THE INTERPLAY BETWEEN GENERATIVE AI AND 5G-ADVANCED TOWARD 6G

# Harnessing the Power of AI-Generated Content for Semantic Communication

Yiru Wang, Wanting Yang, Zehui Xiong, Yuping Zhao, Tony Q. S. Quek, and Zhu Han

## Abstract

Semantic Communication (SemCom) is envisaged as the next-generation paradigm to address challenges stemming from the conflicts between the increasing volume of transmission data and the scarcity of spectrum resources. However, existing SemCom systems face drawbacks, such as low explainability, modality rigidity, and inadequate reconstruction functionality. Recognizing the transformative capabilities of AI-generated content (AIGC) technologies in content generation, this paper explores a pioneering approach by integrating AIGC into SemCom to address the aforementioned challenges. We employ a three-layer model to illustrate the proposed AIGC-assisted SemCom (AIGC-SCM) architecture, emphasizing its clear deviation from existing SemCom. Grounded in this model, we investigate various AIGC technologies with the potential to augment SemCom's performance. In alignment with the SemCom's goal of conveying semantic meanings, we also introduce the new evaluation methods for our AIGC-SCM system. Subsequently, we explore communication scenarios where the proposed AIGC-SCM can realize its potential. For practical implementation, we construct a detailed integration workflow and conduct a case study in a virtual reality image transmission scenario. The results demonstrate the ability to maintain a high degree of alignment between the reconstructed content and the original source information, while substantially minimizing the data volume required for transmission. These findings pave the way for further enhancements in communication efficiency and the improvement of Quality of Service. Finally, we present future directions for AIGC-SCM studies.

## Introduction

The ongoing advancement of artificial intelligence algorithms and the growing support from computational devices have led to the emergence of innovative applications such as autonomous driving, virtual reality (VR), augmented reality (AR) and mixed reality (MR). These applications, in turn, require substantial communication data capacity. At the same time, conventional communication systems are approaching the Shannon limit and the available spectrum resources are becoming increasingly scarce. Therefore, a novel approach to tackle the ever-evolving challenges in communication is greatly needed.

Currently, Semantic Communication (SemCom) has shown its potential as a promising communication paradigm in six-generation (6G), situated at the intersection of artificial intelligence and communication technology. Its primary goal is to convey semantic information from the source to the receiver, focusing on the meaning of the message rather than the precise reception of individual symbols or bits. In contrast to Shannon communication, SemCom adopts the 'extract before coding' approach. By delicately devising a pair of semantic codecs, the semantic information concealed within the highly compressed transmitted messages can be reconstructed at the receiver. This approach proves advantageous in scenarios characterized by communication resource constraints and strict privacy preservation requirements.

At present, the outstanding performance of SemCom has been demonstrated in recent literature [1]. However, there are still some challenges hindering the large-scale implementation of SemCom:

- **Low Explainability:** Traditional SemCom relies on end-to-end models that are hard to be interpreted, making troubleshooting and manual review both time-intensive and expensive.
- **Modality Rigidity:** The uni-modal designs in current SemCom studies are commonplace, but they tend to be insufficient in scenarios involving multiple modalities or cross-modality interactions. This limitation hinders the adaptability of the existing structure to various communication scenarios.
- **Reconstruction Limitation:** Existing methods fall short in recovering high-quality content when the compression rate is reduced to an extreme level. This limitation hinders the adoption of SemCom, especially in scenarios where communication resources are extremely scarce.

Addressing these challenges requires a paradigm shift in SemCom methodologies. In some communication scenarios, the receiver may not require perfect reconstruction of all source information; instead, it prioritizes the accurate presentation of semantic meanings related to the tasks. For example, in vehicular networks, the decisions made by drivers remain unaffected by

Digital Object Identifier:
10.1109/MNET.2024.3420400
Date of Current Version:
16 September 2024
Date of Publication:
28 June 2024

*Yiru Wang is with the School of Electronics, Peking University, Beijing 100871, China, and also with the Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore 487372; Wanting Yang, Zehui Xiong (corresponding author), and Tony Q. S. Quek are with the Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore 487372; Yuping Zhao is with the School of Electronics, Peking University, Beijing 100871, China; Zhu Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul 446-701, South Korea.*

alterations in the coloration of surrounding vehicles. Similarly, when developing a digital twin for a manufacturing facility, the precise texture or hue of machinery may not hold significance for the simulation and analysis of the plant's operational efficiency. Additionally, within a video monitoring system designed for traffic management, variations in the background environment do not influence the system's capability to identify and track vehicles. Therefore, bandwidth for transmitting less-important information of this nature should be conserved. In certain situations, cross-modality communication is also preferred, exemplified by the choice to transmit direct voice alarms instead of complex radar point clouds to nearby drivers in emergency contexts. In these scenarios, AI-generated content (AIGC) technologies demonstrate their versatility in enhancing communication, due to their numerous advantages, including:

- **High-Quality Generation:** AIGC technologies demonstrate remarkable abilities in generating high-quality content in a variety of domains. These capabilities stem from their ability to understand and learn patterns from vast amounts of data during the training process.
- **Multi-Modality Compatibility:** These technologies exhibit remarkable multi-modality capabilities, allowing them to generate and transform content across various data types [2], thereby fostering innovation across a wide range of applications.
- **Personalization:** AIGC technologies offer a remarkable personalization ability by harnessing individual preferences and requirements to create tailored content. These technologies can adapt responses and outputs to meet various demands, thus enhancing engagement and satisfaction [3].

Through seamless integration of AIGC into SemCom, we can address the aforementioned challenges and endow the AIGC-assisted SemCom (AIGC-SCM) with the following features:

- **Separated Functional Components:** AIGC technologies can be effortlessly integrated with Shannon communication, facilitating the establishment of AIGC-SCM in a plug-and-play fashion. This integration not only boosts interpretability but also optimizes the use of existing infrastructure.
- **Multi-Modality Support:** AIGC possesses the capability to generate content in various formats, thereby enhancing SemCom's adaptability to multi-modal tasks. Moreover, by accessing user knowledge bases (KBs),[1] it can learn from user attributes and customize content accordingly, ensuring that each SemCom user receives content tailored to their specific needs.
- **Generation-Assisted Reconstruction:** Leveraging the generative capability of AIGC, AIGC-SCM demonstrates the ability to reconstruct content with high fidelity and semantic consistency, even in cases of heavily compressed transmitted data.

While AIGC has demonstrated its versatility in various applications, integrating it into SemCom requires comprehensive investigations. *Two challenges* arise when using AIGC to enhance SemCom. Firstly, there is a need to establish a prompt extraction model that aligns with the receiver's tasks, striking a balance between service quality and communication costs. Second, the output of AIGC can be somewhat random and unpredictable. However, we expect to maintain the semantic consistency between the source and the destination in SemCom. Hence, it is crucial to consider strategies for controlling the generation process within AIGC-SCM. Although the authors in [4] initiate the integration of SemCom and AIGC, the exploration into AIGC's varied roles within SemCom is not exhaustive. Additionally, the impact of diverse prompts and KBs on reconstructing content is not thoroughly examined, pointing towards areas requiring deeper exploration to fully leverage AIGC's capabilities in SemCom contexts.

AIGC technologies demonstrate remarkable abilities in generating high-quality content in a variety of domains.

To fill in the gap, in this paper, we present a unified approach to provide guidance for constructing AIGC-SCM. The contributions of this paper are summarized as follows:

- We introduce our proposed AIGC-SCM by rebuilding the three-layer model and highlighting how it differs from the existing structure. To elucidate how AIGC can enhance SemCom, we explore various AIGC technologies and showcase their potential roles as extractors, denoisers, and reconstructors in the SemCom process. Subsequently, we introduce evaluation methods for AIGC-SCM and delve into potential applications.
- To address the aforementioned challenges in the practical integration of AIGC into SemCom, we outline step-by-step workflows targeted at different functional aspects. By following these procedures, AIGC-SCM can improve interpretability, achieve multi-modality compatibility, and enhance reconstruction ability.
- In our case study, we follow the proposed workflows and exemplify an application of AIGC-SCM in image transmission. We devise specific prompts to cater to diverse application requirements and examine how variations in KBs affect the reconstruction outcomes. Our study illustrates the effectiveness of AIGC-SCM in minimizing bandwidth usage while preserving semantic similarity.

## Insights Into AIGC-SCM

We present a refined three-layer model of AIGC-SCM, as illustrated in Fig. 1. Different from existing SemCom frameworks, our approach substitutes the semantic representator and semantic encoder with a prompt extractor and prompt encoder at the transmitter. At the receiver, we incorporate an AIGC-based denoiser to mitigate the impact of physical noise. Subsequently, we adapt the semantic decoder and interpreter to function as a prompt decoder and AIGC-assisted reconstructor. In the sequel, we embark on a comprehensive exploration of AIGC, introduce evaluation methods for AIGC-SCM and outline various applications.

[1] Aknowledge base (KB) is a centralized repository designed for the organized and efficient management and retrieval of information. Within the SemCom system, the KB encompasses user preferences, network parameters training datasets, task objectives, and historical communication records.

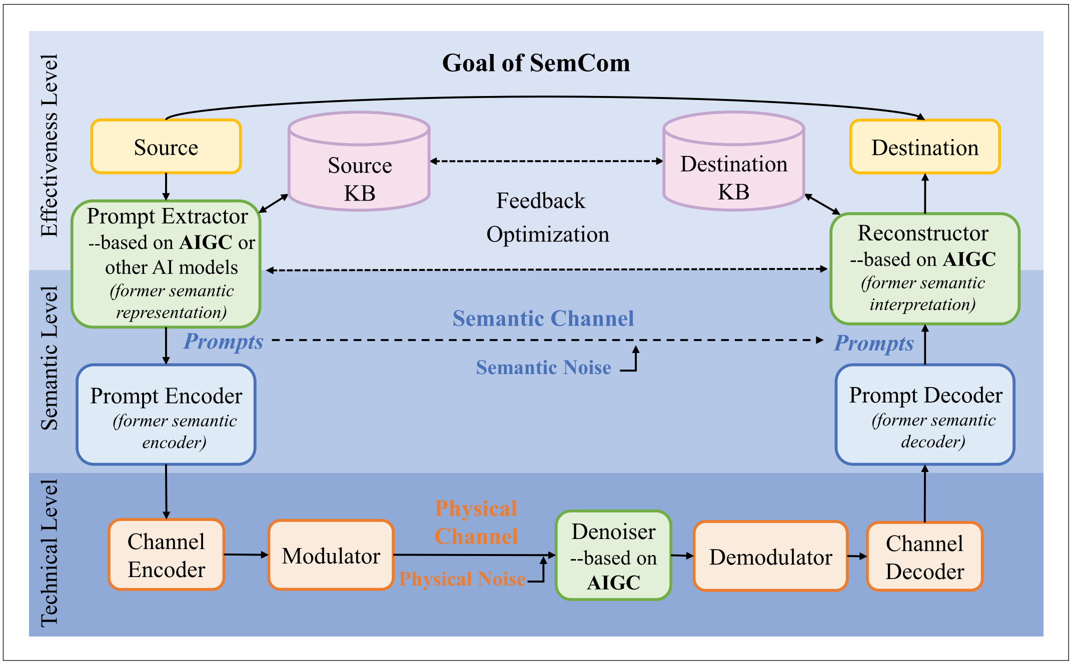
Goal of SemCom
Effectiveness Level
Semantic Level
Technical Level
Source
Source KB
Destination KB
Destination
Feedback
Optimization
Prompt Extractor --based on **AIGC** or other AI models *(former semantic representation)*
Reconstructor --based on **AIGC** *(former semantic interpretation)*
Semantic Channel
Prompts
Prompts
Semantic Noise
Prompt Encoder *(former semantic encoder)*
Prompt Decoder *(former semantic decoder)*
Channel Encoder
Modulator
Physical Channel
Physical Noise
Denoiser --based on **AIGC**
Demodulator
Channel Decoder

FIGURE 1. Three-layer model of the proposed AIGC-SCM. Within this framework, semantic features are identified in the form of prompts. Following the traditional source/channel encoding and modulating steps, the prompt is transmitted through the physical channel to the receiver. After undergoing the corresponding demodulating and channel/source decoding procedures, the prompt is recovered to guide generation at the receiver.

## How AIGC Can Empower SemCom

In the context of our refined three-layer framework, the incorporation of AIGC fulfills diverse roles. It can operate as the prompt extractor, serve as the denoiser, and play a role as the reconstructor at the receiver. Subsequently, we will elaborate on these functions, accompanied by representative examples illustrated in Fig. 2.

1. *For Prompt Extractor:* The goal of the prompt extractor is to distill semantic prompt from the source data. In scenarios where data is high-dimensional, AIGC can efficiently capture and extract the most informative features. As discussed in [2], AIGC possesses the capability to transform the source into an alternative, yet more concise modality, thereby diminishing dimensionality and conserving transmission bandwidth. For example, it can summarize texts from expansive vocabularies, extract descriptive captions from images, and transform lengthy sentences into knowledge graphs.
2. *For Denoiser:* Some AIGC models, such as diffusion models, demonstrate notable capabilities in denoising. By utilizing a progressive denoising approach, they can iteratively enhance content quality. This approach ensures that denoised content remains faithful to the original. This proficiency proves effective across various modal types and noise levels [5]. Therefore, through the application of these models, there is potential to mitigate physical noise across various channel conditions.
3. *For Reconstructor:* In contrast to the extractor, the reconstructor aims to enhance the receiver's quality of experience by generating high-dimensional semantic information based on low-dimensional prompts. In this field, AIGC possesses formidable capabilities and vast applications. For instance, AIGC can also translate text instructions into images [2]. To address the challenges posed by constrained bandwidth resources in video transmission, a proposed solution is to transmit only the changes in target semantic maps. These changes then serve as guidance for the AIGC-assisted reconstructor in generating video frames [6]. To enhance privacy preservation by minimizing users' exposure to cameras, [7] leverages channel state information (CSI) to predict a user's skeleton, which is then employed to instruct the AIGC at the receiver to generate a virtual character.

## How to Evaluate AIGC-SCM Performance

As highlighted by [1], unlike metrics such as bit and symbol error rate in Shannon communications, metrics for assessing SemCom should focus on semantic similarity. In this context, conventional metrics may prove inadequate. To illustrate, we present a summary of two scenarios, each accompanied by its corresponding rationale:

1. *Disparities in Semantic and Representation Similarity* [8]*:* Even when entities share semantic traits, differences in their representation can cause traditional metrics to yield low similarity scores. This happens if the AIGC-assisted reconstructor at the receiver interprets semantic information differently than the transmitter, perhaps through varied sentence expressions or image styles. For example, in image transmission, orchids and

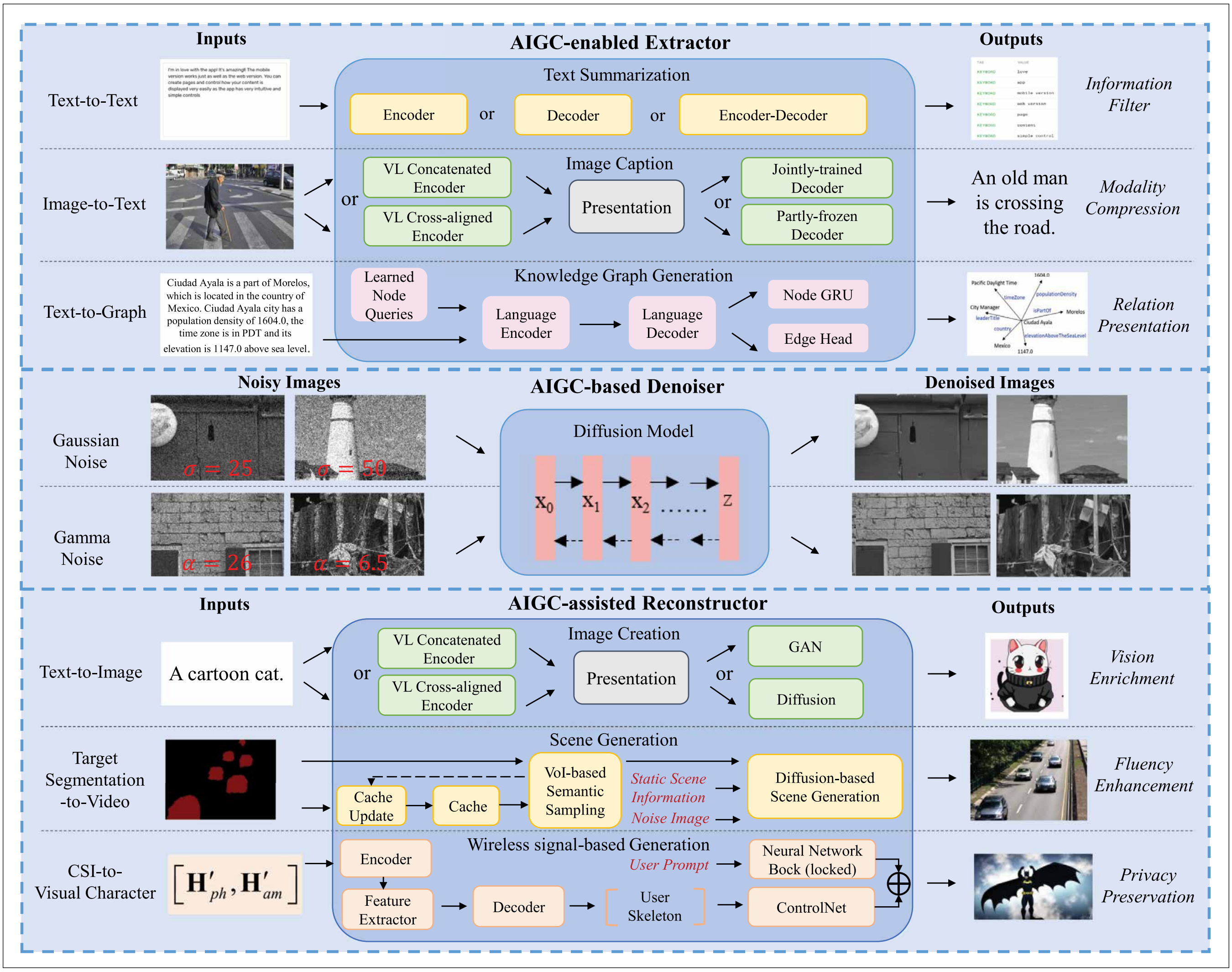

FIGURE 2. Functions and representative technologies of AIGC for enhancing SemCom. In our proposed AIGC-SCM framework, AIGC can be utilized as the prompt extractor, denoiser, and reconstruction. We highlight several representative AIGC technologies capable of effectively fulfilling these functions.

sunflowers share high semantic similarity as flowers, but conventional metrics like peak signal-to-noise ratio (PSNR) rate their similarity lower than that between a deer and a forest, due to emphasis on representation consistency.
2. *Modality Mismatch Between Source and Destination:* Modality transitions are commonplace in AIGC-SCM as they enable data compression and provide personalized services. However, existing SemCom works have not yet developed corresponding metrics for scenarios where the source and destination exist in different modalities.

In this article, we propose to evaluate AIGC-SCM's performance by comparing the vectors in the feature space between the transmitter and the receiver. This choice is motivated by several advantages:
1. Feature-based similarity often captures the semantic relevance of content more effectively;
2. By measuring similarity in the feature space, we can mitigate the effects of variations, noise, and modality mismatching;
3. Features learned from some pre-trained models, like CLIP [9], a pre-trained vision-language (VL) model that learns to map images and text descriptions into a shared embedding space, are more likely to generalize well to diverse datasets. The encoded high-level patterns are transferable across different tasks and domains.

In Fig. 3(b), we introduce evaluation methods using the feature space vector for both uni-modal and cross-modal transmission. For uni-modal transmission, we consider image transmission as an example. In this case, the same feature extraction model, such as an image classification model without the final decision layer, can be employed at both the transmitter and receiver to obtain latent feature vectors. A comparison of these vectors can then be calculated using methods like cosine similarity. In the context of cross-modality transmission, we consider image-text transmission as an example. Here, we can choose a pre-trained VL model and input both the image and text into this model. This process allows us to retrieve their feature vectors in a shared embedding space. Finally, an evaluation can be made using standard similarity calculations.

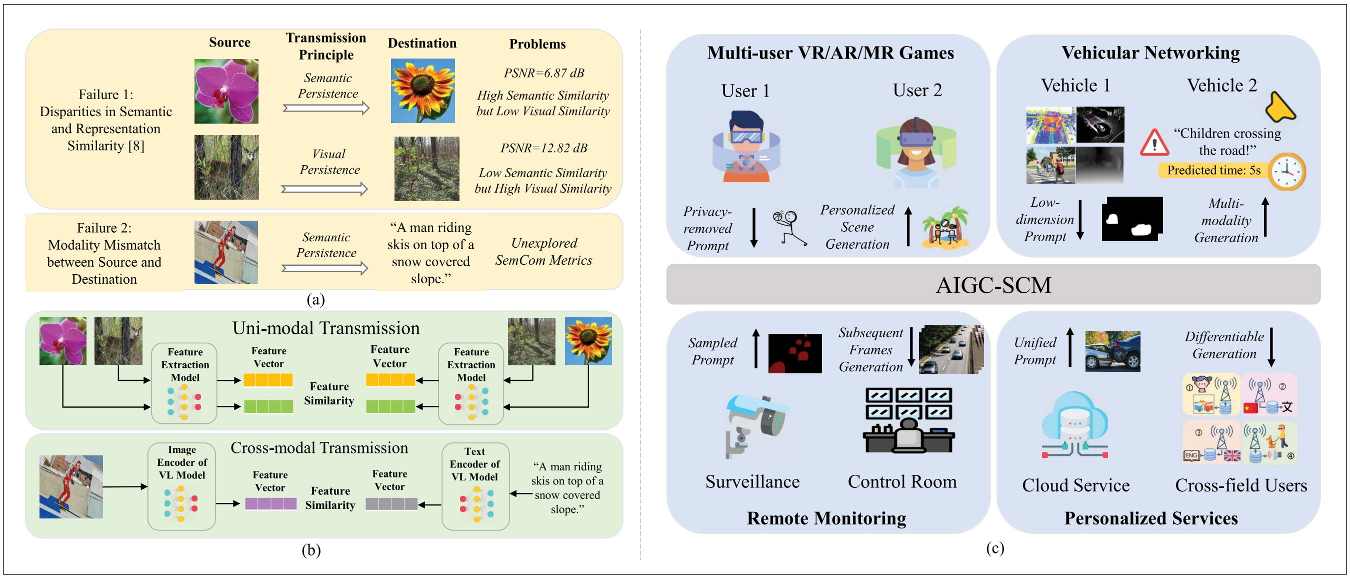

FIGURE 3. Evaluation methods and applications of AIGC-SCM. Certain conventional metrics may fall short in evaluating AIGC-SCM due to disparities in semantic and representation similarity, along with modality mismatches between the source and destination. To address this, we introduce evaluation methods utilizing feature space vectors for both uni-modal and cross-modal transmission. Additionally, we illustrate how AIGC-SCM can enhance 6G in various application scenarios, including multi-user VR/AR/XR games, vehicular networking, remote monitoring, and personalized services.

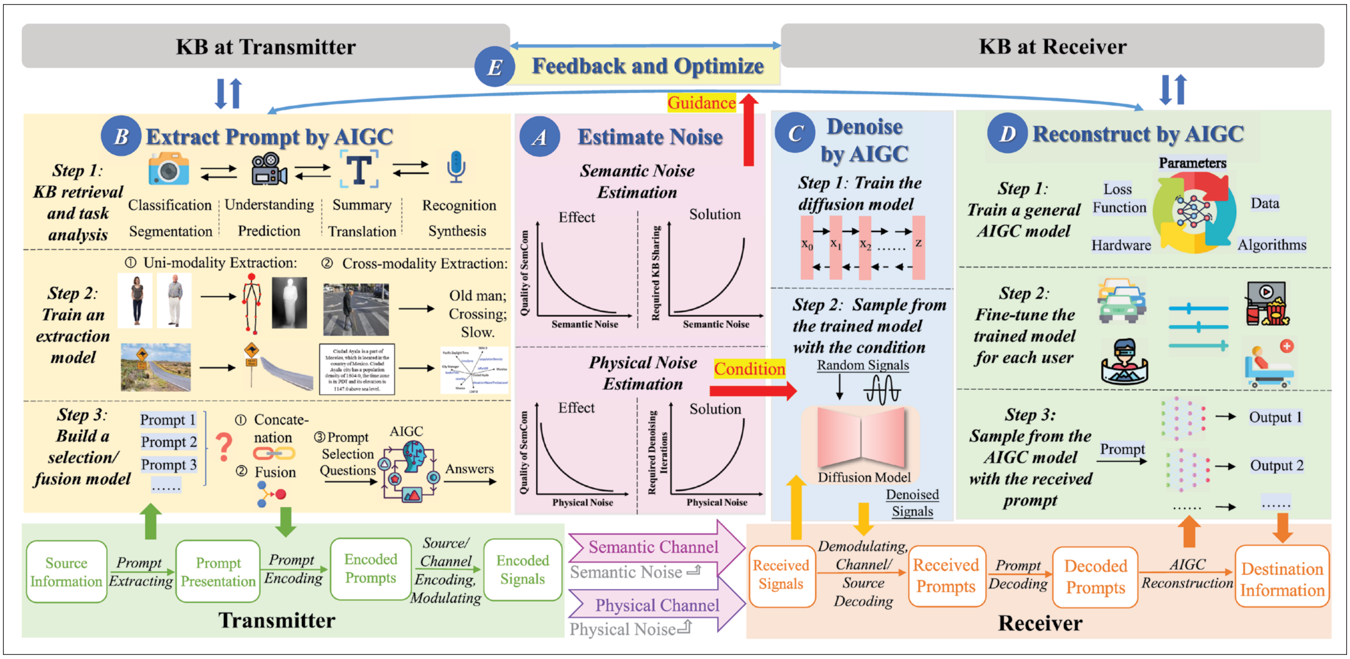

FIGURE 4. Workflow of implementing AIGC-SCM. The AIGC-SCM implementation workflow involves a systematic process, beginning with the estimation of noise as a preparatory step. Subsequently, AIGC is employed in three pivotal stages: extracting prompts, denoising received signals, and reconstructing contents. The final step entails gathering feedback and optimizing the entire system.

## How AIGC-SCM can Empower 6G

As depicted in Fig. 3(c), we illustrate how AIGC-SCM can enhance 6G in four typical application scenarios, described as follows:

1. *Multi-User VR/AR/MR Games:* Several challenges have emerged in achieving the large-scale implementation of VR/AR/MR gaming, including restricted communication resources, privacy concerns, and monotonous content. A promising solution to address these issues is AIGC-SCM, as it can represent virtual scenes within the prompt and generate diverse and personalized content for different users.
2. *Vehicular Networking:* Extensive data exchanges from numerous sensors in vehicular networks exert a considerable burden on transmission systems. To mitigate this, AIGC-SCM can be employed to distill and transmit location-related information. At the receiver end, it can enable the intuitive recognition and display of traffic congestion

and unusual incidents, providing essential information for drivers' decision-making.

3. *Remote Monitoring:* In typical and foreseeable remote monitoring scenarios, AIGC-SCM exhibits the capability to reconstruct the current frame based on received prompts and predict subsequent frames conditioned on it. This capacity offers a solution to reduce the volume of communication data.
4. *Personalized Services:* AIGC-SCM is equipped with a remarkable personalization ability by harnessing individual preferences and behavior data to adapt responses and outputs to cater to specific receivers, enhancing engagement and satisfaction.

## AIGC-SCM Framework Implementation

While the integration of AIGC into the current framework poses potential benefits for SemCom, it is not straightforward. We need to properly control the AIGC generation and ensure that AIGC-SCM meets the receiver's requirements. In this section, we discuss the specific procedures that need to be implemented to ensure the smooth functioning of the entire system.

### Estimate Noise

Both physical noise and semantic noise can impact the reconstruction process and reduce the quality of SemCom. Nonetheless, by accurately estimating their levels, we can optimize the system design to mitigate their effects accordingly. The estimation of physical noise can serve as a condition for the AIGC-assisted denoiser, influencing the determination of denoising iterations. Semantic noise estimation serves as guidance for knowledge sharing. In instances where semantic noise is high, enabling the sharing of data, models, and domain-specific terminology is advisable before engaging in communication.

### Extract Prompt by AIGC

In AIGC-SCM, the prompts are extracted at the transmitter to distill the semantic meanings from the source. Given that both modality and the quantity of prompts impact the system's performance, a decision on the extraction model is crucial. To address this, we provide a general procedure for designing the extraction model and generating effective prompts:

- **Step 1: KB retrieval and task analysis.** In SemCom, the KBs store information about each user's datasets, preferences, and tasks. Accessing KBs enables the identification of the specific tasks for the AIGC model, facilitating the development of suitable extraction models.
- **Step 2: Train an extraction model.** Following task analysis, we can tailor the design of our prompt extraction models accordingly. Traditional deep learning models, including CNNs, RNNs, and Transformers, can be intentionally designed and trained to extract uni-modality semantic features. Furthermore, AIGC models can accomplish modality transformation, providing a method to extract cross-modality prompts and hence reduce data transmission volume, as previously discussed in the section "How AIGC Can Empower SemCom."
- **Step 3: Build a selection/fusion model.** In situations where communication resources are constrained, the determination of final prompts from an extensive set of acquired features becomes crucial. One approach involves building reinforcement learning models to select and concatenate the most critical features. Another option is to leverage deep learning models for feature fusion, creating a more compressed prompt. Additionally, certain AIGC models, functioning as decision-makers, can be utilized to assist in prompt selection.

While the integration of AIGC into the current framework poses potential benefits for SemCom, it is not straightforward.

### Denoise by AIGC

When the signal arrives at the receiver, it has been distorted by the physical channel. As we discuss in the section "How AIGC Can Empower SemCom," some AIGC models can be utilized for denoising communication signals. We take the diffusion model for example and discuss the following steps for denoising:

- **Step 1: Train the diffusion model.** During diffusion model training, there are two key processes: the forward process and the backward process. In the forward process, the diffusion model gradually adds noise to an input data sample. Subsequently, the diffusion model proceeds with the backward process and reverses the degradation. Starting with the most noisy version of the data, the model's objective is to denoise the data iteratively, attempting to recover the original, noise-free input data. In light of the above, the model learns to remove noise and restore data to its original quality.
- **Step 2: Sample from the trained diffusion model with the condition.** After the training, the sampling procedure leverages the trained network to generate a denoised sample. With the condition of the estimated physical noise, we can reduce the discrepancy between the modified and true conditional distributions for each sampling step, thus making the ultimately recovered signal much closer to the original transmitted signal.

### Reconstruct by AIGC

Following the decoding process, the prompt is recovered, serving as a condition for the implementation of an AIGC-based reconstruction. In the single-user scenario, we can directly train a reconstruction model that satisfies the user's requirement. In the multi-user scenario, the sender can likewise transmit only one prompt to enhance the communication efficiency, because the receiver can employ the AIGC model to generate personal content in a variety of presentation formats based on the same semantic information. To ensure that the trained model is adaptable to various user scenarios, we employ an efficient three-step strategy, outlined as follows:

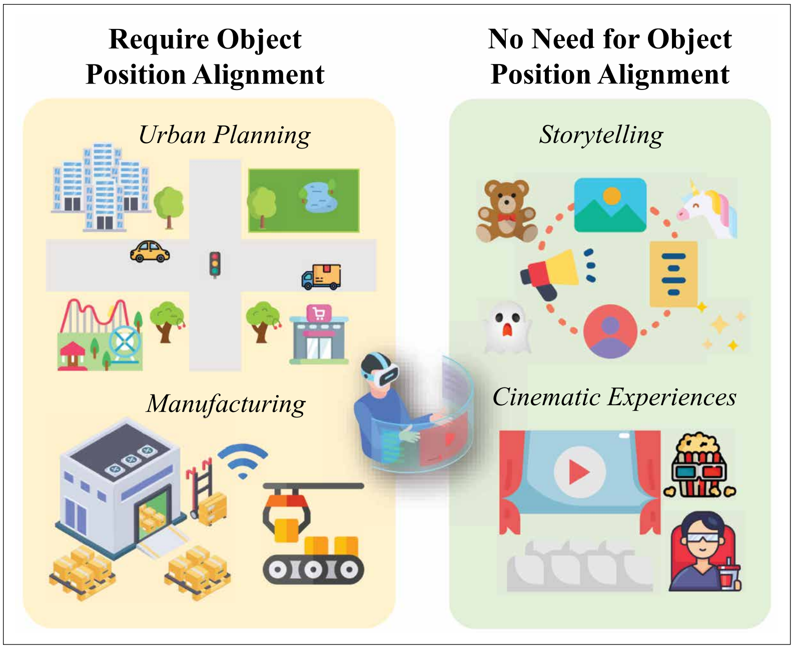
Require Object Position Alignment
No Need for Object Position Alignment
Urban Planning
Storytelling
Manufacturing
Cinematic Experiences

FIGURE 5. VR image transmission scenarios categorized based on object position alignment requirements. The left two applications exemplify VR mapping, demanding high accuracy in object position. In contrast, the misalignment of object position between the transmitter and the receiver does not impact the experience of the right two applications.

- **Step 1: Train a general AIGC model.** First, we develop a general model, like a stable diffusion model [10] for text-to-image transmission. Before training, we should collect the dataset for this model, which is typically extensive and diverse within the modalities and tasks defined. It can enable the basic model to effectively execute different tasks and generalize across various scenarios.
- **Step 2: Fine-tune the trained model.** In a multi-user scenario, receivers often have diverse domains and preferences. To enable the AIGC model to produce personalized content, we need to fine-tune the general model based on personal databases. For instance, in the text-to-image transmission scenario, we fine-tune our trained diffusion model to facilitate image reconstruction in various styles. Even in a single-user scenario, this step proves beneficial when the user's preferences change.
- **Step 3: Sample from the AIGC model with the received prompt.** After training the AIGC generation model, we integrate it into the AIGC-SCM framework. When the decoded prompt is input into the AIGC models of each user, they have the capability to generate more tailored semantic interpretations.

### Feedback and Optimize

The described method initiates the AIGC-SCM model, yet practical use demands ongoing feedback-driven refinement, including KB sharing and fine-tuning between transmitter and receiver. This feedback is vital for the AIGC-SCM's continuous enhancement and its adaptation to users' changing needs and the dynamic communication environment.

## Case Study: AIGC-SCM for Image Transmission

Communication plays a pivotal role in VR applications, serving a multitude of purposes. It enables the facilitation of multiplayer experiences, the establishment of collaborative workspaces, and the enhancement of immersive experiences. However, the substantial data volume and increased latency associated with traditional VR image transmission lead to a diminished VR experience. To address this, we propose to use AIGC-SCM for data transmission, which can efficiently extract and reconstruct semantic information and reduce the required communication resources. As shown in Fig. 5, to differentiate between various VR application requirements and design extraction models accordingly, we classify image transmission scenarios based on whether they require the alignment of object positions between the transmitter and receiver images.

### AIGC-SCM Schemes and KB Differentiation

In our case study, we examine a VR image transmission scenario where Transmitter A intends to multicast its scene to two distinct recipients. To assess the impact of KBs on the reconstruction process, we assume that Receiver B possesses only general knowledge, while Receiver C has additional historical information about the objects created within Transmitter A's scene.

To accommodate scenarios with object position alignment requirements (A) or without object position alignment requirements (I), and to tailor for users with historical (H) knowledge or without historical (N) knowledge, we implement four distinct AIGC-SCM schemes: NI, NA, HI, and HA.

For the prompt extraction processes, we employ the YOLO object detection model [11] to extract labels and bounding boxes in NA and HA schemes. In NI and HI schemes, we utilize the image caption ExpansionNet v2[2] model to extract text descriptions. This extracted text, along with unique identifiers,[3] serves as the prompt for transmission.

For the image reconstruction processes, we employ the stable diffusion v2[4] model for NI scheme and the Dreambooth model [12] for HI scheme, and the GLIGEN [3] model for both NA and HA schemes.

The LAION [13] and COCO2014 [14] datasets are considered as general knowledge, which is stored in both receivers' KBs. The LAION [13] dataset is regarded as the general knowledge due to its role in training the stable diffusion model [10], which is subsequently utilized as the pretrained model for both receivers. Encompassing around 400 million image-text pairs, this dataset encapsulates a broad array of subjects, themes, and visual styles. Another dataset serving as general knowledge is the COCO2014 [14] dataset, which features over 200,000 labeled images across 80 object categories and is utilized in the training of the GLIGEN [3] model for the NA and HA schemes.

Meanwhile, the Dreambooth dataset [12] includes 30 subjects of 15 different classes and contains 4-6 images per subject, which are captured in different conditions, environments and under different angles. In our study,

[2] https://github.com/jcheng-hu/ExpansionNet_v2

[3] In practice, every transmission requires a unique identifier to accurately locate the object mentioned in the prompt within Receiver C's historical data For Receiver B, this identifier can be considered semantic noise and thus disregarded. For the sake of clarity and conciseness, the transmission of these short identifiers is not taken into account in the visual representation of our results or the numerical comparisons presented in this paper.

[4] https://huggingface.co/CompVis/stable-diffusion-v1-2

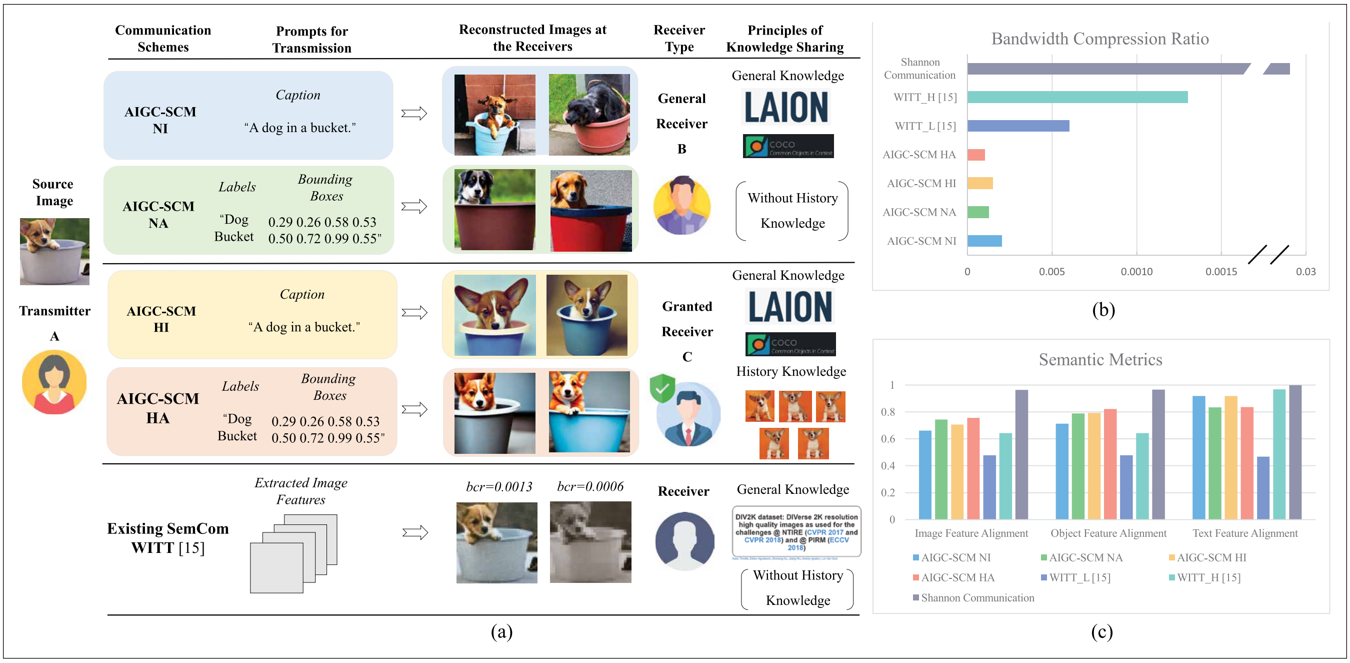

FIGURE 6. Performance comparison of different communication schemes. We compare our proposed AIGC-based SemCom with two benchmarks: 1) Existing AI-based SemCom scheme, where we choose WITT [15] model with a higher (_H) and lower (_L) bandwidth compression ratio. It utilizes the Swin Transformer to manage long-range image dependencies, incorporates a spatial modulation module for adaptability to fluctuating wireless channel conditions, and employs a hierarchical design to streamline high-resolution image processing; 2) Shannon communication, where we use JPEG for direct image compression, ½ LDPC for channel coding, and 16QAM for modulation. All of these schemes undergo testing under the condition of the 10 dB SNR. To enhance the AIGC-SCM's interpretability and ensure its seamless integration with existing communication architectures, the channel-related modules employed in our AIGC-SCM configurations are aligned with those of the Shannon communication scheme.

the images in this dataset represent specific historical information shared by Transmitter A with Receiver C, stored in Receiver C's KB beyond general knowledge. Nonetheless, the application of this dataset diverges between the HI and HA schemes. Within the HI framework, the Dreambooth [12] model is adopted, which is fine-tuned on a pre-trained stable diffusion model [10] using this dataset. When the prompt is decoded, the Dreambooth [12] model can utilize its prior knowledge of the subject and generate class-specific instances according to the prompt's context. Conversely, in the HA configuration, upon prompt decoding at the receiver, a pertinent historical image from the Dreambooth [12] dataset is fetched and integrated into the embedding layers of the GLIGEN [3] model to guide the image reconstruction.

## Visual and Numerical Results

In Fig. 6(a), we present a visual performance comparison between the proposed AIGC-SCM schemes and the existing SemCom scheme WITT [15]. It is evident that all the schemes successfully preserve a certain level of semantic similarity when compared to the source image, as they all depict images of a dog in a bucket. However, by transmitting labels and bounding boxes, we gain better control over the object positions. Since AIGC-SCM NA and NI can only rely on general KBs, while AIGC-SCM HA and HI can condition on historical knowledge, it becomes apparent that the dogs depicted in schemes AIGC-SCM HA and HI exhibit a closer resemblance to the source image. In contrast, the reconstructed images generated by the WITT [15] scheme tend to appear blurry and pale due to its sensitivity to compression, distinguishing its images from the clarity achieved by our proposed schemes under similar compression rates.

In Fig. 6(b), we conduct a comparison involving our four proposed AIGC-SCM schemes, the WITT [15] scheme, and the Shannon communication scheme. This comparison is based on the average bandwidth compression ratio ($k/n$), which is defined as the ratio of the average number of transmission symbols to the original information bits. The primary difference between our proposed AIGC-SCM and WITT [15] lies in the fact that the latter persists in utilizing the features of the complete original figure for transmission, while we opt to eliminate extraneous details in the image through image-to-text modality transformation. Additionally, thanks to the use of the conditional generation mechanism and the high-quality generation capabilities of AIGC, this approach can significantly reduce the amount of data transmitted and still reconstruct content with high fidelity and semantic accuracy.

Finally, we assess the similarity performance between the source and destination images. As discussed in the section "How to Evaluate AIGC-SCM Performance," some traditional metrics like PSNR are not suitable for evaluating AIGC-SCM. Hence, we extract three types of feature vectors by feeding the data from the source and the destination to CLIP [9] and calculate their cosine

This not only demonstrated the model's efficiency but also its potential in revolutionizing the way we approach semantic communication in various high-demand applications.

similarity for each type. The corresponding metrics and evaluation process can be summarized as follows:

1. *Image Feature Alignment:* It is measured by calculating the cosine similarity of the whole image features of the source and destination;
2. *Object Feature Alignment:* This metric entails using an object detector to crop the source and destination images and distill only object-related feature vectors. Subsequently, the cosine similarity of these object-related feature vectors is calculated;
3. *Text Feature Alignment:* Given the value of image captions in enhancing image content comprehension, we also gauge the correspondence between the feature vectors of the generated captions of the source and destination.

In Fig. 6(c), we demonstrate the performance of different communication schemes measured using the abovementioned metrics. The highest similarity performance of Shannon communication can be attributed to its utilization of a significantly larger bandwidth resource. Our proposed AIGC-SCM schemes, consisting of four different approaches, outperform the WITT [15] models in terms of image feature and object feature alignment metrics. This superiority is attributed to AIGC's ability to reconstruct high-quality images with minimal information transmission. Even on the text feature alignment metric, our approach outperforms WITT [15] when the bandwidth compression rate is low. Schemes that leverage historical data attain elevated alignment metrics, attributed to the model's acquisition of object texture knowledge, enabling more precise reconstructions upon receiving pertinent prompts. Aside from the WITT_L scheme, text feature alignment scores of other methods significantly outperform traditional semantic alignment metrics. This discrepancy arises because the captioning process distills image content into text, emphasizing salient features as interpreted by the model, yet this abstraction avoids delving into the local details and nuances of the original image. Consequently, images regenerated from captions align with the broad description but miss the precise visual specifics of the originals. Overall, the results illustrate that AIGC-SCM successfully preserves a high degree of semantic similarity while significantly reducing communication bandwidth consumption.

## Future Directions

### Prompt Selection Based on Semantic Noise Estimation

To mitigate the impact of semantic noise on the quality of SemCom, we can devise prompt selection methods based on the estimated semantic noise. In instances where the noise level is high, we can increase the number of prompts for transmission accordingly, such as transmitting both text description and image features. This approach is analogous to enhancing error-correction ability by expanding code length in traditional channel coding.

### KB Design in AIGC-SCM

The contents in the receiver's KB can guide the AIGC-assisted reconstruction process. To conserve transmission bandwidth, historical or less crucial information can be stored in the receiver's KB during the knowledge-sharing process. For instance, in video transmission, storing the object's appearance and background in the first frame enables the prompt transmission of only the changes in movement. Ultimately, the complete scene can be reconstructed by conditioning on both the stored knowledge and the received prompts.

### Minimizing Latency in AIGC-SCM

In environments where immediacy is critical, it is imperative to strive for the reduction of the overall system latency, for which various strategies can be employed. For example, model optimization can streamline both the extraction and generation processes, thereby reducing processing times to the millisecond scale [2], [11]. Additionally, the implementation of predictive prefetching can enable the anticipation of future requests, facilitating their rapid resolution and further improving system responsiveness.

## Conclusion

This article proposed to use the AIGC-SCM to address challenges in communication systems. We introduced an AIGC-SCM architecture based on a three-layer model and explored some AIGC technologies that can enhance SemCom. Our investigation indicated that thanks to its multi-modality compatibility and high-quality generation capabilities, AIGC-SCM shows promise in enhancing communication across multi-user VR/AR/XR games, vehicular networking, remote monitoring, and the provision of personalized services. We showcased the effectiveness of AIGC-SCM through a case study focused on VR image transmission. The findings from this study illuminated AIGC-SCM's capacity to substantially lower bandwidth requirements while maintaining a high level of semantic similarity to the original information. This not only demonstrated the model's efficiency but also its potential in revolutionizing the way we approach semantic communication in various high-demand applications.

## Acknowledgment

The work was supported  in part by the National Key Research and Development Program of China under Grant 2018YFB1801402 and Grant 2018YFB1801403; in part by the National Research Foundation, Singapore and Infocomm Media Development Authority, under its Future Communications Research and Development Programme; in part by the Ministry of Education, Singapore, under its SMU-SUTD Joint Grant 22-SIS-SMU-048; in part by the SUTD Kickstarter Initiative under Grant SKI 20210204; in part by the China Scholarship Council (CSC) under Grant 202306010154; in part by NSF under Grant CNS-2107216, Grant CNS-2128368, Grant

CMMI-2222810, and Grant ECCS-2302469; in part by the U.S. Department of Transportation, Toyota; and in part by the Amazon and Japan Science and Technology Agency (JST) Adopting Sustainable Partnerships for Innovative Research Ecosystem (ASPIRE) under JPMJAP2326.

## References

[1] W. Yang et al., "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 213–250, 1st Quart., 2023.

[2] Y. Cao et al., "A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT," 2023, *arXiv:2303.04226*.

[3] Y. Li et al., "GLIGEN: Open-set grounded text-to-image generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Vancouver, BC, Canada, Jun. 2023, pp. 22–22511.

[4] L. Xia et al., "Generative AI for semantic communication: Architecture, challenges, and outlook," 2023, *arXiv:2308.15483*.

[5] Y. Xie et al., "Diffusion model for generative image denoising," 2023, *arXiv:2302.02398*.

[6] W. Yang et al., "Semantic change driven generative semantic communication framework," 2023, *arXiv:2309.12775*.

[7] J. Wang et al., "A unified framework for guiding generative AI with wireless perception in resource constrained mobile edge networks," 2023, *arXiv:2309.01426*.

[8] C.-A. Brust and J. Denzler, "Not just a matter of semantics: The relationship between visual similarity and semantic similarity," in *Proc. German Conf. Pattern Recognit.*, Dortmund, Germany, Sep. 2019, pp. 414–427.

[9] A. Radford et al., "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Mach. Learn.*, vol. 139, 2021, pp. 8748–8763.

[10] R. Rombach et al., "Highresolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, New Orleans, LA, USA, Jun. 2022, pp. 10684–10695.

[11] G. Jocher, A. Chaurasia, and J. Qiu. (Jan. 2023). *YOLO by Ultralytics*. [Online]. Available: https://github.com/ultralytics/ultralytics

[12] N. Ruiz et al., "DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Vancouver, BC, Canada, Jun. 2023, pp. 22500–22510.

[13] C. Schuhmann et al., "LAION-400M: Open dataset of CLIP-filtered 400 million image-text pairs," 2021, *arXiv:2111.02114*.

[14] T. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, Zurich, Switzerland, 2014, pp. 740–755.

[15] K. Yang et al., "WITT: A wireless image transmission transformer for semantic communications," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Rhodes Island, Greece, Jun. 2023, pp. 1–5.

## Biographies

Yiru Wang (Graduate Student Member, IEEE) (yiruwang@stu.pku.edu.cn) received the B.S. degree from Xidian University, Xi'an, Shaanxi, China, in 2020. She is currently pursuing the Ph.D. degree with the School of Electronics, Peking University, Beijing, China. She is currently a Visiting Ph.D. Student with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore. Her research interests mainly focus on wireless communications and generative artificial intelligence.

Wanting Yang (Member, IEEE) (wanting_yang@sutd.edu.sg) received the B.S. and Ph.D. degrees from the Department of Communications Engineering, Jilin University, Changchun, China, in 2018 and 2023, respectively. She is currently a Research Fellow with the Singapore University of Technology and Design. Her research interests include wireless semantic communication, generative artificial intelligence, learning, martingale, and predictive resource allocation. She served as a Technical Program Committee Member for flagship conferences, such as WCNC, GLOBECOM, and VTC.

Zehui Xiong (Senior Member, IEEE) (zehui_xiong@sutd.edu.sg) received the Ph.D. degree in computer science and engineering from Nanyang Technological University, Singapore. He was a Researcher with the Alibaba-NTU Joint Research Institute, Singapore. He is currently an Assistant Professor with the Singapore University of Technology and Design. His research interests include wireless communications and edge intelligence.

Yuping Zhao (yuping.zhao@pku.edu.cn) received the B.S. and M.S. degrees in electrical engineering from Northern Jiaotong University, Beijing, China, in 1983 and 1986, respectively, and the Ph.D. and Doctor of Science degrees in wireless communications from Aalto University, Espoo, Finland, in 1997 and 1999, respectively. She is currently a Professor with the State Key Laboratory of Advanced Optical Communication Systems and Networks, School of Electronics, Peking University, Beijing. Her current research interests include wireless communications and signal processing.

Tony Q. S. Quek (Fellow, IEEE) (tonyquek@sutd.edu.sg) received the B.E. and M.E. degrees in electrical and electronics engineering from the Tokyo Institute of Technology in 1998 and 2000, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 2008. He is currently the Cheng Tsang Man Chair Professor with the Singapore University of Technology and Design (SUTD) and an ST Engineering Distinguished Professor. He is also the Director of the Future Communications Research and Development Programme, the Head of ISTD Pillar, and the Deputy Director of the SUTD-ZJU IDEA. His current research interests include wireless communications and networking, network intelligence, non-terrestrial networks, open radio access networks, and 6G. He is a fellow of WWRF and the Academy of Engineering Singapore. He was a member of the Technical Program Committee and the symposium chairs in a number of international conferences. He has been actively involved in organizing and chairing sessions. He was honored with the 2008 Philip Yeo Prize for Outstanding Achievement in Research, the 2012 IEEE William R. Bennett Prize, the 2015 SUTD Outstanding Education Awards—Excellence in Research, the 2016 IEEE Signal Processing Society Young Author Best Paper Award, the 2017 CTTC Early Achievement Award, the 2017 IEEE ComSoc AP Outstanding Paper Award, the 2020 IEEE Communications Society Young Author Best Paper Award, the 2020 IEEE Stephen O. Rice Prize, the 2020 Nokia Visiting Professor, and the 2022 IEEE Signal Processing Society Best Paper Award. He is also an Area Editor of IEEE Transactions on Wireless Communications.

Zhu Han (Fellow, IEEE) (hanzhu22@gmail.com) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 1999 and 2003, respectively. From 2000 to 2002, he was a Research and Development Engineer with JDSU, Germantown, MD, USA. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State University, Boise, ID, USA. He is currently a John and Rebecca Moores Professor with the Electrical and Computer Engineering Department and the Computer Science Department, University of Houston, Houston, TX, USA. His main research interests include novel game-theory related concepts critical to enabling efficient and distributive use of wireless networks with limited resources. His other research interests include wireless resource allocation and management, wireless communications and networking, quantum computing, data science, smart grid, carbon neutralization, security, and privacy. He has been an AAAS Fellow since 2019 and an ACM Fellow since 2024. He was an IEEE Communications Society Distinguished Lecturer from 2015 to 2018 and an ACM Distinguished Speaker from 2022 to 2025. He received the NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the Journal on Advances in Signal Processing in 2015, the IEEE Leonard G. Abraham Prize in the field of Communications Systems (Best Paper Award in IEEE Journal on Selected Areas in Communications) in 2016, the IEEE Vehicular Technology Society 2022 Best Land Transportation Paper Award, and several best paper awards in IEEE conferences. He is also the winner of the 2021 IEEE Kiyo Tomiyasu Award (the IEEE Field Award), for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks." He has been a 1% Highly Cited Researcher since 2017, according to Web of Science.