

Synthetic Data Generation and Evaluation of Positioning, Navigation & Timing Data

Jack C. Kendrick¹, Josh W. Hoff¹, Cooper M. Moody¹, Timeo J. Williams², Brendan B. Cross²,
Hyeyon Y. Bastian¹, and Nathaniel D. Bastian^{1,2}

¹Department of Systems Engineering, United States Military Academy, West Point, New York 10996

²Department of Electrical Engineering & Computer Science, United States Military Academy, West Point, New York 10996

Corresponding author's Email: nathaniel.bastian@westpoint.edu

Author Note: This work was supported in part by the U.S. Army Transformation Decision Analysis Center (TDAC) under Support Agreement No. USMA23053. The views and conclusions expressed in this paper are those of the authors and do not reflect the official policy or position of the U.S. Military Academy, U.S. Army, U.S. Department of War, or U.S. Government.

Abstract: Military capabilities for all-domain sensing are critical for enhancing the military's ability to sense, process, and act on battlefield data across domains while synchronizing kinetic and non-kinetic effects. Essential for the development and assessment of these capabilities is Positioning, Navigation and Timing (PNT) data, which is scarce and expensive to collect/store. This work develops and experiments with an innovative, repeatable methodology using generative artificial intelligence for synthetic PNT data generation, as well as evaluation that combines statistical analysis with machine learning. Our results suggest that a conditional tabular generative adversarial network was the best performing method for synthetic PNT data generation.

Keywords: Synthetic Data Generation, Synthetic Data Evaluation, Machine Learning, Generative Artificial Intelligence, PNT

1. Introduction

The use of Positioning, Navigation, and Timing (PNT) data in military warfighting capabilities is becoming increasingly important across many applications (GPS navigation, drone localization, etc.). The military is a major consumer of PNT data, particularly as capabilities for all-domain sensing have become increasingly prevalent on the modern battlefield. This led to innovative capabilities such as the handheld Drone Buster, which jams signals from unmanned aerial systems Paulsen and Baker (2025). As capabilities relying on PNT data become more interconnected, the ability to assess them using realistic datasets is critical. Assessing capability performance requires large volumes of representative data across varying operational conditions, such as changes in altitude, velocity, and signal degradation. However, collecting sufficient real-world PNT data is difficult because it requires extensive operational trials, specialized sensors, and controlled environments. PNT applications also require diverse datasets because each use case—such as shipping, road navigation, or satellite trajectory analysis—contains different operational characteristics. Collecting datasets across these domains is both time consuming and expensive, limiting the scale at which capabilities can be assessed. As a result, organizations increasingly rely on synthetic data generation to support development and evaluation efforts. This work investigates whether generative artificial intelligence (AI) generated synthetic PNT data can support the development of capabilities that depend on PNT data. Specifically, we present an innovative, repeatable methodology that (1) curates tabular PNT data, (2) generates synthetic datasets using three different generative AI models, and (3) evaluates dataset utility through statistical and machine learning based predictive effectiveness performance.

2. Related Works

Synthetic data generation is increasingly important as organizations confront constraints related to data scarcity, privacy, cost, and accessibility—particularly in environments where real data sharing is restricted. Early statistical and rule-based methods reproduced summary statistics but struggled to capture nonlinear relationships and high-dimensional dependencies (Goyal & Mahmoud, 2024; Thangaraju, 2023). Modern deep generative models address these limitations by learning the underlying data distribution directly. Approaches such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models generate synthetic data through adversarial training, probabilistic latent spaces, or iterative denoising processes, while transformer- and large language model (LLM)-based approaches extend these capabilities by modeling complex feature dependencies (Bauer et al., 2024; Razghandi, Zhou, Erol-Kantarci, & Turgut, 2022).

Across the literature, no single generative AI model consistently outperforms others across all datasets or applications (Hansen, Seedat, van der Schaar, & Petrovic, 2023; Wang, Chukova, Simpson, & Nguyen, 2024). Performance depends on factors such as dataset size, dimensionality, and downstream objectives. GAN and VAE variants often perform well on tabular data but may introduce stability or privacy concerns, while diffusion models can generate high-fidelity samples at increasing computational costs (Kiran & Kumar, 2024; Miletic & Sariyar, 2024). LLM-based approaches provide additional flexibility but raise concerns regarding bias, hallucination, and reproducibility (Guo & Chen, 2024).

Despite progress in synthetic data generation, limited research applies modern generative AI models specifically to PNT datasets. Existing work involving PNT data often relies on simulation-based environments or physics-driven models rather than data-driven generative AI model approaches capable of capturing complex statistical relationships (U.S. Government Accountability Office, n.d.). This creates a gap in understanding how different generative architectures perform on real-world PNT data and how effectively they reproduce realistic PNT behavior.

Evaluation of synthetic data also remains an open challenge, as no standardized framework exists to determine its utility for downstream tasks (Yuan, Liu, & Cheng, 2024). Statistical similarity metrics such as maximum mean discrepancy (MMD), Wasserstein distance, Kolmogorov-Smirnov (KS) tests, and total variation distance (TVD) measure distributional alignment but do not always correlate with machine learning predictive performance (Nafis, Esnaola, Álvaro Martinez-Perez, Villa-Uriol, & Osmani, 2025; Yuan et al., 2024). Prior work suggests neural networks perform to evaluate synthetic tabular datasets (Livieris, Alimpertis, Domalis, & Tsakalidis, 2024), but many studies focus primarily on statistical similarity rather than predictive effectiveness. This study addresses that limitation by evaluating synthetic PNT data using both statistical metrics and downstream machine learning performance within a defined PNT-relevant prediction task.

3. Methodology

This section outlines our analytic pipeline for synthetic PNT data generation and evaluation. It includes PNT data collection and curation, the generation of synthetic datasets using multiple generative AI models, and the evaluation framework to assess statistical similarity and predictive utility. Together, these steps establish a structured, repeatable solution methodology for generating, evaluating, and comparing synthetic tabular PNT data against real-world datasets for a given use case.

3.1. Solution Design

3.1.1. Data Collection & Curation

The solution methodology begins with tabular data input via a CSV file. The dataset has many variables, but only relevant features for a predictive modeling task are retained; for our use case, the variables are speed in three different directions, altitude, and battery percentage. We divide the curated data into fitting and evaluation datasets. We randomly sample the dataset to the desired ratio of fitting data to evaluation data. The fitting dataset fits generative AI models. The evaluation dataset is further split into training and validation datasets for training machine learning models and validating them. We compare the validation dataset against the machine learning model outputs to give information on how well the synthetic and real data match.

3.1.2. Synthetic Data Generation

Our solution design adopts a multi-model synthetic data generation approach centered on a Transformer-based Tabular Variational Autoencoder (TTVAE), Conditional Tabular Generative Adversarial Network (CTGAN), and Denoising Diffusion Probabilistic Model (DDPM). These three models represent fundamentally different generative AI methods—variational, adversarial, and diffusion-based approaches—allowing for a comprehensive comparison of how different architectures capture complex statistical relationships in tabular PNT data. Evaluating these models together enables a more robust assessment of which generative strategy produces the most realistic and useful synthetic datasets. Rather than selecting a single model as a universal solution, the design leverages the complementary strengths of each architecture to create a repeatable and scalable synthetic PNT data generation pipeline. TTVAE serves as a structured latent-space backbone for stable probabilistic modeling of mixed-type tabular PNT variables; CTGAN provides enhanced modeling of nonlinear feature interactions and minority distribution modes; and DDPM enables high-fidelity reconstruction of continuous and high-dimensional signal characteristics. Together, these models represent distinct generative AI approaches, each offering meaningful differences in how synthetic data is learned, represented, and produced. For more information regarding these methods, please refer to Azahad and Hameeda (2023), Wang and Nguyen (2025), and Zhu (2024). By incorporating these three architectures into the methodology, it allows the ability to compare different synthetic data generation methods for the same problem. Evaluating a TTVAE, CTGAN, and DDPM allows us to examine how each model captures the statistical structure, dependencies, and variability inherent in PNT datasets. This comparison is central to our methodology: the goal is not only to generate synthetic data but also to determine which modeling technique is most effective at producing realistic, diverse, and operationally useful tabular PNT outputs.

3.1.3. Verification & Validation

Verification and validation is two steps, statistical similarity evaluation and predictive utility evaluation. The first uses statistical distribution tests, while we determine utility via a use case for machine learning based predictive modeling with synthetically generated data. Algorithms use a percentage of the evaluation data to train the models; we use remaining data to test against the generated data. The validation models are selected with a target dependent variable and supporting features. As the evaluation is done via code, it is repeatable and can be altered with different parameters for other data and use cases.

For all statistical similarity measures utilized, a smaller output means the two distributions compared are more similar to each other, which is one goal of generated data. The KS test compares marginal distributions with cumulative density functions for a continuous variable. Wasserstein distance measures the minimal effort to reconfigure the probability mass of one distribution to become the other distribution. Jensen-Shannon divergence (JS Div) measures the similarity of probability distributions. MMD reproduces probability distributions in kernel Hilbert space then compares their similarity. TVD measures the total amount a probability distribution can change until it is the exact same as another.

Three machine learning predictive models are utilized to reveal how they perform differently when trained using synthetic PNT data generated by CTGAN, TTVAE and DDPM. The generated datasets are compared against each other utilizing the statistical means and effectiveness root mean squared error (RMSE) to determine which generative model and data performs closest to the real data. This returns which data is most useful. The validation models are a fully-connected network (DNN), recurrent neural network (RNN), and a tree-based network (XGBoost or XGB). These three predictive models perform differently based on the data utilized and how they train. These models return a RMSE for the data used to train them and the data used to test them. Predictive models require a use case to evaluate utility; the use case can change based on the data and experiment.

3.2. Solution Experimentation

3.2.1. Data Collection & Curation

Our PNT dataset is collected by scraping it from a Parrot Anafi drone following several hours of flight, where many of the features of the data are specific to the drone's camera. For data curation, we retain the following variables per our predictive use case: battery_percentage, speed_north, speed_east, speed_down, and altitude. The dataset contains 17,445 rows.

3.2.2. Synthetic Data Generation

To evaluate synthetic data generation performance, this study implements a framework utilizing CTGAN, TTVAE, and DDPM models. Within this framework, each model's hyperparameters are tuned according to three structured configurations: a baseline setting, a higher-capacity configuration, and a stability-focused configuration. The baseline establishes a consistent reference point for comparison. The capacity-oriented configuration increases model depth, latent dimensionality, or training iterations to assess improvements in capturing complex nonlinear dependencies and multimodal feature distributions. The stability-focused configuration adjusts learning rates, batch structures, regularization strength, and—where applicable—packing mechanisms to mitigate instability, overfitting, or mode collapse. Together, these three configurations allow us to test how changes in model size, training effort, and stability settings affect the quality of the generated synthetic data, helping us understand which generative AI model architecture produces the most realistic and useful results.

All experiments within our repeatable analytic pipeline maintain identical preprocessing setups, dataset splits, and evaluation procedures to ensure internal validity across architectures. By varying only hyperparameters while holding external conditions constant, the design isolates how optimization dynamics, representational structure, and training strategy influence synthetic data realism and downstream utility. This comparative approach benchmarks performance across CTGAN, TTVAE, and DDPM generative AI models while revealing the structural tradeoffs inherent to each method.

3.2.3. Verification & Validation

To verify the statistical similarity of the generated synthetic PNT data, the previously mentioned statistical distribution tests are performed. There are nine different generated datasets under statistical evaluation. To validate predictive utility of the synthetic data, our evaluation uses the following predictive model based on the following PNT data prediction use case:

$$\text{RemainingBattery}\% = \beta_1 \cdot \text{Speed}_{\text{Down}} + \beta_2 \cdot \text{Speed}_{\text{North}} + \beta_3 \cdot \text{Speed}_{\text{East}} + \beta_4 \cdot \text{Altitude} \quad (1)$$

where the dependent variable is battery percentage and the features include three drone speed measures along with altitude. This predictive utility use case seeks to model how the drone's flight speeds and altitude impacts its remaining battery percentage. Remaining battery percentage is selected because it can be reliably predicted from speed and altitude features.

The prediction validation models are tested based on a ratio of synthetic data utilized to train the validation model. The ratio replaces a random sampling of the training data with synthetically generated data points. The machine learning models are trained and tested for the stability-oriented generative AI model outputs from 0% synthetic training data to 100% synthetic

training data using 10% steps. The percentage of synthetic data randomly replaces real data in the all real training dataset. This demonstrates how well mixed data performs and shows how the different generative models perform against each other.

4. Results & Discussion

Table 1: Statistical Similarity Evaluation across Generative AI Model Hyperparameter Tunings

Model	Tuning	Hyperparameters	KS stat / P-val	Wasserstein	JS Div	MMD	TVD
TTVAE	Baseline	latent_dim=16; batch_size=2048; epochs=30	0.214923 / 4.950745e-129	4.249296	0.354886	0.263943	0.546998
TTVAE	Capacity	latent_dim=32; batch_size=1024; epochs=50	0.255492 / 9.180934e-183	3.296452	0.514843	0.14077	0.718666
TTVAE	Stability	latent_dim=8; batch_size=512; epochs=100	0.19313 / 4.060302e-104	3.095606	0.339498	0.189637	0.531441
CTGAN	Baseline	epochs=100; batch_size=500; pac=10	0.023662 / 0.055804	0.474766	0.006347	0.018207	0.061424
CTGAN	Capacity	epochs=300; batch_size=1000; pac=10	0.039306 / 0.000105	0.644352	0.009923	0.032756	0.095656
CTGAN	Stability	epochs=500; batch_size=768; pac=12	0.039871 / 0.000079	0.940889	0.013098	0.054287	0.085647
DDPM	Baseline	steps=5000; lr=0.0015; weight_decay=1e-4; batch=1024; dim_t=1024; scheduler=cosine	0.352969 / 0	11.327357	0.317936	0.436075	0.414825
DDPM	Capacity	steps=8000; lr=0.0010; weight_decay=1e-4; batch=1024; dim_t=1536; scheduler=cosine	0.349834 / 0	43.905542	0.327004	0.429505	0.410927
DDPM	Stability	steps=10000; lr=0.0008; weight_decay=5e-4; batch=512; dim_t=1024; scheduler=cosine	0.345028 / 0	11.322620	0.309124	0.429339	0.405790

The results in Table 1 demonstrate clear differences in synthetic data generation performance across the three generative AI model architectures—TTVAE, CTGAN, and DDPM—under baseline, capacity-focused, and stability-focused hyperparameter configurations. Across the three evaluated generative models, no single model consistently minimized all statistical divergence metrics, highlighting the importance of comparative evaluation when assessing synthetic data statistical quality.

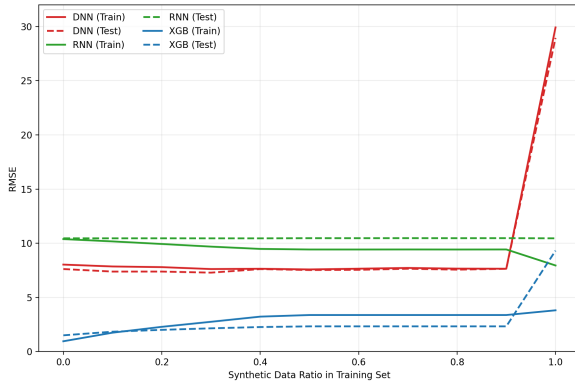
For TTVAE, the stability-focused configuration achieved the lowest KS statistic (0.193) among its configurations, along with reductions in JS divergence (0.339) and MMD (0.189) compared to both baseline and capacity-focused settings. These improvements suggest that reducing latent dimensionality while increasing training epochs improved the model’s ability to capture the statistical structure of the PNT dataset without introducing excessive variance.

CTGAN demonstrated the strongest marginal distribution alignment among the evaluated models, particularly in the baseline configuration, which produced the lowest KS statistic (0.023) across all experiments. Additionally, CTGAN maintained comparatively low Wasserstein distance (0.474) and JS divergence (0.006) values, suggesting strong preservation of feature distributions relative to the real dataset. While both the capacity and stability configurations slightly increased divergence metrics such as JS divergence and MMD, the overall performance of CTGAN remained competitive across evaluation criteria. This indicates that the baseline hyperparameter configuration for CTGAN is already well-suited to the dataset structure, and further increases in model complexity or stability produced only marginal benefits. The results highlight CTGAN’s effectiveness at capturing tabular data distributions, as demonstrated by the convergent validity of all statistical similarity values.

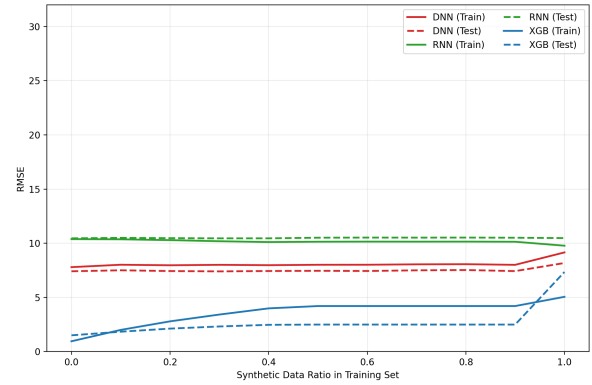
The DDPM produced mixed results, with divergence metrics generally higher than those observed for CTGAN but still competitive with TTVAE under certain configurations. The baseline DDPM configuration achieved moderate KS statistic (0.352) and JS divergence (0.317) values, while the stability-focused configuration slightly improved distributional similarity by reducing the Wasserstein distance relative to the capacity configuration. However, increases in training steps and model capacity did not consistently improve fidelity metrics, suggesting diminishing returns from additional complexity in diffusion-based tabular generation for this dataset. Despite these limitations, DDPM maintained relatively stable TVD across configurations, indicating that DDPM may still capture broader statistical relationships even when marginal distributions diverge slightly.

Overall, these findings reinforce the importance of generative AI model-specific hyperparameter tuning and suggest that while diffusion-based approaches offer promising generative capabilities, GAN-based architectures such as CTGAN currently provide stronger performance for tabular PNT synthetic data generation.

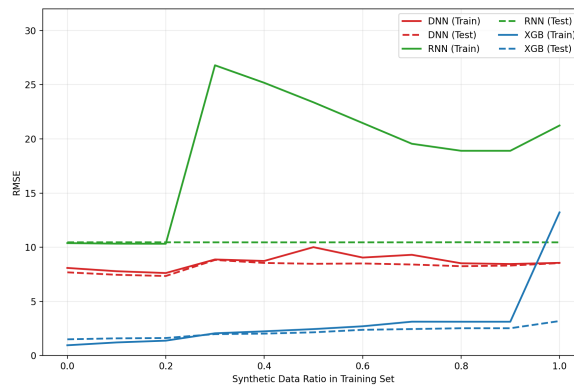
For validation, the machine learning predictive models trained on the stability-oriented data can be seen in Figure 1.



A: TTVAE Predictive Utility Performance



B: CTGAN Predictive Utility Performance



C: DDPM Predictive Utility Performance

Figure 1: Predictive Utility Performance Comparison of Stabilized Generative AI Models

The y-axis measures RMSE, while the x-axis depicts the ratio of synthetic data used to train the predictive models from 0 (all real training data) to 1 (all synthetic training data). The trend is that training with synthetic data typically increases RMSE, thereby decreasing accuracy. The TTVAE and CTGAN models perform similarly, with the DNN and XGB models for the DDPM. The RMSE for training and validating does not deviate greatly until 90% synthetic training ratio, except for the RNN for the DDPM.

Across all the sub-plots in Figure 1, XGBoost maintained the lowest RMSE and should be used for training predictive models for our PNT use case. The DNN is typically in the middle; for the DDPM, it did not deviate at 100% synthetic data. The RNN has the largest RMSE values. TTVAE maintains the closest training and testing RMSE graphs. It performs the best and incurs less change as the ratio of synthetic to real training data increased; however, the TTVAE and CTGAN performs almost the same up to 90% synthetic training ratio. The CTGAN performs best overall, as the validation models trained with CTGAN generated data does not diverge as greatly as the others, meaning it should be utilized for this use case. Pursuing a different use case may return different results regarding training validation models with synthetic-real ratio mixes.

5. Conclusion

This work developed an innovative and repeatable methodology for the synthetic generation and evaluation of tabular PNT data, an area that receives limited focused research despite its growing importance in military operations. By developing a structured analytic pipeline that integrates multiple generative AI model architectures under consistent preprocessing, tuning, and evaluation conditions, this work moves beyond isolated model testing and instead establishes a repeatable experimentation approach to synthetic PNT data generation and evaluation. The methodology emphasizes repeatability, controlled comparison and scalability, ensuring that future research can replicate, adapt, and extend the framework across varying PNT use cases.

Future work will extend this methodology by incorporating additional generative AI approaches such as LLM and

hybrid diffusion–transformer model architectures. The next phase of development includes solution productization through the design and development of a generation and evaluation toolkit that enables users to upload tabular datasets, select generative model(s) and machine learning based predictive model(s), and automatically execute fitting, synthetic generation, verification, and validation workflows. By integrating preprocessing pipelines, hyperparameter configurations, statistical comparison metrics, and downstream validation models into a unified toolkit, this methodology can transition from a research contribution to a deployable capability. Collectively, this work introduces synthetic generation for tabular PNT data as a viable and scalable research direction while establishing a foundation for continued innovation and future operational implementation.

6. References

- Azahad, S., & Hameeda, S. (2023, 01). A deep exposition of gan and its applications. *international journal of engineering technology and management sciences*, 7, 32-37. doi: 10.46647/ijetms.2023.v07i02.005
- Bauer, A., Trapp, S., Stenger, M., Leppich, R., Kounev, S., Leznik, M., ... Foster, I. (2024). *Comprehensive exploration of synthetic data generation: A survey*. Retrieved from <https://arxiv.org/abs/2401.02524>
- Goyal, M., & Mahmoud, Q. H. (2024). A systematic review of synthetic data generation techniques using generative ai. *Electronics*, 13(17). Retrieved from <https://www.mdpi.com/2079-9292/13/17/3509> doi: 10.3390/electronics13173509
- Guo, X., & Chen, Y. (2024). *Generative ai for synthetic data generation: Methods, challenges and the future*. Retrieved from <https://arxiv.org/abs/2403.04190>
- Hansen, L., Seedat, N., van der Schaar, M., & Petrovic, A. (2023). *Reimagining synthetic tabular data generation through data-centric ai: A comprehensive benchmark*. Retrieved from <https://arxiv.org/abs/2310.16981>
- Kiran, A., & Kumar, S. S. (2024). A methodology and an empirical analysis to determine the most suitable synthetic data generator. *IEEE Access*, 12, 12209-12228. doi: 10.1109/ACCESS.2024.3354277
- Livieris, I. E., Alimpertis, N., Domalis, G., & Tsakalidis, D. (2024). *An evaluation framework for synthetic data generation models*. arXiv preprint arXiv:2404.08866. Retrieved from <https://arxiv.org/abs/2404.08866> doi: 10.48550/arXiv.2404.08866
- Miletic, M., & Sariyar, M. (2024). Challenges of using synthetic data generation methods for tabular microdata. *Applied Sciences*, 14(14). Retrieved from <https://www.mdpi.com/2076-3417/14/14/5975> doi: 10.3390/app14145975
- Nafis, N., Esnaola, I., Álvaro Martínez-Perez, Villa-Uriol, M.-C., & Osmani, V. (2025). *Critical challenges and guidelines in evaluating synthetic tabular data: A systematic review*. arXiv preprint arXiv:2504.18544v2 [cs.LG]. Retrieved from <https://arxiv.org/abs/2504.18544v2>
- Paulsen, C. D., & Baker, J. (2025, September 2). *Counter unmanned aerial systems (c-uas) training and implementation at the national training center*. https://www.army.mil/article/288207/counter_unmanned_aerial_systems_c_uas_training_and_implementation_at_the_national_training_center. (Accessed: 2025-10-30)
- Razghandi, M., Zhou, H., Erol-Kantarci, M., & Turgut, D. (2022). Variational autoencoder generative adversarial network for synthetic data generation in smart home. *CoRR*, abs/2201.07387. Retrieved from <https://arxiv.org/abs/2201.07387>
- Thangaraju, D. (2023). Synthetic data generation: Enabling secure use of data for ai, machine learning, and testing. *International Journal of Scientific Research in Engineering and Management (IJSREM)*. Retrieved from <https://ijsrem.com/download/synthetic-data-generation-enabling-secure-use-of-data-for-ai-machine-learning-and-testing/>
- U.S. Government Accountability Office. (n.d.). *Defense navigation capabilities: Dod is developing positioning, navigation, and timing technologies to complement gps* (Technology Assessment No. GAO-21-320SP). U.S. Government Accountability Office. Retrieved from <https://www.gao.gov/assets/gao-21-320sp.pdf>
- Wang, A. X., Chukova, S. S., Simpson, C. R., & Nguyen, B. P. (2024). Challenges and opportunities of generative models on tabular data. *Applied Soft Computing*, 166, 112223. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1568494624009979> doi: <https://doi.org/10.1016/j.asoc.2024.112223>
- Wang, A. X., & Nguyen, B. P. (2025). Tvae: Transformer-based generative modeling for tabular data generation. *Artificial Intelligence*, 340, 104292. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0004370225000116> doi: <https://doi.org/10.1016/j.artint.2025.104292>
- Yuan, Y., Liu, Y., & Cheng, L. (2024). *A multi-faceted evaluation framework for assessing synthetic data generated by large language models*. arXiv preprint arXiv:2404.14445. Retrieved from <https://arxiv.org/abs/2404.14445> (Revised version v2, 24 July 2025) doi: 10.48550/arXiv.2404.14445
- Zhu, J. (2024, 08). Synthetic data generation by diffusion models. *National Science Review*, 11(8), nwae276. Retrieved from <https://doi.org/10.1093/nsr/nwae276> doi: 10.1093/nsr/nwae276