Generative Artificial Intelligence for Future Networking and Communications: Fundamentals, Applications, and Challenges

Zhiyuan Peng, Yuchen Liu *Member, IEEE*, Gaolei Li *Member, IEEE*, Zhaohui Yang *Member, IEEE*, Mingzhe Chen *Member, IEEE*, Dongkuan Xu, Xingqin Lin *Senior Member, IEEE*

***Abstract*—The evolution of generative artificial intelligence (GenAI) marks a pivotal shift in the potential reshaping of technological landscapes. Wireless networks, bolstered by the advent of advanced intelligent technologies, present a promising domain for leveraging GenAI, which could revolutionize the current networking design and communication paradigms. Prior research has extensively reviewed large language models, a significant component of GenAI, and explores their integration into networks. This paper offers a broad introduction to GenAI and delves into its applications within various emerging network technologies. We start by providing an overview of representative GenAI models, including variational autoencoder, Transformer, and diffusion models, elucidating their foundational principles. Subsequently, we spotlight their emergent applications to advanced networking technologies such as digital twins, integrated sensing and communication, and semantic communication. We also underscore crucial challenges and practical considerations, encompassing aspects like data quality, real-time processing capabilities, privacy risks, and security implications. Furthermore, a prospective outlook on research opportunities is taken to surmount these challenges, thereby unlocking the full potential of GenAI in future wireless networks.**

I. Introduction

Generative artificial intelligence (GenAI) has recently risen to prominence, spearheaded by revolutionary applications such as StarryAI and ChatGPT. These advances are propelling various fields towards a new round of intelligent transformation. Fueled by vast amounts of training data, deep learning algorithms, and high-performance computing with graphics processing units (GPUs), GenAI has accelerated the development of large foundation models. It now becomes realistic to synthesize high-fidelity content across various modalities, encompassing images, text, audio, and even video. Pioneering GenAI technologies including Transformer [1] and diffusion models [2] have unlocked a vast potential for innovative content generation. These models have significantly advanced content processing and understanding capabilities, enabling them to generate human-like content in a variety of applications. As the interest in GenAI continues to gain momentum, there is an increasingly urgent need for the seamless integration of GenAI technologies into the realm of networks and communications.

Several existing works [3]–[9] have introduced GenAI into wireless networks, with a heavy concentration on large language models (LLMs), a representative type of GenAI techniques. For instance, [4] offers insights into the current capabilities and limitations of LLMs, with case studies on the resolution of network anomalies, the comprehension of the 3rd Generation Partnership Project (3GPP) specifications, and the development of network models. More technically, [5] summarizes the main challenges of traditional deep offloading, and integrates LLMs into offloading at the mobile edge. Besides the implementation in networks, GenAI brings new challenges and opportunities to cyber security, where GenAI-aided attacks and their respective defense mechanisms are explored in [8], underscoring the need for ongoing research in mitigating potential security threats. In response to the comprehensibility of security approaches, [9] uses LLMs to explain cyber security alerts to non-experts, which has the potential to increase network security by proposing meaningful security measures in an intuitive language from alerts.

While LLMs have garnered immense interest in existing works, they represent merely a fraction of the extensive scope encompassed by GenAI. Importantly, LLMs are computationally intensive, requiring substantial processing power and memory while deploying them on network devices or servers that may necessitate significant hardware upgrades. In practice, GenAI's reach extends beyond text-oriented models, delving into the complex and multi-modal domains of network data. Relatively cost-efficient GenAI models such as variational auto-encoders [10] and diffusion models [2] hold substantial promise and merit continuing exploration, given their vast potential to shape the evolution of next-generation networks and communications. In parallel, we are witnessing concurrent and rapid advancements in emerging networking techniques, including: 1) network digital twins [11], 2) integrated sensing and communication [12], and 3) semantic communication [13], which hold the potential to revolutionize the landscape of next-generation networks, ushering in an era of increased capacity and adaptability. As such, the integration of GenAI with these innovative networking approaches presents a demanding yet promising frontier.

To this end, this paper offers a detailed overview of representative GenAI models, covering their fundamentals, emergent applications, and critical challenges in the aforementioned three emerging networking techniques. We explore the development of network digital twins (DTs) using diffusion models, emphasizing their content generation and data synchronization capabilities with physical networks. We also showcase the application of variational autoencoder and Transformer models in integrated sensing and communication (ISAC), focusing on signal denoising and three-dimensional sensing data transmission. Additionally, we highlight the role of GenAI models like Transformer in semantic communications for efficient semantic information extraction and data recovery. The paper concludes by outlining future research directions and addressing security concerns associated with the use of GenAI models in these networking techniques.

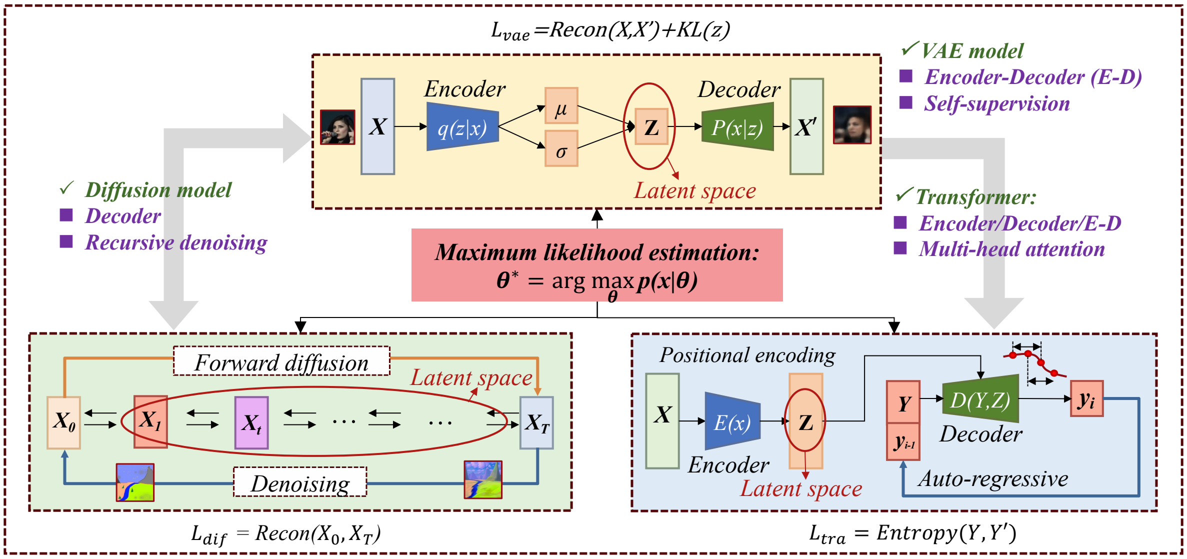

Fig. 1. Illustration of three representative GenAI models. VAE, featured for its simplicity and efficiency, is based on the principles of variational inference. It comprises an encoder that maps data to a latent space and a decoder that maps points from the latent space back to the data space. Diffusion models often referred to as denoising diffusion probabilistic models, are based on the idea of simulating a reverse diffusion process to generate high-fidelity data iteratively. Transformer, on the other hand, refers to an architecture for sequence-to-sequence modeling, primarily using a self-attention mechanism to weigh input elements based on content and position. It allows for parallel processing and has set state-of-the-art performance across various tasks.

II. Preliminaries and Overview

This section provides an overview of three key GenAI models including variational autoencoder, Transformer, and diffusion models for emerging networking and communication applications. The basic mathematics and key components of the GenAI models are described in Fig. 1. These models, while sharing a common objective of addressing the maximum likelihood parameter estimation problem, differ in their modeling and training methodologies.

A. Variational Autoencoder

1) Components: A standard variational autoencoder (VAE) model consists of four important components (see Fig. 1): 1) neural encoder-decoder pair, 2) latent space, 3) loss function, 4) sampling layer. The latent space represents the encoded input data, while the sampling layer generates new data points from this space, aiding VAE in learning a robust and generalizable representation of the input data.

2) Training Process: VAEs typically adopt the negative Evidence Lower Bound (ELBO) as the loss function. It can be divided into two loss terms, namely a reconstruction loss and a Kullback–Leibler (KL) divergence loss. The reconstruction loss measures the difference between the input data and the reconstructed output from the decoder. The KL divergence loss regularizes sample points generated from the latent space to follow a prior, normally standard Gaussian distribution. In training, reparameterization tricks can be leveraged to sample data points from the latent space.

3) Advantages: The versatility and effectiveness of VAEs stem from their ability to generate new data points, learn latent space representations, and incorporate stochastic elements. They serve as powerful models for a wide range of applications, particularly in scenarios where labeled data are scarce or expensive to obtain, such as anomaly detection, drug discovery, art generation, etc.

B. Transformer Model

1) Components: Transformer models have flexible architectures. They can be encoder-only, decoder-only, or encoder-decoder, consisting of critical components like the multi-head self-attention mechanism, positional encoding, residual connections, etc. These components work together, allowing the Transformer to capture the long-term dependencies between words in sentences and generate recursively human-like texts. Notably, ChatGPT, nowadays one of the most prevailing applications, is based on Transformer models.

2) Training Process: Transformer, through large-scale unsupervised pretraining on massive data from diverse sources, learns world knowledge and demonstrates remarkable zero-shot or few-shot learning abilities across a myriad of tasks. Performance can be further boosted by updating the full or partial model parameters on a few task-specific training data.

3) Advantages: Unlike recurrent neural networks (RNNs) that handle sequences in a step-by-step manner, Transformers process entire sequences simultaneously. This inherent parallelization facilitates faster and more efficient training, particularly on hardware like GPUs optimized for parallel computations. Building on the success of Transformer models for texts, e.g., large language models, researchers are now fervently advancing multi-modal Transformers in pursuit of human-equivalent perception abilities.

C. Diffusion-based Generative Model

1) Components: Diffusion models learn the latent structure of data by modeling a process in which data points diffuse through the latent space. They generally consist of forward and reverse processes. The forward process is defined as

Models	Pros & Cons	Existing applications	Typical tools	Future scenarios	Challenges	Risk examples
VAE	✓ *Precise control over latent representations* ✓ *Measurable objective* × *Generated samples are not sharp and blurry*	• *Audio synthesis* • *Image generation* • *Faulty detection*	• *Bing AI (intelligent chatbot and search engine system based on GPT-4)* • *GitHub Copilot (automatic programmer trained on GPT using GitHub code dataset)* • *Stable Diffusion (fast and high-fidelity text-to-image generation)* • *MusicGen (simple and controllable tool for music generation)* • *Runway GEN-2 (generating videos with text, images, or video clips)*	• *ISAC* ✓ *Sensing enhancement* ✓ *Edge decision-making* • *Digital Twin* ✓ *Efficient synchronization* ✓ *Content-based transmission* • *Semantic communication* ✓ *Enriching knowledge base* ✓ *Reducing distortion rate*	◆ *Adversarial attacks* ◆ *Data falsification*	• *ChatGPT produce responses that are seemingly convincing, but are actually **wrong**.* • *Samsung workers unwittingly **leaked secret data** whilst using ChatGPT to help them with tasks.* • *Developers observed in DALL-E 2 that the prompt "a flight attendant" mainly resulted in images portraying women of East Asian. **(Bias)***
Transformer	✓ *The **best** performance in content generation* × *The inference typically takes long time* × *The dimension of latent space is uncompressed*	• *Machine translation* • *Conversational AI* • *Object detection* • *Speech recognition*			◆ *Privacy leakage* ◆ *Prompt injection attack* ◆ *Data poisoning* ◆ *Copyright infringement*	
Diffusion model	✓ *High interpretability* ✓ *Versatile and widely applicable* × *Training is time-consuming and sampling slowly*	• *Text-to-image generator* • *Semantic segmentation* • *Image reconstruction* • *Material design* • *Time series forecasting*			◆ *Cultural discrimination* ◆ *Misleading information* ◆ *Intellectual property* ◆ *To answer off-topic*	

Fig. 2. Comparisons of GenAI models for future wireless networks. VAE models are designed for efficient encoding and synthesizing of data. They offer the advantages of low training cost and small parameter size, making them well-suited for achieving rapid responses in resource-constrained edge ISAC environments. The Transformer excels in extracting deep semantic information and may exhibit emergent effects as the parameter size increases. This characteristic naturally makes it suitable for semantic communications. In contrast, diffusion models are highly versatile and can be used to generate content for more complex scenarios, such as digital twins and metaverse-based mobile networks.

gradually adding Gaussian noise to the data point x_0 in T steps, producing a sequence of noisy data $x_1, \cdots, x_T$. When T goes to infinity, x_T is equivalent to an isotropic Gaussian noise. The second process reverses the forward process by training a neural network to recover x_t from x_{t+1}. Iterating the reverse process for T steps will eventually convert the noise x_T back to the original data x_0. In other words, new data points can be subsequently generated by first sampling from the prior Gaussian distribution, followed by iterating the reverse process.

2) Training Process: Similar to the VAE model, by leveraging the variational Bayes principle, the training objective of diffusion models can be formulated as minimizing the variational lower bound of $x_1, \cdots, x_T$, which eventually turns out to train a neural network to predict the added noise in x_t given the step of t.

3) Advantages: Different from the standard VAEs, a diffusion model can be conceptualized as a *hierarchical* VAE with a fixed encoder. Specifically, the forward process functions as the encoder with no trainable parameters. The reverse process, on the other hand, corresponds to the decoder, which is shared across the T iterative decoding steps. Thus, to some extent, the diffusion model is relatively simpler compared to VAEs and generative adversarial networks as it does not require the training of encoders or additional discriminators. Additionally, the iterative decoding principle contributes to the phenomenal success of diffusion models in generating high-fidelity samples compared to other generative models. Diffusion models have been deployed to address a variety of challenging cross-modal content generation tasks, such as text-to-image, text-to-video, and point cloud generation.

III. GenAI for Network Digital Twins

DT technology is pivotal in bridging the gap between physical and digital realms by creating virtual replicas of physical entities. These replicas, constructed using historical data and real-time mapping, enable meticulous state monitoring, real-time interaction, and predictive validation. Assembled in a network, digital twins optimize adjustments to physical network systems, making them a crucial enabling technology for the forthcoming sixth-generation (6G) era. However, the creation and operation of network DTs require large amounts of online network status data. Generative AI, such as diffusion models, can streamline this process by generating high-fidelity radio maps, aiding replication, simulation, and optimization of physical networks. Additionally, advanced image-to-text models enhance data management and transmission by extracting essential semantic information, reducing communication and storage costs while enhancing privacy and security.

A. Diffusion-based Distributed Twinning

Diffusion models, as illustrated in Sec. II-C, excel in generating high-quality images and audio, characterized by expressive content and meticulous details. This stands in contrast to traditional AI models such as VAE and GAN, which often fall short in terms of content expression and detailed accuracy. The application of diffusion models can also be extended to the creation and deployment of network DTs. For instance, as a high-fidelity data generator, diffusion models can be extremely beneficial to simulate the spatial link quality in a complex network deployment for *efficient twinning* between a digital twin and its physical counterpart. In a physical network, distributed sensors and devices only need to send a short sequence with *compact* data information to the digital space for synchronization. Such a sequence can be converted into a bit stream and transmitted to the DT computing center, on which the original high-dimensional data will be then recovered by using the diffusion models. Efficient data synchronization in turn enables the real-time update and operation of network DTs to manage the state of their physical counterparts. Diffusion models can effectively serve as a data generator that iteratively transforms the actual network scenario into a radio map twin as illustrated in Fig. 3,

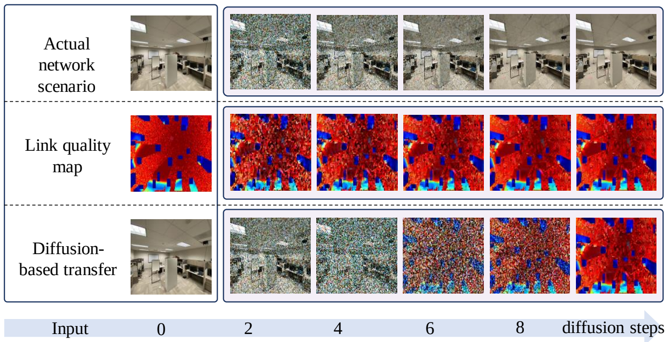

Fig. 3. Sensing and reconstruction using the stable diffusion model.

thereby achieving end-to-end twinning, i.e., from an actual scenario image to a complete radio map through the diffusion-based transfer. Besides, both data generation overhead and the amount of transmitted information can be significantly reduced owing to the high reconstruction accuracy guaranteed by the diffusion models, striking a good balance between twinning accuracy and efficiency.

B. Content-based Transmission for DT Evolution

The update of DTs in runtime demands a large amount of history status data from physical devices and sensors in dynamic environments. These data are collected from end-user devices and transmitted through wireless networks. The high-volume content of data, with large redundancy, brings huge costs for real-time synchronization and exposes the risk of privacy leakage. As a remedy, image-to-text technology driven by GenAI provides a possible solution that can extract essential information from high-dimensional data like images of surroundings, which substantially lowers communication delays, and protects privacy-sensitive details like human faces and environment layouts in the images. To be specific, consider a network planning case in an industry IoT scenario. The optimal positioning of IoT devices can be decided by network DTs given the corresponding images that detail the scenario type, device positions, distribution of objects, and electrical wires. Inevitably, the images contain critical privacy data that may be leveraged by malicious users for illegal access. It is thus undoubtedly forbidden to upload these data to the cloud. A possible workaround is to extract the critical information from images and generate descriptive texts accordingly, which are then uploaded to network DTs for optimization. The texts, treated as the input prompts for the diffusion models, can be converted into images that sketch the prominent features of the network environment for DTs to make decisions for real deployments. Such decisions can be descriptive natural language that explains the reason behind decisions and are understandable by humans, which has the potential of attaining explainable AI-augmented DTs with user privacy protected.

C. Challenges and Opportunities

1) *Quality of Generative Twinning Data*: Despite the promising benefits offered by diffusion models for DTs as discussed in Sec. III-A, the quality of generated twinning data from diffusion models highly relies on the fidelity of the input. In an ideal scenario, minor variations in the input data, such as noise from sensor circuits during data collection or distortion during data transmission, should minimally impact the diffusion models' data generation. However, GenAI models sometimes exhibit sensitivity to such variations, leading to the generation of twinning data that lacks critical features or details in the recovery phase. This sensitivity undermines the reliability of diffusion models for constructing scenes with multiple essential objects and network characteristics in precise spatial arrangements. As a result, there can be a noticeable decline in the quality of service, manifesting as lower-resolution images or videos, distorted signals, or in severe cases, complete loss of data segments. One possible approach to address this challenge is by employing deep Bayesian Gaussian processes to quantify the uncertainty associated with twinning data generation. By incorporating Bayesian principles, these DT models can provide estimates not only of the data itself but also of the uncertainty associated with those predictions on communication attributes.

2) *Accuracy-Latency Trade-off*: Integrating GenAI models into a DT-based network adds system complexity, bringing increased synchronization latency, primarily due to their iterative sampling process. The latency becomes even more pronounced when deploying large GenAI models for high-dimensional data generation, e.g., diffusion models for super-resolution imaging. A general approach to alleviate this latency is to decrease the number of iteration steps. This method essentially balances lower latency with a compromise on accuracy, as exemplified by the adoption of the consistency model [14] . Alternatively, in a wireless network, a GenAI model can be functionally split into several components and distributed to multiple DTs. This involves the coordination among DTs, where distributed DTs can be trained at isolated locations in parallel, with each capturing partial information. Then, a centralized twin aggregates the subordinate DTs' partial output and produces fast global data generation. This demands an efficient twin-to-twin communication protocol to enable multi-DT interactions.

3) *Inadvertent Privacy Leakage during Twinning Process*: Unlike smaller classification models, GenAI models typically feature an extensive number of parameters. The interaction and feedback effect among these parameters can sometimes result in unpredictable behaviors and functionalities. This complexity makes them more susceptible to being exploited by attackers, leading to potential privacy breaches through data poisoning, backdoor attacks, jailbreaking, etc. Without fine-grained authentication and authorization to the GenAI model, private data is likely to be inadvertently disclosed to malicious users during the query and response process. In response, a series of secure countermeasures should be explored, including twinning data anonymization, physical-digital model alignment, and adversarial training, among others.

IV. GenAI for Integrated Sensing and Communication

Integrated sensing and communication (ISAC) is envisioned to be an emerging technique in 6G wireless communications. It combines sensing and communication capabilities within a single framework to enable seamless and efficient integration of data collection (sensing) and data transmission

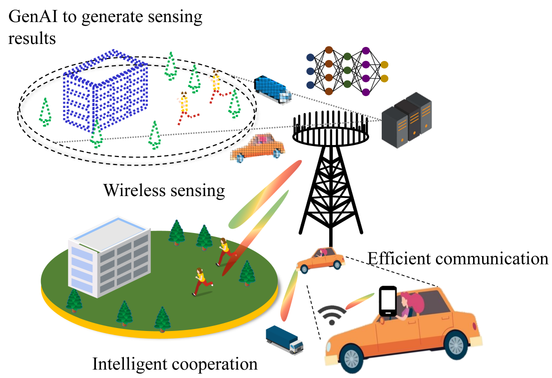

Fig. 4. An example of GenAI-assisted ISAC.

(communication) functions, often in real-time or near-real-time scenarios. This integration allows for enhanced monitoring, control, and decision-making in various applications. To fully leverage ISAC capabilities within dynamic wireless channels, it is crucial to extract valuable information from received wireless signals amidst surrounding noise and interference. In this context, a Transformer-based GenAI framework can be a promising solution. For example, to facilitate immersive communication services, it is typically necessary to transmit 3D sensing results to the receiver's end to enhance the user's quality of experience. Considering the heavy data communication overhead, one approach involves optimizing the transmission of wireless network traffic, where the transmitter sends concise description information, and the receiver employs GenAI techniques to reconstruct detailed results.

A. Sensing Signal Denoising Based on VAE

Wireless signal transmission conveys crucial environmental information, encompassing aspects like location, angle of arrival, and speed. However, the precision of these sensing outcomes is substantially impacted by factors such as non-line-of-sight (NLoS) channel conditions and the properties of multi-path effects. To optimize sensing accuracy, the study in [15] formulates the wireless sensing task as maximizing ELBO and solves it through a VAE model. The decoder of VAE estimates users' location and mobility based on the far-field power beam pattern of reflected echoes from communication users. Training of such a VAE model involves a KL regularization term that enhances the robustness of its estimation *against* environmental noises. By adopting this method, it becomes feasible to deduce the number of grids needed to ensure the establishment of the desired beamforming protocol, ultimately catering to the intended users via an integrated system of sensing, localization, and communication.

B. Sensing Result Rendering Based on Transformer

Given the constraints of limited wireless resources, transmitting high-dimensional sensing results to the receiver poses a significant challenge. To strike a balance between ensuring a high-quality user experience and meeting the necessary transmission requirements, an efficient approach is to leverage the power of GenAI techniques for rendering sensing results, as illustrated in Fig. 4. With the aid of GenAI techniques, such as Transformer-based neural radiance fields (NeRF), it becomes possible to generate high-resolution sensing results based on a low-resolution 3D model and fundamental descriptions. This requires a set of 2D or low-resolution 3D images of a scene or object captured from different viewpoints through the initial sensing process. Then, the 3D scene is modeled as a continuous volumetric function that describes the scene's appearance and geometry. By leveraging Transformer-based neural networks with volume rendering capabilities, this volumetric function can be learned to associate 3D positions in the scene with color and opacity values. Once the model is trained, NeRF can synthesize novel views of the scene from any desired viewpoint. This innovative application of GenAI not only optimizes data transmission but also enhances the overall fidelity and detail of the rendered sensory information, ultimately enriching the user experience.

C. Challenges and Opportunities

1) *Energy efficiency*: Integrating ISAC with GenAI models can be challenging due to the limited computation resources of ISAC and the heavy computation resources required by GenAI models. One potential solution to optimize energy efficiency in ISAC systems is to design energy-efficient waveforms and lightweight on-device GenAI models. These innovations can significantly reduce the computational and energy requirements, making them practical for real-world implementation.

2) *Robustness and Trustworthiness*: Wireless networks are dynamically changing and may experience unexpected scenarios, resulting in errors and uncertainty of sensing results. To meet the quality of user experience, GenAI models should be robust against diverse environmental uncertainty for reliable deployment. Additionally, GenAI models may produce inaccurate rendering results when confronted with insufficient sensing information. For instance, if an attacker manipulates sensors or their raw data, they can trigger GenAI models to generate fabricated information, leading the ISAC system to make incorrect decisions in sensing tasks. One viable solution is to deploy redundant sensors with overlapping sensing capabilities. This redundancy would allow the ISAC system to cross-verify data generated by multiple sensors, enhancing its ability to detect anomalies and ensuring more reliable results even if several sensors are compromised.

V. GenAI for Semantic Communications

Semantic communication allows a transmitter to derive the core meaning of large data and convey the condensed meaning, termed "semantic information", to a receiver. By doing so, it significantly reduces the volume of data transmitted over wireless channels. However, implementing semantic communication for wireless data transmission presents several challenges, including proper knowledge modeling for the transmitter or receiver, accurate semantic information extraction, restricted channel capacity, and security concerns. By addressing these challenges, the potential of semantic communication in wireless systems can be fully realized. GenAI models such as the Transformer will be a promising solution to address these challenges. They can efficiently process large-scale multimodal datasets, including text, audio, and images.

A. Semantic Information Extraction from Text Data

Extracting semantic information from text follows a structured process. Initially, texts are tokenized into word or subword sequences. These tokens are then embedded into vector representations. Crucially, not just the word, but its position in the sequence is vital. The Transformer model addresses this by generating a vector for the word's absolute or relative position. Both word and position vectors have the same dimensionality, and their point-wise addition produces a matrix that serves as the Transformer's primary input. The self-attention mechanism, a hallmark of the Transformer, constructs a correlation matrix emphasizing word interrelationships, spotlighting contextual nuances. The final step involves multiplying the text's matrix representation with this correlation matrix, yielding a vector that encapsulates the text's semantic essence. In essence, Transformers achieve a nuanced understanding of texts by not only focusing on the words themselves but also the intricate web of relationships within the textual context.

B. Semantic Information Extraction from Image Data

To extract semantic information from image data, the first step is to split each image into fixed-size patches. Image patches are treated the same way as words in texts. The vector representation of each patch is the point-wise addition of its linear transformation and its position vector in the image. The input of the Transformer is image matrix representations composed of the vector representations of the patches. Similar to text data, Transformer calculates the correlations between these patches. Then, the vector representation of image semantic information can be obtained by multiplying the image matrix representation with the correlation matrix.

Beyond semantic information extraction, Transformer can also be used for semantic information transmission and original data recovery. For example, one can collaboratively train two Transformers and deploy one Transformer at the transmitter and one Transformer at the receiver. To this end, the receiver can use the pre-trained Transformer and the transmitted semantic information to recover the original text or image data.

C. Challenges and Opportunities

1) Co-Design of Wireless Coding and Transformers for GenAI: As GenAI models transmit semantic information over wireless links, they intersect with signal processing techniques like source coding, channel coding, and quantization. These processes can either introduce or eliminate data redundancies, posing a key question: *How can GenAI seamlessly integrate with signal processing for optimized semantic data transmission?* A critical initial step is crafting a utility function to encapsulate the combined effects of source and channel coding with GenAI on information extraction and relay. Information bottleneck offers potential here, given its capacity to gauge mutual information across a GenAI model's hidden layers and its inputs or outputs. Leveraging this utility function, we can utilize conventional optimization strategies or delve into reinforcement learning to synergistically refine both wireless coding and GenAI parameters.

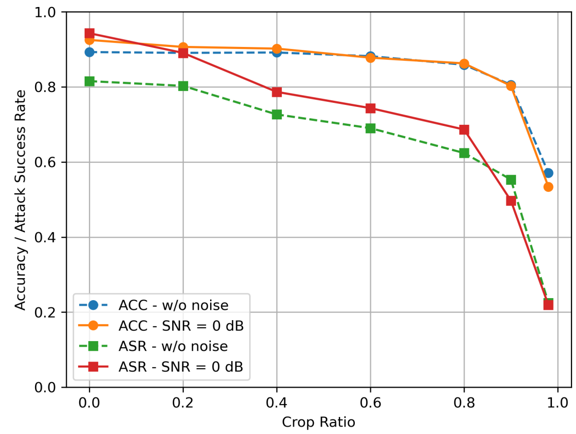

Fig. 5. Comparisons of model accuracy (ACC) and attack success rate (ASR) under different SNRs.

2) Semantic Eavesdropping: Emerging GenAI-based semantic communication is confronted with a potent security vulnerability termed "semantic eavesdropping". This arises because decoder-based receivers, akin to ever-listening microphones, could allow malevolent users to intercept transmitted semantic information if they gain access permissions to the receiver's decoder. In this scenario, an attacker would physically capture the transmitted semantic data, reconstruct the semantic coding for a data category, and subsequently eavesdrop on the entire category instead of isolated data points. This differs from conventional channel eavesdropping observed in electronic warfare as its efficacy is unrelated to signal strength, potentially rendering traditional covert communication defenses obsolete. Future research might pivot to developing strategies for semantic privacy protection and encryption to mitigate this risk.

3) Knowledge Modeling for Transmitters and Receivers: Typical semantic communication assumes that both the transmitter and receiver share the same knowledge database, leading them to interpret semantic information similarly. However, this assumption is not always practical, given that different end-users might possess different knowledge sets. As such, it is essential to model the knowledge database of each user, whether at the transmitter or receiver side. The primary focus is to understand how data samples from one user's knowledge database influence the training and outcomes of GenAI models, specifically observing changes in gradient and loss. Ultimately, with these understandings, we can delve deeper into generating and transmitting semantic information tailored to different knowledge databases.

4) Adversarial Backdoor Attacks: While semantic communication powered by GenAI offers superior resilience against channel distortion and data loss compared to conventional communication methods, it introduces emerging vulnerabilities, particularly to adversarial backdoor attacks. Such attacks are characterized by the subtle manipulation of a minor subset of the data used in establishing a communication system. This manipulation is designed to erroneously trigger specific responses from the system, introducing novel security challenges. To highlight this risk, we carry out range testing to assess the robustness of a GenAI-based semantic

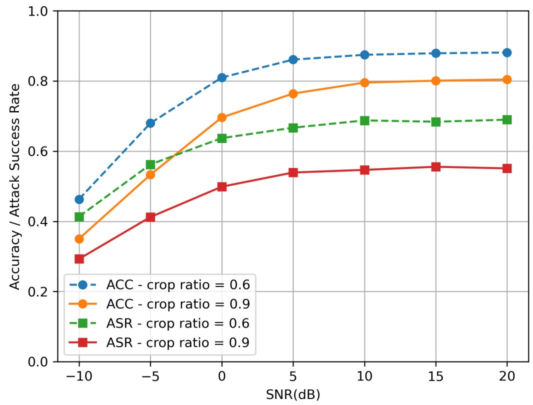

Fig. 6. Comparisons of accuracy (ACC) and attack success rate (ASR) under different crop rates.

communication system. We employ Transformer for semantic feature extraction and use mutual information estimation to simulate channel fading effects. To simulate the attack, the communication system is configured with a backdoor trigger size of 5×5 pixels within images of size 32×32 pixels. 500 out of $5,000$ training images are tainted, establishing a poisoning rate of 0.1. Fig. 5 evaluates the system performance under this attack, focusing on two key metrics: accuracy (ACC) and attack success rate (ASR). Notably, ASR is sensitive to semantic compression (crop ratio). As the compression ratio lessens, the impact on the semantics of backdoor triggers is more profound than on standard semantics. Fig. 6 reveals that increasing solely signal-to-noise ratio (SNR) facilitates the increase of both ACC and ASR, indicating that merely filtering signal interference would not defend against backdoor attacks in semantic communication. Fortifying GenAI-based semantic communications with strategic interference management against these backdoor attacks will be a critical research area moving forward.

VI. Conclusion

This article provides a systematic overview of GenAI and unveils the vast potential of its integration into networks and communication. We first introduce the fundamental knowledge about several representative, cost-effective GenAI models, including variational auto-encoder, Transformer, and diffusion models. Their applications to emerging networking technologies are showcased with three exemplary studies, including digital twins, integrated sensing and communication, and semantic communication. The critical challenges and practical considerations are discussed, encompassing aspects like data quality, real-time ability, privacy leakage, and security concerns. Moreover, a forward-looking vision of the research opportunities in tackling these challenges is provided, aiming to fully harness the power of GenAI for future networks.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
[2] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. the IEEE/CVF Conference on CVPR*, 2022.
[3] L. Bariah, Q. Zhao, H. Zou, Y. Tian, F. Bader, and M. Debbah, "Large generative ai models for telecom: The next big thing?" *IEEE Communications Magazine*, 2023.
[4] A. Maatouk, N. Piovesan, F. Ayed, A. De Domenico, and M. Debbah, "Large language models for telecom: Forthcoming impact on the industry," *arXiv preprint arXiv:2308.06013*, 2023.
[5] L. Dong, F. Jiang, Y. Peng, K. Wang, K. Yang, C. Pan, and R. Schober, "Lambo: Large language model empowered edge intelligence," *arXiv preprint arXiv:2308.15078*, 2023.
[6] H. Zou, Q. Zhao, L. Bariah, M. Bennis, and M. Debbah, "Wireless multi-agent generative AI: From connected intelligence to collective intelligence," *arXiv preprint arXiv:2307.02757*, 2023.
[7] Y. Liu, H. Du, D. Niyato, J. Kang, Z. Xiong, D. I. Kim, and A. Jamalipour, "Deep generative model and its applications in efficient wireless network management: A tutorial and case study," *IEEE Wireless Communications*, 2023.
[8] H. Du, D. Niyato, J. Kang, Z. Xiong, K.-Y. Lam, Y. Fang, and Y. Li, "Spear or shield: Leveraging generative AI to tackle security threats of intelligent network services," *arXiv preprint arXiv:2306.02384*, 2023.
[9] V. Jüttner, M. Grimmer, and E. Buchmann, "Chatids: Explainable cybersecurity using generative ai," 2023.
[10] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *International Conference on Learning Representations*, 2014.
[11] P. Almasan, M. Ferriol-Galmés, J. Paillisse, J. Suárez-Varela, D. Perino, D. López, A. A. P. Perales, P. Harvey, L. Ciavaglia, L. Wong *et al.*, "Digital twin network: Opportunities and challenges," *arXiv preprint arXiv:2201.01144*, 2022.
[12] Y. Cui, F. Liu, X. Jing, and J. Mu, "Integrating sensing and communications for ubiquitous IoT: Applications, trends, and challenges," *IEEE Network*, vol. 35, no. 5, pp. 158–167, 2021.
[13] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor, "Less data, more knowledge: Building next generation semantic communication networks," *arXiv preprint arXiv:2211.14343*, 2022.
[14] Y. Song, P. Dhariwal, M. Chen, and I. Sutskever, "Consistency models," ser. Proceedings of the 40th International Conference on Machine Learning. PMLR, 2023.
[15] A. Adhikary, M. S. Munir, A. D. Raha, Y. Qiao, Z. Han, and C. S. Hong, "Integrated sensing, localization, and communication in holographic MIMO-enabled wireless network: A deep learning approach," *IEEE Transactions on Network and Service Management*, 2023.

Zhiyuan Peng (Member, IEEE) is currently a Post-doc with the Department of Computer Science at North Carolina State University, USA. His current research interests include wireless networking, generative AI, and speech and language processing.

Yuchen Liu (Member, IEEE) is an Assistant Professor with the Department of Computer Science at North Carolina State University, USA. His research interests include wireless networking, generative AI, reinforcement learning, mobile computing, and software simulation.

Gaolei Li (Member, IEEE) is an Assistant Professor in School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include adversarial machine learning and privacy-preserving.

Zhaohui Yang (Member, IEEE) is currently an Assistant Professor with Zhejiang University. His research interests include joint sensing, computing and communication, semantic communication, federated learning, and RSMA.

Mingzhe Chen (Member, IEEE) is currently an Assistant Professor with the Department of Electrical and Computer Engineering and Institute for Data Science and Computing, University of Miami. His research interests include federated learning, reinforcement learning, virtual reality, unmanned aerial vehicles, and the Internet of Things.

Dongkuan Xu (Member, IEEE) is an Assistant Professor with the Department of Computer Science at North Carolina State University, USA. His research focuses on exploring and advancing Landed Generative AI, with particular emphasis on studying the autonomy of intelligent agents, decision/reasoning reliability, and resource efficiency in Generative AI Systems.

Xingqin Lin (Senior Member, IEEE) is a Senior Standards Engineer at NVIDIA, leading 3GPP standardization and conducting research at the intersection of 5G/6G and AI. He is a co-chair of the Alliance for Telecommunications Industry Solutions (ATIS) AI Network Applications working group.