Adam *et al. J Wireless Com Network* (2026) 2026:12
https://doi.org/10.1186/s13638-025-02549-7

EURASIP Journal on Wireless
Communications and Networking

REVIEW **Open Access**

Generative AI for non-terrestrial networks: design, applications, and challenges

Abuzar B. M. Adam[1*], Eva Lagunas[1], Mostafa Samy[1], Almoatssimbillah Saifaldawla[1] and Symeon Chatzinotas[1,2]

*Correspondence:
abuzar.babikir@uni.lu

[1] Interdisciplinary Centre for Security, Reliability and Trust (SnT), University in Luxembourg, 29 Avenue, Luxembourg 1855, Luxembourg
[2] College of Electronics & Information, Kyung Hee University, Yongin-si 17104, Korea

Abstract

Non-terrestrial networks (NTNs) have been identified as a fundamental component of future sixth generation (6 G) mobile networks, providing extended coverage to underserved and isolated regions and enhancing connectivity to fast mobility users (e.g., airplanes, vessels). Artificial intelligence (AI) techniques are set to revolutionize the way we orchestrate and optimize 6 G wireless networks and are especially appealing to anticipate and adapt to the complex NTN environments. In this context, generative AI (GAI) technologies have demonstrated remarkable potential in data generation and decision-making optimization, offering innovative solutions for dynamic and demand-driven network operations. This article investigates the role of GAI in 6 G-NTNs, emphasizing its transformative impact on resource allocation, beamforming, and security. We begin with the motivation and recent advances in GAI, exploring its applications in NTNs, including channel modeling, channel state information (CSI) estimation, intelligent network deployment, semantic communications, image processing, and enhanced network security and privacy. Subsequently, we discuss the fundamental challenges of GAI in NTNs and we propose a hybrid framework combining generative adversarial networks (GANs) and generative diffusion models (GDMs) to enable secure and efficient resource allocation and beamforming. Simulation results validate the effectiveness of the proposed framework.

Keywords: Generative artificial intelligence (GAI), Non-terrestrial networks (NTNs)

1 Introduction

Non-terrestrial networks (NTNs) based on geostationary orbit (GSO) and non-geostationary orbit (NGSO) satellites are envisioned as a vital complement to terrestrial networks, offering the capability to deliver global and ubiquitous coverage for next-generation 6 G wireless networks [1]. This vision is driven by the ever increasing demand to provide reliable data connectivity in an uninterrupted and ubiquitous manner.

6 G-NTN architecture requires novel and efficient methods for performing network planning and orchestration, as NTNs differ significantly from conventional terrestrial networks (TNs) due to the added complexity of the altitude dimension, and the associated degrees of freedom in positioning, mobility management, and beamforming. For instance, NGSO satellites in low Earth orbits (LEO) and medium earth orbits (MEO)

need careful planning to manage the spectrum so as to reduce interference between the satellite beams and the TNs cells.

The third-generation partnership project (3GPP) Release 18, the first 5 G Advanced standard, introduces AI (artificial intelligence) integration into the new radio (NR) air interface to enhance network automation [2]. Key areas include AI-based network management, energy-efficient radio access network (RAN) intelligence, and AI-native air interfaces for tasks such as CSI, beam management, and positioning [3]. In this context, generative artificial intelligence (GAI) offers groundbreaking solutions by leveraging its ability to discern the underlying patterns, features, and structure of input data. Unlike the conventional AI, GAI can generate novel instances that were not present in the original training dataset [4]. By integrating these capabilities with next-generation technologies, GAI holds immense potential to enable advanced wireless intelligence in 6 G networks. Furthermore, many real-world problems lack closed-form solutions or too complex to be solved explicitly by traditional model-based approaches, especially in network architecture like NTNs. This has motivated the adoption of GAI as an effective tools to tackle such complex problems [4]. The transformative technology of GAI is increasingly being applied in NTNs [5, 6], opening new horizons for innovation and additional flexibility.

Unlike conventional AI methods, which depend heavily on labeled datasets and often fail to generalize under the rapid mobility, Doppler variation, and sparse feedback typical of NTNs, GAI introduces a fundamentally different paradigm. By learning the underlying data distribution rather than a single input–output mapping, GAI can synthesize realistic network states and generate missing samples. This data-efficient adaptability makes GAI particularly suited to NTN environments where continuous retraining is impractical and model-based optimization is computationally prohibitive. Furthermore, recent advances in diffusion and energy-based generators allow on-board or federated deployment on satellites and UAVs, providing a cost-effective alternative to centralized learning. Therefore, GAI is not merely an extension of existing AI; it overcomes the limitations of the conventional AI model, while it can act as an enabler for resilient, scalable, and cost-aware NTN intelligence.

Motivated by the above, this article presents the key opportunities and challenges arising from the synergy of the emerging NTNs and GAI, which is expected to trigger the next major evolution in wireless communications. Unlike the work in [7], this work introduces advanced models like score-based generative models (SGMs) and energy-based models (EBMs), alongside hybrid GAN-GDM frameworks, providing innovative solutions to NTN challenges. With tailored design techniques such as deep unfolding, state-space models (SSMs), and federated learning, the magazine delivers NTN-specific applications that enhance scalability, security, and real-time adaptability, offering a more specialized and impactful approach. Table 1 compares our work with existing works in the literature.

2 GAI models and its applications in NTNs

GAI models leverage neural networks and advanced learning approaches, including unsupervised and semi-supervised learning, to transform unlabeled data into foundation models that enable versatile AI systems. Therefore, GAI models have been used in

Table 1 Comparison between representative works and this work

References	Focus and scope	Key contribution
Al Homssi et al. [8]	AI techniques for next-generation massive satellite networks	A comprehensive survey detailing AI applications across LEO satellite networks (channel forecasting, spectrum sensing, ISL optimization, and network security) but not addressing generative AI or model-driven customization
Lin et al. [3]	AI-driven resource and access management in SAGIN	A review of AI methods for SAGIN access and connectivity; synthesizes challenges and trends for large-scale access and orchestration
This work	Opportunities and challenges of applying GAI on NTN	A review on the challenges of applying GAI in NTNs, introducing techniques such as SSMs and deep unfolding, and culminating with a concrete case study (secure joint beamforming/resource allocation via GAN–GDM) that grounds the discussion

optimizing TN and NTNs, and these models include variational autoencoders (VAEs), generative adversarial networks (GANs), transformer-based models (TBMs), and generative diffusion models (GDMs) [7]. In addition to the previously mentioned GAI models, we will also discuss SGMs and EBMs. Figure 1 highlights the architecture of the above models and the potential applications in NTNs.

The architectural diversity of GAI models provides a range of tools for enhancing and optimizing NTNs. VAEs use probabilistic encoders and decoders to map input data to a latent space and reconstruct it, enabling efficient representations. This makes VAEs particularly useful for signal routing, anomaly detection, and signal detection in NTNs [9]. GANs rely on adversarial training between a generator and a discriminator, where the generator creates synthetic data and the discriminator evaluates its authenticity. This architecture excels in generating realistic channel data, aiding NTN channel modeling and simulating radio propagation scenarios for satellite communications [8].

TBMs utilize attention mechanisms to process sequential data, enabling the detection of complex relationships and scalability for large datasets. Their layered architecture supports applications such as semantic communication, where they extract abstract semantic information at the transmitter and reconstruct it at the receiver [10], as well as

(See figure on next page.)

Fig. 1 Taxonomy of GAI models and their primary applications in NTN optimization. Each column summarizes the model architecture and its representative communication-layer applications. This figure illustrates the architectural diversity of various Generative AI (GAI) models and their associated applications in non-terrestrial networks (NTNs). The left portion presents six representative GAI architectures: Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Transformer-Based Models (TBMs), Generative Diffusion Models (GDMs), Score-Based Generative Models (SGMs), and Energy-Based Models (EBMs). For each architecture, the major components are highlighted, such as the encoder–decoder structure in VAEs and TBMs, the generator–discriminator pair in GANs, and the iterative denoising process in GDMs. SGMs are characterized by score function estimation and noise scheduling, while EBMs are depicted with their energy function and contrastive learning process. The right portion lists example NTN-oriented applications supported by these models, including channel modeling, interference management, secure communication, semantic communication, and resource allocation. Arrows and groupings illustrate how model capabilities map to use cases, emphasizing overlaps such as beamforming, signal detection, and anomaly detection. This mapping demonstrates the versatility of GAI architectures and their potential for cross-layer optimization in NTN environments

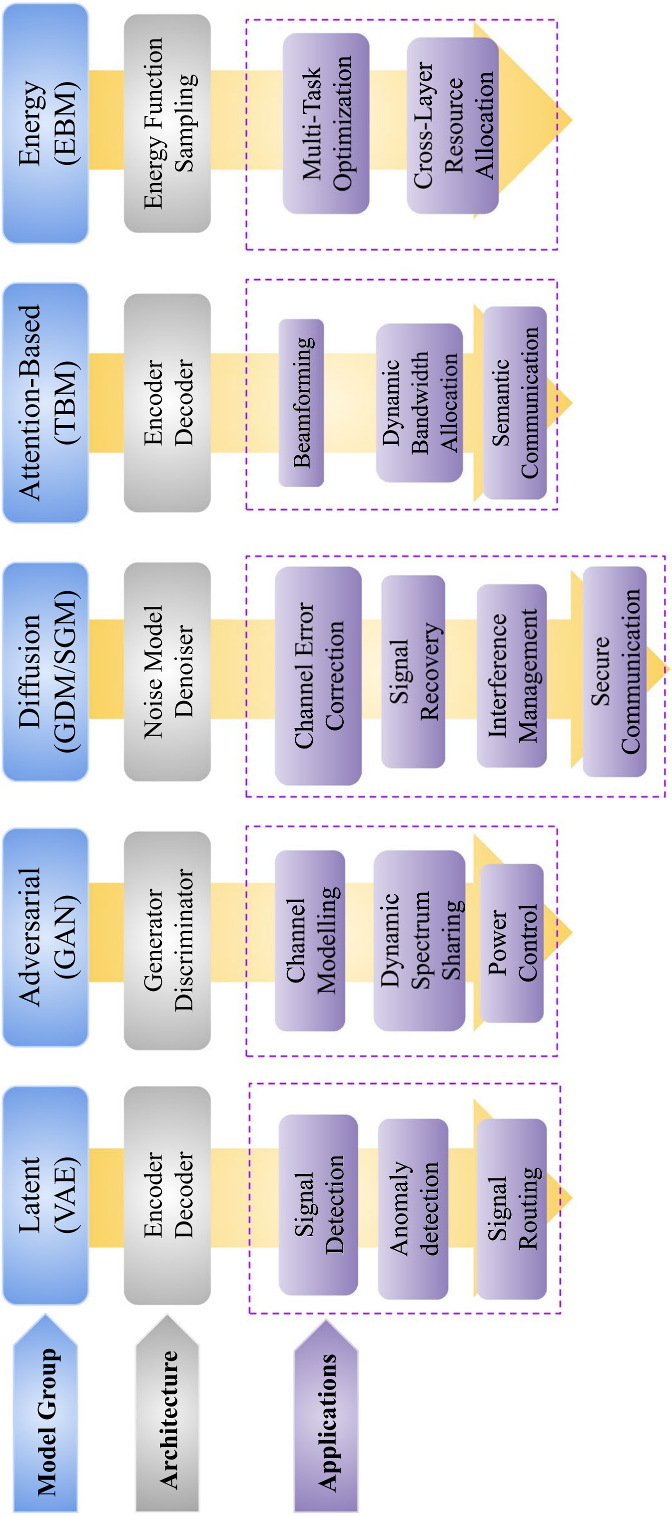

Fig. 1 (See legend on previous page.)

beamforming and interference management [11]. GDMs combine forward and reverse diffusion processes to iteratively add and remove noise, excelling in reconstructing noisy signals, which is vital for NTN error correction [5]. SGMs adopt score matching to estimate gradients of the data distribution, offering a streamlined alternative to diffusion models for synthetic data generation and resource allocation [12]. Finally, EBMs define an energy function over data configurations, leveraging iterative optimization techniques such as Markov Chain Monte Carlo (MCMC) to model complex NTN environments and optimize metrics like rate maximization and energy efficiency [13]. These architectures collectively provide versatile frameworks for addressing the multifaceted challenges of NTNs.

As highlighted in Fig. 1, GAI models possess a broad spectrum of applications within NTNs, with some models displaying considerable overlap in their use cases. However, deriving optimal solutions from these models presents numerous challenges. In the following subsections, we delve into GAI-driven solutions for NTN-specific issues and examine the challenges associated with training and deploying these models effectively.

2.1 Channel modeling and CSI prediction

Channel information in current 5 G systems is achieved by pilot-based estimation. CSI acquisition in NTNs, however, faces challenges such as long propagation delays, Doppler shifts, and rapid mobility. Conventional AI models, despite real-time capabilities, require large labeled datasets and struggle to generalize to dynamic NTN conditions. GAI overcomes this by generating realistic channel environments and adapting to unpredictable conditions with minimal data, beyond the scope of conventional AI. GANs simulate NTN channels, while VAEs compress and reconstruct channel data for rapid CSI updates [7]. TBMs and SGMs enhance accuracy by predicting fast-varying channel coefficients and capturing subtle LoS/NLoS variations, addressing limitations of traditional AI under sparse data. Moreover, GANs and GDMs demonstrated that can generate high-fidelity channel maps and adaptive resource allocations across heterogeneous NTN layers, motivating their adoption for channel estimation and modeling [14].

2.2 Signal detection and interference cancelation

Reliable communication in NTNs depends on effective signal detection and interference cancelation, especially in dynamic, high-mobility environments with shared spectrum. Traditional techniques like energy detectors (EDs) and interference cancelation methods, such as successive interference cancelation (SIC), and minimum mean square error (MMSE), are efficient but struggle with complex or rapidly evolving interference. Unlike traditional AI models and the above method, VAEs and GANs for instance, particularly GAN ensembles, detect anomalies by reconstructing signals, while GDMs iteratively denoise, achieving high accuracy under severe interference [5]. These methods significantly enhance NTN robustness beyond traditional approaches.

2.3 Secrecy and privacy enhancement

Non-terrestrial networks (NTNs) are vulnerable to long-distance interception and jamming [15, 16]. Traditional security models, such as encryption protocols (Advanced Encryption Standard (AES) and Rivest–Shamir–Adleman (RSA)) [17, 18], physical layer security (PLS) [19, 20], and spread spectrum techniques, offer protection but often lack adaptability to dynamic NTN threats [15, 16]. Combining VAEs and GANs to perform data compression and simulation of adversarial attack help in improving NTN security.

On the other hand, GDMs are more robust against adversarial attacks which make them suitable candidates to detect evolving threats, strengthening NTN resilience [21]. GAI models enhance NTN security by compressing data (VAEs) and simulating adversarial threats (GANs). Discriminators refine anomaly detection, improving beamforming and secure routing. GDMs iteratively detect evolving threats, strengthening NTN resilience [21].

2.4 Semantic communication

Satellite constellations generate large Earth observation (EO) data, posing challenges for downlink transmission. Traditional methods struggle in NTNs due to dynamic conditions and limited resources [22]. On the other hand, GAI models like VAEs and GANs prioritize and compress essential data. GDMs denoise information, preserving key features, while TBMs ensure critical data are prioritized for transmission [23]. These models optimize data flow for NTN applications, improving efficiency in remote sensing and disaster monitoring.

In practice, a VAE-based generative NTN encoder maps high-dimensional source data such as sensory and image features into a low-dimensional latent space that preserves semantic meaning rather than raw bit accuracy. Only the latent embedding is transmitted through the satellite and aerial links, significantly reducing bandwidth usage. At the receiver, a semantic decoder reconstructs the intended information by sampling from the learned latent distribution. Performance can be measured by task success rate or semantic similarity between transmitted and reconstructed content, rather than bit-level error rate. This approach aligns with the emerging task-oriented semantic-loss modeling paradigm introduced in [24], which quantifies semantic degradation arising from both source compression and channel imperfections in satellite-based Earth observation systems.

By extending this idea to generative NTN contexts, semantic communication can dynamically prioritize mission-critical meaning over packet precision, ensuring efficient and context-aware data transmission under stringent bandwidth and latency constraints.

2.5 Interference management and joint network optimization

Interference in NTNs stems from inter-satellite links, shared spectrum, and dynamic beam steering. SGMs adapt beamforming by modeling signal gradients, mitigating interference as satellites move. This enhances connectivity and allocates resources by predicting user demand and guiding satellite deployment [12].

EBMs define energy functions for beamforming and spectrum sharing, minimizing interference while maximizing throughput. By modeling satellite mobility and traffic patterns, EBMs ensure efficient joint optimization across NTNs [13].

GDM is trained to generate beamforming vectors or perform power allocation conditioned on given CSI. During inference, the diffusion process iteratively denoises random channel realizations into optimized decisions that maximize SINR while satisfying per-user rate and power constraints. Hence, explicit outputs can be produced, such as user-specific beam patterns and frequency-resource maps, that can be directly evaluated in terms of achievable sum rate, interference suppression ratio, and computational latency.

Hence, the GDM provides a data-driven means of realizing adaptive interference control in complex NTN topologies.

2.6 Routing and connectivity in NTNs

Routing through inter-satellite links in NTNs faces challenges from satellite mobility, unstable inter-satellite links, and latency constraints [25]. LEO and MEO satellites require frequent path updates to manage real-time traffic and IoT applications. Traditional routing struggles with NTN dynamics, while adaptive techniques like RL and DQNs often lack scalability. In this context, GAI enhances NTN routing by generating adaptive [26, 27]. GDMs and VAEs simulate network states to optimize paths, while TBMs capture cross-node dependencies, improving context-aware routing. SGMs refine paths by analyzing link stability and bandwidth, ensuring NTN scalability and efficiency, and addressing traffic diversity beyond traditional routing limits.

3 Challenges of designing and applying GAI models in NTNs and mitigation strategies

GAI models are used in NTN optimization due to their ability to model complex data distributions and adapt to highly dynamic environments. These models provide adaptive, data-driven solutions that account for the unique challenges of NTNs. However, this comes with different challenges. Next, we discuss these challenges and the possible mitigation strategies and practical insights for GAI deployment in NTNs.

3.1 Challenges of designing and applying GAI models in NTNs

3.1.1 Real-time responsiveness and synchronization challenges

Integrating GAI into NTNs demands efficient, low-latency processing to manage their dynamic and distributed nature. Delays can hinder adaptation to rapid environmental changes, affecting applications like beamforming and resource allocation. GDMs, with iterative denoising, and TBMs, with computationally heavy attention mechanisms, risk increasing latency. Synchronizing satellites, UAVs, and ground stations adds further complexity, requiring precise coordination to avoid performance degradation. Fast adaptation, minimal latency, and effective synchronization are essential for GAI deployment in NTNs.

3.1.2 Integration overhead and energy efficiency

The integration of GAI into NTNs introduces substantial computational and communication overhead, alongside significant energy demands. Advanced models like GDMs, TBMs, and VAEs require intensive computational resources, including prolonged GPU hours for training and high processing power for real-time inference. For instance, SGMs, which involve gradient-based refinements, can delay decision-making processes in resource-constrained NTN environments. These challenges are compounded by the limited power resources of satellites, making energy efficiency a critical consideration [28].

In practice, on-board payload processors on LEO and MEO satellites provide limited computational capacity (typically below a few hundred GFLOPS) and operate under power budgets of only 500–2000 W, which constrains the deployment of large generative models containing tens of millions of parameters [8, 29, 30]. Continuous inference for beamforming or channel prediction must also meet sub-10 ms latency targets and strict thermal budgets, creating a direct trade-off between model complexity, accuracy, and energy consumption [8, 31]. Consequently, future GAI–NTN implementations will require lightweight strategies, such as model pruning, quantization, or split learning between satellite and ground nodes, to maintain real-time operation within on-board resource budgets [24, 31]. While these generative models remain relatively heavy for full on-board deployment, we expect ongoing advances in hardware and algorithmic design to progressively bridge this gap. Recent studies have begun exploring the integration of large language and generative models into efficient spiking neural networks (SNNs) for neuromorphic computing platforms [32], which could eventually enable low-power, real-time inference for GAI-enabled NTNs.

3.1.3 Scalability and interoperability across network layers

The diverse architecture of NTNs, comprising satellites, UAVs, and terrestrial stations, presents significant scalability and interoperability challenges for GAI deployment. As the network expands, the computational demands for real-time data processing and decision-making grow. For instance, TBMs and EBMs, which rely on complex dependency modeling, require increasing resources as the network's complexity rises. Interoperability between layers with differing protocols and formats further complicates GAI integration. To maintain performance reliability, new mechanisms of design should be adopted to integrate different models for NTN optimization.

3.1.4 Security, privacy, and robustness risks

While GAI significantly enhances the adaptability and automation of NTNs, it simultaneously introduces new classes of vulnerabilities. Models such as VAEs, TBMs, and GDMs are susceptible to adversarial perturbations, where deliberately or inadvertently manipulated inputs can degrade inference quality or destabilize training. Moreover, training and deployment of GAI models often involve sensitive NTN data, such as user positions, traffic patterns, and beam configurations, which expose them to privacy and integrity risks, particularly when cloud or federated infrastructures are used for large-scale model synchronization. Comprehensive protection mechanisms are therefore essential to safeguard GAI-enabled NTN operations.

GANs, although effective for anomaly detection and synthetic data generation, remain vulnerable to attacks targeting either the generator or discriminator, which can compromise model reliability. Additionally, GAI models may exhibit hallucination effects, producing physically implausible data that disrupts NTN operation. Misleading outputs, such as erroneous interference maps, unrealistic channel realizations, and misestimated user positions, can impair network optimization and autonomous decision making, posing severe risks in mission-critical contexts such as defense, navigation, and disaster response systems.

In practical NTN deployments, such hallucinations or adversarial perturbations can manifest as beam misalignment, spurious interference estimation, or corrupted channel coefficients, ultimately degrading link reliability and QoS. For example, corrupted latent representations may induce inaccurate beam steering or faulty power allocation patterns, while out-of-distribution channel samples can misguide spectrum reuse or scheduling. To mitigate these risks, robustness must be embedded at both the learning and inference stages. Regularization-based stabilization such as spectral normalization or gradient penalties smooths the latent manifold and suppresses unstable training dynamics. Uncertainty quantification, implemented via Bayesian or ensemble-based GAI models, enables the system to assign confidence levels to generated samples and flag unreliable predictions. Furthermore, anomaly detection modules integrated into NTN controllers can identify and reject latent vectors inconsistent with observed propagation states. Together, these mechanisms improve resilience against hallucination-driven and adversarial failures. Finally, higher-level techniques, such as transfer learning, federated adaptation, and retrieval-augmented generation (RAG), discussed in Section 4, serve as complementary system-level strategies that further enhance trustworthiness and robustness in dynamic NTN environments, aligning with the principles of reliable and explainable AI outlined in ITU-T Y.3800 [33].

3.2 Mitigation strategies and practical insights for GAI deployment in NTNs

To make GAI frameworks more feasible for real NTN implementation, several mitigation and optimization strategies can be adopted:

- Enhancing Real-Time Responsiveness: Lightweight model compression techniques such as pruning, quantization, and knowledge distillation can substantially reduce inference latency. LEO satellites and UAV relays can implement asynchronous or event-driven inference to handle dynamic user demands without continuous global synchronization.
- Reducing Energy and Computation Overhead: Hierarchical learning and inference scheduling can distribute workloads intelligently. Heavy generative computations can be offloaded to MEO/GEO nodes or ground stations, while edge devices perform lightweight adaptation. Model sharing among neighboring satellites avoids redundant computation.
- Improving Synchronization and Robustness: Federated or semi-federated training can synchronize model updates less frequently while preserving global coherence. This reduces communication load and aligns with intermittent inter-satellite connectivity.
- Sustaining Security and Reliability: Adversarial and contrastive training during model development enhances resistance to perturbations and false data injection. Integrating energy-aware and uncertainty-penalizing loss terms during training helps maintain stability in both secure and degraded network states.
- Hardware and System-Level Support: The use of AI accelerators, neuromorphic chips, and reconfigurable hardware aboard satellites can enable on-board generative inference at low power, bridging the gap between software-level GAI design and real-time NTN deployment.

These strategies together form a practical roadmap for engineers to address the real-world latency, energy, and synchronization challenges that currently limit the large-scale deployment of GAI-enabled NTNs.

4 Design and customization of GAI models for NTNs

There are various approaches to designing GAI models for NTN optimization, each tailored to address specific challenges and enhance performance. Many studies emphasize conventional design methodologies while focusing on improving the accuracy of GAI models through advanced techniques like transfer learning, federated learning, and RAG, which effectively mitigates hallucination issues, such as the generation of unrealistic interference patterns or erroneous channel conditions that could compromise NTN performance and resource allocation. In the following subsections, we first review conventional design strategies for GAI models and discuss merging designs that integrate multiple GAI models. Finally, we present our proposed advanced model-driven customization techniques, which enable precise tailoring of GAI models for optimized performance in NTNs, as illustrated in Fig. 2.

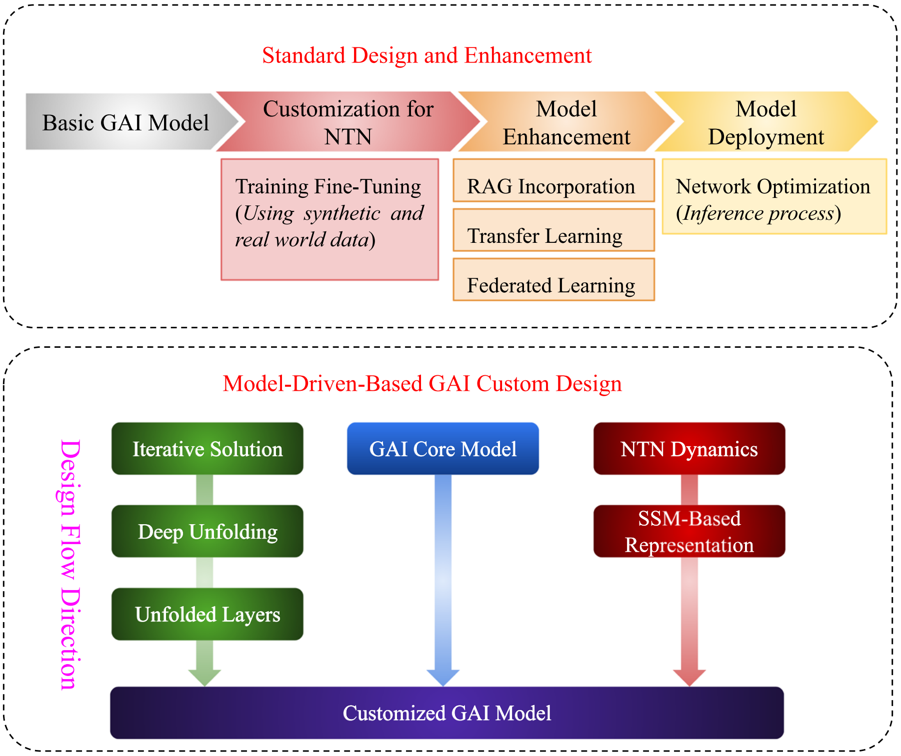

Fig. 2 Hierarchical flow of GAI model design and customization for NTN optimization. Each stage illustrates the progression from base model training to enhancement through RAG and federated learning and finally to model-driven deployment for network-specific optimization. This figure outlines the structured process for designing, enhancing, and customizing GAI models to meet the unique requirements of NTNs. The diagram is divided into two main pathways: the standard design/enhancement flow and the model-driven customization flow. In the standard design track, a basic GAI model undergoes training, fine-tuning, and enhancement through methods such as retrieval-augmented generation (RAG), transfer learning, and federated learning. This process culminates in model deployment for NTN tasks like resource allocation and interference management. The model-driven customization track incorporates advanced techniques such as deep unfolding, embedding iterative optimization algorithms into trainable layers, and State Space Models (SSMs) that encode NTN temporal dynamics. Both tracks feed into a "Customized GAI Model" optimized for network dynamics. The schematic highlights adaptability, scalability, and the integration of physical domain knowledge, emphasizing how these design flows converge to deliver robust, context-aware NTN solutions

4.1 Standard design and enhancement of GAI models for NTN optimization

The standard design and enhancement of GAI models for NTNs involve a structured process that transitions a basic GAI model into a robust, NTN-specific solution. This section outlines the core steps in detail.

4.1.1 Basic steps

The process begins with a basic GAI model, which serves as a generalized framework not yet adapted to the specific requirements of NTNs. This foundation model is then customized through training and fine-tuning to address NTN-specific challenges such as dynamic network topologies, interference, and resource allocation. Fine-tuning leverages domain-specific datasets, enhancing the model's capacity to manage the unique demands of NTNs. This ensures that the GAI model evolves from a generic tool into a domain-adapted solution, capable of addressing complex and heterogeneous NTN environments. The process begins with a basic GAI model, which serves as a generalized framework not yet adapted to the specific requirements of NTNs. This foundation model is then customized through training and fine-tuning to address NTN-specific challenges such as dynamic network topologies, interference, and resource allocation. Fine-tuning leverages domain-specific datasets, including real-world data from operational NTN systems and synthetic data generated by advanced generative emulators like GANs and diffusion models, to ensure robustness across diverse scenarios.

4.1.2 Enhancement via incorporation of RAG mechanism

Advanced enhancement is achieved by incorporating RAG, equipping the GAI model with the ability to access relevant external knowledge during inference [34]. This mechanism reduces hallucination and improves decision accuracy. In scenarios where NTN operators handle classified datasets, federated learning facilitates collaborative model training without exposing raw data, preserving privacy. Transfer learning further boosts adaptability by leveraging pre-trained models, cutting down retraining costs, and increasing efficiency. The combination of RAG, federated learning, and transfer learning allows for enhanced scalability, privacy, and adaptability across various NTN environments.

4.1.3 Augmentation via Generation and Sharing of Synthetic Data

The final step involves deploying the enhanced GAI model for core NTN tasks such as resource allocation, interference management, and traffic routing. Deployment ensures that the model's improvements are effectively integrated into NTN operations, translating into tangible performance enhancements. Real-time adaptive management is achieved by embedding the model seamlessly within NTN environments, ensuring dynamic and efficient network orchestration. This generic design framework is adaptable to various NTN tasks, where the input typically includes CSI, user positions, and interference patterns, and the output consists of optimized resource allocation strategies or beamforming vectors.

4.1.4 Model deployment

The final step involves deploying the enhanced GAI model for core NTN tasks such as resource allocation, interference management, and traffic routing. Deployment ensures that the model's improvements are effectively integrated into NTN operations, translating into tangible performance enhancements. Real-time adaptive management is achieved by embedding the model seamlessly within NTN environments, ensuring dynamic and efficient network orchestration.

4.2 Merging design of GAI models for NTN

The merging design of GAI models for NTN optimization involves combining the strengths of different generative models, such as GANs and GDMs, to address the unique challenges of NTN environments. This approach leverages the adversarial training capabilities of GANs, which excel at generating realistic data distributions and handling data augmentation tasks, alongside the iterative refinement and noise reduction mechanisms of DMs, which are highly effective for denoising and signal recovery in noisy and interference-prone conditions. For example, GANs can be used to simulate realistic interference patterns or rare network scenarios, while DMs can iteratively refine these outputs to ensure they align with physical NTN dynamics and constraints. By combining these models, the resulting hybrid GAI framework offers enhanced flexibility, robustness, and accuracy in tasks like beamforming, resource allocation, and interference management [35]. This merging design not only capitalizes on the complementary strengths of GANs and DMs but also enables adaptive and scalable solutions tailored to the dynamic and heterogeneous nature of NTNs.

4.3 Model driven-based customizing GAI models for NTN

The customization of GAI models for NTN optimization through model-driven techniques introduces tailored methodologies that align the model architecture with the underlying physical and dynamic characteristics of NTNs. Two primary mechanisms are employed to achieve this:

4.3.1 Deep unfolding for iterative solutions

This approach incorporates iterative solution techniques directly into the structure of the GAI model by transforming the iterative algorithms into a sequence of trainable, unfolded layers. Each layer corresponds to one iteration of the solution process, enabling the GAI model to learn and mimic the behavior of optimization or signal recovery algorithms. By embedding deep unfolding, the model benefits from interpretability and efficiency, bridging the gap between data-driven learning and model-driven algorithms. This design is particularly advantageous for solving problems like interference mitigation or beamforming in NTNs, where iterative algorithms are traditionally used.

In NTN resource optimization, iterative convex solvers such as successive convex approximation (SCA) and alternating optimization (AO) are widely used for beamforming and interference management. Deep unfolding reformulates each algorithmic iteration into a neural network layer with learnable parameters that control the update rules. For example, in a LEO-UAV coexistence network, an unfolded structure can infer near-optimal beamforming vectors from noisy CSI within only a few layers, achieving real-time adaptation while maintaining interference suppression accuracy. This transformation bridges classical optimization with data-driven inference, providing lower computational latency and improved robustness against dynamic interference and channel variations in NTNs.

4.3.2 State space models (SSMs) for NTN dynamics

SSMs provide a mathematical framework for describing systems that evolve over time, using a state evolution equation to model dynamic behaviors (e.g., channel conditions or interference) and an observation equation to link these states to measurable outputs (e.g., signal strength or delay). Their ability to capture temporal dependencies and stochastic dynamics makes them ideal for NTNs, which experience constant changes in satellite mobility, channel variability, and interference patterns. SSMs allow for the prediction of system behavior and a comprehensive representation of NTN dynamics, enabling accurate analysis and optimization. When integrated into GAI models, SSMs serve as structural components that embed NTN dynamics as input features or constraints, ensuring the outputs are consistent with real-world properties. This integration supports tasks like signal recovery, where SSMs provide real-time channel updates, and interference mitigation, where they model evolving interference patterns for optimizing beamforming or resource allocation. By embedding SSMs, GAI models combine data-driven adaptability with physical domain knowledge, making them context-aware and aligned with NTN dynamics.

In practical NTN deployments, SSMs enable predictive beam tracking and Doppler-aware control by learning the temporal evolution of satellite or UAV motion. The hidden state can represent slow-varying orbital geometry and atmospheric fading, while the observation model captures instantaneous CSI fluctuations. Such representation allows generative NTN agents to anticipate channel transitions and proactively adjust beam directions or power allocation, thus maintaining link stability and reducing signaling overhead. By embedding SSMs within GAI architectures, NTN controllers achieve real-time adaptability and higher throughput under high-mobility conditions.

As a model-driven deep learning approach, SSMs leverage mathematical representations, such as state and observation equations, to guide the learning process. This hybrid approach combines the interpretability and adherence to physical laws of traditional models with the adaptability of deep learning, enhancing the robustness and accuracy of GAI models for complex NTN environments. SSMs thus exemplify the synergy between classical modeling techniques and modern AI advancements and represent a perfect match for the semi-predictive orbital movements of NTNs.

5 Case study: GAN-GDM framework for secure resource allocation and beamforming in NTN

This section presents a use-case study where a GAN-GDM framework is proposed for secure resource allocation and beamforming in NTNs, combining the complementary strengths of GAN and GDM to address the challenges of dynamic network environments, interference, and security. The motivation behind this combination lies in GANs' ability to generate realistic and diverse scenarios for complex NTN conditions, such as interference patterns and varying user demands, while GDMs excel in iterative refinement and noise reduction, ensuring precision in optimizing resources and beamforming decisions. In this framework (steps of building the framework can be found in [36]), GAN are employed to simulate realistic NTN scenarios, including potential attack vectors or adversarial interference patterns, enabling the system to proactively account for security risks. GDM then processes these outputs, iteratively denoising and refining the resource allocation and beamforming strategies to ensure robustness and reliability under varying conditions. By integrating GAN' scenario-generation capabilities with GDM' fine-tuned optimization, the framework achieves secure, adaptive, and efficient resource allocation and beamforming, enhancing the overall performance and resilience of NTNs. The architecture of the proposed GAN-GDM is illustrated in Fig. 3.

Design rationale and novelty: The proposed hybrid framework is not a simple concatenation of two generative models but a task-driven fusion that aligns the adversarial learning capacity of GANs with the uncertainty modeling capability of diffusion. In our design, the forward diffusion process injects stochastic variations reflecting channel fading, Doppler, interference, and adversarial threats. The reverse diffusion step is amortized through a conditional GAN that directly learns to map noisy latent

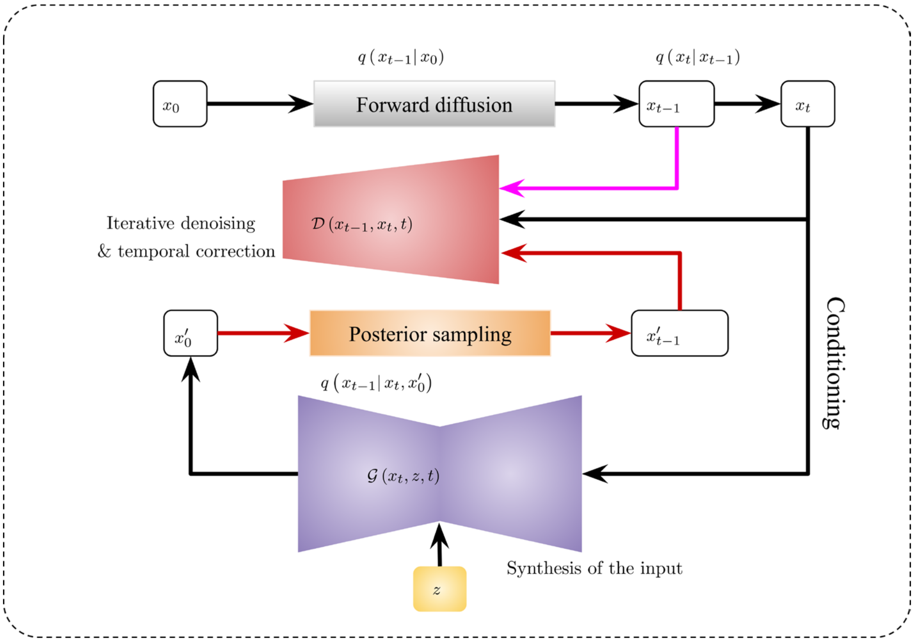

Fig. 3 GAN-GDM for secure resource allocation and beamforming in NTN. This figure presents the architecture of the proposed hybrid GAN–GDM framework tailored for secure and efficient resource allocation and beamforming in NTNs. The left side depicts the forward diffusion process, where Gaussian noise is added to model uncertainties such as channel variations, Doppler shifts, interference, and potential security threats. The right side shows the reverse diffusion process, implemented via a conditional GAN that iteratively denoises and refines the solution. Within the reverse path, the generator integrates NTN-specific inputs, such as imperfect CSI, satellite and user positions, and eavesdropper locations, while the discriminator enforces consistency with realistic NTN dynamics. The diagram also shows the interplay between multimodal denoising, conditional optimization, and adversarial training, resulting in secure carrier assignments, optimized beamforming vectors, and interference mitigation. This architecture leverages the scenario-generation strength of GANs and the refinement capabilities of GDMs to enhance NTN resilience against adversarial threats

states to secure and resource-efficient decisions under NTN conditioning variables (imperfect CSI, satellite/ESIM/eavesdropper geometry). This conditional adversarial training enforces physical realism and convergence stability, producing posterior-like optimized decisions that can operate in real time. This design explicitly couples diffusion-based uncertainty modeling with adversarially guided refinement—an ability not jointly achieved in standalone GAN or diffusion pipelines.

5.1 GAN-GDM architecture and adaptation for NTN dynamics

The proposed GAN-GDM framework optimizes LEO satellite-to-Earth station in motion (ESIM) communications by addressing uncertainties and mitigating security risks posed by static Earth station (ES) eavesdroppers. The main objective is to maximize the sum rate by enhancing beamforming and resource allocation. The system processes imperfect and outdated CSI as input, alongside parameters such as satellite and ESIM positions, eavesdropper locations, and interference patterns. The details of the architecture and the adaptation are given as follows:

- Forward diffusion process: Gaussian noise is injected into the input data to model uncertainties in NTN environments, including channel variations, Doppler shifts, and interference. This process also simulates potential security threats, ensuring the model learns to operate under diverse conditions.
- Reverse diffusion with conditional GAN: A conditional GAN models the reverse diffusion process, refining noisy latent variables into optimized solutions. The generator, guided by a discriminator, iteratively improves beamforming vectors and secure resource allocation by incorporating eavesdropper dynamics and NTN-specific constraints.
- Implicit multimodal denoising generator: The generator integrates NTN-specific inputs to iteratively refine outputs such as carrier assignment, beamforming for ESIM, and satellite uplinks. This ensures robust performance despite imperfect CSI and varying interference levels.

Contrast with single-Model baselines: Existing GAN-only models can generate realistic interference and attack scenarios but lack iterative denoising mechanisms and physics-informed optimization. Conversely, diffusion-only methods effectively remove noise but suffer from slow inference and lack adversarial scenario pressure. By merging the adversarial realism of GANs with the iterative refinement of diffusion, the proposed GAN-GDM achieves both fast convergence and physically consistent decisions, which are critical in NTN environments characterized by Doppler shifts, mobility, and evolving security threats.

The framework reduces the number of required denoising steps, enabling efficient sampling and real-time adaptation to NTN dynamics. By leveraging NTN parameters and adversarial training, the GAN-GDM model enhances security, minimizes signal leakage, and maximizes spectral efficiency across dynamic and hostile NTN environments.

Latency and security implications: The conditional adversarial reverse mapping amortizes the iterative diffusion process, which reduces the number of required denoising steps and consequently lowers inference latency.

During training, the model is exposed to a wide range of uncertain channel realizations and simulated eavesdropping directions, enabling it to learn security-aware allocation and beamforming strategies even without explicit eavesdropper information. Thus, the robustness of the secrecy arises from the architecture and training methodology of the model, where the GAN component enforces adversarial consistency and the diffusion process captures uncertainty, allowing the framework to suppress potential leakage and maintain reliability under unpredictable NTN conditions.

5.2 Secure resource allocation

For secure resource allocation, GAN-GDM leverages the forward diffusion process to simulate diverse NTN scenarios, including the presence of static ES eavesdroppers. The reverse diffusion, guided by the GAN, generates optimized allocation strategies that mitigate security risks while maximizing resource efficiency. For instance, the model ensures that resources are dynamically allocated based on real-time channel conditions and potential eavesdropping threats, effectively reducing the likelihood of intercepted communications. This dynamic approach allows for fair and secure distribution of resources across LEO satellites and ESIMs, ensuring consistent quality of service even in the presence of adversarial threats.

5.3 Secure beamforming at the ESIM

For secure beamforming, the ESIM is responsible for implementing optimized beamforming patterns to ensure secure and reliable communication with LEO satellites. The forward diffusion process introduces variability into channel conditions and potential eavesdropping attempts, while the reverse diffusion refines these inputs into robust beamforming solutions. The GAN component trains the model to account for static ES eavesdroppers, generating beamforming vectors that minimize signal leakage in the direction of potential eavesdroppers. By suppressing interference and directing maximum signal strength toward the intended satellite, the GAN-GDM framework ensures secure and high-quality signal reception at the ESIM, safeguarding NTN communications against security threats.

5.4 Joint optimization algorithm

This subsection summarizes the end-to-end operation of the proposed GAN–GDM framework, which performs joint beamforming and resource allocation under secrecy constraints. The workflow couples the forward diffusion process, conditional adversarial refinement, and multimodal denoising generator into a unified training and inference pipeline, as detailed below.

Algorithm 1 Joint Secure Beamforming and Resource Allocation via GAN–GDM

```
Require: Channel states $\mathbf{H}$, NTN context $c_t$ (satellite/ESIM geometry, Doppler, interference), and diffusion noise schedule $\{\beta_t\}_{t=1}^{T}$
Ensure: Optimized beamforming $\mathbf{w}^*$ and resource allocation $\mathbf{r}^*$
    Forward Diffusion: Add Gaussian noise via $q(x_t|x_{t-1}) = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}\epsilon$ to simulate uncertainty in CSI, interference, and propagation.
    Conditional GAN Training: The generator $G(x_t, z, \theta_G)$ outputs provisional $(\mathbf{w}, \mathbf{r})$ conditioned on NTN context $c_t$, while the discriminator $D$ enforces adversarial realism between synthetic and true solutions.
    Diffusion Refinement: The generator parameters are updated by minimizing $\mathcal{L}_{\text{total}} = \lambda_1\mathcal{L}_{\text{GAN}} + \lambda_2\mathcal{L}_{\text{diff}} + \lambda_3\mathcal{L}_{\text{sec}}$, where $\mathcal{L}_{\text{sec}}$ captures secrecy-rate regularization between legitimate and eavesdropping links.
    Inference: For unseen $\mathbf{H}$, amortized reverse diffusion is performed through $G$ using $T' < T$ denoising steps to yield optimized beamforming $\mathbf{w}^*$ and resource allocation $\mathbf{r}^*$ with reduced latency.
```

This algorithm jointly optimizes secure resource allocation and beamforming by merging diffusion-driven uncertainty modeling with adversarial learning. The diffusion process exposes the generator to diverse noisy channel states, while the adversarial discriminator stabilizes and constrains the learned mapping to realistic NTN dynamics. The resulting hybrid model achieves robust secrecy-aware decisions and efficient real-time inference.

5.5 Performance evaluation

Simulation Setup: Following the guidelines of 3GPP TR 38.811 [15] and 3GPP TR 38.821 [37] for NTN evaluation, we consider a LEO satellite operating at an altitude of 600 km and a Ka-band carrier frequency of 20 GHz. The satellite employs $N_b = 8$ spot beams, each covering a footprint of approximately 300 km in diameter, to serve $U = 10$ ESIMs randomly distributed within the coverage area. Each ESIM follows a random waypoint mobility model with velocities uniformly distributed between 70 and 120 km/h, representing vehicular user terminals (UTs) in accordance with the mobility profiles described in 3GPP TR 38.821 [37]. The satellite uses a uniform planar array (UPA) of 64×64 antennas, while each ESIM employs a UPA of 16×16 elements. Channel coefficients are generated using the composite LoS/NLoS model defined in 3GPP TR 38.811 [15], incorporating free-space path loss, shadowing, and Doppler effects consistent with the ITU-R P.618 [38] propagation model and atmospheric attenuation defined in ITU-R P.676 [39].

The GAN-GDM framework was trained using 8000 samples generated via exhaustive search optimization. The number of diffusion steps T was 100, batch size 64, and learning rate 2×10^{-4} with the Adam optimizer (β_1=0.5, β_2=0.999). Training lasted 1000 epochs. For comparison, standalone GAN and GDM models were trained using identical datasets and hyperparameters. All performance metrics, including secrecy rate and computational latency, were averaged over 20 Monte Carlo runs for statistical reliability.

The proposed GAN-GDM framework is trained using optimal data generated through exhaustive search. Similarly, standalone GAN and GDM frameworks are trained on the same optimal dataset for comparison. The test results, as shown in Fig. 4, demonstrate that the proposed GAN-GDM framework achieves performance closer to the optimal solution (exhaustive search-based solution) and surpasses all other frameworks such as the GAN, GDM, successive convex approximation (SCA) + Matching, and the Benchmark Scheme (which maximum ratio combining (MRC) with random resource

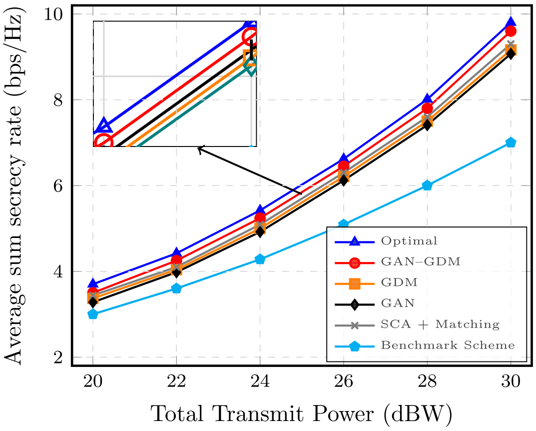

Fig. 4 Average sum secrecy rate versus total transmit power. This figure plots the performance comparison between the proposed GAN–GDM framework, standalone GAN, standalone GDM, the SCA+Matching approach, a benchmark scheme, and the optimal exhaustive search solution. The x-axis represents the total transmit power in dBW, while the y-axis shows the average sum secrecy rate in bits/s/Hz. The optimal curve serves as the upper performance bound. The GAN–GDM framework achieves results closest to the optimal solution across the transmit power range, significantly outperforming the other methods. The performance gain is especially notable in mid-to-high power regimes, where the GAN–GDM curve shows both high secrecy rates and robustness against degradation. Standalone GAN and GDM methods perform moderately well but lag behind the hybrid approach, while the benchmark scheme exhibits the lowest secrecy rates. This result validates the proposed framework's ability to balance security and spectral efficiency while being substantially faster (approximately 68×) than SCA+Matching in computation time

allocation). Additionally, it is noteworthy that, on average, the GAN-GDM framework is approximately 68 times faster than the SCA+Matching solution.

Across the tested transmit power range (Fig. 4), the hybrid GAN-GDM framework maintains secrecy rate performance close to the exhaustive search bound while achieving approximately 68× lower runtime than SCA+Matching. This demonstrates its latency advantage and its enhanced robustness against eavesdroppers, confirming the effectiveness of the task-coupled, conditionally amortized refinement for real-time NTN operation. Specifically, the proposed GAN-GDM achieves a median inference latency of about 9.86 ms per decision (95th percentile 11.2 ms) on an RTX A6000 GPU, whereas the SCA+Matching baseline requires a few seconds per iteration on the same hardware, depending on user load and channel dynamics. This confirms that the GAN-GDM framework satisfies the sub-10 ms responsiveness target for NTN beam-control loops and offers more improvement in reconfiguration speed over conventional optimization-based schemes.

5.6 Integration of the GAN–GDM framework into the 3GPP AI/ML architecture

Recent 3GPP Releases 18 and 19 have introduced standardized AI/ML integration for New Radio (NR) and NTNs through functional entities such as the Network Data Analytics Function (NWDAF), Model Management Service (MMS), and the AI/ML Model Repository [40–42]. Within this architecture, the proposed GAN–GDM framework can be deployed as an AI/ML inference service co-located with the NTN node such a LEO satellite, ESIM, and ES and orchestrated via the NWDAF or Operation, Administration, and Maintenance (OAM) entities.

Training of the GAN and GDM components is performed offline at terrestrial ground stations using synthetic datasets generated under 3GPP-compliant channel models. Validated models, along with metadata and performance indicators, are then uploaded to the MMS repository, enabling model exchange and version control consistent with the 3GPP AI/ML lifecycle.

During on-orbit operation; inference triggers, such as beam-quality degradation, Doppler-induced SINR drops, and eavesdropping detection events, are activated through NWDAF subscription interfaces, prompting the GAN-GDM module to re-optimize beamforming and resource allocation decisions in real time.

Performance reports and latent space reliability metrics are continuously returned via feedback loops defined in [40], allowing model health monitoring and incremental adaptation. This integration pathway ensures interoperability between terrestrial and NTN segments, supports trustworthy model management, and aligns the proposed GAI-enabled NTN functions with the broader 3GPP Release 18/19 vision for intelligent network automation. By situating the GAN-GDM framework within this standardized AI/ML pipeline, the proposed system bridges generative research models and industry-grade NTN implementations, enhancing both practicality and compliance.

6 Future Directions of GAI in NTNs

This section highlights some key areas for future development in the implementation and advancement of GAI-enabled NTNs scenarios.

6.1 On-board multi-beam resource management

Future GAI models should emphasize lightweight, on-board inference to efficiently manage multi-beam systems across satellites and aerial platforms. Instead of relying on static beam scheduling, GAI-enabled payloads can dynamically optimize power and bandwidth allocation based on user mobility, interference levels, and service demand. Generative diffusion and transformer-based models can predict spatio-temporal traffic variations and synthesize adaptive resource allocation maps that meet latency and energy constraints. Furthermore, integrating neuromorphic or reconfigurable AI accelerators aboard satellites will enable near-real-time inference within tight power envelopes, ensuring computational efficiency and rapid decision making without violating on-board energy budgets.

6.2 Real-time positioning and adaptive beamforming

A critical direction for future GAI research in NTNs lies in real-time positioning and adaptive beamforming. By leveraging Two-Line Element (TLE) data and Keplerian dynamics, diffusion-based and state-space generative models can learn satellite motion, Doppler evolution, and channel variation patterns, enabling predictive beam steering that anticipates rather than reacts to motion. In parallel, GAI-based sensor fusion can combine GNSS, inertial, and channel state data to enhance user localization accuracy even under partial satellite visibility or GNSS blockage. Transformer-driven GAI agents can then translate these spatio-temporal estimates into optimized beamforming vectors, allowing real-time beam tracking and interference mitigation under dynamic conditions. In the future, such GAI-embedded beam controllers may autonomously perform direction-of-arrival estimation, phase adaptation, and power allocation, achieving sub-10 ms reconfiguration latency and robust link maintenance under mobility and atmospheric variation.

6.3 NTNs handover strategies

The integration of GAI for NTN handover management offers a promising pathway toward seamless connectivity across multi-layer 3D networks involving LEO, MEO, GEO, and aerial nodes. Generative models can learn the joint statistical distribution of user mobility, beam coverage, and satellite trajectories to predict link degradation and initiate proactive handovers. By employing GANs or diffusion transformers, networks can synthesize potential future states and generate optimal transition decisions that minimize service disruption. Reinforcement-guided generative policies can further refine these decisions, optimizing for latency, outage probability, and stability during high-mobility scenarios. Additionally, federated GAI architectures will allow distributed satellites to share learned handover patterns without exposing raw data, improving scalability and data privacy. This predictive and probabilistic reasoning capability enables GAI to maintain continuity, energy efficiency, and service quality during frequent inter-satellite transitions.

6.4 LLMs and AI agents applications in NTNs

Large language models (LLMs) and autonomous AI agents will redefine NTN control and management by introducing semantic reasoning and explainable decision making. Domain-adapted LLMs trained on satellite telemetry, control logs, and mission data can interpret natural language network states and generate optimization actions such as beam retuning, interference mitigation, or power redistribution. When coupled with GAI backbones, multi-agent architectures can coordinate decisions among satellites, UAVs, and ground stations, enabling collaborative and self-healing operations. This synergy between LLM-based reasoning and generative adaptability will give rise to intelligent, self-optimizing NTN ecosystems that evolve autonomously, responding to dynamic mission objectives, environmental variations, and user demands in real time.

7 Conclusion

This article investigated the integration of GAI technologies into NTNs. We began by providing an overview of NTNs, encompassing their ground, aerial, and space-based network components. Subsequently, we introduced key GAI technologies, including VAEs, GANs, GDMs, TBMs, SGMs, and EBMs. We then discussed critical issues, solutions, and potential challenges in implementing GAI-enabled NTN systems. To improve the quality of service and enhance security in NTNs, we proposed a GAN–GDM framework for secure resource allocation and beamforming, leveraging the complementary strengths of GANs for realistic scenario generation and GDMs for iterative refinement. Simulation results validated the effectiveness of the proposed framework.

Beyond demonstrating the feasibility of GAI integration, this work also emphasizes the importance of developing trustworthy and explainable GAI models for NTN applications. By explicitly considering robustness, reliability, and data-protection aspects, the presented analysis aligns with the principles of trustworthy AI defined in ITU-T Y.3800. Future efforts should aim to translate these principles into concrete design guidelines, enabling certification-ready, energy-efficient, and interpretable GAI solutions for next-generation satellite and aerial networks. Such a direction will not only ensure performance and security but also promote interoperability across emerging 6 G standards and space–air–ground network architectures.

Abbreviations

3GPP	3Rd Generation Partnership Project
6 G	Sixth Generation
AES	Advanced Encryption Standard
AI	Artificial Intelligence
CSI	Channel State Information
EBMs	Energy-Based Models
EDs	Energy Detectors
EO	Earth Observation
ES	Earth Station
ESIM	Earth Station in Motion
GAI	Generative AI
GANs	Generative Adversarial Networks
GDMs	Generative Diffusion Models
GSO	Geostationary Orbit
GUs	Ground Users
LEO	Low Earth Orbits
LLMs	Large Language Models
MCMC	Markov Chain Monte Carlo
MEO	Medium Earth Orbits
ML	Machine Learning
MMSE	Minimum Mean Square Error
MRC	Maximum Ratio Combining
NGSO	Non-Geostationary Orbit
NR	New Radio
NTNs	Non-Terrestrial Networks
PLS	Physical Layer Security
RAN	Radio Access Network
RL	Reinforcement Learning
RSA	Rivest–Shamir–Adleman
SCA	Successive Convex Approximation
SIC	Successive Interference Cancelation
SGMs	Score-Based Generative Models
SNN	Spiking Neural Network
SSMs	State Space Models
TBMs	Transformer-Based Models
TLE	Two-Line Element
TNs	Terrestrial Networks
VAEs	Variational Autoencoders

Author contributions

A.B.M. Adam conceived the main idea, led the writing, and developed the GAN-GDM framework, performance evaluations and simulations. E. Lagunas contributed to the technical supervision. M. Samy assisted with the literature review and validation. A. Saifaldawla contributed to the article revision. S. Chatzinotas provided strategic guidance and overall project direction. All authors read and approved the final manuscript.

Funding

This work was supported by Luxembourg National Research Fund (FNR) under the Project SmartSpace under Grant C21/IS/16193290 and from the project "Physics-based wireless AI providing scalability and efficiency (PASSIONATE)". PASSIONATE [CHIST-ERA] has been funded by MICIU/AEI/10.13039/501100011033 and the European Union under Grant CHIST-ERA-22-WAI-04.

Data availability

The datasets generated and analyzed during the current study will be made available on the project's website.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication
All authors agree to publish the research in this journal.

Conflict of interest
The authors declare that they have no conflict of interest.

Received: 13 August 2025 Accepted: 24 November 2025

References

1. G. Geraci, D. López-Pérez, M. Benzaghta, S. Chatzinotas, Integrating Terrestrial and Non-Terrestrial Networks: 3D Opportunities and Challenges. IEEE Commun. Mag. **61**(4), 42–48 (2023). https://doi.org/10.1109/MCOM.002.2200366
2. Lin, X.: Artificial Intelligence in 3GPP 5G-Advanced: A Survey (2023). https://arxiv.org/abs/2305.05092
3. L. Zhi, F. Zimo, G. Kefeng, N. Ali, N. Dusit, W. Jiangzhou, AI-Driven Seamless and Massive Access in Space–Air–Ground Integrated Networks. IEEE Wireless Commun. **32**(3), 72–79 (2025). https://doi.org/10.1109/MWC.2025.3563021
4. A. Celik, A.M. Eltawil, At the Dawn of Generative AI Era: A Tutorial-cum-Survey on New Frontiers in 6G Wireless Intelligence. IEEE Open J. Commun. Soc. **5**, 2433–2489 (2024). https://doi.org/10.1109/OJCOMS.2024.3362271
5. Adam, A.B.M., Samy, M., García, C.E., Lagunas, E., Chatzinotas, S.: Diffusion Model-Based Signal Recovery in Coexisting Satellite and Terrestrial Networks. In: in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), pp. 1–5 (2024). https://doi.org/10.1109/WCNC57260.2024.10570763
6. Lin, Z., Guo, M., Han, C., Sun, Y., Ma, R., An, K., He, Y., Al-Dhahir, N.: Wireless endogenous security for SAGINs: achieving ubiquitous access and secure communication in symbiosis. IEEE Netw. **39**(6), 155–163 (2025). https://doi.org/10.1109/MNET.2025.3565745
7. Zhang, R., Du, H., Niyato, D., Kang, J., Xiong, Z., Jamalipour, A., Zhang, P., Kim, D.I.: Generative AI for Space-Air-Ground Integrated Networks. IEEE Wireless Commun., 1–11 (2024) https://doi.org/10.1109/MWC.016.2300547
8. B. Al Homssi, K. Dakic, K. Wang, T. Alpcan, B. Allen, R. Boyce, S. Kandeepan, A. Al-Hourani, W. Saad, Artificial Intelligence Techniques for Next-Generation Massive Satellite Networks. IEEE Commun. Mag. **62**(4), 66–72 (2024). https://doi.org/10.1109/MCOM.004.2300277
9. A. Saifaldawla, F. Ortiz, E. Lagunas, A.B.M. Adam, S. Chatzinotas, GenAI-Based Models for NGSO Satellites Interference Detection. IEEE Trans. Machine Learn. Commun. Netw. **2**, 904–924 (2024). https://doi.org/10.1109/TMLCN.2024.3418933
10. Meng, S., Wu, S., Zhang, J., Cheng, J., Zhou, H., Zhang, Q.: Semantics-Empowered Space-Air-Ground-Sea Integrated Network: New Paradigm, Frameworks, and Challenges. IEEE Commun. Surv. Tutor., 1–1 (2024) https://doi.org/10.1109/COMST.2024.3416309
11. Zhang, S., Zhang, S., Yuan, W., Quek, T.Q.S.: Transformer-empowered Predictive Beamforming for Rate-splitting Multiple Access in Non-Terrestrial Networks. IEEE Trans. Wireless Commun., 1–1 (2024) https://doi.org/10.1109/TWC.2024.3486673
12. Dockhorn, T., Vahdat, A., Kreis, K.: Score-Based Generative Modeling with Critically-Damped Langevin Diffusion. in Proc. Int. Conf. Learn. Represent. (ICLR) (2022)
13. Y. Du, I. Mordatch, *Implicit generation and modeling with energy-based models* (Curran Associates Inc., Red Hook, NY, USA, 2019)
14. R. Zhang, H. Du, D. Niyato, J. Kang, Z. Xiong, A. Jamalipour, P. Zhang, D.I. Kim, Generative ai for space–air–ground integrated networks. IEEE Wireless Commun. **31**(6), 10–20 (2024). https://doi.org/10.1109/MWC.016.2300547
15. 3GPP: Study on New Radio (NR) to Support Non-Terrestrial Networks (Release 15). Technical Report TR 38.811, V15.4.0, 3rd Generation Partnership Project (3GPP) (2020). https://www.3gpp.org/DynaReport/38811.htm
16. H. Guo, X. Lu, Y. Chen, Y. Wang, P. Popovski, Security and privacy in integrated satellite–terrestrial networks: A survey. IEEE Commun. Surv. Tutor. **25**(2), 1234–1262 (2023). https://doi.org/10.1109/COMST.2023.3242810
17. Daemen, J., Rijmen, V.: The Design of Rijndael: AES — The Advanced Encryption Standard. Springer, ??? (2002)
18. R.L. Rivest, A. Shamir, L. Adleman, A method for obtaining digital signatures and public-key cryptosystems. Commun. ACM **21**(2), 120–126 (1978)
19. S. Arum, B. Clerckx, P. Popovski, A survey of physical layer security in wireless communications for 5g and beyond. IEEE Commun. Surv. Tutor. **23**(1), 246–281 (2021). https://doi.org/10.1109/COMST.2020.3034893
20. Zhou, X., Song, L., Zhang, Y.: Physical Layer Security in Wireless Communications. CRC Press, ??? (2019)
21. Zhao, C., Du, H., Niyato, D., Kang, J., Xiong, Z., Kim, D.I., Shen, X., Letaief, K.B.: Generative AI for Secure Physical Layer Communications: A Survey. IEEE Trans. Cogn. Commun. Netw., 1–1 (2024) https://doi.org/10.1109/TCCN.2024.3438379
22. P. Jiang, C.-K. Wen, X. Li, S. Jin, G.Y. Li, Semantic satellite communications based on generative foundation model. IEEE J. Sel. Area Commun. **43**(7), 2431–2445 (2025). https://doi.org/10.1109/JSAC.2025.3559113
23. Grassucci, E., Park, J., Barbarossa, S., Kim, S.-L., Choi, J., Comminiello, D.: Generative AI Meets Semantic Communication: Evolution and Revolution of Communication Tasks (2024). https://arxiv.org/abs/2401.06803
24. Nguyen, T.T., Le, T.-D., Ha, V.N., Chou, H.-F., Eappen, G., Tran, D.-D., Nguyen-Kha, H., Thiruvasagam, P., Garces-Socarras, L.M., Gonzalez-Rios, J.L., Merlano-Duncan, J.C., Chatzinotas, S.: A semantic-loss function modeling framework with task-oriented machine learning perspectives. In: 2025 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN), pp. 1–6 (2025). https://doi.org/10.1109/ICMLCN64995.2025.11140499
25. Rao, Z., Zhu, Z., Niyato, D., Yao, Y., Xu, Y., Cheng, Y.: Dynamic leo satellite routing approach based on deep graph attention and incremental evolutionary reinforcement learning. IEEE Internet Things J., 1–1 (2025) https://doi.org/10.1109/JIOT.2025.3607492

26. Q. Guo, F. Tang, N. Kato, Routing for space-air-ground integrated network with gan-powered deep reinforcement learning. IEEE Trans. Cognitive Commun. Network **11**(2), 914–922 (2025). https://doi.org/10.1109/TCCN.2024.3522579
27. R. Zhang, H. Du, Y. Liu, D. Niyato, J. Kang, Z. Xiong, A. Jamalipour, D.I. Kim, Generative ai agents with large language model for satellite networks via a mixture of experts transmission. IEEE J. Sel. Area Commun. **42**(12), 3581–3595 (2024). https://doi.org/10.1109/JSAC.2024.3401159
28. Zhi, L., Hehao, N., Yuanzhi, H., Kang, A., Xudong, Z., Zheng, C., Pei, X.: Self-powered absorptive reconfigurable intelligent surfaces for securing satellite-terrestrial integrated networks. China Commun. **21**(9), 276–291 (2024). https://doi.org/10.23919/JCC.fa.2023-0437.202409
29. E.S. Agency, Artificial intelligence for space systems: Onboard processing and autonomy. ESA Bull. **176**, 1–9 (2023)
30. K. Scarbrough, J.L. Gonzalez-Rios, S. Chatzinotas, On-board artificial intelligence for non-terrestrial networks: Capabilities and power constraints. IEEE Open J. Comm. Soc. **3**, 2194–2206 (2022). https://doi.org/10.1109/OJCOMS.2022.3195823
31. W. Shi, S. Dustdar, The promise of edge ai for 6g wireless networks. IEEE Commun. Mag. **59**(11), 106–112 (2021). https://doi.org/10.1109/MCOM.001.2100156
32. Zhengzheng, T., Zhu, E.: BrainGPT: A Brain-Inspired SNN-Based Large Language Model (2025). https://openreview.net/forum?id=uXytIlC1iQ
33. ITU-T: Framework of Artificial Intelligence Systems and Trustworthiness. Technical Report Recommendation Y.3800, International Telecommunication Union, Telecommunication Standardization Sector (ITU-T) (2023). https://www.itu.int/rec/T-REC-Y.3800-202303-I
34. Lin, X., Kundu, L., Dick, C., Galdon, M.A.C., Vamaraju, J., Dutta, S., Raman, V.: A Primer on Generative AI for Telecom: From Theory to Practice. arXiv preprint arXiv:2408.09031 (2024)
35. Guo, K., Wu, M., Li, X., Lin, Z., Tsiftsis, T.A.: Joint trajectory and beamforming optimization for federated DRL-aided space-aerial-terrestrial relay networks with RIS and RSMA. IEEE Trans. Wireless Commun. **23**(12), 18456–18471 (2024). https://doi.org/10.1109/TWC.2024.3468298
36. Xiao, Z., Kreis, K., Vahdat, A.: Tackling the generative learning trilemma with denoising diffusion GANs. In: Int. Conf. Learning Represent (ICLR) (2022)
37. 3GPP: Solutions for NR to Support Non-Terrestrial Networks (Release 16). Technical Report TR 38.821, V16.2.0, 3rd Generation Partnership Project (3GPP) (2021). https://www.3gpp.org/DynaReport/38821.htm
38. ITU-R: Propagation Data and Prediction Methods Required for the Design of Earth-Space Telecommunication Systems. Technical Report Recommendation P.618-14, International Telecommunication Union, Radiocommunication Sector (ITU-R) (2019). https://www.itu.int/rec/R-REC-P.618-14-201908-I
39. ITU-R: Attenuation by Atmospheric Gases. Technical Report Recommendation P.676-13, International Telecommunication Union, Radiocommunication Sector (ITU-R) (2019). https://www.itu.int/rec/R-REC-P.676-13-201908-I
40. 3GPP: Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR and NG-RAN. Technical Report TR 23.700-91, 3rd Generation Partnership Project (3GPP) (2023). https://www.3gpp.org/DynaReport/23700-91.htm
41. 3GPP: Management and Orchestration; AI/ML Model Management Service. Technical Report TS 28.104, 3rd Generation Partnership Project (3GPP) (2023). https://www.3gpp.org/DynaReport/28104.htm
42. 3GPP: Management and Orchestration; AI/ML Model Management; Framework and Procedures. Technical Report TS 28.105, 3rd Generation Partnership Project (3GPP) (2023). https://www.3gpp.org/DynaReport/28105.htm

Publisher's Note