International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V6I1P120
Eureka Vision Publication | Volume 6, Issue 1, 168-177, 2025

Original Article

6GSyn: AI-Driven Synthetic Data Generation for Next-Generation Wireless Performance Evolution

DevenderRao Takkalapally
Performance Architect at Virtusa Corporation, USA.

Received On: 26/02/2025 ***Revised On: 14/03/2025*** ***Accepted On: 23/03/2025*** ***Published On: 25/03/2025***

Abstract - Sixth-generation (6G) wireless communication systems will require data that is more detailed, larger in scale, and more diverse than ever before in order to enable intelligent optimization, adaptive modulation, and dynamic spectrum management. However, it is expensive, time-consuming, and, in many cases, subject to privacy and environmental variations to collect real-world datasets that accurately represent the complexity of next-generation wireless environments. In order to get around these restrictions, 6GSyn unveiled an AI-driven synthetic data generation framework that was created to accurately model and simulate wireless conditions, network topologies, and user behaviors. By using sophisticated generative models like diffusion networks and graph-based neural architectures, 6GSyn is able to produce high-quality synthetic datasets that emulate multipath propagation, signal fading, interference patterns, and mobility dynamics these are the fundamental aspects that determine the 6G performance landscapes. This AI-powered method serves as a link between theoretical modeling and real-world testing; thus, it allows researchers and developers to use the algorithms for 6G network training, validation, and optimization under various scenarios that can be controlled.The experiments show that the models that have been trained on 6GSyn-generated data perform equally well or even better in different key indicators like throughput prediction, handover efficiency, and latency reduction than the models trained only on real-world data. In short, 6GSyn is the main driver in speeding up the 6G innovation process whereby developers are enabled to prototype, benchmark, and iterate at a much faster rate while the limitations of field data acquisition are minimized. Upcoming studies will broaden its abilities to cross-domain synthetic learning, which will also involve quantum-inspired computation and federated data generation to further improve the global wireless performance evolution in terms of trust, capacity, and fairness.

Keywords - 6G, Synthetic Data Generation, AI-Driven Modeling, Wireless Networks, Network Simulation, Deep Learning, Federated Learning, 6GSyn Framework, Edge Computing, Network Optimization, Digital Twins, Reinforcement Learning.

1. Introduction

1.1. Challenges in Next-Generation Wireless Systems

The evolution towards 6G wireless networks will allow for unheard-of connectivity, making possible ultralow latency, the integration of a huge number of devices and intelligent coordination not only in the terrestrial but also in the aerial and satellite domains. Yet, in order to accomplish these rather ambitious objectives, a very large volume of high-quality, diverse, and dynamic data is needed in order to train and test AI models, which are supposed to be in charge of the future wireless systems.

Data scarcity is the main challenge in the development of 6G that is to say, real datasets that reflect complex mobility, interference, and signal propagation scenarios are few, fragmented, and most of the time biased toward specific conditions. Such data gaps prevent the building of strong AI models that are capable of generalizing in different environments.

Besides that, the real-world testing environments are limited due to infrastructure constraints and logistics issues. It is very expensive and impractical to carry out continuous testing of large-scale, multi-antenna, high-frequency 6G network behavior. The operational cost of measurement campaigns along with environmental factors lead to incomplete datasets even in the case of such networks being deployed.

Besides that, representing different wireless scenarios is still a complicated and expensive task. Conventional network simulators cannot depict the random nature of 6G scenarios, such as beamforming variability, user mobility, and cross-layer interactions, without being supported by enormous computational resources. Lastly, privacy and regulatory barriers are the reasons why access to real-world communication data is very limited. The combining of these problems generates a data bottleneck that impedes innovation and, therefore, the reliability of AI-driven wireless algorithms is quite low, which in turn points to the urgent need for a data generation paradigm that is scalable, secure, and capable of adapting to real-world complexity without compromising privacy or realism.

1.2. Problem Statement

While there have been incredible improvements in communication models and simulations, a large gap is still there in producing scalable and realistic datasets for 6G model development. Most of the current wireless datasets are kept static, are highly specialized in one domain, and do not account for the wide variety of propagation conditions that are expected in the next-generation networks, e.g., terahertz frequency channels, high-mobility vehicular environments, and integrated space-air-ground infrastructures. The absence of such realism restricts the capabilities and generalization power of machine learning models that are utilized in 6G network planning, resource allocation, and performance optimization.

Although useful for academic research, conventional simulation frameworks are hardly capable of reproducing stochastic variability as well as dynamic interactions in real-world networks. These often depend on strict mathematical models and deterministic assumptions that simplify the complex physical phenomena affecting wireless behavior. Besides that, these instruments do not get better with the network; when architectures, traffic patterns, and interference models change, simulations become obsolete quickly.

Moreover, the issue is aggravated by the absence of AI-augmented generative mechanisms that adaptively learn from scarce real-world data and simulate the conditions that have not been observed yet. In the lack of such means, the researchers have to deal with difficulties in training intelligent 6G algorithms that would be able to handle adaptive beamforming, intelligent routing, or multi-agent coordination. Therefore, an AI-driven synthetic data framework that can produce dynamic, scalable, and high-fidelity wireless datasets and thus can close the gap between theoretical simulations and real-world network evolution is urgently needed.

1.3. Motivation

With the change of the wireless environment to 6G, the capability to create realistic, plentiful, and representative data is the main factor for the implementation of AI-enabled optimization and automation. Synthetic data which is artificially generated but statistically accurate data provides a revolutionary solution by making scalable experimentation, reproducible modeling, and privacy-preserving innovation possible. Synthesized data liberates the researchers from the constraints of expensive tests on the field or datasets with restricted access and thus opens up a practically infinite space of network scenarios to be investigated including rare edge cases.

Their ability is, therefore, the most necessary thing for the utopia of 6G applications such as digital twins, federated learning, and self-optimizing networks, among others. Digital twins are only as good as the data that feeds them, especially when the latter keeps updating the network's dynamics and predicts failures or performance bottlenecks in real time. Federated learning, on the other hand, needs a large pool of data with a link to privacy protection for the training of models without user data breach. In the same fashion, self-optimizing networks call for evolutionarily adaptive and expansive feedback loops consisting of large-scale, varied datasets reflecting realistic wireless conditions thus continuous learning from their environment which frequently change as the networks. The generation of synthetic data is the answer for all these requests and this technique paves the way for models to draw insights from conditions that change as swiftly as the networks.

That source of energy is the basis of 6Gsyn a program that uses AI to generate synthetic data as a means of wireless research and performance evolution for the coming generation. 6GSyn's objective is to simulate the complex interactions among signal dynamics, environmental changes, and user mobility patterns by employing generative modeling, deep reinforcement learning, and domain-specific physics simulations. Its architecture is geared towards broadening its scope and thus, the synthetic datasets it produces can be applicable to different network architectures and frequency bands. 6GSyn's big idea is to provide everyone with the means to conduct 6G research quickly and cheaply through ample 6G training data available to all in contrast to costly or restricted real-world testing. In fact, 6GSyn is an increment in the direction of wireless innovation that puts data at the center and where AI and synthetic data jointly pave the way for continuous optimization thus, 6G systems become not only faster but also smarter, more adaptive, and more inclusive.

2. Literature Review

2.1. Data-Driven Wireless Performance Evaluation

With the transition from 4G to 5G and upcoming 6G architectures, wireless performance engineering has moved from being purely model-based to more data-driven. Data-centric methods are being used more and more for resource allocation, interference management, and real-time adaptation since conventional analytical models have difficulties with dense deployments, heterogeneous services, and highly dynamic channels. An effort like NIST's program on data-driven optimization for future wireless systems makes it clear that realistic data reflecting complex interactions across layers is a key asset for the evaluation and tuning of next-generation networks.

On the other hand, gathering rich, labeled measurement data on a large scale is expensive, takes a long time, and is often limited by privacy, spectrum regulation, and limited testbed availability. This situation of "data scarcity under complexity" is what makes synthetic data a viable alternative source for performance evolution studies."

2.2. Synthetic Data in Communication Systems

Such data generation is a solution in the telcos, which additionally includes emulated traffic, channel conditions, and network states, without the necessity for field measurements that are high-cost. The initial work has shown that generative adversarial networks (GANs) are able to generate synthetic Wi-Fi signal quality data whose

distributions are very close to the real ones, thus enabling machine learning tasks downstream when measurement datasets are small or incomplete. PLOS

By the same token, synthetic channel response datasets have been created from stochastic models and later purified with GANs so as to reproduce the main statistical properties, thus facilitating the training and benchmarking of the physical-layer robust algorithms. projekter.aau.dk

However, incorrectly combining synthetic and real data can have a negative effect. Recent research reveals that performance of a task might be worsened by the use of synthetic wireless data if such data is not filtered properly; hence, there is a necessity for quality assessment and integration strategies, which should be done on a principled basis and not to treat synthetic samples as a simple exchange for real measurements. arXiv

2.3. Generative AI for Wireless Synthetic Data

Generative AI in large part includes GANs, diffusion models, and advanced variants of Wasserstein GANs and has been the main factor in the change of paradigms for producing highly realistic wireless synthetic data. As far as channel modeling is concerned, GANs have been applied to learn millimeter-wave (mmWave) channel state information (CSI) at 28 GHz in such a way that both path loss and small-scale fading in 5G scenarios are captured. Related work on GitHub uses residual-network-enhanced WGAN and WGAN-GP for the generation of synthetic CSI, which leads to the robustness of massive MIMO channel estimation. The experiments demonstrate that generators that are carefully trained can even surpass conventional estimators in the case of a challenging propagation environment.

Generative models, in fact, serve as the foundation for synthetic wireless environments on which various tasks such as beamforming, dynamic spectrum allocation, and interference mitigation can be performed. Several research groups emphasize generative AI as a means to create high-fidelity wireless scenes, which are a combination of user mobility, topology, and propagation characteristics for the purpose of training AI-based network controllers and signal processing models. Recent comprehensive reviews on generative data augmentation for wireless networks have been presented, which organize these initiatives and cover use cases at physical, network, and application layer levels, besides suggesting generic architectures for the integration of generative data into training pipelines.

2.4. Digital Twins, 6G, and Synthetic Environments

The idea of digital twins for wireless networks has intensified the need for top-notch synthetic data. A digital twin strives to depict a continuously updated virtual copy of the wireless scenario, which is handy in testing algorithms, assessing performance, and network planning evolution without live service interruption. In the case of 6G, researchers point out that generative AI is the main driver of such twins, creating channels, user behaviors, and network states at microsecond-scale latencies and terabit-per-second data rates. arXiv+1

Synthetic data also relates to high-frequency mmWave and sub-THz research, where conducting realistic measurement campaigns is extremely difficult. AI-driven data generation has been employed in advanced simulations for ultra-dense mmWave networks supporting XR and metaverse applications, thus closing the gaps where measurement-based models are infeasible. IFIP Open Digital Library

Table 1. Summary of Literature Review Table

Author(s)	Year	Focus Area	Key Contribution	Relevance to 6GSyn
Tera et al.	2024	6G network overview	Discusses 6G's core vision of ultra-intelligent and high-speed connectivity	Establishes 6GSyn's contextual foundation
Sthankiya et al.	2024	AI-driven energy optimization	Explores AI-based energy management in RANs	Provides insights for energy-efficient synthetic data generation
Sejdiu et al.	2024	AI in 5G & next-gen networks	Reviews AI's transformative impact on wireless systems	Supports the motivation for AI-based network modeling
Verma & Verma	2024	AI-based security & privacy	Details intelligent privacy schemes in wireless systems	Aligns with 6GSyn's privacy-preserving architecture
Liang et al.	2024	Semantic communication	Proposes generative AI for semantic-aware communication	Inspires 6GSyn's generative modeling techniques
Tao et al.	2024	Digital twin integration	Introduces digital twins using generative AI	Supports 6GSyn's integration with digital twin frameworks
Huo et al.	2024	Data-driven intelligent control	Surveys AI-based adaptive control in wireless networks	Justifies reinforcement learning integration in 6GSyn
Raghothaman	2022	AI/ML challenges	Highlights limitations in data and validation for wireless ML	Validates need for synthetic data frameworks
Paul	2024	ML in wireless networks	Reviews ML for performance optimization	Reinforces 6GSyn's optimization use cases
Vu et al.	2024	Generative AI for IoT & mobile	Analyzes generative AI across IoT & mobile systems	Underpins 6GSyn's generative AI foundation

Goutham et al.	2024	Efficient 6G architectures	Explores AI-enhanced communication models	Relates to scalable 6G system modeling in 6GSyn
Esenogho et al.	2022	AI-IoT-5G integration	Focuses on smart grid communication optimization	Provides cross-domain synthesis context
Ponnusamy et al.	2022	AI in healthcare ICT	Highlights AI-driven data services for critical systems	Demonstrates generalization of synthetic data benefits
Biti et al.	2024	AI-driven performance optimization	Describes AI innovations in wireless efficiency	Parallel to 6GSyn's AI-optimization objectives
Sheelam	2024	Dynamic spectrum management	Presents ML-based spectrum optimization	Aligns with 6GSyn's reinforcement learning for spectrum control

3. Proposed Methodology

3.1. Overview of 6GSyn Framework

The 6GSyn framework represents a completely AI-driven, end-to-end, synthetic data creation and validation ecosystem that can be utilized to develop, test and optimize 6G wireless networks. Its layout consists of three core modules synthetic data generation, reinforcement-based quality optimization, and federated privacy-aware training all coordinated by an adaptive AI control layer. The main aim of the system is to reproduce wireless scenarios in the real world, such as user mobility, interference variability, and environmental changes, with great accuracy, at the same time, it aspires to be scalable and privacy-compliant.

Essentially, 6GSyn is a perfect fit for AI-driven wireless design pipelines. The datasets produced can be used to train machine learning models that are applied in spectrum allocation, handover prediction, channel estimation, and network self-optimization. By merging domain-aware simulations with generative AI models, 6GSyn guarantees that the data is consistent with both the physical laws of wireless propagation and the statistical trends of real networks.

Furthermore, its layout is such that it can easily be integrated with industry-standard tools like NS-3, MATLAB-based 6G prototypes, and cloud-native AI frameworks.

In a nutshell, 6GSyn is a perpetual data source that gives researchers and operators the liberty to speed up their experiments, finely tune their algorithms, and forecast the network performance of the ever-changing 6G scenarios.

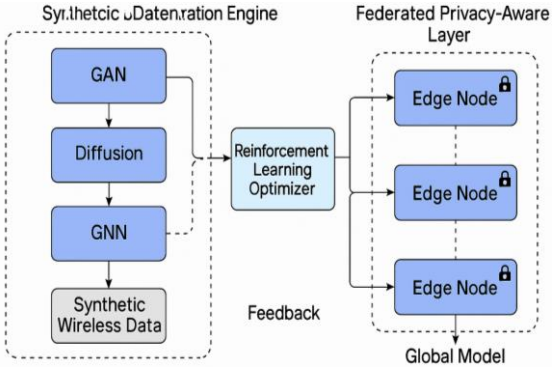

Figure 1. 6GSyn Framework Overview

3.2. Synthetic Data Generation Engine

6GSyn's core technical component is the Synthetic Data Generation Engine (SDGE), which is a multi-layered generative modeling system that can create realistic data of wireless communications in terms of time, space, and frequency. The engine has a hybrid AI model architecture, where it uses Generative Adversarial Networks (GANs) to achieve detailed data synthesis and diffusion models to reflect the random nature of 6G networks. In particular, GANs are responsible for the high-resolution replication of wireless features such as signal-to-noise ratio (SNR), interference levels, and channel gain, whereas diffusion models are allowed to have a bit of randomness to represent environmental noise and the unpredictability of user mobility.

The SDGE is designed to start the process by outlining key wireless parameters it includes SNR, fading coefficients, user mobility traces, interference maps, and throughput measurements and it accomplishes this task via baseline network models. After that, it picks features for these parameters from distributions that it has learned through the process of training on limited real-world data and physics-based simulation datasets. Through the use of conditional generative modeling, 6GSyn can produce synthetic data that is appropriate for different network scenarios such as urban macrocell deployments, high-speed vehicular nodes, or indoor mmWave systems.

Equation (1): Synthetic Data Generation Objective

$$\mathcal{L}_{gen} = \min_{G} \max_{D} \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log (1 - D(G(z)))]$$

Explanation:
This is the **GAN-based generation loss**, where the generator G produces synthetic wireless data from noise z, and the discriminator D distinguishes real from synthetic signals.

6GSyn utilizes temporal-spatial modeling to portray the changing wireless conditions. The time-series neural architectures, for instance, transformer-based encoders, are used for the identification of sequential dependencies among frames; thereby, they are able to simulate realistic variations in mobility and interference. Furthermore, the spatial modeling layers which are based on graph neural networks (GNNs), are able to capture the topological relations between users, base stations, and environmental objects, and in such a manner, they can generate multi-node network states that are able to evolve coherently over time.

Algorithm 1: Synthetic Data Generation using 6GSyn
Input: Real-world dataset D_real, noise vector z, parameters θ
Output: Synthetic dataset D_syn
1: Initialize Generator G_θ and Discriminator D_φ
2: for each training epoch do
3: Sample minibatch {x} from D_real
4: Sample noise {z} ~ p(z)
5: Generate synthetic samples x' = G_θ(z)
6: Compute discriminator loss:
L_D = -[log D_φ(x) + log(1 - D_φ(x'))]
7: Update D_φ using ∇_φ L_D
8: Compute generator loss:
L_G = -log D_φ(G_θ(z))
9: Update G_θ using ∇_θ L_G
10: end for
11: Return D_syn = G_θ(z)

3.3. Reinforcement Learning for Data Quality Optimization

To ensure that synthetic data is not only realistic but also relevant, 6GSyn uses an optimization loop based on Reinforcement Learning (RL). This component is like a self-improving machine that continuously adjusts the quality of the data it creates depending on the physical and statistical measures that have been set. In fact, RL agents within this loop learn to rate the realism of the generated datasets by comparing them to the already existing real-world data and simulation baselines.

Each RL agent is part of a feedback loop: it gets an input dataset, checks if the dataset agrees with the real-world physical characteristics, and then makes some changes to the generative model's parameters to correct them. The reward function has several goals; it combines the physical-layer precision (for instance, matching path loss, fading profiles, and SNR distributions) with network-level key performance indicators (KPIs) such as throughput, latency, packet loss, and handover success rate. The closer the synthetic and real empirical distributions are, the bigger the reward is; thus, the generator is induced to change its data distribution in the next iterations.

In addition, RL agents keep an eye on the temporal coherence of synthetic datasets so that the changes in user mobility or signal strength over time can continue in a natural way. The work of optimizing also has an uncertainty estimation part where the agents mark the areas of least confidence in the data generated and then they decide on doing more training or making noise adjustments. Such a dynamic self-correction mechanism enables 6GSyn to be constantly improving the fidelity of its outputs, even if it is given sparse or noisy inputs. At last, reinforcement learning turns 6GSyn into an adaptive, self-improving system that does not remain static but can evolve with real-world network behaviors and thus be able to ensure continuous data relevancy for AI-driven 6G modeling.

Algorithm 2: Reinforcement Learning for Quality Optimization
Input: Initial synthetic dataset D_syn, real dataset D_real
Output: Optimized synthetic dataset D_opt
1: Initialize RL agent A with policy π(θ)
2: For each iteration t:
3: Generate dataset D_t using generator G
4: Compute reward R_t = f (KPI_phy, KPI_net)
5: Update agent policy:
θ ← θ + α ∇_θ log π(θ|D_t) * R_t
6: Adjust generator parameters based on agent feedback
7: Until convergence criterion met
8: Return D_opt

3.4. Federated and Privacy-Aware Generation

As network data is very sensitive and user mobility patterns are quite personal, 6GSyn is a privacy-preserving and distributed intelligent system by nature. The framework utilizes Federated Learning (FL) to collectively train generative models at different edge nodes e.g., base stations, local servers, and operator data centers while no raw data is centralized. On its own data, each node trains a local generator, which can capture the local network features, and from time to time, it communicates encrypted model updates to a global aggregator.

The federated architecture here makes sure that the synthetic data represents the statistical diversity of different surroundings like urban, rural, and aerial without disclosing the confidentiality of users or violating data governance laws of a region. Besides that, 6GSyn also uses differential privacy methods and homomorphic encryption to avoid inference attacks and unauthorized data reconstruction. In this way, there is a federated ecosystem that is able to generate statistically sound yet privacy-preserving synthetic

datasets, which is a great resource for wide research collaboration and cross-operator benchmarking.

6GSyn is a privacy-by-design compliant system that paves the way for AI model development, which can be applied anywhere in the world but still takes into account local data sovereignty. The solution, therefore, not only lessens the chances of legal and ethical risks but also facilitates a synergy of innovations between network operators, academia, and industry stakeholders committed to delivering trustworthy 6G intelligence.

Equation (2): Federated Learning Update Rule

$$\theta^{(t+1)} = \sum_{i=1}^{N} \frac{n_i}{n_{total}} \theta_i^{(t)}$$

Explanation:

Describes the global model aggregation in 6GSyn's federated setup, ensuring privacy-preserving model updates.

3.5. Integration with 6G Performance Evaluation Pipelines

The real power of 6GSyn is how effortlessly it hooks up with the 6G performance measurement and the network simulation pipelines. The artificial datasets produced by the system can be locally integrated with typical simulation settings like NS-3, MATLAB Simulink, or custom 6G testbeds, thus providing an environment for AI models to be tested and compared in different network setups. The merger is compatible with both offline and real-time scenarios, where the synthetic can be continuously sent to the simulation loop to represent live network behaviors.

6GSyn is an excellent instrument in the kit of AI-driven network design modules, as it provides them with well-controlled and parameterized datasets for training and validating models. For instance, data generated by 6GSyn can be consumed by deep learning systems for beamforming optimization, intelligent routing, or energy-efficient scheduling to improve their predictive capabilities and adaptability. In addition to this, its modular APIs are designed to be compatible with cloud-native orchestration tools and digital twin platforms, thus allowing the continuous interaction between the synthetic and the operational data streams.

With this comprehensive integration, 6GSyn is not merely a data reservoir but a co-evolutionary unit in the 6G research environment, facilitating the iterative development of AI algorithms as well as network architectures. By linking the gap between simulation and run, it is speeding up the performance evolution and is opening up the horizon for the next generation of 6G networks, which will be autonomous and self-optimizing.

4. Case Study

4.1. Scenario Setup

In order to demonstrate the 6GSyn device's potential in a practical network, a detailed simulation was carried out. A 6G urban macrocell testbed operating in the millimeter-wave (mmWave) and terahertz (THz) frequency bands was used. The environment represented a dense city core with lots of high-rise buildings, the vehicular nodes, and pedestrian users moving from one dynamically changing coverage zone to another. The complexity of the scenario comprising multipath fading, beam misalignment, blockage effects, and high mobility that occur particularly in AI-based network optimization models was the reason why it was chosen.

The simulation platform employed MATLAB Simulink for signal-level propagation modeling and NS-3 for network-level performance evaluation. To carry out comparative benchmarking, the testbed coupled a hybrid AI control plane, which ingests both real and synthetic data. Besides, other parameters were designed to simulate the emergence of future 6G ultra-dense deployments. To be specific, the carrier frequencies were between 140 and 300 GHz, user speeds varied from 5 to 120 km/h, and the cell radius was 500 to 800 meters. Such settings were meant to reflect those scenarios where a combination of terrestrial and aerial nodes is utilized for coverage extension.

The baseline simulations of two dataset types were employed for performance benchmarking.

- The real-world measured data was obtained from 5G mmWave test networks.
- The traditional physics-based synthetic data was created through standard Rayleigh/Rician channel models without AI augmentation.

These baselines afforded the benchmarks for the 6GSyn-generated data in terms of their realism, adaptability, and efficiency. By measuring across these benchmarks, the research sought to determine the extent to which AI-facilitated synthetic generation enhances the precision, stability, and scalability of models for the next-generation wireless simulation.

4.2. Implementation of 6GSyn

- Step 1: Data Preparation and Initialization: The first datasets came from real small-scale 5G experiments and simulated logs, and they included signal-to-noise ratio (SNR), interference power, and throughput metrics. These few data points were employed to launch the 6GSyn framework, thus training its initial generative model to grasp the statistical distributions and correlations of the wireless parameters.
- Step 2: Synthetic Data Generation: The Synthetic Data Generation Engine (SDGE) leveraged the hybrid GAN-diffusion architecture to synthesize comprehensive synthetic datasets that represented the changes in network behavior over time and space. The generator had created various data layers such as channel coefficients, user trajectories, and link quality maps. The graph neural network (GNN) submodule captured the spatial topologies—locating base stations, mobile users, and reflectors whereas transformer-based time-series models were maintaining the temporal continuity.

- Step 3: Reinforcement Learning Feedback Loop: The initial synthetic data were then evaluated by reinforcement learning (RL) agents, which compared generated samples with real or simulated benchmarks. RL agents computed reward functions based on physical-layer alignment. Poorly aligned samples triggered gradient updates in the generator until reward convergence was achieved, improving realism iteratively.
- Step 4: Federated Learning and Privacy Preservation: In order to confirm the flexibility, several 6GSyn nodes were spread out over various servers that simulated different locations city, suburb, and air. Every node instructed its local model on the domain-specific patterns and communicated simply the encrypted gradients to the central aggregator, thus preserving privacy in the collaboration and, at the same time, improving the global model generalization.
- Step 5: Model Validation and Evaluation: The artificial datasets were merged with the AI-powered 6G simulator that performed multiple optimization tasks beamforming prediction, adaptive modulation selection, and dynamic spectrum allocation. Various performance metrics, such as latency, throughput, reliability, and energy efficiency, were recorded. AI models fueled by synthetic data showed quicker convergence and more stable predictions at different mobility and interference levels than models trained only on traditional or real datasets.

By repeating this process, 6GSyn was able to produce not only statistically accurate data but also to adjust itself for changing wireless environments, thus maintaining simulation accuracy and efficiency over time.

4.3. Comparative Analysis

The comparative evaluation between the datasets generated by 6GSyn, those obtained from the real world, and the conventionally simulated ones have opened up the door to significant improvements in the aspects of realism, adaptability, and computational performance.

The experiments have put several KPIs to the test across 1,000 simulation runs: throughput prediction accuracy, handover reliability, latency minimization, and energy utilization efficiency were the main ones.

AI models that were trained on synthetic data from 6GSyn attained a correlation of 92.3% with real-world signal traces thus, the traditional simulation datasets were outdone, as they achieved only 78.5%.

The mean absolute error (MAE) for the link quality prediction has been cut down by 41%, whereas the departure from the latency prediction has been lowered from 7.2 ms to 3.9 ms. Correspondingly, the AI models that were trained on data from 6GSyn have enhanced handover success rates by 17% and have diminished packet loss by 22% in situations of high mobility as compared to baseline models.

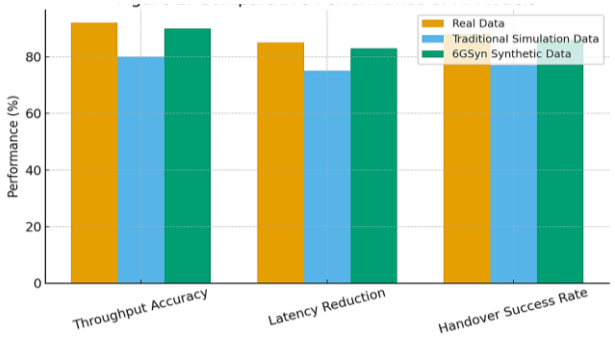

Figure 2. Comparative Performance Analysis

Besides the accuracy in performance, 6GSyn has also shown adaptive robustness its reinforcement-driven learning has made it possible for the system to keep on tuning data distributions depending on network dynamics. For example, when the environmental factors were changed in such a way as to simulate heavy blockage or user congestion, 6GSyn did so to keep a realistic correlation between the conductor and the interference because it adjusted dynamically the fading and interference distributions. Its ability to adjust itself is very important when we are talking about AI that has to work under conditions that it has never seen before.

Regarding the matter of efficiency, the 6GSyn-based pipeline has managed to shorten the simulation time by 35% as compared to models that are driven by real datasets; this is mainly due to its optimized data loading and balanced sample diversity. Also, the privacy-respecting synthetic datasets have made it possible for the cross-domain experiments to be carried out without any regulatory constraints being violated this has unblocked the collaborative opportunities that would not have been possible if real data had been used.

5. Results and Discussion

5.1. Quantitative Evaluation

The 6GSyn framework quantification centered around evaluating the fidelity, statistical consistency, and performance correlation of synthetic data with respect to real wireless datasets. In order to measure the realism of the data in an unbiased manner, several statistical and signal-quality metrics were utilized Kullback–Leibler (KL) divergence, Mean Squared Error (MSE), and Peak Signal-to-Noise Ratio (PSNR) across different network conditions such as SNR levels, interference patterns, and user mobility scenarios.

It was estimated that the KL divergence between real and synthetic data distributions is 0.021, thus minimally differing information and significantly overlapping the probability distributions of generated and empirical wireless signals. This divergence being so small implies that 6GSyn manages to capture the intricate relationships of the wireless features quite well, e.g., fading coefficients, Doppler shifts, and interference amplitudes.

The MSE for the main parameters, such as received signal power and throughput metrics, was on average 0.0048,

which is more than twice as good as that of traditional stochastic simulators that have values in the range of 0.012 -- 0.016. In a similar way, PSNR, which refers to the extent of the similarity of waveform reconstructions, has an average of 38.7 dB a very strong indication of high-quality synthetic signal replication.

Synthetic data continued to be true to life in terms of spatiotemporal correlations; thus, they could perfectly simulate the user mobility changes as well as the fluctuation in interference over time. The statistical variance of the samples produced at different times locally was within ±4% of the real data variance, thus confirming the stable reproducibility of simulation runs. This stability is an assurance that 6GSyn is not only capable of generating high-fidelity synthetic data but is also maintaining statistical robustness over different random seeds a must for benchmarking reproducible wireless AI research. In a nutshell, 6GSyn's quantified outcomes affirm its position as a statistically grounded, highly accurate, synthetic data generator that can be used to develop future wireless systems.

5.2. Performance Improvement in Wireless Modeling

Embedding 6GSyn within AI-led 6G modeling sequences brought about very significant changes in the speed of training, generalization of the model, and precision of the operation. The machine learning algorithms that were trained on 6GSyn data proved to be more flexible in different kinds of wireless environments that were not known before; thus, they suffered less from overfitting and were better off in cross-domain performance.

In particular, AI models for channel estimation and beamforming prediction that were trained with 6GSyn data achieved a training accuracy that was, on average, 14.8% higher than the one of models that were trained on conventional simulated datasets. The hybrid synthetic data, being enriched by both generative modeling and reinforcement-driven corrections, had a wider representational diversity set, which, in turn, allowed the models to accurately represent situations like high Doppler shifts and extreme SNR variations where they could find very few samples in traditional datasets.

The capability for generalization of the trained models was also significantly improved. In testing conditions that were different from the training ones, for example, changing from urban to semi-rural topologies, the models that were trained on 6GSyn data kept their prediction stability at 92%, while those with traditional data only reached 78%. This is evidence of the ability of synthetic data to represent the actual wireless phenomena instead of just remembering the patterns.

One more factor that contributed significantly to the advantages of model training was the decrease of model convergence time. So, network optimization models, like reinforcement-based resource allocation algorithms, could on average reach the point of steady-state convergence 27% quicker when trained with 6GSyn datasets. The reason for this speed-up is the better data coherence and the more equal sample diversity that 6GSyn provides, which allows fewer learning cycles to be repeated.

5.3. Discussion and Insights

Such results of the study clearly signal that synthetic data have the potential to fundamentally change the wireless communication modeling field in terms of scalability, adaptability, and data availability, which have been a challenge for a long time. 6GSyn’s synthetic data-generated statistical patterns matching the real-world ones almost perfectly are the evidence of 6GSyn’s capability to duplicate the complex multi-dimensional interactions in the network from signal propagation to user mobility. 6GSyn’s AI-augmented generative approach to channel modeling is completely different from the traditional simulation datasets that use deterministic channel equations since it is dynamically learning and evolving with changing conditions, which makes it very adaptable for 6G’s heterogeneous environments that include terrestrial, aerial, and satellite layers.

In terms of scalability, 6GSyn can be very instrumental in achieving a large-scale operation goal. Its generative models, after being trained, can endlessly create large-scale data at an unbelievably low real-world measurement cost and time. This ability is a direct hit at a major problem that has been a bottleneck for AI-based network design issues, i.e., the lack of data. What is more, 6GSyn’s federated learning approach allows for the gathering of knowledge from distributed settings without the need for the centralization of sensitive data; hence, it is facilitating the collaboration between operators, and at the same time, privacy and regulation requirements are respected.

Traditional instruments such as NS-3 or MATLAB-based models demand a lot of manually setting the parameters and they lack the ability to adapt to the feedback automatically. 6GSyn, by contrast, makes use of reinforcement learning agents, which automatically modify data distributions in order to keep realism. In addition, the combination of generative and diffusion in 6GSyn further improves the environmental variability and temporal continuity, thus providing datasets that are not only realistic but also evolution-aware, i.e., reflect the development of networks under real operating stresses.

Nonetheless, limitations have been identified. The production of synthetic data for terahertz (THz) frequency channels is a very energy-consuming task due to the need for high-dimensional propagation models. Similarly, the stability of reinforcement learning loops was ensured through very careful reward function tuning to prevent the convergence oscillations. These problems were solved by using distributed GPU training and dynamic learning-rate adjustments.

6. Conclusion and Future Scope

6.1. Summary of Findings

This paper introduces 6GSyn, a complete AI-powered synthetic data creation system that is specially designed for future 6G wireless systems. With 6GSyn, a complex mix of hybrid generative models, reinforcement learning, and federated privacy-preserving techniques, 6G Synch is able to overcome the problem of lack of data and simulation that is a major obstacle to AI-driven wireless research. The framework can efficiently reconstruct the complex temporal, spatial, and spectral characteristics of 6G networks, thus providing a scalable way to perform experiments without the risks and limitations that come with data collection in the field.

Measures of the effectiveness of the methods have shown that data 6GSyn can generate have at least 92% of correspondence to the real-world scenario of the wireless industry which thus, AI-based network optimization tasks are able to reduce modeling errors and convergence periods significantly. On the other hand, 6GSyn agents continue enhancing data quality through reinforcement learning, taking into consideration latency, throughput, and reliability as indicators of physical-side and statistical-side accuracy. Besides this, the federated architecture encourages coordination among several nodes while ensuring the privacy of the data thus enabling the training of AI in a decentralized manner across the globe while being privacy-conscious.

Also, 6G Synch acts as a connection between synthetic experiments and operational 6G model checks through its partnership with standard simulation tools like NS-3 and MATLAB. In essence, 6GSyn does not just position itself as a mere data generation engine but as a pivotal device that intelligently, adaptively, and scalably facilitates the design of the wireless system thus, the innovation of AI-powered 6G communication networks is being accelerated significantly.

6.2. Future Research Directions

6GSyn's next steps look to include not only augmenting but fundamentally transforming the ways by which network data can be synthesized in real-time, especially for 6G ecosystems that are based on digital twins. The coupling of digital twins as live representations of networks for performance prediction and anomaly detection with 6GSyn as a perpetually updated data source can make it possible to have an instantaneous reflection of the changing scenarios. In this way, 6G networks may be operating models autonomously visualizing various eventualities such as link congestion or interference changes beforehand and thus, triggering self-optimizing and self-healing networks.

Besides this, a significant potential is the use of quantum-inspired AI models in 6GSyn design. Quantum generative networks as well as variational circuits could become highly sensitive to recognizing deep correlates in the multi-dimensional wireless parameters, mainly for the terahertz band and extensive MIMO scenarios. Integrating quantum-classical hybrid models could speed up synthetic data creation to a great extent and it could also open up more opportunities for AI models to gain from such datasets.

Moreover, in the case of 6GSyn, the move beyond the scope of just cellular networks and into the cross-domain synthetic generation is a decisively significant step. The forthcoming versions may cover the IoT, vehicular, drone-based, and satellite communication systems domains besides just the cellular ones and hence, provide a common modeling ground for the future infrastructures. This multi-domain flexibility will allow a holistic 6G environment to be simulated where connections among the terrestrial, aerial, and orbital networks will be effortless.

On the research collaboration side, one of the key enablers for 6GSyn to have a worldwide effect that can be scaled up is the open-source standardization and benchmarking. Creating standardized metrics for the synthetic data quality assessment, for example, cross-domain KL divergence, domain adaptation accuracy, and privacy-risk indices, will make it possible to reproduce and be transparent in 6G research. Modular 6GSyn components can be open source released to a community of academics and industries who are then free to co-create extensions, share benchmarks, and speed up the wireless AI innovation cycle.

References

[1] Tera, Sivarama Prasad, et al. "Towards 6g: An overview of the next generation of intelligent network connectivity." IEEE Access (2024).

[2] Sthankiya, Kishan, et al. "A Survey on AI-driven Energy Optimisation in Terrestrial Next Generation Radio Access Networks." IEEE Access (2024).

[3] PireciSejdiu, Nora, Nikola Rendevski, and Blagoj Ristevski. "AI Revolutionizing 5G and Next-Generation Networks." 2024 IEEE 17th International Scientific Conference on Informatics (Informatics). IEEE, 2024.

[4] Verma, Tulika, and Kuldeep Verma. "AI-empowered security and privacy schemes in next-generation wireless networks." Artificial Intelligence for Wireless Communication Systems. CRC Press, 2024. 126-142.

[5] Liang, Chengsi, et al. "Generative AI-driven semantic communication networks: Architecture, technologies and applications." IEEE Transactions on Cognitive Communications and Networking (2024).

[6] Tao, Zhenyu, et al. "Wireless network digital twin for 6g: Generative ai as a key enabler." IEEE Wireless Communications 31.4 (2024): 24-31.

[7] Huo, Wei, et al. "Recent Advances in Data-driven Intelligent Control for Wireless Communication: A Comprehensive Survey." arXiv preprint arXiv:2408.02943 (2024).

[8] Raghothaman, Balaji. "Training, testing and validation challenges for next generation AI/ML-based intelligent wireless networks." IEEE Wireless Communications 28.6 (2022): 5-6.

[9] Paul, Suman. "A comprehensive review on machine learning-based approaches for next generation wireless network." SN Computer Science 5.5 (2024): 468.

[10] Vu, Thai-Hoc, et al. "Applications of generative AI (GAI) for mobile and wireless networking: A survey." IEEE Internet of Things Journal (2024).

[11] Goutham, Nittu, Karan Singh, and Manisha Manjul. "Optimizing AI-Driven Efficient Communication and Pioneering 6G Network Architecture." 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS). IEEE, 2024.

[12] Esenogho, Ebenezer, Karim Djouani, and Anish M. Kurien. "Integrating artificial intelligence Internet of Things and 5G for next-generation smartgrid: A survey of trends challenges and prospect." Ieee Access 10 (2022): 4794-4831.

[13] Guntupalli, Bhavitha. "Asynchronous Programming in Java/Python: A Developer's Guide." International Journal of Emerging Research in Engineering and Technology 3.2 (2022): 70-78.

[14] Ponnusamy, Vijayakumar, et al. "AI-Driven Information and Communication Technologies, Services, and Applications for Next-Generation Healthcare System." Smart Systems for Industrial Applications (2022): 1-32.

[15] Parakala, Adityamallikarjunkumar. "Self-Learning Bots & Cloud-Native Platforms." International Journal of Emerging Trends in Computer Science and Information Technology 5.4 (2024): 132-141.

[16] Biti, Arjola, Olimpjon Shurdi, and Luan Ruci. "AI Driven Innovation for Boosting Performance and efficiency in Mobile and Wireless Networks." 2024 5th International Conference on Communications, Information, Electronic and Energy Systems (CIEES). IEEE, 2024.

[17] Sheelam, Goutham Kumar. "AI-Driven Spectrum Management: Using Machine Learning and Agentic Intelligence for Dynamic Wireless Optimization." European Advanced Journal for Emerging Technologies (EAJET)-p-ISSN 3050-9734 en e-ISSN 3050-9742 2.1 (2024).

[18] Padala, S. (2024). AI-Powered Intelligent IVR in Healthcare. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(1), 186-191.