

Data Center Electric Power Estimation based on Generative AI Synthetic Data for System Design and Grid Integration

Grant M. Fischer
SPARK Lab, Pigman Eng. College
University of Kentucky
Lexington, KY, USA
grant.fischer@uky.edu

Aron Patrick
Research and Development
PPL Corporation
Allentown, PA, USA
alpatrick@pplweb.com

Rosemary E. Alden
Electrical Systems Department
ABB US Corporate Research Center
Raleigh, NC, USA
rosemary.alden@ieee.org

Dan M. Ionel
SPARK Lab, Pigman Eng. College
University of Kentucky
Lexington, KY, USA
dan.ionel@ieee.org

Abstract—Innovative strategies and infrastructure adaptations can support the power grid and its increasing number of data center facilities with their highly variable, high-power load profiles. A machine learning (ML) Wasserstein generative adversarial network with gradient penalty (WGAN-GP) model is trained utilizing real data center server-level power profiles under ML training workloads. Realistic synthetic profiles are produced with satisfactory loss convergence, visual similarity, and performance metric calculations including an RMSE below 0.1pu. A case study is conducted to aggregate the real and synthetic power profiles and test multi-megawatt data center loads on an IEEE 9-bus benchmark transmission system, providing methods for future studies for additional energy storage and uninterruptible power supply (UPS) support and design for power electronics to manage larger power fluctuations. These simulations for distributed and concentrated data center load cases, that are enabled by the proposed data generation methods, exemplify data center location studies and reduced voltage violation potential when distributed across the power grid.

Index Terms—Artificial intelligence (AI), generative AI, WGAN-GP, synthetic data, variable power fluctuations.

I. INTRODUCTION

Advanced power delivery strategies designed to accommodate dynamics, including hybrid energy storage systems and advanced control methods, may be able to support the power grid and its increasing inclusion of large-scale data centers, particularly those powering AI workloads with highly variable second-scale load swings [1], [2]. Complex power architectures incorporating optimized power converters and integrated uninterruptible power supply (UPS) systems may be employed to enhance data center reliability and improve stability under dynamic loading conditions [3], [4]. These developments highlight the growing importance of resilient and dynamically responsive power electronics converters and

energy storage architectures for supporting high power AI data centers with variable power fluctuations.

A previous study proposed a method for synthetically generating profiles with a generative adversarial network (GAN) machine learning model trained on the MLPerf Power ML training power profiles, described in further detail in this paper, to generate more power profiles without requiring tests with AI server hardware [5]. The development of a Wasserstein GAN with gradient penalty (WGAN-GP) can address training instabilities found in the previous GAN model and provide a more meaningful loss function to improve the generator's synthetic results [6].

This paper addresses the limited availability of public data center power profile data, which constrains the accurate modeling of data centers for grid reliability studies, electromagnetic transient simulations, and energy management studies for economic generation and energy storage dispatch. The generated data represents server-level power profiles that are aggregated to the facility level, enabling the creation of realistic synthetic datasets from limited available data without requiring extensive AI accelerator computing hardware for producing more data.

The contributions of this paper include the development of a WGAN-GP model to accurately generate synthetic machine learning training workload power profiles with varying lengths. A case study is conducted to aggregate the real data and the generated synthetic data for multi-megawatt scale data center loads for testing on an IEEE 9-bus benchmark transmission system. Also included is a proposal and discussion for future hardware-in-the-loop (HIL) studies evaluating data center penetration scenarios involving energy storage, UPS, and power electronics support.

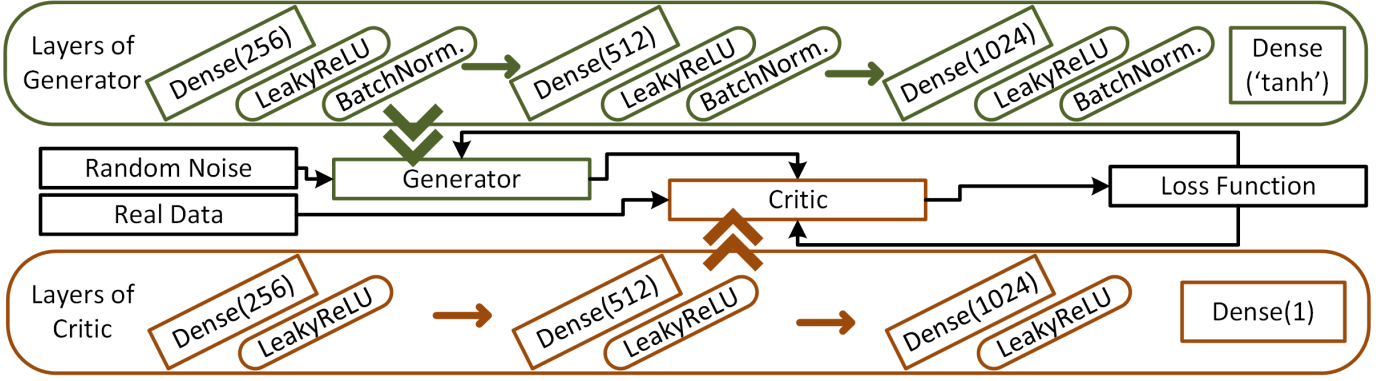


Fig. 1. Structure overview for the developed WGAN-GP model including layers of the generator and critic. The loss function encompasses the Wasserstein-1 distance between the real and generated data distributions and includes gradient penalty calculations based on the critic score outputs.

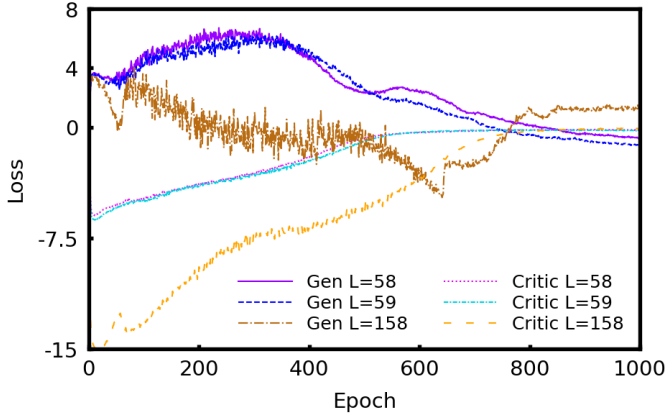


Fig. 2. Generator and critic losses recorded at each epoch of the WGAN-GP model training for the groups with signal lengths 58, 59, and 158. Satisfactory convergence of the generator and critic losses is displayed for all three of the signal groups.

II. METHODS FOR WASSERSTEIN GAN WITH GRADIENT PENALTY

The data center facility aggregated load data for the IEEE 9-bus benchmark transmission system may be produced utilizing generative AI. A WGAN-GP model was developed to produce realistic synthetic signals based on the limited public data center server-level power profiles. The Wasserstein GAN improves upon a previously developed GAN model by replacing the discriminator with a critic and using a loss function based on the Wasserstein-1 distance between the real and generated data distributions, which can improve training stability [7]. In WGAN-GP, a gradient penalty is added to the critic loss to enforce the Lipschitz constraint by penalizing large changes in critic score caused by small variations in the real or generated input signals to further promote the model's stability [6]. The generator and critic's loss with gradient penalty applied can be calculated with the equations:

$$L_C = \mathbb{E}[D(\tilde{x})] - \mathbb{E}[D(x)] + \lambda \mathbb{E} \left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2 \right] \quad (1)$$

$$L_G = -\mathbb{E}[D(\tilde{x})] \quad (2)$$

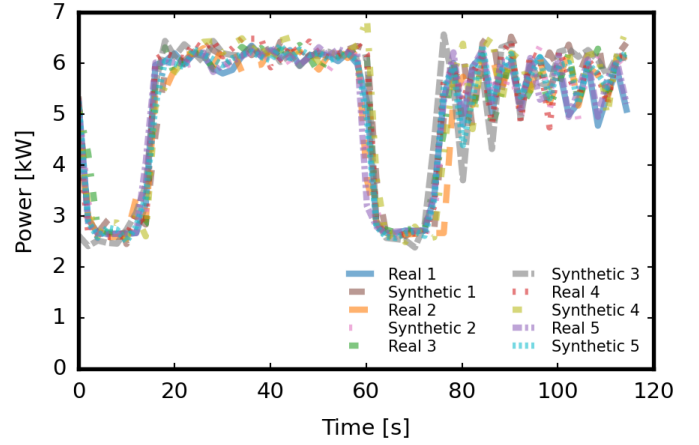


Fig. 3. Five real signals withheld from model training and five synthetically generated signals are plotted following WGAN-GP model training from the 58-length signal results. There is significant visual similarity between the real and synthetic signals.

where $D(\tilde{x})$ represents the critic score for the synthetic signals from the generator, $D(x)$ is the critic score for the real data samples, λ is the gradient penalty coefficient, and $\nabla_{\hat{x}} D(\hat{x})$ represents the gradient of the critic output with respect to the sample interpolated between the real and synthetic samples, which is used to calculate the gradient penalty.

The WGAN-GP model was trained using power profiles from servers performing machine learning training workloads from the MLPerf Power open-source database [9]. This dataset consists of machine learning training power profiles from models DLRM-DCNv2, GPT-3, Llama 2 70B LoRA, ResNet, SSD, 3D U-Net, and BERT. The machine learning profiles used in this study have a high resolution with samples every 2.005 seconds, and the profiles show significant transient power fluctuations as high as 50%. The dataset was cleaned and categorized by signal length and whether the measurement was AC or DC power. The DC signals ranged in length from 40 to 1702 sample length. The signal lengths of 58, 59, and 158 samples were the largest subsets with 304, 202, and 458 signals recorded for each length, respectively. For each signal

TABLE I
WASSERSTEIN GENERATIVE ADVERSARIAL NETWORK WITH
GRADIENT PENALTY TRAINING PERFORMANCE METRICS

Signal Length	Wasserstein Dist.	Mean Diff.	STD Diff.	Classifier Acc.	JS Div.	RMSE [pu]
58	0.014499	0.000572	0.000141	$0.543375 \pm 0.046817^*$	0.022181	0.089284
59	0.013435	0.000018	0.004263	$0.469062 \pm 0.072618^*$	0.021352	0.098213
158	0.010350	0.000386	0.000735	$0.668089 \pm 0.062973^*$	0.025474	0.021862
Excellent / Ideal Thresholds [8]	< 0.15	< 0.05	< 0.10	≤ 0.60	< 0.05	< 0.10
Good / Tolerable Thresholds [8]	< 0.25	< 0.10	< 0.20	≤ 0.65	< 0.10	< 0.20

*To provide a representative classifier accuracy considering the real variability between signal groups, the STD across 200 repeated tests for each group is included.

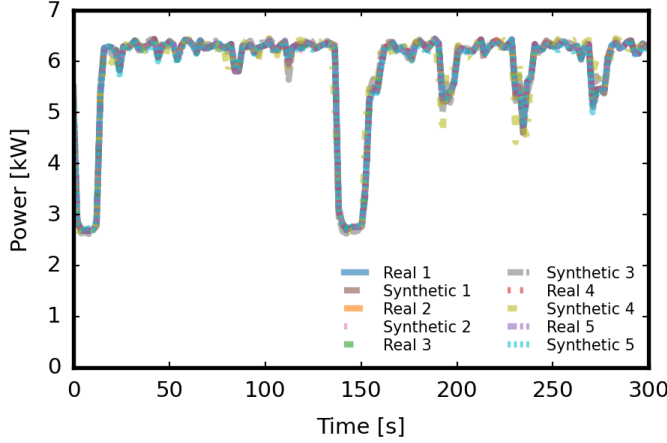


Fig. 4. Real signals withheld from model training and synthetically generated signals are plotted following WGAN-GP model training from the 158-length signal results. Despite comparatively larger generator loss variability, the longer 158-length generated signals still show significant visual overlap with the real data withheld from model training.

group, 20% of the signals were held out from training and reserved for evaluation.

A structural overview is provided in Fig. 1, illustrating the layers and hyperparameters of both the critic and generator. The architecture employs mirrored dense layers of sizes 256, 512, and 1024 with the LeakyReLU($\alpha=0.2$) activations. The generator also utilizes BatchNormalization(momentum=0.8). Some of the initial hyperparameters were adapted from Kaneva *et al.* and subsequently tuned to improve model performance [8]. These include a latent noise dimension of 40 for the 58- and 59-length signals and 105 for the 158-length signals, keeping a latent noise dimension to sample length ratio close to 67%. A critic-to-generator update ratio of 10:1, a batch size of 128, and 1000 training epochs were adapted from the reference architecture [8], while a typical value of 10 was set for the gradient penalty coefficient [6].

III. EXAMPLE STUDIES AND MODEL VALIDATION

In this study, the WGAN-GP model was trained using the three largest DC signal groups, each with quantities above 100. The 58- and 59-length signals have very similar power profiles, and each total approximately two minutes of data. The 158-length signal power profiles are distinct from the 58- and 59-length profiles, and they contain approximately five minutes of data. The model was designed to account for

these two different time scales, and future improvements to this WGAN-GP model can refine the hyperparameters, adjusting for varying signal lengths to expand the model's generalization to different workloads.

The generator and critic losses for all three WGAN-GP trainings are visualized in Fig. 2 and show satisfactory convergence and stabilization over 1000 epochs. For the 58- and 59-length signals, both losses exhibit an initial transient phase before steadily converging toward a small residual gap near zero, consistent with the correlation between loss convergence and training stability [7]. Despite the comparatively noisier generator loss, 158-length signal losses still converge to a slightly larger residual gap and stabilize by the end of the WGAN-GP model training.

An overlay comparison between the generated signals of the WGAN-GP model for the 58-length signals and signals from the 20% real data withheld from training is shown in Fig. 3 with significant visual overlap. The 59-length signals, which are not depicted, share significant visual similarity with the 58-length signals. A comparison between 158-length generated signals and their real counterparts is included in Fig. 4, which also displays significant visual overlap.

To numerically quantify the performance of the model, six performance metrics in Table I were computed utilizing the normalized data of the real and synthetic signals based on the metric calculations performed in Kaneva *et al.* used to validate WGAN-GP models [8]. The calculations for average Wasserstein distance, mean difference, standard deviation difference (STD Diff.), and Jensen-Shannon Divergence (JS Div.) used 304, 202, and 458 generated signals for the groups with signal length 58, 59, and 158, respectively, matching the corresponding real signal counts. The classifier accuracy and root mean square error (RMSE) computations used the set of withheld real data, making up 20% of the real datasets, which was compared against an equivalent number of generated synthetic profiles.

The excellent average Wasserstein distance, mean difference, STD difference, JS divergence, and RMSE performance metrics; the significant visual similarity between the withheld real data and the generated profiles; and the convergence of the generator and critic losses over 1000 epochs all support the conclusion that the WGAN-GP model produces realistic synthetic profiles (Table I). The classifier accuracy model results yielded a wide variability of reported results depending on the random sample of generated data compared.

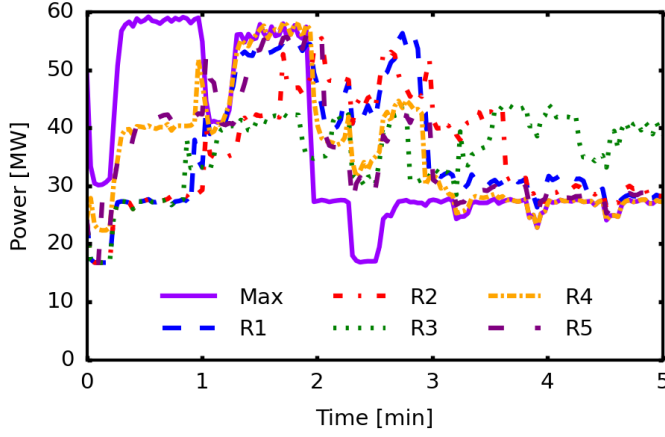


Fig. 5. Aggregated loads including real and generated synthetic signal power profiles to represent large data centers. Six batches of machine learning training loads, including base loads, show load variability with multi-megawatt changes in seconds.

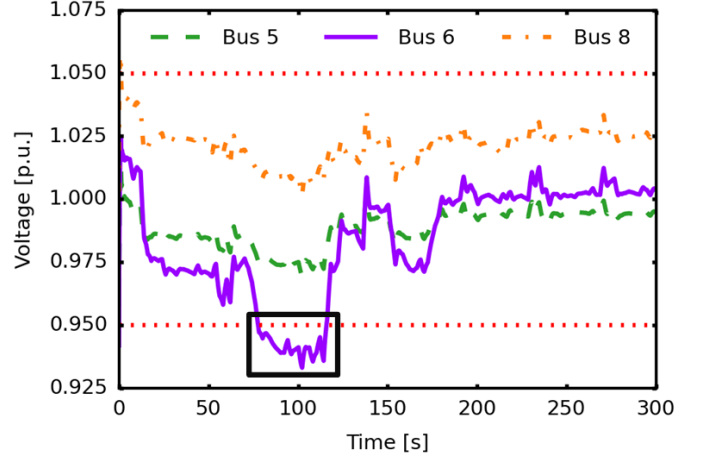


Fig. 7. Concentrated case where per-unit voltages at buses five, six, and eight are recorded with aggregate loads Max, R1, R4, and R5 connected to bus six and R2 and R3 connected to bus eight. A 38s 0.95pu voltage violation occurs during overlapping peak periods on bus six.

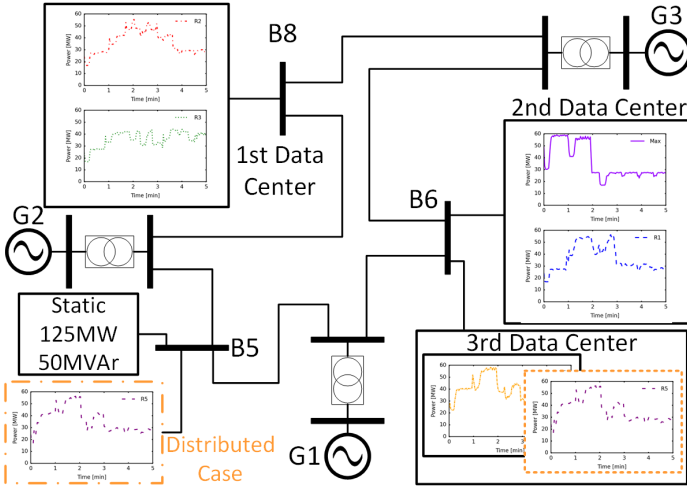


Fig. 6. Diagram of the IEEE 9-bus transmission system with six aggregated loads paired together for three data centers connected as loads at buses eight and six. In the distributed case, the R5 load in the orange dashed box was relocated from bus six to bus five in the dashed-dotted orange box to more evenly distribute the data center loads over the grid.

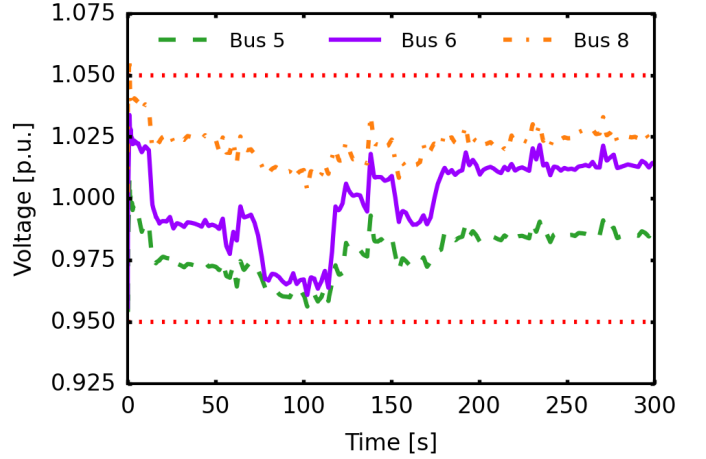


Fig. 8. Distributed case with aggregate loads Max, R1, and R4 connected to bus six; R2 and R3 connected to bus eight; and R5 connected to bus five. This simulation distributed load R5, relocating it to bus five clearing the 0.95pu voltage violation at bus six.

To better characterize the variability of the results, classifier accuracy was computed over 200 repeated samplings, each drawing a subset of generated profiles equal in size to 20% of the corresponding real dataset, from a pool of 2,100 generated profiles. Given the inconsistent performance of the classifier, the other metrics are the best identifiers for WGAN-GP model performance. Overall, out of the six metrics tabulated, five indicate high synthetic data quality. Future improvements could test different approaches for potentially improving the classifier accuracy computation and results, such as withholding a larger portion of data before training.

IV. POWER FLOW WITH HIGHLY VARIABLE DATA CENTER LOADS

A case study was conducted to aggregate the combined real and new synthetic server data to produce multi-megawatt

scale data center loads of different nominal sizes and assess the impacts on transmission systems, specifically the IEEE 9-bus benchmark [10]. All real data and approximately 2,300 generated signals for all three signal groups were included in the aggregation. The data center load shapes account for a power usage effectiveness (PUE) of 1.33, where a quarter of the load consists of a base load to account for all non-IT loads such as cooling and lighting, calculated based on the average of the corresponding aggregated IT power profile. This falls within the projected PUE range for data center facility profiles of 1.15-1.35 [11] and provides validation for the assumptions and aggregation approach. This study assumes that the non-IT loads have a constant power demand across time, as the focus of this study is primarily on the AI server power draw data.

Six example load cases are visualized in Fig. 5, including a max case where multiple batches start simultaneously and

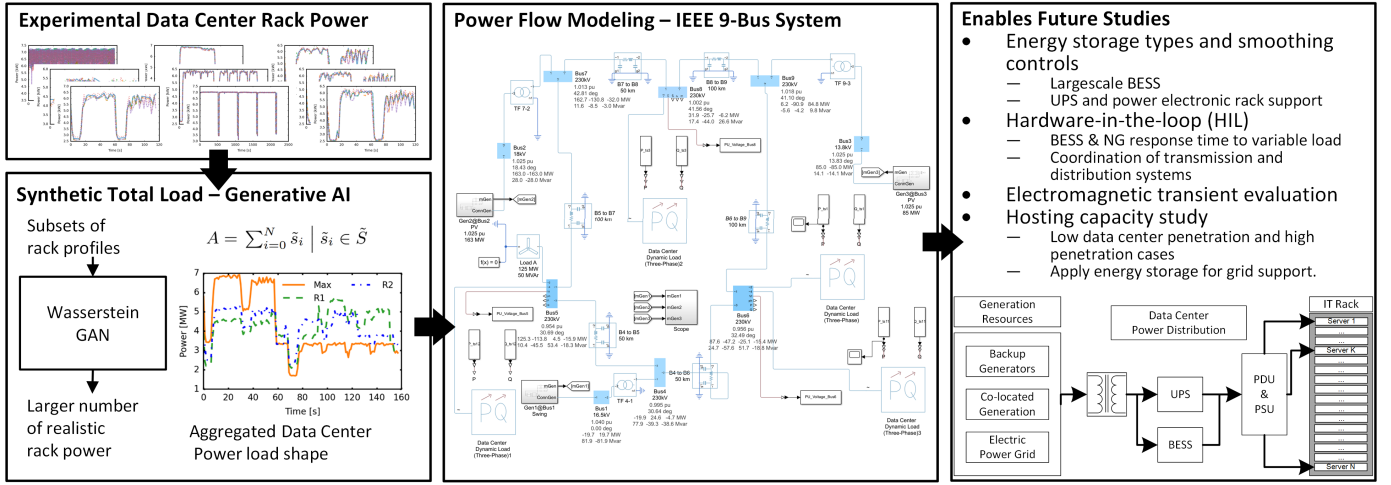


Fig. 9. Flowchart for the proposed future studies including an overview of synthetic aggregate data center load generation employing a WGAN-GP model, the IEEE 9-bus benchmark transmission system, a list of proposed studies and an example diagram for data center server power support. The proposed methods support future studies on power electronics design, transient effects of large load fluctuations, and energy storage for data center power support.

five other cases with random distributions across time. Large dynamic power fluctuations were captured, including an approximate 30MW loss in demand over a few seconds in the max case where the two minute data ends and the servers could be sitting idle for the next large job.

A diagram of the IEEE 9-bus transmission system with connection locations for the six aggregated profiles is included in Fig. 6. In the first power flow simulation study, referred to as the concentrated case, aggregate loads Max, R1, R4, and R5 were connected to bus six and R2 and R3 were connected to bus eight. These loads were adjusted to have a power factor of 0.95 and were simulated with time steps of 10ms. The per-unit voltages at buses five, six, and eight are visualized for this five minute simulation in Fig. 7 where a 38s 0.95pu voltage violation occurs during overlapping peak periods between loads connected at bus six.

A second simulation distributes the data center loads of bus six as an example of varied grid impact based on location. The violation is mitigated without additional energy storage support. The per-unit voltages shown in Fig. 8 display that the distributed data center loads successfully avoided the 0.95pu voltage violation and showcase how the proposed synthetic data generation method enables grid hosting and impact studies. The high variability of data center loads and the high parallelization of training profiles can cause large swings in load shape and increase the likelihood of voltage violations. Distributing the data centers across different areas of the power grid could help integration in regions lacking energy storage support.

V. DISCUSSION

The power flow simulation comparing between the concentrated and distributed data center cases shows that there is real application at higher sampling rate for data generated through the WGAN-GP model. The overview included in Fig. 9 shows proposals for broader applications of this paper's methods to

HIL, energy storage, transient analysis and mitigation, and hosting capacity studies which could support the development of energy storage and UPS support and the design for power electronics systems.

Future work could study optimal power support for reducing the effects of fluctuating data center loads on both the facility and the grid. This includes power electronics and short-term energy storage, such as UPS and supercapacitor systems, as well as battery energy storage systems (BESS) for longer-term energy management. Energy dispatch, HIL, and hosting capacity studies could use the generated aggregate profiles to evaluate grid response to variable data center loads, including the response time of grid resources, coordination between transmission and distribution systems, and the effects of varying levels of data center penetration.

Improvements to the model and its performance metric calculations could further support its accuracy, such as improving the classifier accuracy computation, including frequency domain or spectral comparison of signals, and further optimizing the model for differing length signals and step sizes. The implementation of a Conv1D-WGAN-GP could potentially further improve the model [8]. Further model improvements and more comprehensive validation could expand its applicability to datasets with different sampling rates.

The model optimized for higher resolution datasets could support longer term studies using newly released data sources if more is made publicly available, such as 0.1s resolution profiles and year long data sets from the National Laboratory of the Rockies (NLR), formerly NREL [12]. The release of microsecond resolution data could support studies for electromagnetic transient (EMT) analysis for data center AI training load fluctuations.

VI. CONCLUSION

An advanced generative AI model, a Wasserstein generative adversarial network with gradient penalty (WGAN-GP), was

developed to generate synthetic data of varying signal lengths from real server load profiles measured under AI training workloads, achieving satisfactory performance with RMSE below 0.1pu. A novel case study aggregated the real and synthetic data to form multi-megawatt data center load shapes for evaluating data centers on the IEEE 9-bus benchmark transmission system.

A comparison of concentrated and distributed data center load cases showed reduced voltage violation potential when data centers are distributed across the grid, demonstrating that the synthetic data can support grid-level power flow analysis. The methods proposed in this paper applied to higher resolution data could support the specification and design of uninterruptible power supplies and power electronic converters at the subsystem level and of energy storage at the grid level, enabling the integration of data center transient loads, specialized controls, and electromagnetic transient evaluation.

ACKNOWLEDGMENT

The support of University of Kentucky L. Stanley Pigman Chair in Power Endowment, ABB Inc., and the PPL Corporation is gratefully acknowledged. Any findings and conclusions expressed herein are those of the authors and do not necessarily reflect the views of the sponsor organizations.

REFERENCES

- [1] Y. Li and Y. Li, "AI load dynamics—a power electronics perspective," *arXiv preprint arXiv:2502.01647*, 2025. [Online]. Available: <https://arxiv.org/abs/2502.01647>
- [2] Y. Li, M. Mughees, Y. Chen, and Y. R. Li, "The unseen AI disruptions for power grids: LLM-induced transients," 2024. [Online]. Available: <https://arxiv.org/abs/2409.11416>
- [3] Q. Lin, R. Burgos, D. Dong, X. Li, H. Zheng, and M. N. Aalam, "Architecture analysis and stability evaluation of high voltage DC-powered server racks," in *2024 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2024, pp. 1053–1058.
- [4] S. O. Athar and M. N. Aman, "Machine learning-enabled power converter optimization for next-generation data centers: A review," in *2023 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2023, pp. 6323–6330.
- [5] G. M. Fischer, R. E. Alden, D. D. Lewis, A. Patrick, and D. M. Ionel, "Data center developments for flexible generation dispatch, advanced infrastructure, and ultra-fast digital twins," in *2025 14th International Conference on Renewable Energy Research and Applications (ICRERA)*, 2025, pp. 1984–1989, doi: 10.1109/ICRERA66237.2025.11283760.
- [6] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," 2017. [Online]. Available: <https://arxiv.org/abs/1704.00028>
- [7] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017. [Online]. Available: <https://arxiv.org/abs/1701.07875>
- [8] T. Kaneva, I. Valova, K. Gabrovska-Evstatieva, and B. Evstatiev, "A data-driven approach for generating synthetic load profiles with GANs," *Applied Sciences*, vol. 15, no. 14, 2025. [Online]. Available: <https://www.mdpi.com/2076-3417/15/14/7835>
- [9] A. Tschand *et al.*, "MLPerf power: Benchmarking the energy efficiency of machine learning systems from μ Watts to MWatts for sustainable AI," in *2025 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 2025, pp. 1201–1216.
- [10] MathWorks, "IEEE 9-Bus System," MATLAB & Simulink Documentation, 2026, [Online]. Available: <https://www.mathworks.com/help/sps/ug/ieee-9-bus.html>.
- [11] A. Shehabi *et al.*, "2024 United States data center energy usage report," Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA, Tech. Rep. LBNL-2001637, Dec. 2024. [Online]. Available: <https://eta.lbl.gov/publications/2024-united-states-data-center-energy-usage-report>
- [12] R. Vercellino, J. Willard, G. Campos, W. da Silva Pereira, O. Hull, M. Selensky, and J. Mueller, "Measurement of generative AI workload power profiles for whole-facility data center infrastructure planning," 2026. [Online]. Available: <https://arxiv.org/abs/2604.07345>