Studying Utility, Privacy, and Fairness in GAN-generated Synthetic Data

Rudraksh Kapil
Department of Computing Science
University of Alberta
rkapil@ualberta.ca

Bharathvaj Kumba Mothilal
Department of Computing Science
University of Alberta
kumbamot@ualberta.ca

Katyani Singh
Department of Computing Science
University of Alberta
katyani@ualberta.ca

Abstract

Synthetic datasets can be generated using generative adversarial networks (GANs) to train machine learning models. The utility, privacy, or fairness of such models is often studied in isolation. However, given the proliferation of automated decision-making systems for critical tasks, it has become imperative to study the relationships between all three factors in order to develop ML models that can have a positive impact on society. This can be accomplished by ensuring that all three factors are maximized. To this end, we study the performance of three different GAN architectures – Tabular GAN, CTGAN and PATE-GAN – in the context of tabular data generation. Moreover, we vary the privacy budget in the differentially private PATE-GAN to determine its effect on utility and fairness. Our experiments on the Adult Census-Income dataset highlight the trade-offs that exist between these factors. We observe that different GAN architectures perform differently with respect to these factors due to their inherent architectural differences. In our study, better utility is attained by the Tabular GAN-generated dataset, whereas the CTGAN-generated dataset showcases more fairness. PATE-GAN is differentially private and hence provides strict privacy guarantees. Our study reveals that achieving all three factors simultaneously requires carefully selecting the value of privacy budget ϵ, which likely depends on the dataset and classification model under consideration.

1 Introduction

Many useful machine learning (ML) tasks require models to be trained on datasets containing sensitive information. For example, electronic health records have been used to train a model that predicts whether a patient is at risk of pancreatic cancer [1]. Even if the sensitive attributes of the patients are masked, an adversary can carry out attacks such as membership inference [2] to determine which records were present in the training dataset. This is an example of a privacy risk. One of the techniques to mitigate such risks is to train ML models on synthetic dataset derived from the original dataset. The intuition is that if the model has never seen the original training dataset, it cannot leak the exact sensitive information of any individual present in the dataset.

However, privacy is not the only concern when training ML models on synthetic datasets. The trained model also needs to be useful for the given task. The performance of ML models is highly dependent on both the quantity and quality of their training dataset [3]. This means that in order to achieve good utility, the synthetic dataset needs to have a large number of diverse records, and also the distribution of the data records need to be similar to those in the original dataset. Moreover, the synthetic dataset should also preserve the correlation between the features as in the original dataset.

Apart from privacy and utility, another crucial aspect is that the outcomes of these predictive models should be fair. Based on different contexts, there are various definitions of fairness of ML models available in the literature. This fairness can be understood in terms of group fairness that measures the degree of dissimilar treatment for different social groups (e.g., female vs. male), or individual fairness that emphasizes similar treatment for similar individuals [4]. To ensure this, the synthetic dataset should ideally neither replicate the biases present in the original dataset nor inadvertently introduce any new biases.

Satisfying just one or even two of utility, privacy, or fairness is not enough. For example, consider a synthetic dataset with two columns, one for age and the other for income. Assume that the income column has the same value (e.g. \$500) for every generated record, whereas the age column values are randomly sampled from a uniform distribution $\mathcal{U}(1, 100)$. This dataset contains no personally identifiable information, so privacy is preserved. The income is the same regardless of age, so fairness is ensured as well. However, this synthetic dataset would be ineffective for learning the relationship between income and age. Therefore, all three aspects need to be simultaneously satisfied, and this is possible only when the relationships and trade-offs between all three are analyzed within the context of synthetic datasets. This forms the motivation for our study, which focuses on addressing the following research questions:

1. How does GAN-generated synthetic data perform with respect to each of the three dimensions: utility, privacy, and fairness?
2. What relationships and trade-offs exist among the three aspects?

The rest of this paper is organized as follows. Section 2 summarizes existing research related to our work, and is followed by Section 3 where we have outlined our contributions. Section 4 discusses the architectures we use for generating synthetic data and our methodology for assessing the utility, privacy, and fairness of such data. Section 5 details our experimental setup and provides an analysis of our results. Section 6 concludes our study with potential future works.

2 Related work

Synthetic data generation In the recent past, generative modelling has been extensively used to create new, representative data from the original training data. For this task of synthetic data generation, GAN [5] and its variants have been used in the literature due to their ability to generate more realistic data over other methods such as variational auto-encoders [6]. For instance, continuous recurrent neural networks-GAN [7], which is suitable for continuous sequential data, has been used to generate high-quality music. Temporal GAN [8] can learn a semantic representation of unlabeled videos to generate new ones. A cascade of convolutional networks within a Laplacian pyramid framework that uses GANs [9] has been employed for generating natural images. Deep convolutional GANs [10] have been used for unsupervised representation learning. Improved GAN techniques have also been proposed to generate high quality images that passed a visual Turing test [11].

With respect to textual data, GANs have been used to effectively synthesize representative data of time-series data [12] as well as tabular binary data, thereby obscuring the private information in it. Wasserstein GAN [13] has been used in generating sparse and heterogeneous synthetic Electronic Health Records [14]. GANs have also been applied to real-world historical data to generate synthetic consumers' credit data [15]. All these GANs employ a generator and a discriminator in an adversarial setting to generate new records.

In particular, we employ three different GANs – Tabular GAN [16], conditional tabular GAN (CTGAN) [17], and private aggregation of teacher ensembles-GAN (PATE-GAN) [18] – for synthetic data generation. The reasoning behind our choice of using Tabular GAN and CTGAN lies in their ability to perform particularly well for the tabular data, which we will be using in this work. Further, the choice of using PATE-GAN over other differentially private GANs like DPGAN [19] comes naturally from the observation that the PATE-GAN consistently outperforms DPGAN when the utility (AUROC and AUPRC) of the synthetic datasets are measured for a range of different epsilon values [18]. Moreover, this PATE mechanism provides tighter differential privacy guarantees, meaning that when the differential privacy guarantee is fixed, PATE-GAN generates higher quality synthetic data than DPGAN, both in terms of better preservation of feature-label relationships and the production of realistic samples. This PATE-GAN has also been empirically proved to scale well for larger tasks while providing both high utility and ensuring very strong privacy [20]. Furthermore, all three GAN architectures can be used in datasets containing categorical as well as numerical features. They also seem to work well in the presence of multiple sensitive attributes.

Studying utility, privacy, and fairness Much of the existing works which study the factors such as utility, or privacy, or fairness of machine learning models trained on GAN-generated synthetic datasets focus their analysis only on the trade-off between any two of them. For instance, the privacy implications in the GAN generated samples and their robustness to membership inference attacks have been studied [21] previously, without taking fairness into the account. To analyze this further, a conditional table generative adversarial network (CTAB-GAN) [22] has been developed, which can achieve good utility while preserving privacy. Further, bias amplification in differentially private synthetic data generation schemes has been studied in the literature [23] to understand the trade-off between privacy and fairness, leaving out the utility. To study the interplay between utility and fairness, the concept of fairness-aware learning has been developed to prevent discrimination in the training data using FairGAN [24] while simultaneously achieving good data utility. However, the effect of privacy has not been considered.

These papers fail to look into the collective effects and their interdependencies in a comprehensive manner, which is crucial in order to develop predictive models that benefit society. The concept of synthetic data generation overcomes a few challenges in achieving privacy and fairness simultaneously, such as removing class imbalance, providing access to unlimited training data, and providing access to datasets without having to worry about privacy issues. However, it is necessary to get insights on the maximum utility that can be derived from these predictive models without compromising on other measures. This void in the research literature motivates us to dive deeper into understanding the dynamics of these various measures.

3 Contribution

Addressing these concerns, and to answer our research questions, (1) we study the relationship between all three aspects – utility, privacy, and fairness – and report the variations in these relationships for different GAN architectures, and (2) we study the effect of privacy budget ϵ on the model's utility and fairness to find out the optimal range of ϵ for generating a synthetic dataset that achieves high utility and maintains fairness while also preserving privacy.

4 Methodology

Figure 1 shows the flow for the followed methodology. Our first step involves the generation of synthetic datasets from an original dataset of real records. We employ three popular GAN architectures for this task – Tabular GAN, CTGAN, and the differentially private PATE-GAN – which are described below. An ML model is then trained on these synthetic datasets. The performance of the trained model is assessed in terms of utility, privacy, and fairness using a set of evaluation metrics.

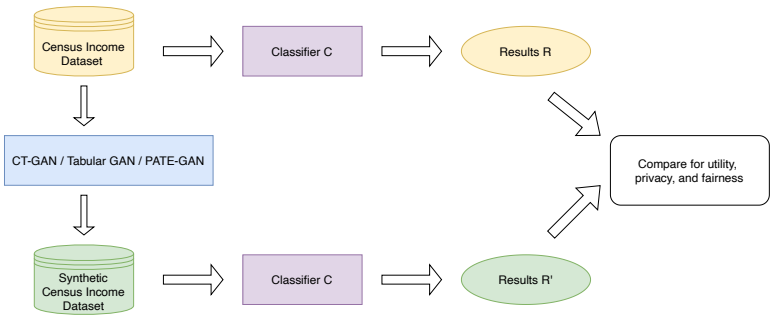


Figure 1: Flow for studying utility, privacy and fairness of model trained on synthetic data.

4.1 Generation of synthetic data

In this subsection, we explain the working of the three GAN architectures used in this study and how they can be applied to generate synthetic data.

Tabular GAN The key idea of Tabular GAN is that the marginal distribution for each column is learned by minimizing the KL Divergence with the real data column. This architecture handles categorical and numerical columns differently. Categorical columns are encoded in a one-hot representation, whereas numerical columns are clustered using a Gaussian mixture model. The generator relies on a long short-term memory (LSTM) architecture, so the correlation between variables is captured by using the recurrent neural network structure. The columns are generated in their original order in the table. The discriminator concatenates the columns of a record together and uses a multilayer perceptron to determine whether the record is real or has been generated.

CTGAN In CTGAN, synthetic rows conditioned on one of the discrete columns are generated. The one-hot vectors of all discrete columns are merged, but the vectors are reproduced only with the specification of the selected category. In the training phase, the conditional generator produces multiple sets of one-hot discrete vectors. However, it is forced to generate a discrete one-hot column equal to the mask vector to penalize its loss by adding the cross-entropy loss within them, averaged over all the instances of the batch. A critic then evaluates the output generated by the conditional generator by approximating the distance between the learned conditional distribution and the conditional distribution of the original data. CTGAN utilizes a training-by-sampling method in which the conditional vector and training data are sampled according to the log-frequency of each category. This allows CTGAN to explore all the potential discrete values.

PATE-GAN PATE-GAN generates synthetic data using a modified version of the PATE framework introduced by [25, 20]. Here, the dataset is divided into a number of disjoint subsets. Then, different classifiers are trained separately on these partitions, assigning differentially private noise-aggregated class labels with the help of majority voting. As these classifiers train directly on the sensitive data, they are used as teachers to train the student-discriminator model over the generator's samples without looking into the teachers' parameters. This setting helps in overcoming the problem of non-differentiability of the noise-aggregated outputs with respect to the parameters of the generator. On the other hand, the generator takes random noise from the uniform distribution as the input and outputs generated samples, while trying to minimize the minimax loss function in order to fool the student-discriminator. By doing so, the problem of learning a differentially private generator is changed to learning a differentially private discriminator using the post-processing theorem [26].

Here, the teacher-discriminators are being trained to improve their loss with respect to the generator. At the same time, the generator is being trained to improve its loss with respect to the student using backpropagation, which in turn is being trained to improve its loss with respect to the teacher-discriminators. This process is repeated until our privacy constraint, ϵ, has been reached.

Interestingly, there exists a trade-off here. If a large number of teacher-discriminators is used, fewer data samples will be used to train each of them, which may make the output meaningless, even though the noise has a smaller effect. On the other hand, for a small number of teacher-discriminators, the noise may be too large and thus rendering the output meaningless. Finding the right balance between them is crucial to efficiently training PATE-GAN.

4.2 Assessment of synthetic data

In this subsection, we discuss the evaluation metrics used in our assessment of synthetic data with respect to three dimensions. The first dimension, utility, evaluates if the synthetic dataset can be used as a valid representative of the original data. The second dimension, privacy, looks into the nearest neighbor distances within and between the original and synthetic datasets. The third dimension, fairness, explores whether the outcomes of models trained using synthetic data are unbiased.

Utility To address whether synthetic data can be used as a representative of the original data for training ML models, we train separate random forest classifiers using synthetic and original data. The performances of the trained models are measured in terms of accuracy, area under the ROC curve (AUROC), and area under the precision-recall curve (AUPRC) scores. We compare these scores for the model trained on synthetic data against the model trained on original data for a binary classification task.

Privacy To assess the preservation of privacy in synthetic data, we employ distance-based metrics. Since our study involves both non-differentially private synthetic data generators (CTGAN and Tabular GAN) and a differentially private synthetic data generator (PATE-GAN), we resort to distance-based metrics for privacy assessment instead of differential privacy. Since differential privacy mechanisms have to be built into the architecture, by using these distance-based metrics, we get an even ground for comparing the three GAN architectures.

We use the VirtualDataLab [27] library's distance-based privacy assessment tests. The tests involve two metrics designed to evaluate privacy guarantees on datasets - distance to closest record (DCR) [22] and nearest neighbor distance ratio (NNDR) [28]. The DCR calculates the Euclidean distance between a synthetic record and its closest corresponding neighbor in the original dataset. A higher value corresponds to a reduced risk of privacy breach. The NNDR calculates the ratio of the Euclidean distance between a synthetic record and its closest neighbor, to the Euclidean distance between the same synthetic record and its second-closest neighbor. This ratio lies between 0 and 1, with a higher ratio corresponding to a lower privacy risk.

Privacy guarantee is evaluated by dividing the original data into a reference set and a holdout set. These distance metrics are then calculated between the synthetic data and the reference set. The same metrics are also calculated between the holdout set and the reference set. To pass the privacy test, the value of the distance metrics between the synthetic data to the reference set should be smaller than the value of the distance metrics between the holdout and the reference set.

Fairness In our study, we evaluate group fairness using demographic parity difference (DPD) [29] and equality of odds (EoD) [30]. DPD is the absolute difference between the true positive rates for each subgroup. Ideally, this value should be the same for each subgroup. EoD takes the maximum of the absolute difference between the true positive rates of different subgroups and the absolute difference between the false positive rates of different subgroups. For a fair classification, the true positive rates or the false positive rates of different subgroups should be the same. We use the Fairlearn [31] library's built-in functions for calculating these fairness metrics.

4.3 Effect of privacy budget on fairness and utility.

One of the major focuses of our study involves analyzing the effect of the privacy budget ϵ on the model's fairness as well as utility. Figure 2 shows the flow for studying the effect of privacy budget on utility and fairness. We generate separate datasets using the PATE-GAN model corresponding to ϵ

values of 2, 4, 6 and 8. We evaluate these datasets using our evaluation metrics for utility and fairness as mentioned in Subsection 4.2.

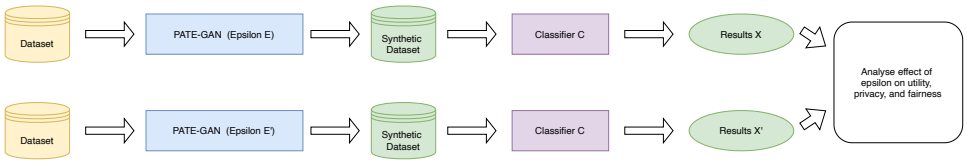


Figure 2: Flow for studying the effect of privacy budget on utility and fairness.

5 Experimental setup, results and analysis

In this section we elucidate our experimental setup and our analysis of the results. The code to replicate our results can be found on the project repository [1].

5.1 Dataset

Our study employs the Adult Census-Income dataset available in the UCI Repository of Machine Learning Databases [32]. This dataset is based on the 1994 US census. The column headers correspond to the questions asked during the census, such as those about age, marital status, and occupation. Each person's response to these questions constitutes one record in the dataset. The target column is the person's annual income, classified as being either $>$ \$50K or $\leq$ \$50K. The primary task for this dataset is binary classification. It involves predicting the income category of a person. This dataset is popularly used for research purposes as the true value of the target variable is known, unlike in other similar datasets such as the German Credit Data available in the same repository. Moreover, the protected attribute such as race or gender is known beforehand, which makes this dataset ideal for our study.

5.2 Experiment

The original dataset contains disjoint training and testing sets. The training set is used for generating synthetic records and the testing set is used for performance evaluation. Before generating synthetic records, we drop the *'fnlwgt'* feature as this column does not provide any meaningful information for the classification task. Since all the three GAN architectures can generate synthetic records from both discrete as well as continuous columns, we do not discard any other columns. Rows with missing values, denoted by a '?' character, are dropped from the dataset. We perform preliminary data exploration on the original data by plotting the correlation matrix between the columns in the dataset. This pre-processed training dataset is then used to generate the synthetic dataset using the three GAN architectures. Each GAN architecture is implemented and trained with the same specifications provided in its respective paper. For PATE-GAN, we encode the values in categorical columns using a multi-label encoder prior to the synthetic data generation. The values in the generated dataset are then decoded. For our experimentation purpose, we fix the number of teacher-discriminators in PATE-GAN to 10, and vary the ϵ values ranging from 2 to 8 with an incremental step of 2.

5.3 Results and analysis

Table 1 summarizes the performance of the trained random forest classifiers on the original dataset and the synthetic dataset in terms of our evaluation metrics. A synthetic dataset can be deemed to have higher quality if the values of the utility metrics are similar to that of the original dataset. From our experiments, we observe that the dataset generated by Tabular GAN attains utility scores closest to the original dataset (accuracy of 0.787 compared to 0.85). Probing into the possible reasons for this, we infer that Tabular GAN preserves the underlying relationships among features present in

[1] https://github.com/rudrakshkapil09/CMPUT622-Project

the original dataset, without much alteration. This is evident from the observation that the synthetic dataset generated by Tabular GAN has a similar data distribution to the original dataset, which can be attributed to the LSTM-based generator that uses recurrent connections to capture the correlation between columns. This shows that Tabular GAN-generated synthetic dataset can be a good proxy for training classification models when the original dataset contains sensitive information about individuals. This recurrent nature is missing from the other two GANs, so they do not perform as good as Tabular GAN. This shows that the best utility can be attained if we apply Tabular GAN for synthetic data generation over other type of GANs.

Table 1: Performance of real and synthetic data on the utility, fairness and privacy evaluation metrics

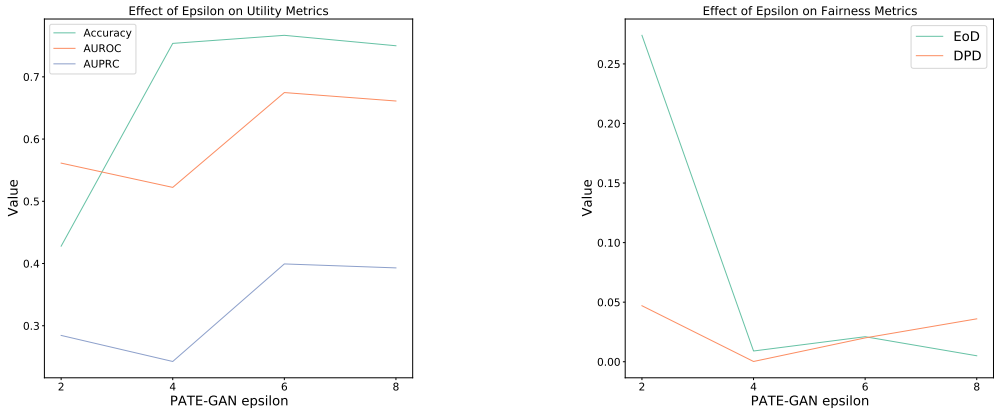
Dataset	Accuracy	AUROC	AUPRC	EoD	DPD	DCR	NNDR
Original	0.850	0.896	0.763	0.075	0.171	-	-
CTGAN	0.763	0.666	0.372	**0.008**	0.006	Passed	Passed
Tabular GAN	**0.787**	**0.830**	**0.609**	0.165	0.198	Passed	Passed
PATE-GAN ϵ=2	0.427	0.561	0.284	0.274	0.047	Passed	Passed
PATE-GAN ϵ=4	0.753	0.522	0.242	0.009	**0.000**	Passed	Passed
PATE-GAN ϵ=6	0.766	0.674	0.399	0.021	0.020	Passed	Passed
PATE-GAN ϵ=8	0.749	0.661	0.392	0.005	0.036	Passed	Passed

We evaluate group fairness on the basis of the sensitive attribute *'sex'*. A fair classifier should result in smaller values of our fairness metrics. Considering EoD, CTGAN achieves a value of 0.008. This value is much lower than that of the original dataset value of 0.075. CTGAN results in a DPD value of 0.006, which is again much lower than the original dataset DPD value of 0.171. Interestingly, PATE-GAN with $\epsilon = 4$ reaches a DPD score approximately equal to 0. Overall, these values indicate that the models trained on synthetic records generated by CTGAN and PATE-GAN (with certain values of ϵ) reduce the unfair outcomes and disparate treatment across different subgroups. We observe a class imbalance in the original dataset with respect to males and females earning more than \$50K. However, the synthetic datasets generated by CTGAN and PATE-GAN mitigate this class imbalance, meaning that both genders are treated equally, thereby achieving more fairness. Unfortunately, this class imbalance still exists for the Tabular GAN-generated synthetic dataset.

In terms of privacy evaluation, all the synthetic datasets passed the DCR and NNDR privacy tests provided by the VirtualDataLab library. This indicates that all three GAN architectures indeed generate synthetic records from scratch, rather than simply adding random noise or removing/adding new records to the original dataset.

Figure 3 shows the effect of privacy budget ϵ in PATE-GAN on utility and fairness metrics. A smaller value of ϵ corresponds to higher privacy. From Figure 3a, it is evident that as the value of ϵ is decreased, the utility gets compromised. This shows that privacy and utility share a somewhat inverse relationship with each other. Looking at the effect of privacy budget ϵ on fairness metrics in Figure 3b, we find that the fairness values somewhat decrease with a decrease in values of ϵ signifying more fairness. However, at an extremely small value of ϵ equal to 2, the bias is quantified. This can be validated from the fact that when we try to achieve higher privacy, the model compromises on the utility, which in turn produces more misclassifications. These misclassifications may result in different true positive rates or false positive rates for different social groups, thereby introducing bias into the model.

These results show the trade-off between the three dimensions – utility, privacy, and fairness. To achieve better utility and fairness, without compromising privacy, the values for ϵ in PATE-GAN should neither be too low nor too high. For this particular dataset and classification model, considering an ϵ value in the range 4 to 6 in PATE-GAN achieves good performance on all the three dimensions.



(a) Privacy vs Utility

(b) Privacy vs Fairness

Figure 3: Effect of privacy budget ϵ in PATE-GAN on utility and fairness metrics

6 Conclusion and future work

With the growing applications of GAN-generated synthetic data, our study assesses its performance along the three dimensions – utility, privacy, and fairness. Our study reports the variations in these relationships for different GAN architectures. The effect of privacy budget ϵ on the model's utility and fairness has been studied using the differentially private PATE-GAN. Our results can be utilized in practical applications to understand the compromises that must be made in one of the factors to maximize another.

Due to computation and time constraints, our study is limited to a narrow range of ϵ values in PATE-GAN. Further, we have been able to consider only one classification model. We have studied the three GAN architectures only in the context of tabular data as well. Considering these limitations, it would be interesting to study how our identified relationships are affected when other classification models are used, or when totally different tasks such as regression are required. A potential extension of our work can be directed towards understanding the effect of other factors such as the number of teachers in PATE-GAN. Also, our work provides a platform to extend this study on other synthetic data generation techniques, or even on other datasets. In future, our study will also help prove if there exists an impossibility condition among these three factors.

Acknowledgements

We express our gratitude towards Dr. Nidhi Hedge, Associate Professor, Computing Science Department, University of Alberta, for sharing her valuable feedback with us during the course of this research project.

References

[1] Di Zhao and Chunhua Weng. Combining PubMed knowledge and EHR data to develop a weighted bayesian network for pancreatic cancer prediction. *Journal of biomedical informatics*, 44(5):859–868, 10 2011.

[2] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 3–18. IEEE, 2017.

[3] Valerie Sessions and Marco Valtorta. The effects of data quality on machine learning algorithms. In *Proceedings of the 11th International Conference on Information Quality, , 2006*, pages 485–498. MIT, 2006.

[4] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, page 214–226. Association for Computing Machinery, 2012.

[5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, 2014.

[6] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2014.

[7] Olof Mogren. C-rnn-gan: Continuous recurrent neural networks with adversarial training. *arXiv preprint arXiv:1611.09904*, 2016.

[8] Masaki Saito, Eiichi Matsumoto, and Shunta Saito. Temporal generative adversarial nets with singular value clipping. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2830–2839, 2017.

[9] Emily L Denton, Soumith Chintala, Arthur Szlam, and Rob Fergus. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

[10] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *4th International Conference on Learning Representations*, 2016.

[11] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in Neural Information Processing Systems*, 29:2234–2242, 2016.

[12] Nicole Bridgland. Obscuring and analyzing sensitive information with generative adversarial networks. *World Wide Technology*, 2019.

[13] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, pages 214–223. PMLR, 2017.

[14] Kieran Chin-Cheong, Thomas Sutter, and Julia E Vogt. Generation of heterogeneous synthetic electronic health records using gans. In *Workshop on Machine Learning for Health (ML4H) at the 33rd Conference on Neural Information Processing Systems*, 2019.

[15] Nari Park, Yeong Hyeon Gu, and Seong Joon Yoo. Synthesizing individual consumers' credit historical data using generative adversarial networks. *Applied Sciences*, 11(3):1126, 2021.

[16] Lei Xu and Kalyan Veeramachaneni. Synthesizing tabular data using generative adversarial networks. *arXiv preprint arXiv:1811.11264*, 2018.

[17] Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. Modeling tabular data using conditional GAN. *arXiv preprint arXiv:1907.00503*, 2019.

[18] James Jordon, Jinsung Yoon, and Mihaela Van Der Schaar. PATE-GAN: Generating synthetic data with differential privacy guarantees. In *International Conference on Learning Representations*, 2018.

[19] Liyang Xie, Kaixiang Lin, Shu Wang, Fei Wang, and Jiayu Zhou. Differentially private generative adversarial network. *arXiv preprint arXiv:1802.06739*, 2018.

[20] Nicolas Papernot, Shuang Song, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Úlfar Erlingsson. Scalable private learning with PATE. In *6th International Conference on Learning Representations*, 2018.

[21] Zinan Lin, Vyas Sekar, and Giulia Fanti. On the privacy properties of gan-generated samples. In *International Conference on Artificial Intelligence and Statistics*, pages 1522–1530. PMLR, 2021.

[22] Zilong Zhao, Aditya Kunar, Hiek Van der Scheer, Robert Birke, and Lydia Y. Chen. CTAB-GAN: Effective table data synthesizing. *arXiv preprint arXiv:2102.08369*, 2021.

[23] Aman Gupta, Deepak Bhatt, and Anubha Pandey. Transitioning from real to synthetic data: Quantifying the bias in model. *arXiv preprint arXiv:2105.04144*, 2021.

[24] Depeng Xu, Shuhan Yuan, Lu Zhang, and Xintao Wu. FairGAN: Fairness-aware generative adversarial networks. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 570–575, 2018.

[25] Nicolas Papernot, Martín Abadi, Úlfar Erlingsson, Ian J. Goodfellow, and Kunal Talwar. Semi-supervised knowledge transfer for deep learning from private training data. In *5th International Conference on Learning Representations*, 2017.

[26] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

[27] MOSTLY AI. Virtual data lab. `https://github.com/mostly-ai/virtualdatalab`, 2021.

[28] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[29] Sahil Verma and Julia Rubin. Fairness definitions explained. In *Proceedings of the International Workshop on Software Fairness*, page 1–7. Association for Computing Machinery, 2018.

[30] Moritz Hardt, Eric Price, and Nathan Srebro. Equality of opportunity in supervised learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, page 3323–3331, 2016.

[31] Fairlearn. Fairlearn. `https://github.com/fairlearn/fairlearn`, 2021.

[32] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.