FinDEx: A Synthetic Data Sharing Platform for Financial Fraud Detection

Fabian Karst
University of St.Gallen
fabian.karst@unisg.ch

Mahei Manhai Li
University of Kassel
mahei.li@uni-kassel.de

Jan Marco Leimeister
University of Kassel; University of St.Gallen
janmarco.leimeister@unisg.ch

Abstract

The rising number of financial frauds inflicted in the last year more than 800 billion USD in damages on the global economy. Although financial institutions possess advanced AI systems for fraud detection, the time required to accumulate a sufficient volume of fraudulent data for training models creates a costly vulnerability. Combined with the inability to share fraud detection training data among institutions due to data and privacy regulations, this poses a major challenge. To address this issue, we propose the concept of a synthetic data-sharing ecosystem platform (FinDEx). This platform ensures data anonymity by generating synthesized training data based on each institution's fraud detection datasets. Various synthetic data generation techniques are employed to rapidly construct a shared dataset for all ecosystem members. Using design science research, this paper leverages insights from financial fraud detection literature, data sharing practices, and modular systems theory to derive design knowledge for the platform architecture. Furthermore, the feasibility of using different data generation algorithms such as generative adversarial networks, variational auto encoder and Gaussian mixture model was evaluated and different methods for the integration of synthetic data into the training procedure were tested. Thus, contributing to the theory at the intersection between fraud detection and data sharing and providing practitioners with guidelines on how to design such systems.

Keywords: Synthetic Data, Data Sharing Platform, Data Ecosystem, Financial Services, Fraud Detection, Data Scarcity, Hybrid Intelligence

1. Motivation

The proliferation of digital financial services in recent years has been a crucial enabler in the effort to reduce poverty and promote economic growth. By providing low-cost, faster, and more secure financial services, digital platforms have expanded access to previously unbanked populations (Pazarbasioglu et al., 2020). This trend is particularly pronounced in developing countries, where double-digit increases in digital account ownership have been recorded as it is the only way for individuals to access banking infrastructure (Demirgüç-Kunt et al., 2022).

However, with a bigger number of individuals and companies using digital financial services and a higher reliance on those systems for critical transactions, these platforms get increasingly targeted by malicious actors, resulting in a steep increase in financial fraud in recent years (e.g.: card fraud losses were $32.34 billion in 2022, up 13.8% from 2020 (Nilson, 2022)). Due to the volume of transactions and the complexity of fraud schemes, manual fraud detection methods have been replaced by automated systems for uncovering malicious transactions (Mandal et al., 2016). However, the capability of these tools is often limited by the fast-changing nature of fraud schemes and the limited amount of data available to each financial institution (Al-Hashedi & Magalingam, 2021; Hilal et al., 2022).

Amplified by the limited exchange of transaction data between financial institutions, this poses two societal challenges. First, smaller datasets at each institution restrict the potential of fraud detection systems based on sophisticated supervised deep learning models, as these models require large amounts of data. Second, new fraud schemes cannot be tackled effectively and can spread among banks as they independently need to catch up, giving malicious actors more time to defraud users. Due to the complexity of securely anonymizing interconnected transaction data, a method to share and generate anonymized fraudulent data from various institutions can enhance welfare. By using this shared data to train proven supervised machine learning models, we can rapidly expand fraud detection capabilities across multiple institutions. Thus, our research goal is to conceptualize the design of a platform that allows different financial institutions to share training data across organizations to train fraud detection models more efficiently. This extends the existing literature in two ways. Firstly, it transfers synthetic data sharing into a previously unexplored domain characterized by network transactional data under privacy constraints.

URI: https://hdl.handle.net/10125/106897
978-0-9981331-7-1

Secondly, it enriches the realm of synthetic data generation by benchmarking different architectures against each other.

While many financial institutions already possess both capabilities and resources to share data in-house, current data privacy requirements and regulations prevent them from sharing data with other organizations and, thus, hinder them from taking full advantage of current technological potential. One reason for this is missing guidance in terms of architecture and procedure on how to establish such a system. To address this, our research is guided by the following research question (RQ):

RQ: *How to design a financial transaction data sharing platform (FinDEx) for fraud detection based on synthetic data generation?*

To address the RQ, this paper is structured as follows: In Section 2, we present an overview of privacy and performance in information systems, synthetic data generation and collaboration in financial fraud detection. We then outline our research methodology in Section 3. In Section 4, we diagnose the problem and the design requirements (DR). Before design principles (DP) are derived in the first cycle a new system architecture based on these DPs is proposed in Section 5. In the second elaborated action design research (eADR) cycle, the feasibility of different synthetic data generation and integration methods are evaluated (Section 6). Finally, in Section 7, we discuss the findings and provide a perspective for future work.

2. Theoretical Background

2.1 Privacy and Performance in Information Systems

The increased availability of information in modern societies has driven the performance of machine learning (ML)-based systems (Brynjolfsson & McAfee, 2014) and sparked discussions on data privacy implications (Bélanger & Crossler, 2011). Hereby the ML performance refers to the ability to predict missing information using available data through regression or classification (Agrawal et al., 2018). However, the pursuit of data privacy, enabling individuals or institutions to control their data, often competes with this goal (Bélanger & Crossler, 2011). This is especially true for financial transaction data, which is considered sensitive in both its relationship to customer privacy as well as its importance as a source of proprietary advantage to banks (Y. Wang et al., 2018). Enhancing algorithmic performance by providing ML algorithms with more extensive and diverse data poses heightened risks to data privacy, particularly concerning personal or sensitive business data (B. Liu et al., 2021). Conversely, ensuring high data privacy may hamper algorithmic performance and thus increase losses due to financial fraud.

2.2 Synthetic data generation and its application

Synthetic data generation can be described as a statistical process in which information is extracted from a real data set and transformed into a set of synthetic data that shares the statistical characteristics of the real data but protects its privacy (Raghunathan, 2021). This approach enables widespread dissemination of the valuable information contained within the original data set while mitigating the risk of inadvertent or malicious exposure of sensitive details about the data source/s (Raghunathan, 2021). With new technological capabilities due to the introduction of deep neural networks, synthetic data has been applied in a variety of fields, where it is primarily used to facilitate more efficient and effective development of AI solutions (Lu et al., 2023). Nevertheless, the significance of synthetic data for sharing is resurfacing as privacy concerns intensify in various domains owing to regulatory pressure and customer expectations, alongside the growing necessity for extensive datasets to support cutting-edge ML models (Hittmeir et al., 2019).

2.3 Financial fraud detection

A variety of different financial fraud detection methods exist, with most (81.3%) focusing on bank and insurance fraud. With supervised algorithms, namely, support vector machines and random forests showing superior performance compared to un- or semi-supervised models, they are most frequently employed (Al-Hashedi & Magalingam, 2021). However, they require a substantial amount of labelled training data to deliver optimal performance, which some of the other methods, namely unsupervised methods or expert-based systems can circumvent (Richhariya, 2012).

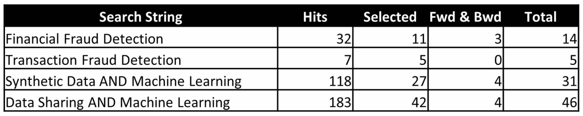
2.4 Data sharing for financial fraud detection

With supervised models delivering superior performance and the capability for sharing even large amounts of data due to technological progress, different approaches have emerged focusing on the improvement of fraud detection through sharing data and/or models. One is the sharing of local data with other institutions, usually on an aggregated level, to exchange recently emerging fraud patterns (Chiu & Tsai, 2004). Another approach is sharing the fraud detection model itself (Dai et al., 2016) or collaboratively training new models (Yang et al., 2019). However, while the first steps are made no widespread implementation of financial data sharing for fraud detection exists. As seen in the previous paragraphs, synthetic data can be used as a substitute for private data and shared without privacy concerns, however, in financial fraud detection, this has not yet been investigated.

3. Research Approach

This paper aims to develop and evaluate DPs for a financial transaction data sharing platform for fraud detection. These DPs with their DRs and MRs, as a nascent design theory, capture a general solution in a class of artefacts (Baskerville et al., 2018), which can be used to guide actions in a wider range of problems, in particular systems where entity-based time series data needs to be shared under privacy restrictions (Hevner et al., 2004). They contribute to the theoretical advancement of the Information Systems (IS) community and provide valuable guidance for practitioners in designing similar artefacts (Baskerville et al., 2018; Sein et al., 2011). This study follows the eADR method (Mullarkey & Hevner, 2019), which enhances the building-intervention-evaluation process, dividing it into distinct cycles of diagnosis, design, implementation, and evolution. Within each of these cycles, various activities are conducted iteratively, thus allowing for continuous refinement of the artefact's design based on real-world settings (Mullarkey & Hevner, 2019; Sein et al., 2011). Since the eADR approach requires integration into an organisational context, the project was conducted in collaboration with a major bank (5 million customers, $15 billion assets under management) from a developing country, which rapidly scaled its digital transaction infrastructure and is now looking for new ways to tackle transaction fraud.

In the next paragraph, the activities in each cycle are introduced. First, the ADR project starts with a problem-centred diagnosis cycle, focusing on stakeholder requirements. This was done by conducting a systematic literature review on data sharing, synthetic data and financial fraud detection in the AIS basket of eight, proceedings of prominent IS conferences (ICIS, ECIS, HICSS) as well as the journal ACM Computing Surveys, resulting in the selection of papers described below:



Search String	Hits	Selected	Fwd & Bwd	Total
Financial Fraud Detection	32	11	3	14
Transaction Fraud Detection	7	5	0	5
Synthetic Data AND Machine Learning	118	27	4	31
Data Sharing AND Machine Learning	183	42	4	46

Figure 1: Results of systematic literature search

By conducting four individual semi-structured interviews with employees at different levels at our partner bank, who are engaged in data sharing initiatives or data analytics and machine learning projects, we gained further insights. Next, we iterated the first round of the cycle. In the design phase, we formulated the initial set of DPs. These principles were translated into a system architecture during the implementation phase, specifying the material properties like algorithms and interaction layers. Subsequently, an evaluation was conducted, involving feedback from academics and industry experts. The outcomes helped evaluate the feasibility of the initial design and led to the refinement of selected DPs in the second iteration. In cycle 2 we conducted a literature review identifying suitable algorithms for synthetic financial transaction data creation and based on them instantiated a prototype which was subsequently evaluated on a publicly available credit card transaction dataset. Throughout the eADR cycles, we iteratively abstracted the requirements, DPs, and system features. Thus, our main theoretical contributions lie in the abstracted artefacts, particularly the DPs, which are elaborated in section 5.1 and continuously refined throughout the paper.

4. Diagnosis

The diagnosis phase consists of two tasks, understanding the problem and solution domain and defining the requirements of the platform. First, we positioned our eADR project within the domain of fraud detection for financial transactions. After conducting a literature review on financial fraud detection, the limited availability of data was identified as the most prominent challenge in the field. The reason for this is the data’s sensitivity, which makes it subject to laws in different jurisdictions that prevent it from leaving the country or being shared at all (Ryman-Tubb et al., 2018). Even if institutions are able and willing to share such data it needs to be guaranteed that the original data cannot be recreated or imputed which is difficult (Shokri, 2015). This lack of available data is further aggravated by the highly imbalanced nature of datasets (large datasets needed for a sufficient number of samples in the minority class) as well as the fast-changing nature of fraudulent patterns (Hilal et al., 2022; Ryman-Tubb et al., 2018).

Looking at potential solutions, we combined our knowledge from the financial fraud detection literature with insights about the data sharing barriers and requirements in healthcare, an area with very active research in sharing sequential data as well as similar data properties regarding structure, privacy and volume (Fang et al., 2017; Martínez et al., 2022; Saenyi & Ademaj, 2022). By consolidating these perspectives alongside the information obtained from interviews with our project partners, we formulated two meta-requirements (MR) that any solution must adhere to. MR1 emphasizes the ease of data sharing between financial institutions, encompassing both technical and legal aspects. The imperative for technical ease of use was informed by insights drawn from the medical field, where challenges related to tool availability and varying data standards were identified as hindrances to data sharing (van Panhuis et al., 2014). Considering the legal dimension in platform usability was primarily motivated by the literature highlighting diverse regulatory requirements across jurisdictions, as observed in exploring existing approaches to sharing financial transaction data (Blake et al., 2019). MR2 highlights the necessity of improved fraud detection performance as a result of sharing data. This requirement emanated from discussions with our partner regarding their goal of establishing a data-sharing platform and from our literature review, which revealed the consensus that existing fraud detection approaches could benefit from improved data availability (Hilal et al., 2022; Ryman-Tubb et al., 2018). Next, we refined the MRs into more specific design requirements (DRs), drawing from literature as well as our project partners.

To incentivize users to participate in data-sharing, costs, setup as well and reoccurring, need to be as low as possible, which is reflected in MR1 and propagates into DR1 and DR2. With different data structures and standards at different banks (Major & Mangano, 2020), a data platform needs to be flexible enough to accommodate various input data structures (DR1). This is particularly important when considering that data should be regularly updated and the cost for these updates needs to be as low as possible. Furthermore, data privacy standards imposed by regulators and internal policies must be upheld. Our interviews revealed, that in the context of our partner institutions, this means, that all real data must be processed locally within the financial institution (DR2). From a data-centric perspective, the performance of ML methods can be enhanced by increasing the volume of training data available (Sun et al., 2017). Thus, MR2 can be achieved through the data platform by enabling the combination of data from multiple sources, making it accessible as a unified data source (DR3). Given the absence of a dominant fraud detection algorithm in the literature, and the insight from our interviews that banks prefer their own custom solutions, the data platform must support diverse types of algorithms (DR4). Additionally, the imbalanced nature of fraud data necessitates tools on the platform to address data imbalances through filtering, oversampling, and undersampling (DR7), as most ML algorithms perform better on balanced datasets (Longadge & Dongre, 2013). As fraud patterns change quickly when discovered, the timely integration of recent fraud patterns into fraud detection algorithms is crucial (Zhu et al., 2021). As this is utterly important two DRs were dedicated to achieving this. Firstly, institutions should have the capability to automatically update the data, ensuring that the dataset incorporates the most recent fraud patterns (DR5). This not only aligns with MR1 by enhancing user convenience and reducing the need for frequent user inputs but also guards against model drift. However, even with automatic updates, the dataset may still be dominated by outdated fraud patterns, posing a risk to the algorithms (Paleyes et al., 2023). Therefore, users should be able to incorporate pattern-based artificial data into the platform (DR6). Allowing the data platform to benefit from expert domain knowledge which is not yet reflected in the data (Richhariya, 2012). After having defined the problem as well as the solution space and outlined our requirements, we can now commence the first design, implementation and evaluation cycle.

5. Cycle 1: DPs and system architecture for synthetic data sharing

5.1 Design

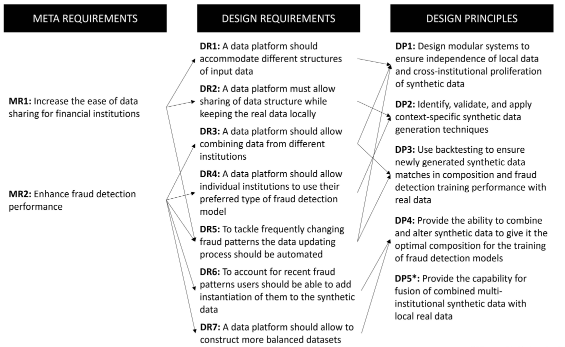
In our first design phase, our primary emphasis was on identifying the foundational DPs. Building on the DRs derived in the previous section and following the recommendations of Chandra et al. (2015), we created DPs that "provide prescriptive knowledge about action and an artefact's material properties in terms of both form and function". Furthermore, to ground these artefacts in practical relevance, expert interviews with our partners were conducted to justify the DPs derived from the literature. Figure 2 depicts which MRs and DRs influenced which DP can be seen at the end of the section.

DP1 - Modular systems design to ensure independence of local data and cross-institutional proliferation of synthetic data. To address DR1 and DR3, the data platform must possess the capability to process data from diverse sources, while enabling the integration of this data for synthetic data generation. Drawing upon the principles of modular systems theory (Tiwana et al., 2010), institutions are granted flexibility in designing their module structures while adhering to a standardized representation, thereby ensuring that the data can be exchanged with the platform. Additionally, once the initial setup is complete, enabling automated data updating becomes straightforward, as all computations can be performed locally, without the need for sensitive data to be transmitted outside the local system. This capability fulfils the requirements outlined in DR5.

DP2 - Apply generative adversarial networks (GANs) to generate synthetic transaction data: The requirement for fraud detection algorithms to be trained on transaction-level data (Hilal et al., 2022) and DR4, which requires users to train different types and variations of algorithms the platform must provide the user with this low-level data. However, sharing transaction-level data poses challenges due to regulatory constraints (Blake et al., 2019) and internal policies mandating its local storage (DR2). As anonymization is not able to preserve both data utility and privacy (Loukides et al., 2010), we propose to solve this challenge by using GANs due to their unique ability to learn patterns in data and generate synthetic data nearly indistinguishable from the original data (Walia et al., 2020). This enables us to preserve real data locally while sharing only the privacy-preserving GAN-generated data within the data sharing ecosystem. This data can then be merged with synthetic data from other institutions and allows the training of fraud detection models on the combined dataset. This approach ensures the confidentiality of sensitive data while empowering the ecosystem to enhance fraud detection capabilities by training algorithms with substantial volumes of high-quality data.

DP3 - Use back-testing to ensure newly generated synthetic data matches in composition and fraud detection training performance with real data: To facilitate the seamless integration of data from multiple institutions (DR3) and enable frequent system updates without human intervention (DR5), it is essential to establish a robust quality control mechanism. This mechanism serves to uphold the integrity of the data introduced into the ecosystem, as only a few bad data points can have tremendous effects on ML models (Chakravarty et al., 2020). One approach to achieve this is by implementing a back-testing procedure, whereby the performance of the generated synthetic data is evaluated against the corresponding local real data. This ensures that the synthetic data accurately captures the underlying patterns (Dankar et al., 2022). Furthermore, exploring the implementation of a consensus mechanism among datasets could be beneficial. Such a mechanism would enforce consistency and coherence in the synthetic data generated across the ecosystem, enhancing the overall quality of the shared data.

DP4 - Provide the ability to combine and alter synthetic data to give it the optimal composition for the training of fraud detection models: To further enhance model performance, a data-sharing platform should be designed to provide users with the ability to alter and extend the existing data to create the right data for their use case. Especially in fraud detection class balance is a challenge, resulting in the requirement, that a data platform should be able to provide more balanced datasets (DR7). This can be accomplished by equipping users with advanced filtering options or enabling them to manipulate the existing data through techniques such as under or over-sampling. This can be accomplished by equipping users with advanced filtering options or enabling them to manipulate the existing data through techniques such as under or over-sampling (Lopez-Rojas & Axelsson, 2012).


META REQUIREMENTS
DESIGN REQUIREMENTS
DESIGN PRINCIPLES
MR1: Increase the ease of data sharing for financial institutions
MR2: Enhance fraud detection performance
DR1: A data platform should accommodate different structures of input data
DR2: A data platform must allow sharing of data structure while keeping the real data locally
DR3: A data platform should allow combining data from different institutions
DR4: A data platform should allow individual institutions to use their preferred type of fraud detection model
DR5: To tackle frequently changing fraud patterns the data updating process should be automated
DR6: To account for recent fraud patterns users should be able to add instantiation of them to the synthetic data
DR7: A data platform should allow to construct more balanced datasets
DP1: Design modular systems to ensure independence of local data and cross-institutional proliferation of synthetic data
DP2: Identify, validate, and apply context-specific synthetic data generation techniques
DP3: Use backtesting to ensure newly generated synthetic data matches in composition and fraud detection training performance with real data
DP4: Provide the ability to combine and alter synthetic data to give it the optimal composition for the training of fraud detection models
DP5*: Provide the capability for fusion of combined multi-institutional synthetic data with local real data

* DP was added in iteration 2

Figure 2: Relationship between MRs, DRs and DPs

5.2 Implementation

Based on the DRs, and DPs, we present a multi-layered platform architecture for a synthetic transaction data-sharing platform. While the local processing layer is implemented at every institution, the synthetic data generation layer as well as the fraud detection layer are centralized. An overview of this architecture mapped with corresponding DPs can be seen in Figure 3.

Local Processing Layer: The local processing layer is modular and situated at every financial institution (DP1). Here the GAN models are trained on sensitive transaction data to produce accurate synthetic representations of this data (DP3). Furthermore, the conversion of a local data format to the data standard the synthetic data needs to conform to is enforced. Moreover, back-testing is done to ensure data quality while guaranteeing that the real data never leaves the local environment (DP2).

Global Data Layer: Contrary to the previous layers, the synthetic data layer is not situated at a specific institution. Instead, this layer is where synthetic data is merged and modifications to the data composition through the addition of pattern-based data generators or the artificial rebalancing of different classes can be achieved (DP4).

Fraud Detection Layer: This layer is accessible to any participating company allowing them to access the synthetically generated data and modify it to fit their models by providing capabilities to subsegment and alter data, making it optimal for their custom fraud detection models.

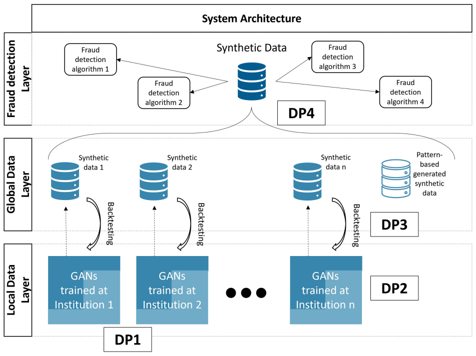

System Architecture
Fraud detection Layer
Global Data Layer
Local Data Layer
Synthetic Data
Fraud detection algorithm 1
Fraud detection algorithm 2
Fraud detection algorithm 3
Fraud detection algorithm 4
DP4
Synthetic data 1
Synthetic data 2
Synthetic data n
Pattern-based generated synthetic data
Backtesting
Backtesting
Backtesting
DP3
GANs trained at Institution 1
GANs trained at Institution 2
GANs trained at Institution n
DP2
DP1

Figure 3. System Architecture

5.3 Evaluation

After deriving the system architecture from our DPs, we presented both to three experts from different departments of our partner institution as well as five academics in the field. The feedback gathered from the experts was overall positive and especially the use of modular system design to ensure reduced complexity of the eco-system and complete control of the local layer by the single institutions was highly appreciated. Furthermore, the proposed architecture was seen as a good first outline to create a prototype, only the computational resources required to train the synthetic data generation models for frequent updates were raised as a concern. When discussing the proposed DPs as well as architecture with academic experts from the field of design science research, data sharing and fraud detection, DP2 was criticized for multiple reasons. Firstly, the limitation to a single technology for data generation (GANs) was seen as being too restrictive and limiting the system's adaptability to different domains. Furthermore, concerns emerged about the feasibility of generating financial transaction data from limited local data and the utility of synthetic data to benefit fraud detection performance.

6. Cycle 2: Synthetic financial transaction data generation

6.1 Design

To address the expert feedback from design cycle one. The second design cycle focuses on the refinement and extension of DP2. Based on the comments and thus it was adjusted to:

DP2 - Identify, validate, and apply context-specific synthetic data generation techniques so that it is no longer restricted to a single method for generating the data and includes the necessary validation of selected techniques to obtain optimal data generation and in return fraud detection performance. To validate DP2 and identify a suitable method to generate synthetic financial transaction data, a rigorous literature review following vom Brocke et al. (2009) was conducted. In the first step top publications regarding synthetic data generation were reviewed, resulting in the following search string: *("synthetic data generation" OR "artificial data generation") AND ("transaction data" OR "time series data").* Next, this search string was used to identify English journal- and conference papers published after 2018 in the following databases: ScienceDirect, Ebscohost, SpringerLink, IEEE Xplore and Aisnet. This resulted in 289 hits for which title, abstract and keywords were evaluated. During the review, papers without generated data or lacking a description of the generation method were excluded, leaving 47 papers. After further analysis, 8 additional papers were included through forward and backward search, bringing the total to 55 papers. From these papers, 46 distinct algorithms were extracted and grouped by their underlying algorithm type based on referenced papers. Consequently, GANs emerge as the primary underlying mechanism (used by 55.3% of methods), in generating synthetic transaction data. GAN models work by creating two neural networks that learn by competing in synthesizing and identifying synthetic data and thus, once trained, are able to generate synthetic data that is indistinguishable from real one (Goodfellow et al., 2014). However, different implementations exist. To allow for variations between the algorithms tested and address the high degree of similarity between the different GAN architectures, we decided to only include two of them in our comparison.: CTGAN (L. Xu et al., 2019), which was the most mentioned algorithm and is a representative of GANs taking only dependencies between attributes, but not samples, into account and TimeGAN (Yoon et al., 2019) (ranked third by mentions) which takes the temporal dimension of the data into account. To tackle the criticism from cycle one, we extend our overview beyond GAN-based architectures. The most frequently mentioned implementations using other algorithm types were, Gaussian mixture models, which learn the distribution for each attribute and then generate new samples by drawing from these (S. Xu et al., 2021) and TVAE (Ishfaq et al., 2023), a variational autoencoder (VAE), which works by learning to compress and decompress data into a low-dimensional space and then use the decompress module to synthesize new data. The literature predominantly focuses on applying these algorithms to health records (Xing et al., 2022), with limited exploration in other domains such as traffic data (S. Xu et al., 2021) and IoT data (X. Liu et al., 2019), however, none of the papers identified has examined the application of these methods to financial transaction data. Furthermore, while (Weldon et al., 2021), synthetic data alone suffices to achieve performance gains, others, such as (Frid-Adar et al., 2018), show that adding real-world data will improve performance. Thus, the optimal algorithm for financial transactions and the necessity of combining synthetic with real data remain unclear.

6.2 Implementation

In this section, we operationalized the derived DPs into a prototype system in Python using the synthcity library (Qian et al., 2023). Looking at the system architecture from design cycle one, the local and global data layer was implemented, resulting in a platform that allows data ingestion, synthetic data generation and data sharing. Only the fraud detection layer and graphical user interface are still missing from the platform. Furthermore, the platform was implemented in a way that allows to easily switch between different synthetic data generation methods, thus allowing an easy evaluation of the most suitable algorithm for financial transaction data generation, contributing to the further refinement of DP2. Lastly, the current prototype of the platform already instantiates the first version of DP4 allowing for different combinations of synthetic- and real data.

6.3 Evaluation

In the second evaluation, the different synthetic data generation approaches outlined before will be compared. However, as financial transaction data from our partner institutions is not available to us yet, the credit card transaction dataset from the IEEE-CIS Kaggle competition will be used. This dataset was chosen as credit card transactions are not only financial data but also represent the spending patterns of users which make them comparable to financial transactions. Furthermore, this dataset was the only one identified, which allowed matching transactions to users, allowing for models expecting time series data to be trained. However, choosing this dataset also comes with limitations, such as the limited observation period (6 months), a large number of obscured features as well, and the ability to only identify senders of payments but not receivers. As we aim to analyze the benefits of sharing synthetic data among financial institutions, we split the dataset by credit card provider, creating four distinct datasets. A client distribution analysis for each provider revealed marked differences, consistent with expected variations in multi-institutional bank datasets. After obtaining a suitable dataset, we define our evaluation criteria. For this, the fraud detection model, specifically a random forest classifier (commonly used in fraud detection as per Al-Hashedi & Magalingam (2021)), trained on either real data, synthetic data or a combination of both, will be assessed using the ROC AUC score on a holdout dataset. The ROC AUC score was chosen as it provides a comprehensive evaluation of the classifier's performance across different levels of sensitivity and specificity and is frequently used in the literature (Sun et al., 2023). Furthermore, the evaluation will be conducted in two stages with the first one covering the performance of individual synthetic data generation algorithms, thus helping us to validate DP2 and the second looking at the overall benefit of the proposed fraud detection platform. In the first stage, the focus is on the performance of the different algorithms on financial transaction data. Taking a closer look, GMMs (ROC AUC score: 0.52), as well as TimeGANs (ROC AUC score: 0.5), did not perform well, which can be explained by the composition of the data. While GMMs struggled with the high dimensionality of the data, TimeGAN had problems with short transaction chains (below 2 transactions per user) due to the short observation period. While CTGAN (ROC AUC score: 0.59) performed a little better, TVAE (ROC AUC score: 0.89) was able to deal with these difficult data conditions and performed particularly well in situations where little training data was available (the datasets for "Discover" and "American Express" each contained less than 10.000 transactions). Thus, confirming that the selection of the right algorithm is crucial and therefore validating DP2. The second-stage evaluation assessed the advantage of training on shared synthetic data versus isolated real data. Figure 4 compares the performance of models trained on isolated real data, isolated synthetic data, shared synthetic data as well as shared synthetic data combined with isolated real data. Models trained solely on synthetic data from one source underperformed compared to those trained on real data. Yet, combining synthetic data from multiple sources led to a further performance drop, likely due to varying fraud cases across providers, which dilutes relevant patterns. However, merging synthetic with real data for each institution boosted performance, increasing the ROC AUC score by 1%.

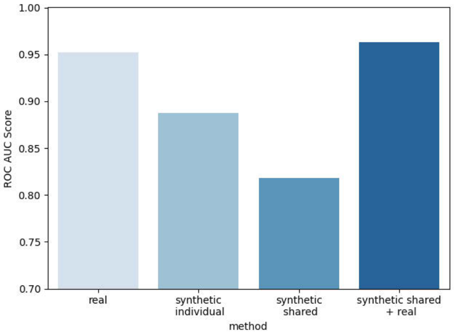


Figure 4: Comparison between synthetic and real data combinations

To better understand the impact of this improvement we can look at the recall, or what percentage of fraudulent cases are identified. Using synthetic and real data combined, we find that 2.14% more true positives are detected. Combining this with an estimated number of 24.16 million fraudulent card transactions per year (European Central Bank, 2021), the improved model would have detected about half a million additional transactions. Thus, showing the benefit of our platform. However, this fusion of shared synthetic data with local real data is not yet reflected in any design principle, however, the evaluation showed it to be a critical principle of our proposed design. Thus, a new DP:

DP5 - Provide the capability for fusion of combined multi-institutional synthetic data with local real data was created, incorporating this important design criterion. Based on this the proposed system architecture was revised, which can be seen below:

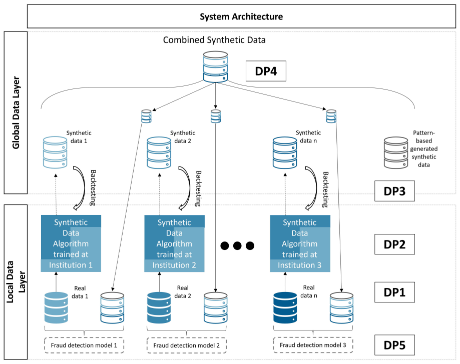


Figure 5: Updated System Architecture

7. Discussion

This research paper aims to create DPs for a synthetic data sharing platform that allows financial institutions to exchange transaction data to increase fraud detection performance while protecting client privacy. To create this artefact, we followed the process of eADR, with this paper covering the first two iterations. Starting in the diagnosis stage our study contributes to descriptive knowledge concerning the problem space by identifying data scarcity in combination with the inability to share data as a major hurdle for financial fraud detection. Furthermore, we contribute towards the exploration of the solution space by identifying two main dimensions, fraud detection performance and ease of use (from an interface as well as a legal perspective), which were derived from literature, interviews, and insights from more mature fields. During our first iteration cycle, our research contributed to the field of IS research by generating prescriptive knowledge concerning the solution space by offering a set of DPs for designing a synthetic data sharing platform. While DP1 contributes to the design of data sharing platforms by extending the local–global layer logic seen in federated learning (Yang et al., 2019) toward synthetic data sharing, DP2 contributes to the literature on synthetic data generation (Pathare et al., 2023) by transferring existing algorithms to financial transaction data. Furthermore, DP3 is an instantiation of an efficient mechanism to ensure data quality in a multi-party data sharing scenario (Freudiger et al., 2014). Furthermore, during the implementation stages, a blueprint architecture was created and refined, providing guidance on how to implement a system based on the constructed DPs to practitioners. By abstracting these implementation cycles, our initial DPs were further refined, and an additional design principle was added to incorporate the newly gained insights about the necessity of integrating shared synthetic data to the local real data context for optimal performance. Thus, the MRs, DRs, and DPs developed in our study are the primary contributions, serving as a nascent design theory. They not only deepen the understanding of the solution domain but also offer practical guidance for addressing challenges where data needs to be shared with privacy restrictions. Thus, extending beyond the financial domain and tackling challenges in many data sharing communities (Susha et al., 2019). However, some limitations need to be addressed in future studies. Firstly, the evaluation of our synthetic data sharing approach is limited by data availability, necessitating evaluation with additional and more comprehensive datasets to increase the statistical significance. Secondly, a separate design cycle is needed to analyze platform usability, which was beyond the scope of this paper. While the focus of this research was on the performance and thus viability of sharing synthetic data future research should consider legal challenges of novel technologies (Dickhaut et al. 2023), organizational incentive structures or novel value-driven data structures for model training (Li et al. 2018).

8. Conclusion

With digital financial fraud becoming more prevalent and the shift of critical transactions towards digital financial services providers, better fraud detection solutions are needed. Having identified synthetic data generation as a potential solution, for data scarcity when building fraud detection models, this paper investigates how a synthetic data sharing platform for financial transaction data needs to be designed. Thus, a set of DPs was developed and evaluated through practitioners, researchers, and experiments. Our study's findings are both feasible and practical, with the potential to make a tangible impact on our partner companies as well as society as a whole by combating financial fraud.

9. References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press.

Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review*, *40*, 100402.

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design Science Research Contributions: Finding a Balance between Artifact and Theory. *Journal of the Association for Information Systems*, *19*(5).

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, *35*(4), 1017–1042.

Blake, M., McWaters, J., & Galaski, R. (2019). *The Next Generation of Data-Sharing in Financial Services* (p. 33) [White Paper]. World Economic Forum.

Chakravarty, S., Demirhan, H., & Baser, F. (2020). Fuzzy regression functions with a noise cluster and the impact of outliers on mainstream machine learning methods in the regression setting. *Applied Soft Computing*, *96*, 106535.

Chandra, L., Seidel, S., & Gregor, S. (2015). Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions. *2015 48th Hawaii International Conference on System Sciences*, 4039–4048.

Chiu, C.-C., & Tsai, C.-Y. (2004). A Web services-based collaborative scheme for credit card fraud detection. *IEEE International Conference on E-Technology, e-Commerce and e-Service, 2004. EEE '04. 2004*, 177–181.

Dai, Y., Yan, J., Tang, X., Zhao, H., & Guo, M. (2016). Online Credit Card Fraud Detection: A Hybrid Framework with Big Data Technologies. *2016 IEEE Trustcom/BigDataSE/ISPA*, 1644–1651.

Dankar, F. K., Ibrahim, M. K., & Ismail, L. (2022). A Multi-Dimensional Evaluation of Synthetic Data Generators. *IEEE Access*, *10*, 11147–11158.

Demirgüç-Kunt, A., Klapper, L., Singer, D., & Ansar, S. (2022). *The Global Findex Database 2021—Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19*. International Bank for Reconstruction and Development / The World Bank.

European Central Bank. (2021). *Seventh report on card fraud. 2021*.

Dickhaut, E., Janson, A., & Leimeister, J. M. (2022). Conceptualizing Design Knowledge in IS Research – A Review and Taxonomy of Design Knowledge Properties. Hawaii International Conference on System Sciences.

Fang, R., Pouyanfar, S., Yang, Y., Chen, S.-C., & Iyengar, S. S. (2017). Computational Health Informatics in the Big Data Age: A Survey. *ACM Computing Surveys*, *49*(1), 1–36.

Freudiger, J., Rane, S., Brito, A. E., & Uzun, E. (2014). Privacy Preserving Data Quality Assessment for High-Fidelity Data Sharing. *Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security*, 21–29. https://doi.org/10.1145/2663876.2663885

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). Synthetic data augmentation using GAN for improved liver lesion classification. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 289–293.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Networks* (arXiv:1406.2661). arXiv.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105.

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. *Expert Systems with Applications*, *193*, 116429.

Hittmeir, M., Ekelhart, A., & Mayer, R. (2019). On the Utility of Synthetic Data: An Empirical Evaluation on Machine Learning Tasks. *Proceedings of the 14th International Conference on Availability, Reliability and Security*, 1–6.

Ishfaq, H., Hoogi, A., & Rubin, D. (2023). *TVAE: Triplet-Based Variational Autoencoder using Metric Learning* (arXiv:1802.04403). arXiv. http://arxiv.org/abs/1802.04403

Li, M., & Peters, C. (2018). Reconceptualizing Service Systems – Introducing Service System Graphs.

Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F., & Lin, Z. (2021). When Machine Learning Meets Privacy: A Survey and Outlook. *ACM Computing Surveys*, *54*(2), 31:1-31:36.

Liu, X., Iftikhar, N., Huo, H., Li, R., & Nielsen, P. S. (2019). Two approaches for synthesizing scalable residential energy consumption data. *Future Generation Computer Systems*, *95*, 586–600.

Longadge, R., & Dongre, S. (2013). *Class Imbalance Problem in Data Mining Review* (arXiv:1305.1707). arXiv.

Lopez-Rojas, E. A., & Axelsson, S. (2012). *Money Laundering Detection using Synthetic Data*. 8.

Loukides, G., Gkoulalas-Divanis, A., & Shao, J. (2010). Anonymizing Transaction Data to Eliminate Sensitive Inferences. In P. G. Bringas, A. Hameurlain, & G. Quirchmayr (Eds.), *Database and Expert Systems Applications* (pp. 400–415). Springer.

Lu, Y., Wang, H., & Wei, W. (2023). *Machine Learning for Synthetic Data Generation: A Review* (arXiv:2302.04062). arXiv.

Major, T., & Mangano, J. (2020). *Modernising Payments Messaging: The ISO 20022 Standard*. Reserve Bank of Australia.

Mandal, P., Mahata, A., Biswas, B., Pal, U., Sarfaraj, M., & Barman, S. (2016). A complete literature review on financial fraud detection applying data mining techniques. *International Journal of Trust Management in Computing and Communications*, *3*, 336.

Martínez, A. L., Pérez, M. G., & Ruiz-Martínez, A. (2022). A comprehensive review of the state of the art on security and privacy issues in Healthcare. *ACM Computing Surveys*, 3571156.

Mullarkey, M. T., & Hevner, A. R. (2019). An elaborated action design research process model. *European Journal of Information Systems*, *28*(1), 6–20.

Nilson, H. S. (2022). Card Fraud Losses Worldwide. *The Nilson Report*, *1232*.

Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2023). Challenges in Deploying Machine Learning: A Survey of Case Studies. *ACM Computing Surveys*, *55*(6), 1–29.

Pathare, A., Mangrulkar, R., Suvarna, K., Parekh, A., Thakur, G., & Gawade, A. (2023). Comparison of tabular synthetic data generation techniques using propensity and cluster log metric. *International Journal of Information Management Data Insights*, *3*(2), 100177.

Pazarbasioglu, C., Mora, A. G., Uttamchandani, M., Natarajan, H., Feyen, E., & Saal, M. (2020). *Digital Financial Services* (p. 54). World Bank Group.

Qian, Z., Cebere, B.-C., & van der Schaar, M. (2023). *Synthcity: Facilitating innovative use cases of synthetic data in different data modalities* (arXiv:2301.07573). arXiv.

Raghunathan, T. E. (2021). Synthetic Data. *Annual Review of Statistics and Its Application*, *8*(1), 129–140.

Richhariya, P. (2012). A Survey on Financial Fraud Detection Methodologies. *International Journal of Computer Applications*, *45*.

Ryman-Tubb, N. F., Krause, P., & Garn, W. (2018). How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. *Engineering Applications of Artificial Intelligence*, *76*, 130–157.

Saenyi, B., & Ademaj, G. (2022). Systematic Literature Review Data Standardization in Health Information Systems

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. *MIS Quarterly*, *35*(1), 37–56.

Shokri, R. (2015). Privacy Games: Optimal User-Centric Data Obfuscation. *Proceedings on Privacy Enhancing Technologies*, *2015*(2), 299–315.

Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). *Revisiting Unreasonable Effectiveness of Data in Deep Learning Era*. 843–852.

Sun, C., van Soest, J., & Dumontier, M. (2023). Generating synthetic personal health data using conditional generative adversarial networks combining with differential privacy. *Journal of Biomedical Informatics*, 104404.

Susha, I., Grönlund, Å., & Van Tulder, R. (2019). Data driven social partnerships: Exploring an emergent trend in search of research challenges and questions. *Government Information Quarterly*, *36*(1), 112–128.

Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Research Commentary—Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. *Information Systems Research*, *21*(4), 675–687.

van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J., Heymann, D., & Burke, D. S. (2014). A systematic review of barriers to data sharing in public health. *BMC Public Health*, *14*(1), 1144.

Walia, M., Tierney, B., & McKeever, S. (2020). *Synthesising Tabular Data using Wasserstein Conditional GANs with Gradient Penalty*.

Wang, Y., Adams, S., Beling, P., Greenspan, S., Rajagopalan, S., Velez-Rojas, M., Mankovski, S., Boker, S., & Brown, D. (2018). Privacy Preserving Distributed Deep Learning and Its Application in Credit Card Fraud Detection. *2018 17th IEEE TrustCom/BigDataSE*, 1070–1078.

Weldon, J. C., Ward, T., & Brophy, E. (2021). Generation of Synthetic Electronic Health Records Using a Federated GAN. *ArXiv*.

Xing, X., Wu, H., Wang, L., Stenson, I., Yong, M., Del Ser, J., Walsh, S., & Yang, G. (2022). *Non-Imaging Medical Data Synthesis for Trustworthy AI: A Comprehensive Survey* (arXiv:2209.09239). arXiv.

Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). *Modeling Tabular data using Conditional GAN* (arXiv:1907.00503). arXiv.

Xu, S., Marwah, M., Arlitt, M., & Ramakrishnan, N. (2021). STAN: Synthetic Network Traffic Generation with Generative Neural Models. In G. Wang, A. Ciptadi, & A. Ahmadzadeh (Eds.), *Deployable Machine Learning for Security Defense* (pp. 3–29). Springer International Publishing.

Yang, W., Zhang, Y., Ye, K., Li, L., & Xu, C.-Z. (2019). FFD: A Federated Learning Based Method for Credit Card Fraud Detection. In K. Chen, S. Seshadri, & L.-J. Zhang (Eds.), *Big Data – BigData 2019* (pp. 18–32). Springer International Publishing.

Yoon, J., Jarrett, D., & van der Schaar, M. (2019). Time-series Generative Adversarial Networks. *Advances in Neural Information Processing Systems*, *32*.

Zhu, X., Ao, X., Qin, Z., Chang, Y., Liu, Y., He, Q., & Li, J. (2021). Intelligent financial fraud detection practices in post-pandemic era. *The Innovation*, *2*(4), 100176.