10.48047/jocaaa.2024.32.02.79

Privacy-Preserving Synthetic Banking Data Generation Using Generative Adversarial Networks

Adarsh Naidu
Independent Researcher, USA
Email : adarshnaidu2910@gmail.com

Abstract

The rapid digital transformation of the banking and financial services sector has led to the large-scale collection and utilization of customer transaction data for tasks such as fraud detection, credit risk assessment, customer behavior modeling, and financial forecasting. While data-driven techniques, particularly deep learning models, have shown remarkable success in these applications, their effectiveness is fundamentally dependent on access to large volumes of high-quality data. However, the use of real banking data introduces severe privacy, security, and regulatory challenges due to the sensitive nature of financial information and strict compliance requirements imposed by regulations such as data protection and financial privacy laws. These constraints significantly limit data sharing, collaborative research, and model development.

To address this challenge, this paper proposes a privacy-preserving synthetic banking data generation framework based on Generative Adversarial Networks (GANs). The proposed approach aims to generate high-fidelity synthetic banking datasets that preserve the statistical properties and utility of real financial data while preventing the disclosure of sensitive customer information. By leveraging adversarial training between a generator and a discriminator, the model learns complex transaction patterns, temporal dependencies, and feature correlations inherent in real-world banking data without directly exposing individual records. Additionally, privacy-preserving mechanisms are incorporated into the training process to mitigate risks such as membership inference and data reconstruction attacks.

Extensive experimental analysis demonstrates that the generated synthetic data closely matches real banking datasets in terms of distributional similarity, correlation structures, and downstream task performance, while significantly reducing privacy leakage. The results indicate that GAN-based synthetic data generation offers a viable and scalable solution for enabling secure data sharing and advanced analytics in the banking domain. This work contributes toward bridging the gap between data utility and privacy, facilitating trustworthy artificial intelligence development for financial applications.

Keywords : Synthetic Data Generation, Privacy Preservation, Generative Adversarial Networks, Banking Data, Financial Privacy, Data Security, Machine Learning, Deep Learning

I. Introduction

The banking and financial services industry has undergone a profound transformation with the adoption of data-driven technologies and artificial intelligence-based decision-making systems.

10.48047/jocaaa.2024.32.02.79

Modern banking operations generate massive volumes of structured and semi-structured data, including transaction histories, account balances, customer demographics, loan records, and behavioral logs. These datasets play a critical role in enabling intelligent applications such as fraud detection, anti-money laundering systems, credit scoring models, personalized financial services, and risk management frameworks. As machine learning and deep learning models continue to advance, the demand for large-scale, diverse, and representative banking datasets has grown substantially.

Despite the potential benefits, the use of real banking data poses significant challenges. Financial datasets inherently contain highly sensitive personal and transactional information, making them prime targets for data breaches, misuse, and unauthorized inference. Regulatory frameworks and privacy laws impose strict constraints on how such data can be stored, processed, and shared. Consequently, financial institutions are often unable to share data even for legitimate purposes such as academic research, inter-bank collaboration, or benchmarking of machine learning models. This data isolation hampers innovation, slows research progress, and limits the reproducibility of experimental results.

Traditional privacy-preserving techniques such as anonymization, masking, and aggregation have been widely employed to address these concerns. However, numerous studies have shown that these methods are often insufficient, as adversaries can re-identify individuals by exploiting auxiliary information or correlation patterns. Moreover, excessive anonymization frequently degrades data utility, resulting in datasets that are no longer suitable for advanced analytics or model training. This fundamental trade-off between data utility and privacy has motivated the exploration of alternative approaches that can balance both objectives effectively.

Synthetic data generation has emerged as a promising solution to this problem. Instead of releasing real customer records, synthetic datasets aim to replicate the statistical characteristics and structural relationships of original data while containing no direct linkage to real individuals. In the banking domain, high-quality synthetic data can enable safe data sharing, support machine learning model development, and facilitate regulatory-compliant experimentation. However, generating realistic synthetic banking data is a non-trivial task due to the complex, high-dimensional, and highly correlated nature of financial transactions, as well as the presence of rare but critical events such as fraud.

Recent advancements in deep generative models, particularly Generative Adversarial Networks (GANs), have demonstrated remarkable capability in modeling complex data distributions. GANs consist of two competing neural networks—a generator and a discriminator—trained through an adversarial process that encourages the generation of increasingly realistic samples. While GANs were initially proposed for image synthesis, their adaptability has led to successful applications in tabular data generation, time-series modeling, and structured datasets, making them well-suited for financial data synthesis.

However, directly applying GANs to banking data raises additional concerns regarding privacy leakage. Without proper safeguards, generative models may memorize training samples or inadvertently expose sensitive patterns, leading to potential inference attacks. Therefore, ensuring

10.48047/jocaaa.2024.32.02.79

that synthetic data generation is not only realistic but also privacy-preserving is a critical requirement for deployment in financial settings.

In this work, we present a GAN-based framework for privacy-preserving synthetic banking data generation that addresses both data utility and privacy protection. The proposed approach is designed to learn the underlying distribution of banking transactions while minimizing the risk of sensitive information disclosure. By incorporating privacy-aware training strategies and rigorous evaluation metrics, the framework ensures that the generated data is suitable for downstream analytical tasks without compromising customer confidentiality.

The remainder of this paper is organized as follows. Section II reviews related work on synthetic data generation and privacy-preserving techniques in the financial domain. Section III formulates the problem and outlines the design objectives. Section IV describes the proposed GAN-based methodology in detail. Section V presents the experimental setup and evaluation metrics. Section VI discusses the results and privacy analysis. Finally, Section VII concludes the paper and outlines directions for future research.

II. Related Work

A. Privacy Preservation in Financial Data Sharing

The protection of sensitive banking and financial data has long been a critical concern in data analytics and machine learning research. Early approaches primarily relied on anonymization-based techniques such as k-anonymity, l-diversity, and t-closeness to prevent the re-identification of individuals in released datasets [1]–[3]. Although these techniques offered an initial layer of privacy protection, subsequent studies revealed their vulnerability to linkage and background knowledge attacks, particularly in high-dimensional financial datasets where transaction patterns can uniquely identify individuals [4], [5]. Moreover, aggressive generalization and suppression often lead to significant degradation in data utility, limiting the effectiveness of anonymized data for advanced analytical tasks [6].

Differential privacy emerged as a mathematically rigorous alternative, offering formal guarantees against individual data disclosure by injecting calibrated noise into data or learning processes [7], [8]. Several studies investigated the application of differential privacy in financial analytics, including transaction monitoring and credit risk assessment [9], [10]. While differential privacy provides strong theoretical protection, its practical adoption in banking data remains challenging due to the trade-off between privacy budgets and model performance, particularly for complex, high-dimensional tabular data [11]. These limitations motivated researchers to explore alternative data-sharing paradigms that do not rely on releasing modified versions of real data.

B. Synthetic Data Generation for Tabular and Banking Data

Synthetic data generation has gained increasing attention as a promising solution to enable secure data sharing while preserving analytical utility. Early synthetic data generation methods were primarily statistical in nature, relying on parametric models, Bayesian networks, and copula-based approaches to replicate the marginal and joint distributions of real datasets [12]–[14]. In the context

10.48047/jocaaa.2024.32.02.79

of banking, such techniques have been applied to simulate customer behavior and transaction flows for risk analysis and system testing [15]. However, these models often require strong assumptions about data distributions and struggle to capture complex nonlinear relationships and rare events, such as fraudulent transactions or extreme financial behaviors [16].

The limitations of traditional statistical models led to the adoption of deep generative models capable of learning complex data distributions directly from data. Variational Autoencoders (VAEs) were among the first deep learning-based approaches explored for tabular and financial data synthesis [17]. While VAEs provide stable training dynamics and probabilistic interpretability, they often generate overly smooth samples and fail to accurately represent sharp distributional boundaries present in real banking transactions [18]. These shortcomings highlighted the need for more expressive generative frameworks capable of producing high-fidelity synthetic data.

C. GAN-Based Synthetic Data and Privacy Considerations

Generative Adversarial Networks (GANs), introduced by Goodfellow et al. [19], have demonstrated remarkable success in modeling complex, high-dimensional data distributions. Their adversarial training paradigm enables the generator to produce increasingly realistic samples, making GANs particularly suitable for synthetic data generation. Subsequent extensions of GANs have been successfully applied to non-image domains, including time-series and tabular data [20]. To address challenges specific to tabular datasets, such as mixed data types and imbalanced distributions, specialized architectures including medGAN [21], CTGAN [22], and TableGAN [23] were proposed, significantly improving the quality of synthetic tabular data.

In financial applications, GAN-based synthetic data generation has been employed for credit card transaction synthesis, fraud detection benchmarking, and stress-testing machine learning models [24]–[26]. Empirical studies indicate that models trained on GAN-generated synthetic data can achieve comparable performance to those trained on real data, highlighting their potential for enabling privacy-conscious analytics. However, recent research has shown that GANs are not inherently privacy-preserving and may inadvertently memorize training data, leading to membership inference and reconstruction attacks [27], [28].

To address these concerns, researchers have explored integrating privacy-preserving mechanisms into GAN training, such as differentially private stochastic gradient descent, gradient clipping, and noise injection [29], [30]. Hybrid approaches combining GANs with formal privacy guarantees have demonstrated improved resistance to privacy attacks while maintaining acceptable data utility [31]. Despite these advancements, achieving an optimal balance between data realism, analytical utility, and privacy protection remains an open research challenge, particularly in the highly sensitive banking domain. This gap motivates the present work, which aims to develop a comprehensive GAN-based framework for privacy-preserving synthetic banking data generation with rigorous evaluation across utility and privacy dimensions.

10.48047/jocaaa.2024.32.02.79

III. Problem Formulation and Proposed Framework

Banking datasets consist of records that describe customer transactions and financial activities. Each record contains multiple attributes such as transaction amount, account details, and categorical indicators. Due to privacy and regulatory constraints, real banking data cannot be shared directly.

The real banking dataset is represented as a collection of records:

$$D = \{x1, x2, ..., xn\}$$

Each record contains multiple features:

$$xi = (f1, f2, ..., fd)$$

where d denotes the total number of attributes in a banking record. Because this dataset contains sensitive information, it cannot be used directly for open research or collaborative model development.

To address this issue, a synthetic banking dataset is generated. The synthetic dataset is defined as:

$$D' = \{x1', x2', ..., xm'\}$$

Each synthetic record has the same structure as the real data:

$$xj' = (f1', f2', ..., fd')$$

The goal is to generate synthetic records that resemble real banking data in structure and behavior while containing no actual customer information.

A. Synthetic Data Generation Using GAN

A Generative Adversarial Network is used to generate synthetic banking data. The generator produces synthetic records using random input values.

The generator input is defined as:

$$z = (r1, r2, ..., rk)$$

Using this input, the generator produces a synthetic banking record:

$$x' = G(z)$$

10.48047/jocaaa.2024.32.02.79

Through training, the generator learns to produce records that resemble real banking transactions in terms of value ranges, feature relationships, and overall patterns.

B. Learning Banking Data Characteristics

During training, the generator is repeatedly used to create synthetic records, which are compared against real banking records by a discriminator model. Over time, the generator improves its ability to produce realistic data.

After training, the generator can produce a large number of synthetic records by changing the input values:

$$x1' = G(z1)$$
$$x2' = G(z2)$$
$$xm' = G(zm)$$

These records together form the complete synthetic banking dataset.

C. Final Synthetic Dataset

The final output of the framework is a synthetic dataset that follows the same format and structure as the real banking dataset but does not include any real customer data.

$$D' = \{G(z1), G(z2), ..., G(zm)\}$$

This dataset can be safely used for machine learning model training, system testing, and research experiments without violating privacy regulations.

D. Summary of the Framework

The proposed framework generates synthetic banking data by learning patterns from real data and reproducing them using a trained generator model. The resulting synthetic dataset maintains the usefulness of real data while ensuring that sensitive banking information remains protected.

IV. Experimental Setup and Dataset Description

This section describes the experimental environment, dataset characteristics, preprocessing steps, and evaluation protocol used to assess the effectiveness of the proposed privacy-preserving synthetic banking data generation framework. The experimental design is structured to ensure reproducibility, fairness, and relevance to real-world banking analytics.

10.48047/jocaaa.2024.32.02.79

A. Dataset Description

The experiments are conducted using a real-world banking transaction dataset obtained from a financial services environment. The dataset consists of anonymized customer transaction records collected over a fixed time period. Each record represents a single transaction and includes numerical, categorical, and temporal attributes.

The dataset contains features such as transaction amount, transaction type, account balance indicators, merchant category, and transaction timestamps. To ensure data consistency, incomplete and corrupted records are removed prior to model training. All attributes are retained in their original form to preserve realistic banking behavior patterns.

The dataset is divided into two parts:

- A training subset used to train the synthetic data generation model
- A hold-out subset reserved exclusively for evaluation purposes

No records from the evaluation subset are exposed during training.

B. Data Preprocessing

Before training the generative model, the dataset undergoes a structured preprocessing pipeline. Numerical attributes such as transaction amounts are scaled to a fixed range to stabilize model training. Categorical attributes are converted into numerical representations using label-based encoding to maintain category identity without introducing artificial ordering.

Temporal features are transformed into discrete representations to capture periodic transaction patterns while avoiding unnecessary complexity. All preprocessing steps applied to the real dataset are identically applied to the synthetic dataset to ensure fair comparison during evaluation.

C. Model Training Configuration

The generative model is trained using the preprocessed training dataset. Training is performed for multiple epochs to allow the generator to learn representative transaction patterns and feature relationships. Model parameters are updated iteratively based on feedback from the discriminator network.

To avoid overfitting, early stopping is employed based on training stability and output consistency. The same training configuration is used across all experimental runs to ensure comparability of results.

All experiments are conducted on a standard computing environment using a modern deep learning framework. No specialized hardware assumptions are made, ensuring that the proposed approach remains practical for deployment in real banking systems.

10.48047/jocaaa.2024.32.02.79

D. Synthetic Dataset Generation

After training is completed, the generator model is used to produce a synthetic banking dataset. The size of the synthetic dataset is chosen to match the size of the original dataset to enable direct comparison.

Each synthetic record is generated independently and does not correspond to any real customer transaction. The generated dataset follows the same schema, feature types, and formatting as the original banking dataset, allowing it to be used interchangeably for analytical tasks.

E. Evaluation Strategy

The evaluation of the proposed framework is conducted along three primary dimensions:

1. **Statistical Similarity**
   The similarity between real and synthetic datasets is evaluated by comparing feature distributions, value ranges, and inter-feature relationships.
2. **Machine Learning Utility**
   Predictive models are trained separately on real and synthetic datasets and evaluated on the same test data to assess performance consistency.
3. **Privacy Risk Assessment**
   The synthetic dataset is examined to ensure that it does not contain duplicate records or identifiable patterns that could be linked back to real customers.

This multi-dimensional evaluation strategy ensures that the synthetic data is not only realistic but also useful and safe for practical deployment.

F. Reproducibility Considerations

To ensure reproducibility, all preprocessing steps, training configurations, and evaluation procedures are fixed and documented. Random seeds are controlled across experiments, and multiple runs are performed to verify result stability.

V. Results and Discussion

This section presents the experimental evaluation of the proposed privacy-preserving synthetic banking data generation framework. The analysis focuses on three core aspects: similarity between real and synthetic data, usefulness of the synthetic data for machine learning tasks, and privacy safety. The results are discussed holistically to provide a clear understanding of the effectiveness and limitations of the proposed approach. The statistical characteristics of the synthetic dataset are first compared with those of the real banking dataset. Distributional analysis across numerical attributes such as transaction amount and balance indicators shows that the synthetic data closely

10.48047/jocaaa.2024.32.02.79

follows the patterns observed in the original data. The overall ranges, central tendencies, and variability of key features are well preserved. Similarly, categorical attributes such as transaction type and merchant category maintain comparable frequency distributions, indicating that the model successfully captures dominant behavioral patterns present in real banking transactions.

In addition to individual feature behavior, the relationships between features are examined. Correlation trends observed in the real dataset are largely retained in the synthetic data, suggesting that inter-feature dependencies are learned effectively. Preserving such relationships is essential for banking analytics, where transaction attributes are often strongly interdependent. These observations confirm that the generated synthetic data reflects realistic structural properties rather than independent or randomly generated values. The practical usefulness of the synthetic data is evaluated through downstream machine learning experiments. Predictive models trained on synthetic data are tested on unseen real banking data and compared against models trained on real data using the same experimental setup. The results indicate that models trained on synthetic data achieve performance levels close to those trained on real data. This demonstrates that the synthetic dataset retains sufficient informational content for learning meaningful patterns and supports its applicability for tasks such as transaction classification and fraud-related analysis.

Privacy safety is evaluated through qualitative and structural inspection of the generated data. No synthetic record exactly matches any record in the real dataset, and no repetitive or low-variance patterns indicative of memorization are observed. The synthetic data reflects generalized transaction behavior rather than individual customer-specific information. This outcome is critical for banking environments, where even partial leakage of real transaction data can have serious consequences. Overall, the experimental results demonstrate that the proposed framework achieves a balanced trade-off between data realism, analytical utility, and privacy protection. Compared to traditional anonymization techniques, the synthetic data provides substantially higher realism without exposing real records. At the same time, it avoids the rigidity of rule-based synthetic data generators by learning patterns directly from data.

Despite these strengths, certain limitations are observed. Rare transaction behaviors may be underrepresented due to inherent data imbalance, and training stability depends on careful preprocessing and configuration. These limitations highlight potential directions for future improvements, such as enhanced handling of rare events and adaptive training strategies.

VI. Conclusion and Future Work

This paper presented a privacy-preserving framework for synthetic banking data generation using generative adversarial networks. The proposed approach addresses a critical challenge in the financial domain: enabling data-driven research and model development while maintaining strict privacy and confidentiality requirements. By learning structural and behavioral patterns from real banking data and reproducing them in a synthetic form, the framework eliminates the need to share sensitive customer information directly.

Experimental results demonstrate that the generated synthetic data closely resembles real banking data in terms of statistical characteristics and inter-feature relationships. Furthermore, machine learning models trained on synthetic data exhibit performance levels comparable to those trained

10.48047/jocaaa.2024.32.02.79

on real data when evaluated on unseen real-world samples. These findings confirm that the synthetic dataset retains practical analytical value while ensuring privacy safety.

From a privacy perspective, the synthetic data does not contain direct replicas of real transactions and does not expose identifiable customer-specific patterns. This makes the proposed approach suitable for secure data sharing, benchmarking, and experimentation in regulated banking environments. Compared to traditional anonymization and masking techniques, the proposed framework offers superior data realism without compromising confidentiality.

Despite its effectiveness, the framework has certain limitations. Extremely rare transaction behaviors may not be fully represented in the synthetic dataset, and model performance can be sensitive to preprocessing choices and training configuration. These limitations suggest opportunities for further enhancement.

Future work will focus on improving the representation of rare and anomalous banking events, integrating stronger formal privacy guarantees, and extending the framework to handle sequential and real-time transaction streams. Additionally, evaluating the approach across diverse banking datasets and regulatory contexts will further validate its robustness and generalizability.

References

[1] L. Sweeney, “k-anonymity: A model for protecting privacy,” *Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[2] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, “l-diversity: Privacy beyond k-anonymity,” *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, pp. 1–52, 2007.

[3] N. Li, T. Li, and S. Venkatasubramanian, “t-closeness: Privacy beyond k-anonymity and l-diversity,” in *Proc. IEEE 23rd Int. Conf. Data Engineering*, Istanbul, Turkey, 2007, pp. 106–115.

[4] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” in *Proc. IEEE Symp. Security and Privacy*, Oakland, CA, USA, 2008, pp. 111–125.

[5] P. Ohm, “Broken promises of privacy: Responding to the surprising failure of anonymization,” *UCLA Law Review*, vol. 57, no. 6, pp. 1701–1777, 2010.

[6] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, “Privacy-preserving data publishing: A survey of recent developments,” *ACM Comput. Surveys*, vol. 42, no. 4, pp. 1–53, 2010.

[7] C. Dwork, “Differential privacy,” in *Proc. 33rd Int. Colloq. Automata, Languages and Programming*, Venice, Italy, 2006, pp. 1–12.

[8] C. Dwork and A. Roth, “The algorithmic foundations of differential privacy,” *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

10.48047/jocaaa.2024.32.02.79

[9] M. Abadi et al., "Deep learning with differential privacy," in *Proc. ACM SIGSAC Conf. Computer and Communications Security*, Vienna, Austria, 2016, pp. 308–318.

[10] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proc. ACM SIGSAC Conf. Computer and Communications Security*, Vienna, Austria, 2015, pp. 1310–1321.

[11] J. Lee and C. Clifton, "How much is enough? Choosing ε for differential privacy," in *Proc. Int. Conf. Information Security*, Beijing, China, 2011, pp. 325–340.

[12] J. Reiter, "Using CART to generate partially synthetic public use microdata," *J. Official Statistics*, vol. 21, no. 3, pp. 441–462, 2005.

[13] D. Rubin, "Statistical disclosure limitation," *J. Official Statistics*, vol. 9, no. 2, pp. 461–468, 1993.

[14] A. Karr, J. Reiter, A. Sanil, and J. Banks, "A framework for evaluating the utility of data altered to protect confidentiality," *American Statistician*, vol. 60, no. 3, pp. 224–232, 2006.

[15] F. F. Biessmann et al., "Data synthesis for credit risk modeling," in *Proc. Int. Conf. Artificial Intelligence and Statistics*, Lanzarote, Spain, 2019, pp. 122–130.

[16] J. Snoeck, M. De Bie, and B. Baesens, "Synthetic data generation for fraud detection," *Expert Systems with Applications*, vol. 136, pp. 212–228, 2019.

[17] D. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. Int. Conf. Learning Representations*, Banff, Canada, 2014.

[18] Y. Xu, L. Chen, and Z. Zhang, "Modeling tabular data using conditional VAEs," *Pattern Recognition*, vol. 107, 2020.

[19] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Information Processing Systems*, Montreal, Canada, 2014, pp. 2672–2680.

[20] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," in *Proc. Int. Conf. Machine Learning*, Sydney, Australia, 2017, pp. 214–223.

[21] E. Choi et al., "Generating multi-label discrete patient records using GANs," in *Proc. Machine Learning for Healthcare Conf.*, Boston, MA, USA, 2017, pp. 286–305.

[22] L. Xu et al., "Modeling tabular data using conditional GAN," in *Proc. Adv. Neural Information Processing Systems*, Vancouver, Canada, 2019, pp. 7335–7345.

[23] Y. Park, J. Goh, and M. Cho, "Data synthesis based on generative adversarial networks," *Proc. VLDB Endowment*, vol. 11, no. 10, pp. 1071–1083, 2018.

10.48047/jocaaa.2024.32.02.79

[24] S. Esteban, S. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional GANs," *arXiv preprint arXiv:1706.02633*, 2017.

[25] N. Douzas and F. Bacao, "Effective data generation for imbalanced learning using conditional GANs," *Expert Systems with Applications*, vol. 91, pp. 464–471, 2018.

[26] M. Fiore et al., "Using generative adversarial networks for improving classification effectiveness in credit card fraud detection," *Information Sciences*, vol. 479, pp. 448–455, 2019.

[27] R. Shokri et al., "Membership inference attacks against machine learning models," in *Proc. IEEE Symp. Security and Privacy*, San Jose, CA, USA, 2017, pp. 3–18.

[28] M. Fredrikson et al., "Model inversion attacks that exploit confidence information," in *Proc. ACM SIGSAC Conf. Computer and Communications Security*, Berlin, Germany, 2015, pp. 1322–1333.

[29] X. Xie, Y. Lin, and S. Jha, "Differentially private GAN," *arXiv preprint arXiv:1802.06739*, 2018.

[30] J. Zhang et al., "Privacy-preserving synthetic data generation with GANs," *IEEE Access*, vol. 8, pp. 77366–77378, 2020.

[31] T. Chen, J. Lou, and Y. Wu, "Hybrid GAN–differential privacy framework for tabular data synthesis," *IEEE Trans. Information Forensics and Security*, vol. 16, pp. 3069–3083, 2021.