Hallucination in Generative AI-Enabled Wireless Communications: Concepts, Issues, and Solutions

Xudong Wang, Jiacheng Wang, Lei Feng, Dusit Niyato, Ruichen Zhang, Jiawen Kang, Zehui Xiong, Hongyang Du, and Shiwen Mao

Abstract

Generative AI (GenAI) is driving the intelligence of wireless communications. Due to data limitations, random generation, and dynamic environments, GenAI may generate channel information or optimization strategies that violate physical laws or deviate from actual real-world requirements. We refer to this phenomenon as hallucination in GenAI-enabled wireless communications, which results in invalid channel information, spectrum wastage, and low communication reliability but remains underexplored. To address this gap, this article provides a comprehensive concept of wireless hallucinations in GenAI-driven communications, focusing on hallucination mitigation. Specifically, we first introduce the fundamental, analyze its causes based on the GenAI workflow, and propose mitigation solutions at the data, model, and post-generation levels. Then, we systematically examine representative hallucination scenarios in GenAI-enabled communications and their corresponding solutions. Finally, we propose a novel integrated mitigation solution for GenAI-based channel estimation. At the data level, we establish a channel estimation hallucination dataset and employ generative adversarial networks (GANs)-based data augmentation. Additionally, we incorporate attention mechanisms and large language models (LLMs) to enhance both training and inference performance. Experimental results demonstrate that the proposed hybrid solutions reduce the normalized mean square error (NMSE) by 0.19, effectively reducing hallucinations in wireless communications.

Introduction

Generative AI (GenAI) technology has garnered widespread attention and research interest in recent years. By learning from vast amounts of training data, GenAI captures patterns and structures to generate content with similar styles or information. Notably, thanks to its powerful ability to model complex scenarios and dynamically adapt, GenAI is increasingly being applied in wireless communications, such as channel estimation, interference management, and resource allocation. By leveraging its generative capabilities, GenAI-driven models can facilitate efficient network optimization, enhance adaptability to dynamic wireless environments, and improve spectrum utilization through predictive and generative reasoning [1]. The integration of GenAI into wireless networks enables autonomous modeling of complex wireless environments, paving the way for intelligent wireless networks capable of self-learning and adaptive decision-making.

Despite significant progress, the application of GenAI in wireless communications still faces challenges, including the emergence of hallucinated outputs that we term "wireless hallucination" for clarity.[1] Hallucination is a widely discussed issue in GenAI applications especially large language models (LLMs), referring to the generation of false, inaccurate, or entirely fabricated content by GenAI models [2]. With subtle differences from errors, which typically stem from computational mistakes or incorrect input data, hallucinations involve the model producing seemingly plausible but unverifiable or entirely fictional information. While errors can often be corrected through debugging or data refinement, hallucinations require more advanced mitigation techniques, such as fact-checking or retrieval-augmented generation. The causes of hallucination are primarily associated with the model's training methodology, limited training data, and the probabilistic nature of generative models [3]. This phenomenon also manifests in GenAI-enabled wireless communications, where domain-specific hallucinations can lead to resource inefficiency, degraded user quality of experience (QoE), and potential security threats. For example, in RF signal generation, existing GenAI models struggle to accurately capture the time-frequency characteristics and hardware-specific traits of the devices, leading to RF signals that appear structurally reasonable but

[1] While we adopt the term "wireless hallucination" to describe the domain-specific phenomenon of GenAI-enabled wireless communications, we clarify that it is a contextual extension of the general hallucination behavior known in generative models.

Digital Object Identifier: 10.1109/MCOM.001.2500124

Xudong Wang and Lei Feng (corresponding author) are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, China; Jiacheng Wang, Dusit Niyato, and Ruichen Zhang are with the College of Computing and Data Science, Nanyang Technological University, Singapore; Jiawen Kang is with the School of Automation, Guangdong University of Technology, China; Zehui Xiong is with the Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore; Hongyang Du is with the Department of Electrical and Electronic Engineering, University of Hong Kong, Pok Fu Lam, Hong Kong SAR, China; Shiwen Mao is with the Department of Electrical and Computer Engineering, Auburn University, USA.

0163-6804/26/$25.00 © 2026 IEEE

deviate from real-world physical constraints. Unlike conventional errors, these wireless hallucinations introduce fictitious spectral components or unrealistic distortions that do not correspond to any valid transmission, thereby weakening the robustness of classifiers and reducing wireless sensing capabilities [4]. Actually, existing work on GenAI-empowered wireless communications has not systematically addressed this critical challenge, nor proposed comprehensive solutions to tackle it.

To this end, this article is the first to investigate hallucinations in GenAI-enabled wireless communications and propose mitigation strategies. We begin by reviewing hallucinations in GenAI models, defining wireless hallucination, categorizing its types, and identifying its causes. Then, we propose mitigation strategies at the data, model, and post-generation levels. Additionally, we analyze existing GenAI-enabled wireless communication schemes, focusing on their potential hallucination causes and employed mitigation methods. Since most studies adopt a single mitigation approach with limited performance, we introduce a GenAI-enabled channel estimation scheme as a case study and propose an integrated mitigation solution to enhance estimation accuracy. The main contributions are summarized as follows:

- We provide a comprehensive overview of generalized hallucination in GenAI models and define wireless hallucination from the perspective of communications. By systematically analyzing the workflow of GenAI-enabled communications, we reveal its causes and propose solutions at the data, model, and post-generation levels.
- We review existing GenAI-enabled wireless communication paradigms from the physical layer, data link layer, and network layer, such as radio map construction and GenAI-enabled semantic communication. Special focus is given to identifying the causes of wireless hallucination and exploring mitigation strategies.
- We design an efficient integrated strategy to reduce wireless hallucination in GenAI-enabled uplink channel estimation. More specifically, generative adversarial networks (GANs) are used to balance channel data distribution, while an attention mechanism enhances the reverse denoising of the diffusion model. Additionally, an LLM-enhanced mixture of experts (MoE) is developed to select optimal experts based on user information. Simulation results show that the proposed integrated approach reduces normalized mean squared error (NMSE) by 0.19, while effectively reducing wireless hallucination rate (WHR).

Hallucination in Wireless Communications

Overview of Basic Hallucination in GenAI Models

The hallucinations and their underlying causes in common GenAI models are discussed based on their principles, including Variational Autoencoders (VAEs), GANs, generative diffusion models (GDMs), and Transformers:

Variational Autoencoders: VAEs [5] compress data into latent space representations using an encoder, then decode samples from the latent space to generate new data resembling the original input. However, over-simplification of the latent space and random sampling can introduce ambiguity and detail loss, leading to hallucinations. A typical example is contradictory sentences in descriptive text generated by VAEs.

Generative Adversarial Networks: GANs [6] consist of a generator and a discriminator that compete to realistically mimic data distributions. Mode collapse and generator reliance on shortcuts can result in unrealistic or repetitive outputs. For instance, hallucinations may manifest as faces with multiple pupils or missing ears in generated images.

Generative Diffusion Models: GDMs [7] gradually add noise to data and learn to reverse the process, generating high-quality new samples. However, error accumulation and an imbalance between global structure and fine details during iterative denoising can produce artifacts. Such hallucinations may appear as sudden pitch changes in generated music.

Transformers: Transformers [4] use self-attention mechanisms to model global relationships in input data, generating new sequences by predicting the next element. However, over-reliance on learned correlations and attention biases can result in outputs that are plausible but factually incorrect or inconsistent. This can lead to hallucinations such as motion trajectories in generated videos that defy physical laws.

We summarize the principles of different GenAI models, and illustrate the potential causes and manifestations of hallucinations within each model, shown as Fig. 1.

Definition of Wireless Hallucination

While the term hallucination is widely used in natural language and vision domains, in this work, we extend this concept to wireless communications. We use the term wireless hallucination to describe its domain-specific manifestations, where GenAI models generate outputs that appear plausible but are physically invalid or contextually inappropriate under real-world wireless conditions.

In wireless communications, such hallucinations manifest as fabricated, distorted, or logically flawed representations of channels, signals, protocols, or resource allocation schemes, potentially degrading system performance or rendering solutions infeasible.

To make the notion of wireless hallucination operational, we propose a more rigorous signal-theoretic definition. For the given input x and randomness z, the GenAI output $\hat{y}$ is said to exhibit wireless hallucination if it either

- Violates the physical feasibility set C_T,
- Its distortion exceeds a task-dependent threshold τ_T, with the corresponding mathematical expression shown in Fig. 1.

Wireless hallucinations is categorized into the following types:

- **Constraint Violation Wireless Hallucination:** This occurs when a GenAI model generates content that violates theoretical or technical constraints in wireless communication. For instance, if the total spectrum is 20 MHz, but the model allocates 25 MHz to a user, it exceeds the available bandwidth.
- **Fabricated Output Wireless Hallucination:** This occurs when a GenAI model generates seemingly realistic but entirely unsuitable optimization solutions or content. For example, it

The hallucinations and their underlying causes in common GenAI models are discussed based on their principles, including Variational Autoencoders (VAEs), GANs, generative diffusion models (GDMs), and Transformers.

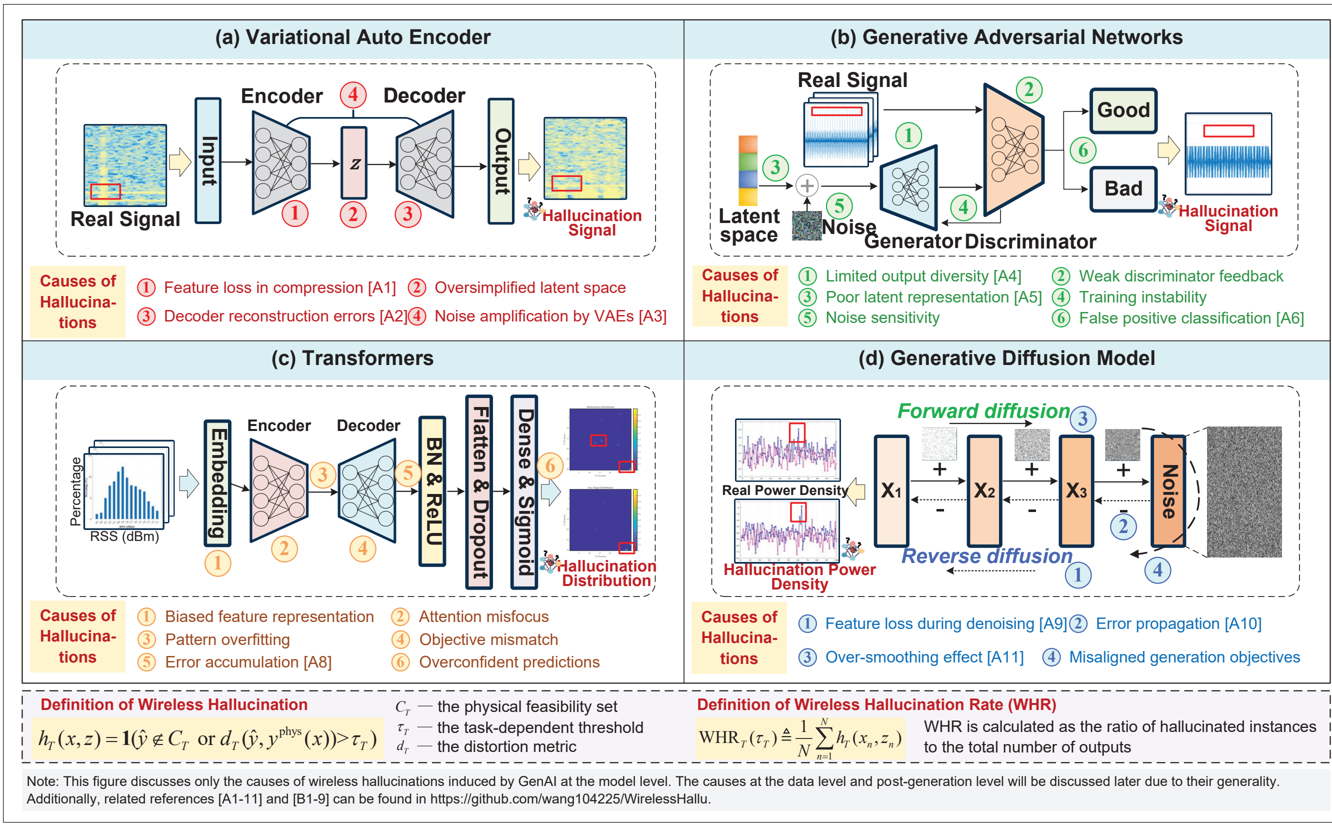

FIGURE 1. A summary of different types of GenAI models, potential causes and manifestations of hallucination.

may create a new channel fading model for high-density urban environments that does not exist in academic literature or industry standard.

- **Context Detachment Wireless Hallucination:** This occurs when GenAI generates content deviating from task objectives, problem context or the actual requirements. For instance, when simulating urban channels, GenAI may incorrectly use rural parameters, resulting in overly long propagation distances and low multipath density.

Causes of Wireless Hallucination

Before analyzing the causes of wireless hallucinations, we review the GenAI workflow in wireless communications. First, the system collects and preprocesses real-time data, including Channel State Information (CSI), network traffic, and user locations. This data then undergoes preprocessing and feature extraction to meet the GenAI model's input requirements. Next, based on the problem's needs, the system feeds the data into a suitable GenAI model for training, guiding the model to converge towards predefined optimization goals, such as throughput or latency. Finally, the trained model generates decisions or content based on real-time input signals or network status. From this workflow, we identify key factors contributing to wireless hallucinations.

Insufficient or Biased Data: Insufficiently diverse training data causes GenAI models to learn narrow features and fail on unseen wireless scenarios [4, 8], while biased data further skews inference [9]. Unlike classical models grounded in explicit physical constraints, GenAI relies on implicit data mappings and lacks domain priors, making it prone to infeasible or non-physical outputs under underrepresented conditions.

Inherent Randomness in GenAI Models: GenAI models, especially GDMs, begin with random noise and gradually denoise it through a reverse diffusion process to restore the target data. Variations in the initial noise can cause differences in the generated images or signals, and each denoising step may introduce randomness. As a result, even with the same input and constraints, GenAI models can produce varied outcomes, leading to some results meeting the requirements while others are unreasonable or impractical [10].

Absence of Validation Procedure: If the validation procedure is missing, the results from GenAI may not be compared or validated against actual requirements and constraints. The generated solutions may overlook these constraints, leading to outputs that violate real-world communication rules or fail to adapt to the actual environment, thus producing unrealistic results and causing wireless hallucinations [1, 6].

To systematically capture the underlying causes of wireless hallucination, we formalize its threat model as a triplet $\mathcal{H} = (D, G, V)$, where D denotes data bias, G, and V captures the absence of validation mechanisms. This formalized threat model offers a unified framework that supports the structured design of mitigation strategies presented in the next section.

Optimization Objectives for Wireless Hallucination

We summarize the evaluation metrics used to assess wireless hallucinations, including both wireless performance metrics and hallucination-aware metrics, as detailed below:

- **Wireless Performance Metric:** Wireless performance metrics quantify the task-specific impact of hallucinations in GenAI-enabled communications. These metrics reflect how hallucinated outputs degrade system behavior. For instance, Bit Error Rate (BER) measures decoding accuracy and is sensitive to non-physical signal distortions. Spectral Efficiency (SE) evaluates bandwidth utilization, where hallucinated resource allocations may cause inefficiencies. Normalized Mean Squared Error (NMSE), used in channel estimation, captures deviations from real channel states.

- **Hallucination-Aware Metric:** We introduce the Wireless Hallucination Rate (WHR) to quantify the proportion of generated outputs that deviate significantly from expected performance. Wireless hallucination rate is defined as the proportion of test samples for which the generative wireless model produces outputs that are judged to be hallucinations, where an output is considered hallucinated when its relative error exceeds a predefined threshold compared to the reference or ground truth. The computation of WHR is shown in Fig. 1, where N denotes the number of test instances and τ_T represents the threshold. WHR offers a unified metric to assess hallucination frequency across wireless tasks.

Mitigation Solutions for Wireless Hallucination

Data-Level Solutions

Data Augmentation: Data augmentation techniques help expand training datasets, addressing wireless environment variability and reducing hallucinations. For instance, injecting noise, time shifts, and frequency offsets enhances signal diversity. Simulation platforms such as NS-3 can generate rich datasets, modeling the effects of mobility patterns, user distributions, device types, and weather conditions on network performance. Additionally, semi-automated annotation tools, such as labeling platforms, can improve the accuracy of network data annotations.

Data Cleaning: Anomaly detection can effectively identify data deviations caused by sensor faults, network issues, or external interference. Techniques like Z-score and interquartile range (IQR) detect outliers, while clustering methods such as K-means analyzes multi-dimensional wireless data (e.g., signal strength, latency, throughput) for anomalies. Additionally, hash-based duplicate detection minimizes computational overhead from redundant transmissions.

Model-Level Solutions

Dynamic Feedback and Reinforcement Learning (RL): RL enables online decision-making by training an agent through trial-and-error interactions with the wireless environment. This closed-loop mechanism allows the agent to adjust generation strategies in real time based on dynamic network states such as interference, fading, and mobility. As a result, RL constrains generative outputs within physically valid and context-appropriate boundaries [1]. Unlike static supervised learning, RL discourages hallucinations by penalizing unrealistic or suboptimal actions, mitigating constraint violations and enhances robustness in non-stationary environments.

Attention Mechanism: An attention mechanism can dynamically focus on different parts of the input data, enabling more effective information extraction and modeling, particularly in the presence of high noise or incomplete data, thereby reducing the occurrence of hallucinations. For example, in frequency-selective fading channels, attention can help the model learn frequency-domain correlations, leading to more stable estimation results and minimizing the likelihood of hallucinated channel responses.

Adversarial Training: Adversarial training involves generating adversarial examples to train the model, enabling it to better handle complex, anomalous, or disruptive environments, thereby improving the model's robustness and reducing hallucinated outputs. Generative Adversarial Networks have the ability to generate perturbation signals or data, and using these adversarial examples during training can effectively help the GenAI model enhance its adaptability to irregular data.

Post-Generation Solutions

Interactive AI (IAI): IAI verification integrates wireless expertise with machine automation, enabling real-time feedback and adjustments to generated results. LLMs, with their strong natural language understanding and generation capabilities, serve as effective tools for interpreting and validating communication protocols and constraints. LLMs can identify issues in generated network management policies and configuration, e.g., signal interference, congestion, and even network attacks. As such, real-time feedback can be achieved to detect, prevent, and correct hallucination.

Mixture of Experts: The MoE architecture enhances accuracy and adaptability by integrating multiple expert models, selecting the most suitable one for specific tasks while maintaining efficiency. Through a gating mechanism, MoE ensures that GenAI-generated solutions are validated across the appropriate expert domains. Additionally, when GenAI proposes multiple implementation approaches, the MoE framework evaluates and optimizes them by leveraging the appropriate expert, ultimately identifying the best solution and mitigating wireless hallucinations.

Wireless Hallucination and Mitigation in GenAI-Enabled Communications

We summarize the principles, hallucination expressions, and causes of hallucination in GenAI-enabled wireless communications from the physical layer, data link layer, and network layer, as illustrated in Table 1.

Physical Layer Perspective

GenAI supports wireless data augmentation, signal detection, and time-series prediction, but still struggles to generate accurate RF signals because it cannot fully model complex-valued, time-varying properties. These limitations lead to hallucinations such as distorted spectra, phase errors, and over-smoothed frequency content. To mitigate this,the authors in [4] introduced a time-frequency diffusion model with a hierarchical diffusion transformer, improving spectral fidelity through joint time-frequency noise injection and separate optimization of different noise types. For channel estimation, hallucinations are measured via SNR, and the method reduces them by 77.5% compared with VAE-based approaches in 5G FDD

GenAI supports wireless data augmentation, signal detection, and time-series prediction, but still struggles to generate accurate RF signals because it cannot fully model complex-valued, time-varying properties.

Type	Use Case	Principles	Hallucination Expressions	Causes of Hallucination	Paper
Physical Layer	Signal generation [4]	GDMs gradually generate results approximating real RF signals through the processes of noise addition and removal.	• Neglecting or distorting critical characteristics of RF signals; • Over-smoothing or excessive noise in the generated signal spectrum	• Insufficient diversity in training data [B1]; • Lack of temporal modeling capability in GDMs; • Difficulty in capturing the complex-domain characteristics of RF signals [B2];	[4], [B1]–[B2]
	Radio map construction [9]	GenAI models leverage deep learning to learn signal propagation patterns from sparse data and generate high-resolution radio maps.	• Fails to accurately reflect multipath effects and signal attenuation; • RSS values violate physical constraints;	• Insufficient diversity in training data; • Data bias; • Ignoring physical laws [B3]; • Lack of real data validation;	[9], [B3]
Data Link Layer	Semantic communication [10]	GenAI enhances background knowledge by generating context-aware samples to improve the accuracy of semantic communication.	• The encoding process may include irrelevant or ambiguous information; • The decoder's output may deviate from its original intent;	• Irrelevant semantic redundancy [B4]; • Limited number of generated samples; • Lack of explicit semantic information modeling in the model [B5];	[10], [B4]–[B5]
	NOMA [6]	GANs generate the multi-dimensional sparse codewords approximating the distribution through adversarial training.	• Generated codebooks do not meet design requirements; • Excessive decoding delay of the generated codebooks;	• Mode collapse in GANs; • Gradient vanishing issue during SCMA codebook training process [B6]; • Limitations in discriminator feedback;	[6], [B6]
Network Layer	Wireless QoE improvement [13]	GDMs generate resource optimization strategies through iterative denoising, ensuring personalized AIGC service provisioning.	• Generates images that misalign with user input intentions or preferences; • Produces content with logical inconsistencies;	• Lack of global optimization in GDMs; • Absence of real-time perception and feedback mechanisms for user personalization;	[13], [B7]
	Security communication [12]	GDMs enhance interference resistance by learning channel noise distribution and recovering the original signal during denoising.	• Misclassifying legitimate signals as attacks; • Significant channel capacity reduction in the pursuit of enhanced security;	• Lack of effective handling and optimization for multiple objectives simultaneously; • Limited capability in efficiently processing high-dimensional and complex data.	[12], [B8]–[B9]

TABLE 1. The principles, hallucination manifestations, and causes of hallucination in the application of GenAI to wireless communications from the perspectives of the physical layer, data link layer, and network layer.

systems. In radio-map construction, hallucinations stem from inadequate propagation modeling. The *SpectrumNet* dataset in [8] combines ray tracing, terrain data, and ITU climate models to enhance generalization and strengthen GenAI robustness under diverse propagation conditions.

Data Link Layer Perspective

At the data link layer, GenAI enables semantic communication and non-orthogonal multiple access (NOMA) but remains vulnerable to hallucinations from contextual ambiguity and unstable training. In semantic communication, misaligned background knowledge can cause semantic distortion. To address this, the authors in [9] proposed the Gen-SC framework as a model-level solution, which integrates adversarial training with a context regulator to better align encoding and decoding, reducing hallucinations by 9.76% based on semantic similarity. In NOMA systems such as sparse code multiple access (SCMA), GenAI may generate impractical codebooks due to mode collapse and domain mismatch. The method in [6] employs a Wasserstein GAN with an autoencoder, where the Wasserstein loss stabilizes training and the encoder improves codebook structure and decodability, enhancing performance in 6G SCMA settings.

Network Layer Perspective

At the network layer, GenAI aids QoE-aware optimization and secure communication but remains prone to hallucinations due to limited feedback and generalization. For QoE optimization, GenAI may generate content misaligned with user preferences under dynamic network conditions. To improve adaptability, the authors in [1] integrated reinforcement learning with an LLM-based user simulator that models user traits and generates QoE scores, achieving a 10% $\sim$ 15% QoE gain. In security tasks, incomplete training data can trigger hallucinations such as false threat detection. To address this, the authors in [11] proposed a mixture-of-experts GenAI model that decomposes objectives like secrecy rate and energy efficiency into expert sub-models, and used dynamic gating to select suitable experts, enhancing security adaptability and resource efficiency.

Case Study: Wireless Hallucination Mitigation Method for GDM-Based Channel Estimation

Design Overview

We present a case study on uplink channel estimation using GDMs and propose an integrated solution to mitigate hallucinations in estimation models. Uplink estimation is chosen for its critical role in precoding and resource allocation and its sensitivity to errors under low-SNR and NLoS conditions. Specifically, GDM follows the denoising diffusion probabilistic model (DDPM) architecture, with a U-Net-like convolutional backbone conditioned on diffusion step embeddings, tailored to reconstruct high-resolution angular-do-

The interface pipeline includes prompt generation from pilot metadata, LLM-driven expert selection, index resolution, and GDM reverse denoising.

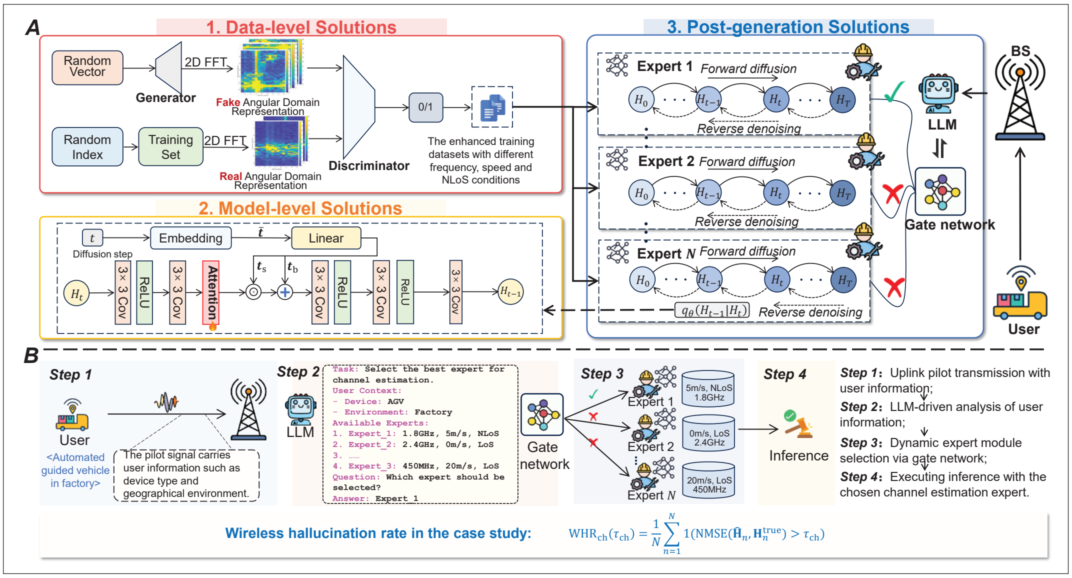

FIGURE 2. Part A illustrates the proposed integrated solution for reducing wireless hallucination in channel estimation. At the data level, a GAN enhances uneven training data. At the model level, an attention mechanism captures latent channel distribution features. At the post-generation level, an LLM-enhanced MoE selects models that align with user requirements for for channel estimation. Part B outlines the workflow: Upon receiving pilot signals, the LLM analyzes and infers user requirements, subsequently selecting the suitable expert for channel estimation.

main channel matrices from noisy observations [12]. And the GDM is trained using a denoising score-matching loss derived from the variational lower bound. This study provides a practical example of integrating data-level, model-level, and post-generation solutions into a unified framework, an approach that has not been comprehensively explored in previous research. The proposed solution and workflow of channel estimation are shown in Fig. 2. In the case study, the WHR is computed as the ratio between the number of instances where the deviation between the generated channel $\hat{\mathbf{H}}$ and the true angle-domain channel $\mathbf{H}^{\mathrm{true}}$ exceeds the threshold $\tau\mathrm{au}_{\mathrm{ch}}$, and the total number of test instances, with the expression shown in Fig. 2. Also, the NMSE can also reflect the degree of hallucination exhibited by the GenAI model. Note that we evaluate WHR with a hallucination threshold of 0.01 on 1000 test samples, using NMSE to capture both reconstruction accuracy and physical plausibility under varying SNR conditions.

NTEGRATED MITIGATION STRATEGY FOR WIRELESS HALLUCINATION

Data-level Solution: First, we construct a channel dataset *A* using QuaDriGa [13], incorporating various carrier frequencies, mobility speeds, and LoS/NLoS conditions, totaling 10,000 samples. Notably, NLoS channel data are harder to obtain, leading to fewer NLoS samples than LoS samples. To enable a fair comparison, we also create a mixed channel dataset *B*, which does not differentiate data types but maintains the same 10,000-sample size as a baseline. To address data imbalance, we employ GANs, where the generator and discriminator undergo adversarial training to synthesize realistic NLoS channel data. Specifically, we adopt a residual-block generator and a PatchGAN discriminator [14], and train the GAN with WGAN-GP loss to synthesize environment-aware angular-domain channel data. This data augmentation strategy enhances dataset diversity, improving the generalization ability of the channel estimation model across different scenarios. The released datasets can be found in https://github.com/wang104225/WirelessHallu.

Model-Level Solution: To further reduce hallucinations caused by the randomness in GenAI-generated outputs, we design an attention-enhanced diffusion model to improve channel estimation accuracy. Specifically, we incorporate a Transformer-style multi-head self-attention (MHSA) module between the first convolutional block and subsequent layers in the reverse diffusion U-Net to capture long-range dependencies across input channel features. The feature map is flattened and linearly projected into query, key, and value matrices, and attention scores are computed using scaled dot-product attention with multiple heads, following the structure of TransUNet. By integrating the attention mechanism, the model enhances robustness by capturing spatial and angular interdependencies, particularly in challenging wireless environments.

Post-generation Solution: We propose a novel LLM-enhanced Mixture of Experts architecture that dynamically selects the most suitable diffusion expert for channel estimation based on user context. Each expert is a pre-trained GDM model optimized for specific wireless conditions such as LoS or NLoS, operating frequency, and mobility speed. Inspired by the Switch Transformer design [15], our gating mechanism avoids exhaustive evaluation by selectively activating one expert at inference time. Instead of a learned softmax gate, we deploy a 7B LLaMA 3 model as a gating controller on the edge. User context such as device type and deployment environment, along with expert profiles, is encoded into structured prompts, as shown in Fig. 2b. The LLM outputs an expert label such as Expert_1, which is parsed and mapped to an index, then used to activate the corresponding GDM expert. The interface pipeline includes prompt generation from pilot metadata, LLM-driven expert selection, index resolution, and GDM reverse denoising. This modular and interpretable routing enables flexible expert switching without retraining.

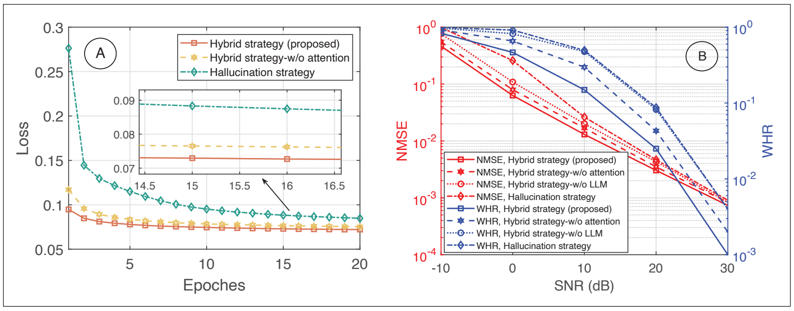

FIGURE 3. Experimental results: a) the training loss curves of the proposed integrated mitigation strategy and other baselines over epochs; b) NMSE and WHR performance of the proposed integrated mitigation strategy and other baselines over SNR.

Experimental Results

Experimental Configuration: Our experiment is implemented using Torch 2.5.1 and Python 3.12.8. The proposed model is built on the GDM framework and trained on an Ubuntu server equipped with a 2.10 GHz Intel Xeon Gold 5218R CPU (40 cores, 503 GB RAM) and an NVIDIA A100 80 GB GPU. The number of diffusion steps is set to 60. And Llama 3 LLMs with 7 billion parameters are deployed to support the gate network.

To systematically evaluate the contribution of each mitigation component, we adopt the following baselines: **Hallucination Strategy**, a diffusion-based channel estimation method without any hallucination mitigation; **Integrated Strategy-w/o Attention**, which employs GAN-based data augmentation training but excludes the attention mechanism; **Integrated Strategy-w/o LLM**, which uses random expert selection during inference.

Experimental Results: Figure 3a presents the training loss of the proposed integrated strategy for hallucination alongside baselines. Due to data augmentation and the attention mechanism, the proposed strategy converges to a lower loss value than the hallucination strategy, demonstrating its effectiveness in reducing hallucinations. Additionally, it outperforms the integrated strategy without attention, confirming that the attention mechanism in GDMs enhances the model's ability to capture fine-grained details. We further evaluate the NMSE and WHR of the proposed strategy across different SNR levels, as shown in Fig. 3b. Note that Fig. 3b reflects the level of wireless hallucination because WHR quantifies how often the generated channel estimates exceed the hallucination error threshold, while NMSE indicates the severity of their deviation from the physical ground truth. The integrated strategy achieves lower NMSE and WHR across all SNR levels, with the NMSE reduced from 0.29 to 0.10 and the WHR reduced from 0.21 to 0.015 at 0 dB SNR compared to the hallucination strategy. This improvement stems from the combined effect of the three mitigation components. Specifically, compared to the hallucination strategy, the hybrid strategy without attention reduces the NMSE from 0.29 to 0.08 and the WHR from 0.21 to 0.06 at 0 dB, indicating that data-level balancing improves generalization and alleviates hallucination caused by underrepresented scenarios. Additionally, the strategy without attention performs better than the hallucination baseline but worse than the full integrated strategy, indicating the attention layer helps suppress noisy features and enhances spatial-frequency modeling. Finally, compared to the hybrid strategy without LLM, the proposed hybrid strategy reduces the NMSE from 0.12 to 0.06 and the WHR from 0.045 to 0.015 at 0 dB, highlighting the role of context-aware inference of LLMs in preventing model-task mismatch and reducing hallucination. Notably, the proposed method demonstrates a greater hallucination reduction effect in low-SNR regions than in high-SNR regions. This is because noise has a stronger impact on model generation in low-SNR scenarios, making noise-guided generation more prone to unrealistic channel estimates.

Conclusion

In this article, we investigated wireless hallucinations in GenAI-enabled communications. We first defined wireless hallucinations and analyzed the GenAI workflow to identify their causes. Then, we explored mitigation strategies at the data, model, and post-generation levels. Additionally, we reviewed existing cases of wireless hallucinations and corresponding countermeasures. Finally, we proposed an integrated strategy for reducing hallucinations in GDM-based channel estimation. Specifically, we published a hallucination dataset and use GANs to enhance training data. Attention mechanisms and LLMs are incorporated into both training and inference, reducing NMSE-based hallucinations by 0.19.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant 62571057, in part by the Beijing Natural Science Foundation of China under Grant L254061, in part by the Fundamental Research Funds for the Beijing University of Posts and Telecommunications under Grant 2025TSQY03, in part by the BUPT Excellent Ph.D. Students Foundation under Grant CX20253001; this work is also supported in part by Seatrium New Energy Laboratory, Singapore Ministry of Education (MOE) Tier 1 (RT5/23 and RG24/24), the Nanyang Technological University (NTU) Centre for Computational Technologies in Finance (NTU-CCTF), and the Research Innovation and Enterprise (RIE) 2025 Industry Alignment Fund — Industry Collaboration Projects (IAF-ICP) (Award I2301E0026), administered by Agency for Science, Technology and Research (A*STAR).

References

[1] H. Du *et al.*, "User-Centric Interactive AI for Distributed Diffusion Model-Based AI-Generated Content," arXiv preprint arXiv:2311.11094, 2023.

[2] N. Chakraborty, M. Ornik, and K. Driggs-Campbell, "Hallucination Detection in Foundation Models for Decision-making: A Flexible Definition and Review of The State of The Art," *ACM Computing Surveys*, 2025.

[3] L. Huang *et al.*, "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," *ACM Trans.n Information Systems*, 2024.

[4] G. Chi *et al.*, "RFDiffusion: Radio Signal Generation via Time-Frequency Diffusion," *Proc. 30th Annual Int'l. Conf. Mobile Computing and Networking*, 2024, pp. 77–92.

[5] M.-H. Wu *et al.*, "Intelligent Reflecting Surface-Assisted Millimeter Wave Communications: Cross Attention-aided Variational Autoencoder-based Precoding Design," *IEEE Trans. Cognitive Communications and Networking*, 2025, pp. 1–1.

[6] L. Miuccio, D. Panno, and S. Riolo, "A Flexible Encoding/Decoding Procedure for 6G SCMA Wireless Networks via Adversarial Machine Learning Techniques," *IEEE Trans. Vehicular Technology*, vol. 72, no. 3, 2023, pp. 3288–3303.

[7] X. Wang *et al.*, "Generative AI Enabled Matching for 6G Multiple Access," arXiv preprint arXiv:2411.04137, 2024.

[8] S. Zhang *et al.*, "Generative AI on SpectrumNet: An Open Benchmark of Multiband 3D Radio Maps," *IEEE Trans. Cognitive Communications and Networking*, 2024, pp. 1–1.
[9] F. Zhao *et al.*, "Enhancing Reasoning Ability in Semantic Communication Through Generative AI-Assisted Knowledge Construction," *IEEE Commun. Letters*, vol. 28, no. 4, 2024, pp. 832–36.
[10] H. Zhang *et al.*, "The Emergence of Reproducibility and Consistency in Diffusion Models," *41st Int'l. Conf. Machine Learning*, 2024.
[11] C. Zhao *et al.*, "Enhancing Physical Layer Communication Security through Generative AI with Mixture of Experts," *IEEE Wireless Commun.*, 2025, pp. 1–9.
[12] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," Advances in Neural Information Processing Systems, vol. 33, 2020, pp. 6840–51.
[13] F. H. H. Institute, 2025, QuaDRiGa — Quasi Deterministic Radio Channel Model; https://quadriga-channel-model.de/
[14] P. Isola *et al.*, "Image-to-Image Translation With Conditional Adversarial Networks," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2017, pp. 1125–34.
[15] W. Fedus, B. Zoph, and N. Shazeer, "Switch Transformers: Scaling to Trillion Parameter Models With Simple and Efficient Sparsity," *J. Machine Learning Research*, vol. 23, no. 120, 2022, pp. 1–39.

Biographies

Xudong Wang (xdwang@bupt.edu.cn) is currently pursuing the Ph.D. degree at Beijing University of Posts and Telecommunications, China.

Jiacheng Wang (jiacheng.wang@ntu.edu.sg) is a research fellow with the College of Computing and Data Science at Nanyang Technological University, Singapore.

Lei Feng (fenglei@bupt.edu.cn) is currently an Associate Professor with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China.

Dusit Niyato (dniyato@ntu.edu.sg) is a professor at the College of Computing and Data Science, Nanyang Technological University, Singapore.

Ruichen Zhang (ruichen.zhang@ntu.edu.sg) is a research fellow with the College of Computing and Data Science at Nanyang Technological University, Singapore.

Jiawen Kang (kavinkang@gdut.edu.cn) is a professor with Guangdong University of Technology.

Zehui Xiong (zehui_xiong@sutd.edu.sg) is a professor with Queen's University Belfast.

Hongyang Du (duhy@eee.hku.hk) is an assistant professor with The University of Hong Kong.

Shiwen Mao (smao@ieee.org) is a professor and Earle C. Williams Eminent Scholar, and director of the Wireless Engineering Research and Education Center at Auburn University.