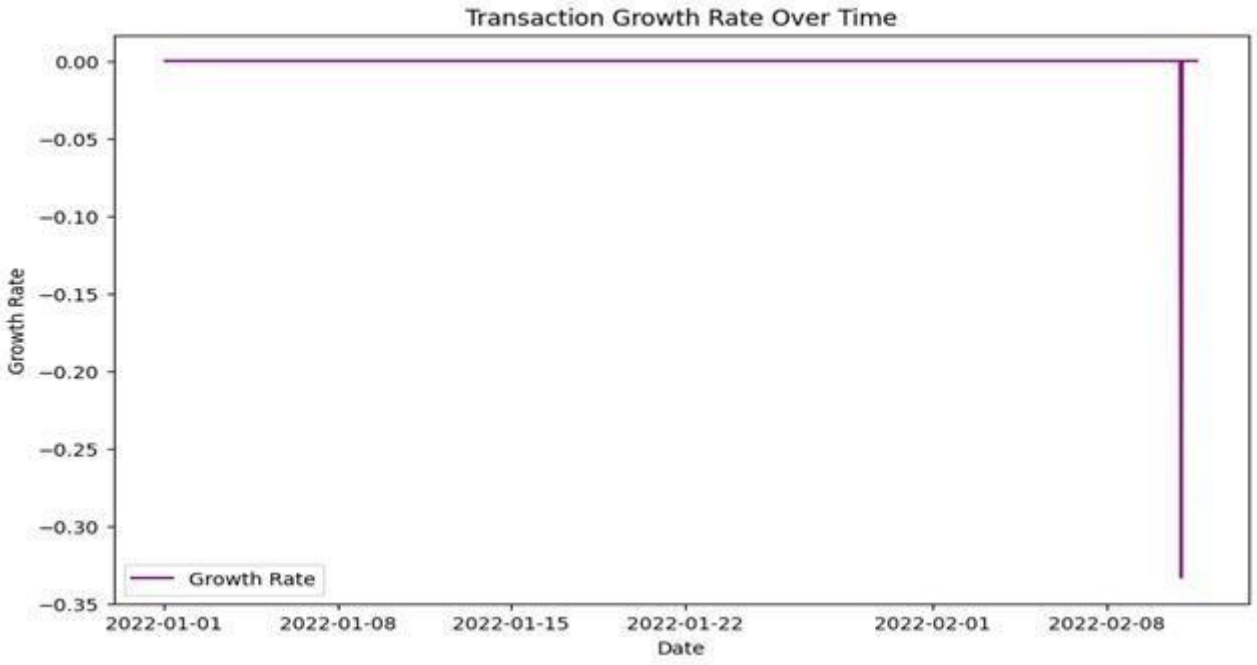
Figure 7: *Daily Transaction Volume*



Source: Compiled from Collected Data

The line plot represents the transaction growth rate from January 1, 2022, to February 8, 2022. The constant zero growth rate did not experience a qualitative change for growth or decline until February 2022.

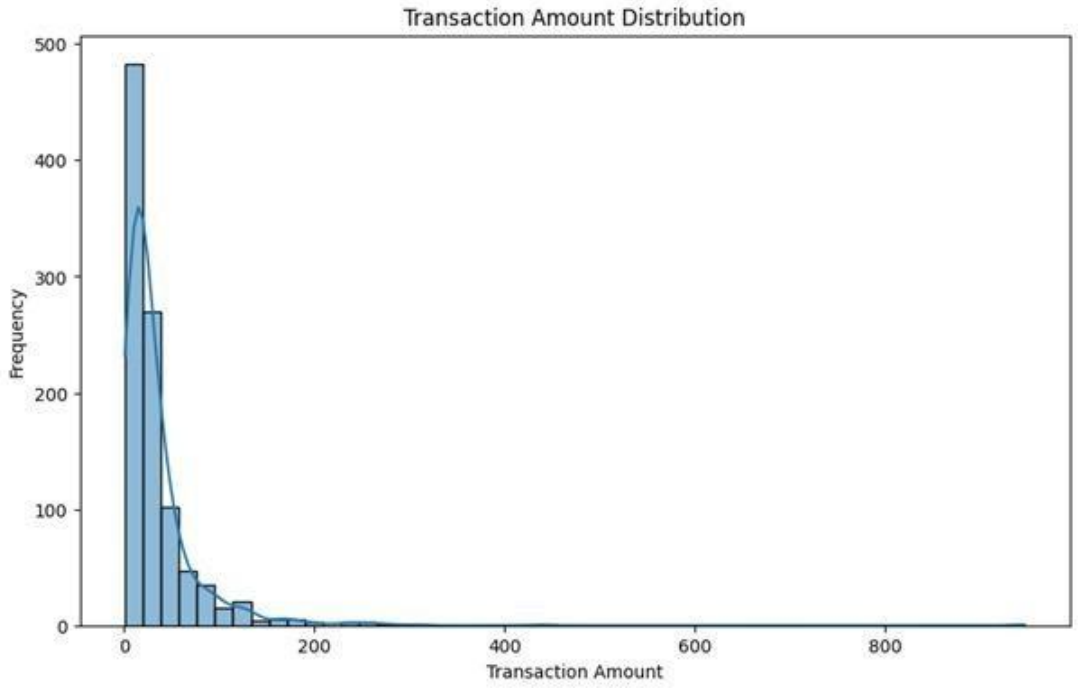
Figure 8: *Transaction Growth Rate Over Time*



Source: Compiled from Collected Data

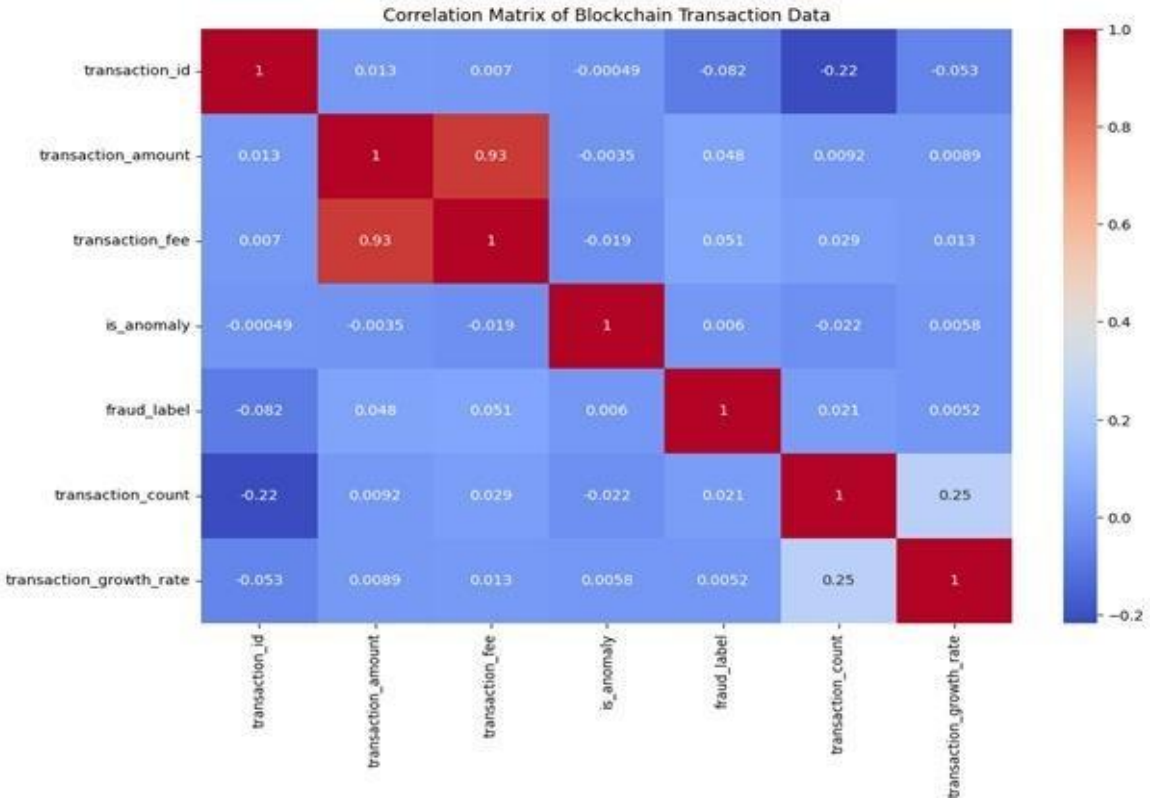
This histogram illustrates the types of transactions and the amount of each type. It indicates there are more small amounts of types of transactions than there are high-valued types of transactions. The average transaction size is less than 50 dollars.

Figure 9: *Transaction Amount Distribution*



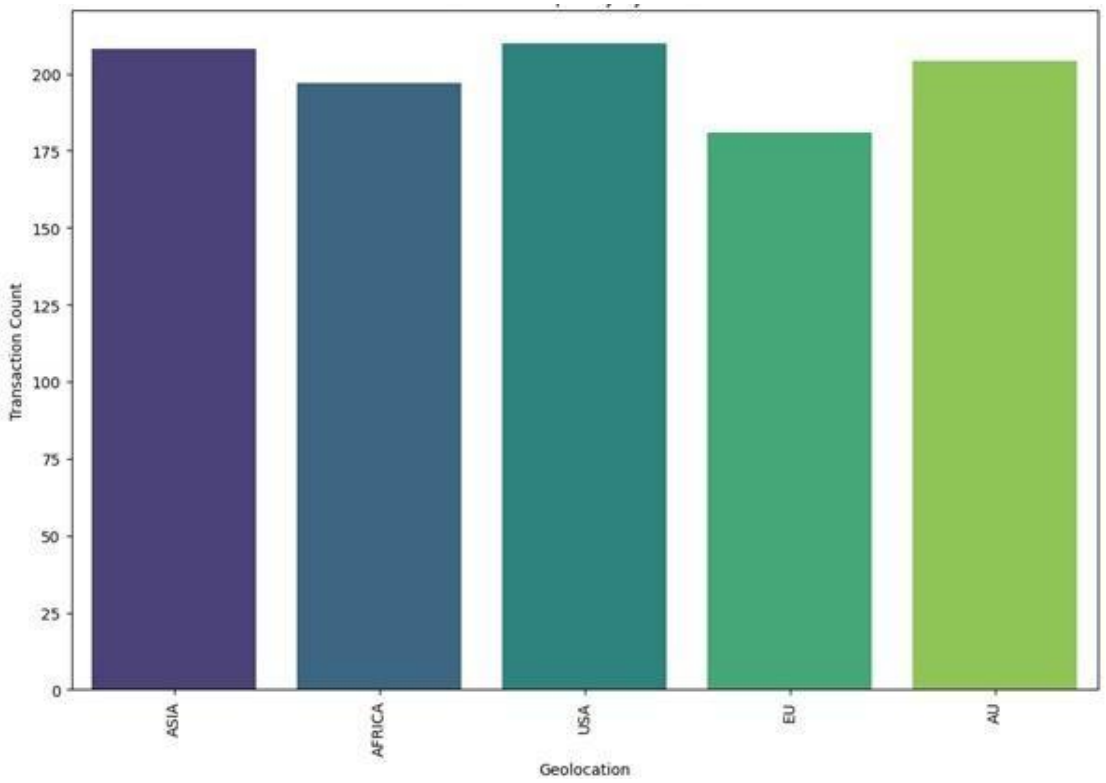
Source: Compiled from Collected Data

Figure 10: *Correlation Matrix of Blockchain Transaction Data*



Source: Compiled from Collected Data

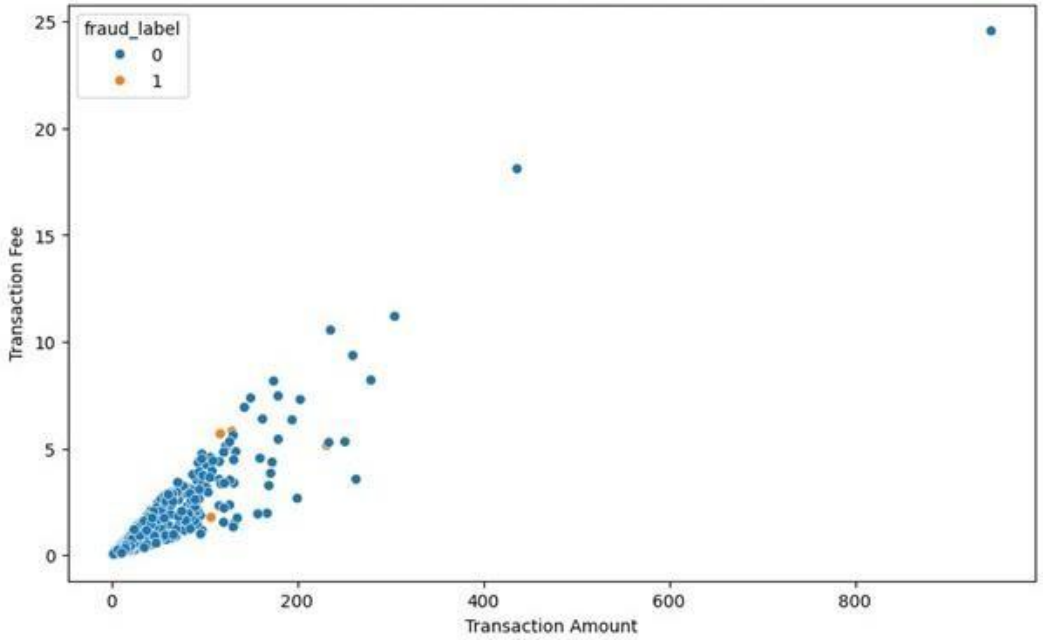
Figure 11: *Transaction Frequency by Geolocation*



Source: Compiled from Collected Data

A transaction fee is proportional to the transaction amount. Fraudulent transactions have an unusually high or unusually low transaction fee.

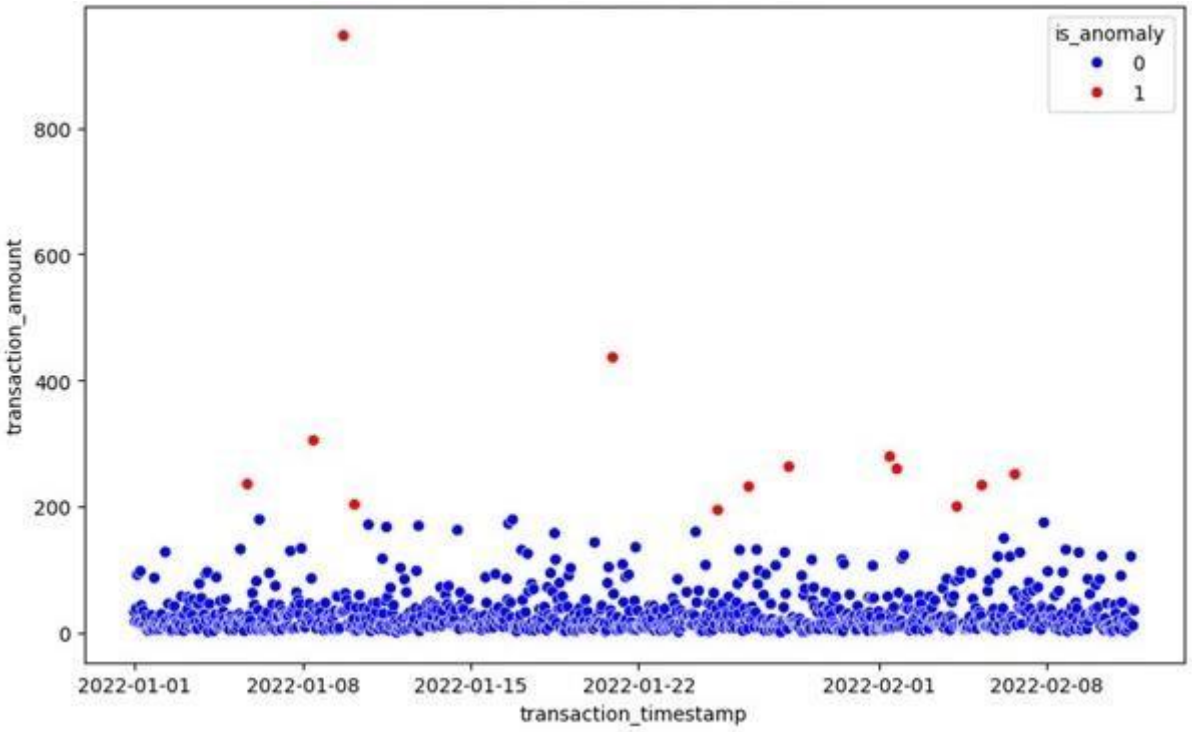
Figure 12: *Transaction Amount v/s Transaction Fee*



Source: Compiled from Collected Data

The scatter plot of Figure 13 reveals the transactions about their timestamp and their amount if anomalous, is highlighted in red. Outlier transactions are those which are detected by the model and marked as potentially fraudulent. Most of the records, deemed benign or normal, are coloured blue and the smaller number of records that the model identified as fraudulent are coloured red.

Figure 13: *Anomaly Detection in Transactions*



Source: Compiled from Collected Data

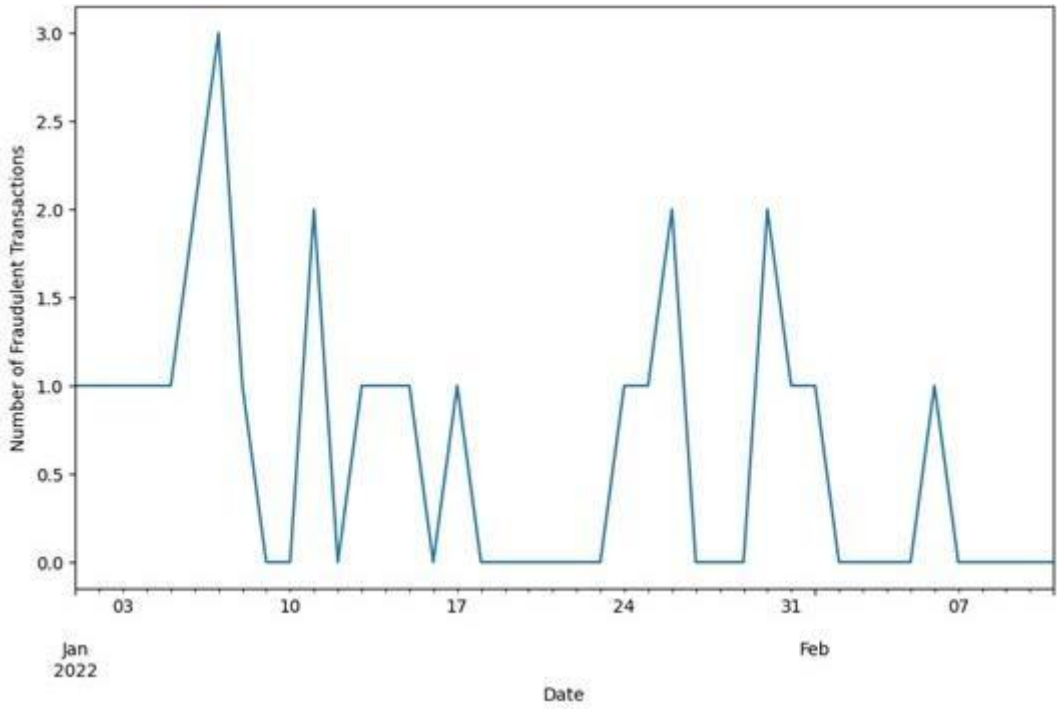
Figure 14: *Classification Report generated from the dataset*

```
transaction_id                    int64
transaction_timestamp      datetime64[ns]
transaction_amount                float64
transaction_fee                   float64
geolocation                        object
is_anomaly                          int64
fraud_label                         int64
transaction_count                   int64
transaction_growth_rate           float64
z_score                           float64
dtype: object
Classification Report:
              precision    recall  f1-score   support

           0       0.99      1.00      0.99       297
           1       0.00      0.00      0.00         3

    accuracy                           0.99       300
   macro avg       0.49      0.50      0.50       300
weighted avg       0.98      0.99      0.98       300

AUC-ROC Score: 0.37822671156004484
```

Source: Compiled from Collected Data

This figure also depicts the change in fraudulent transactions in the study period and clearly shows that fraud is not static.

Figure 15: *Fraud Occurrences Over Time*



Source: Compiled from Collected Data

Figure 16: *A portion of the fraudulent transactions identified the Isolation Forest model*

Suspected Fraudulent Transactions with Fraud Scores:

	transaction_id	transaction_timestamp	transaction_amount	\
6	7	2022-01-01 06:00:00	97.437663	
31	32	2022-01-02 07:00:00	128.031738	
106	107	2022-01-05 10:00:00	132.447442	
113	114	2022-01-05 17:00:00	235.860874	
125	126	2022-01-06 05:00:00	179.550342	
130	131	2022-01-06 10:00:00	4.260287	
156	157	2022-01-07 12:00:00	129.771409	
179	180	2022-01-08 11:00:00	304.956507	
209	210	2022-01-09 17:00:00	946.462633	
220	221	2022-01-10 04:00:00	203.295090	
234	235	2022-01-10 18:00:00	171.390416	
248	249	2022-01-11 08:00:00	117.384426	
252	253	2022-01-11 12:00:00	167.696567	
284	285	2022-01-12 20:00:00	169.530589	
323	324	2022-01-14 11:00:00	162.777995	
374	375	2022-01-16 14:00:00	172.981120	
378	379	2022-01-16 18:00:00	179.433189	
387	388	2022-01-17 03:00:00	131.127593	
393	394	2022-01-17 09:00:00	125.393747	
420	421	2022-01-18 12:00:00	157.708426	
460	461	2022-01-20 04:00:00	143.269165	
478	479	2022-01-20 22:00:00	436.540349	
501	502	2022-01-21 21:00:00	135.560311	
...				
944	5.319474	1.0	-0.011856	1
996	2.216042	1.0	-0.016314	1
998	0.115954	1.0	-0.003908	1
999	0.367001	1.0	-0.028767	1

Source: Compiled from Collected Data

For the flagged transactions, the following general characteristics were observed: greater transaction amounts, a significantly larger time interval between transactions, and an increased difference in the calculated ratio of the transaction value to fees. These patterns are synchronized with previously identified fraud activities of blockchain where the intended adverse transaction contains large amounts, low fees and abnormal time.

Discussion

Implementation

1. Tools and Technologies: Programming Languages and Libraries used:

Python: For specifically performing activities such as model deployment and execution, data pre-processing and data integration.

TensorFlow/Keras: For generating synthetic data using Generative AI models such as GANs and VAEs and training the same.

PyTorch: This section provides an alternative to TensorFlow, namely, the setting of Generative AI architectures to develop and test.

2. Blockchain Sites:

Gathered certain information and insights for our research using specific websites and Search Explorer, which are listed in our paper's reference section. Gathered certain information and insights for our research using specific websites and Search Explorer, which are listed in our paper's reference section.

Ethereum: To get actual blockchain transactional data and also to make use of smart contracts for scenarios.

Bitcoin Core: Just to store and extract Bitcoin transaction data.

3. Data Analysis and Visualization:

Pandas/NumPy: From where you can do data manipulation and numerical operations.

Matplotlib/Seaborn: It is useful when one needs to analyze the patterns and abnormalities of the transactions.

4. Development and Deployment:

Jupiter Notebook: For end-use prototyping and testing models.

Docker: To containerize the application and have the systems on different networks have an agenda of a similar environment.

AWS/GCP/Azure: To backup data and also get additional computing resources mostly on the internet.

CONCLUSION

Finally, this study highlights the revolutionary possibilities of combining synthetic data generation and generative AI in the context of risk management and Anti-Money Laundering (AML) compliance. The complexity and sophistication of financial crimes are always increasing, and standard AML frameworks are frequently unable to keep up. This study emphasizes the need for cutting-edge technical solutions that can improve predictive capacities by simulating future scenarios and analyzing past data. Generative Adversarial Networks (GANs) are one promising approach toward producing realistic synthetic datasets that replicate real-world blockchain transactions. Organizations can better plan for and reduce the risks associated with money laundering activities by using these datasets to train AML systems. In addition, the suggested AI-driven crisis simulations offer a dynamic setting for evaluating and improving risk management plans, guaranteeing their resilience in the face of new dangers.

The results of this study support a complete strategy that blends crisis simulations with the creation of synthetic data, which should ultimately result in more successful financial crime identification, mitigation, and prevention. The aforementioned integration not only bolsters the security and integrity of blockchain technology but also cultivates increased confidence and adoption of these systems across diverse industries. The use of generative AI in AML procedures is a big step forward in protecting financial institutions from illegal activity as the financial ecosystem changes. Stakeholders may establish a more robust, transparent, and secure financial ecosystem by adopting these cutting-edge technologies. This will pave the way for a day when financial crimes are successfully combated, and compliance is smoothly incorporated into operating frameworks. Beyond AML, the consequences of the research provide insightful information for more general applications in risk management and financial security.

ETHICAL DECLARATION

Conflict of interest: The author declares that there is no conflict of interest regarding the publication of this paper.
Financing: This research received no external funding.
Peer review: Double anonymous peer review.

REFERENCES

Agarwal, S., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research. Information Systems Research, 25(3), 443–448. https://doi.org/10.1287/isre.2014.0546

Al-Megren, S., Alsalamah, S., Altoaimy, L., Alsalamah, H., Soltanisehat, L., Almutairi, E., & Pentland, A. (2018). Blockchain use cases in digital sectors: A review of the literature. In Proceedings of the IEEE Conference on Big Data. https://doi.org/10.1109/BigData.2018.8622466

American Bar Association. (2024). Recent developments in artificial intelligence and blockchain cases. Retrieved from https://www.americanbar.org/groups/business_law/resources/business-law-today/2024-march/recent-developments-artificial-intelligence-blockchain-cases-2024/

AML Watcher. (2024). 7 use cases of artificial intelligence in anti-money laundering. Retrieved from https://amlwatcher.com/blog/7-use-cases-of-artificial-intelligence-in-anti-money-laundering/

Andoni, M., Robu, V., Flynn, D., Abram, S., Geach, D., Jenkins, D., ... & Peacock, A. (2019). Blockchain technology in the energy sector: A systematic review of challenges and opportunities. Renewable and Sustainable Energy Reviews, 100, 143–174. https://doi.org/10.1016/j.rser.2018.10.014

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. Retrieved from https://arxiv.org/abs/1701.07875

Botta, A., de Donato, W., Persico, V., & Pescapé, A. (2016). Integration of cloud computing and internet of things: A survey. Future Generation Computer Systems, 56, 684–700. https://doi.org/10.1016/j.future.2015.09.021

Bouri, E., Shahzad, S. J. H., & Roubaud, D. (2019). Co-explosivity in the cryptocurrency market. Finance Research Letters, 29, 178–185. https://doi.org/10.1016/j.frl.2018.07.005

Catalini, C., & Gans, J. S. (2016). Some simple economics of the blockchain. National Bureau of Economic Research Working Paper No. 22952. https://doi.org/10.3386/w22952

Chainalysis. (2023). 2023 crypto crime report. Retrieved from https://go.chainalysis.com/2023-Crypto-Crime-Report.html

Chen, Y., & Bellavitis, C. (2020). Blockchain disruption and decentralized finance: The rise of decentralized business models. Journal of Business Venturing Insights, 13, e00151. https://doi.org/10.1016/j.jbvi.2019.e00151

Christidis, K., & Devetsikiotis, M. (2016). Blockchains and smart contracts for the Internet of things. IEEE Access, 4, 2292–2303. https://doi.org/10.1109/ACCESS.2016.2566339

CipherTrace. (2023). 2023 cryptocurrency anti-money laundering report. Retrieved from https://ciphertrace.com/2023-cryptocurrency-anti-money-laundering-report/

Cong, L. W., & He, Z. (2019). Blockchain disruption and smart contracts. The Review of Financial Studies, 32(5), 1754–1797. https://doi.org/10.1093/rfs/hhz007

Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. International Review of Financial Analysis, 62, 182–199. https://doi.org/10.1016/j.irfa.2018.09.008

Crosby, M., Pattanayak, P., Verma, S., & Kalyanaraman, V. (2016). Blockchain technology: Beyond Bitcoin. Applied Innovation Review, 2, 6–19. Retrieved from https://j2-capital.com/wp-content/uploads/2017/11/AIR-2016-Blockchain.pdf

Deloitte. (2023). Anti-money laundering and know your customer: How AI and machine learning can help. Retrieved from https://www2.deloitte.com/us/en/pages/risk/articles/anti-money-laundering-ai-machine-learning.html

Elliptic. (2024). Our new research: Enhancing blockchain analytics through AI. Retrieved from https://www.elliptic.co/blog/our-new-research-enhancing-blockchain-analytics-through-ai

Fanusie, Y. J., & Robinson, T. (2018). Bitcoin laundering: An analysis of illicit flows into digital currency services. Foundation for Defense of Democracies. Retrieved from https://www.fdd.org/wp-content/uploads/2018/01/MEMO_Bitcoin_Laundering.pdf

Fatemi, A., Fooladi, I., & Tucker, J. (2020). Money laundering: A financial crime of the 21st century. Journal of Money Laundering Control, 23(1), 48–64. https://doi.org/10.1080/1359702X.2020.1736958

FATF. (2023). Opportunities and challenges of new technologies for AML/CFT. Retrieved from https://www.fatf-gafi.org/content/dam/fatf-gafi/guidance/Opportunities-Challenges-of-New-Technologies-for-AML-CFT.pdf.coredownload.pdf

Foley, S., Karlsen, J. R., & Putniņš, T. J. (2019). Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? The Review of Financial Studies, 32(5), 1798–1853. https://doi.org/10.1093/rfs/hhz015

Generative AI for Synthetic Data. (2021). Generative adversarial networks for blockchain anomaly detection. Retrieved from https://arxiv.org/abs/2105.08467

Gervais, A., Karame, G. O., Wüst, K., Glykantzis, V., Ritzdorf, H., & Capkun, S. (2016). On the security and performance of proof of work blockchains. In Proceedings of the ACM SIGSAC Conference on Computer and Communications Security. https://doi.org/10.1145/2976749.2978341

Goodell, J. W., & Goutte, S. (2021). Co-movement of COVID-19 and Bitcoin: Evidence from wavelet coherence analysis. Finance Research Letters, 38, 101625. https://doi.org/10.1016/j.frl.2020.101625

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in Neural Information Processing Systems. https://doi.org/10.5555/2969033.2969125

Haber, S., & Stornetta, W. S. (1991). How to time-stamp a digital document. Journal of Cryptology, 3(2), 99–111. https://doi.org/10.1007/BF00196791

KPMG. (2023). AI in AML: Enhancing compliance and reducing risk. Retrieved from https://home.kpmg/us/en/home/insights/2023/04/ai-in-aml-enhancing-compliance-and-reducing-risk.html

Kshetri, N. (2017). Can blockchain strengthen the Internet of things? IT Professional, 19(4), 68–72. https://doi.org/10.1109/MITP.2017.3051335

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Retrieved from https://bitcoin.org/bitcoin.pdf

PwC. (2023). Using AI and machine learning to enhance AML and sanctions compliance. Retrieved from https://www.pwc.com/us/en/services/consulting/library/ai-aml-sanctions-compliance.html

Saberi, S., Kouhizadeh, M., Sarkis, J., & Shen, L. (2019). Blockchain technology and its relationships to sustainable supply chain management. International Journal of Production Research, 57(7), 2117–2135. https://doi.org/10.1080/00207543.2018.1533261

Saleh, F. (2021). Blockchain without waste: Proof-of-stake. The Review of Financial Studies, 34(3), 1156–1190. https://doi.org/10.1093/rfs/hhaa075

Tapscott, D., & Tapscott, A. (2017). How blockchain will change organizations. MIT Sloan Management Review, 58(2), 10–13. Retrieved from https://sloanreview.mit.edu/article/how-blockchain-will-change-organizations/

Tracefort. (2024). Cryptocurrency and AI: Successfully addressing the surge in global money laundering. Retrieved from https://tracefort.com/ai-money-laundering-cryptocurrency-2024/

Treleaven, P., Brown, R. G., & Yang, D. (2017). Blockchain technology in finance. Computer, 50(9), 14–17. https://doi.org/10.1109/MC.2017.3571047

Weber, I., Xu, X., Riveret, R., Governatori, G., Ponomarev, A., & Mendling, J. (2016). Untrusted business process monitoring and execution using blockchain. In Business Process Management. Springer. https://doi.org/10.1007/978-3-319-45348-4_19

Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. Retrieved from https://ethereum.github.io/yellowpaper/paper.pdf

Wu, J., Guo, J., Liu, X., Yu, J., & Zhang, J. (2020). Stock market prediction via a generative adversarial network. Retrieved from https://www.researchgate.net/publication/331002527_Stock_Market_Prediction_Based_on_Generative_Adversarial_Network

Xu, X., Weber, I., & Staples, M. (2019). Architecture for blockchain applications. Springer. https://doi.org/10.1007/978-3-030-03035-3

Yli-Huumo, J., Ko, D., Choi, S., Park, S., & Smolander, K. (2016). Where is current research on blockchain technology? A systematic review. PLoS One, 11(10), e0163477. https://doi.org/10.1371/journal.pone.0163477

Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017). An overview of blockchain technology: Architecture, consensus, and future trends. In Proceedings of the IEEE International Congress on Big Data. https://doi.org/10.1109/BigDataCongress.2017.85