Generative AI-Driven Semantic Communication Networks: Architecture, Technologies, and Applications

Chengsi Liang, *Student Member, IEEE*, Hongyang Du, *Graduate Student Member, IEEE*, Yao Sun, *Senior Member, IEEE*, Dusit Niyato, *Fellow, IEEE*, Jiawen Kang, *Senior Member, IEEE*, Dezong Zhao, *Senior Member, IEEE*, and Muhammad Ali Imran, *Fellow, IEEE*

***Abstract*—Generative artificial intelligence (GAI) has emerged as a rapidly burgeoning field demonstrating significant potential in creating diverse content intelligently and automatically. To support such artificial intelligence-generated content (AIGC) services, future communication systems must fulfill stringent requirements, including high data rates, throughput, and low latency, while efficiently utilizing limited spectrum resources. Semantic communication (SemCom) has been deemed as a revolutionary communication scheme to tackle this challenge by conveying the meaning of messages instead of bit reproduction. GAI algorithms serve as the foundation for enabling intelligent and efficient SemCom systems in terms of model pre-training and fine-tuning, knowledge base construction, and resource allocation. Conversely, SemCom can provide AIGC services with low latency and high reliability due to its ability to perform semantic-aware encoding and compression of data, as well as knowledge- and context-based reasoning. In this survey, we break new ground by investigating the architecture, wireless communication schemes, and network management of GAI-driven SemCom networks. We first introduce a novel architecture for GAI-driven SemCom networks, comprising the data plane, physical infrastructure, and network control plane. In turn, we provide an in-depth analysis of the transceiver design and semantic effectiveness calculation of end-to-end GAI-driven SemCom systems. Subsequently, we present innovative generation level and knowledge management strategies in the proposed networks, including knowledge construction, update, and sharing, ensuring accurate and timely knowledge-based reasoning. Finally, we explore several promising use cases, i.e., autonomous driving, smart cities, and the Metaverse, to provide a comprehensive understanding and future direction of GAI-driven SemCom networks.**

***Index Terms*—Semantic communication, AIGC, generative AI, intelligent wireless networks, knowledge management.**

Manuscript received 22 December 2023; revised 23 May 2024; accepted 19 July 2024. Date of publication 29 July 2024; date of current version 7 February 2025. This work was supported by UK Department for Science, Innovation & Technology (DSIT) Towards Ubiquitous 3D Open Resilient Network (TUDOR) Project. The associate editor coordinating the review of this article and approving it for publication was Z. Qin. *(Corresponding author: Yao Sun.)*

Chengsi Liang, Yao Sun, Dezong Zhao, and Muhammad Ali Imran are with the James Watt School of Engineering, University of Glasgow, G12 8QQ Glasgow, U.K. (e-mail: Chengsi.Liang@glasgow.ac.uk; Yao.Sun@glasgow.ac.uk; Dezong.Zhao@glasgow.ac.uk; Muhammad.Imran@glasgow.ac.uk).

Hongyang Du is with the Department of Electrical and Electronic Engineering, University of Hong Kong, Pok Fu Lam, Hong Kong (e-mail: duhy@eee.hku.hk).

Dusit Niyato is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: dniyato@ntu.edu.sg).

Jiawen Kang is with the School of Automation, Guangdong University of Technology, Guangzhou 510006, China (e-mail: kavinkang@gdut.edu.cn).

Digital Object Identifier 10.1109/TCCN.2024.3435524

I. Introduction

GENERATIVE artificial intelligence (GAI), regarded as one of the most significant advancements in the field of artificial intelligence (AI), has recently achieved remarkable progress in many areas such as natural language processing (NLP) and multimedia content synthesis [1]. Intrinsically, GAI learns from external knowledge, categorizes data, analyzes examples, and comprehends patterns to produce human-like artifacts in digital content creation. The revolution brought by GAI is now reshaping industries and ushering in novel economic markets. From McKinsey's estimation, GAI could add the equivalent of $2.6 trillion to $4.4 trillion annually, significantly increasing the impact of all AI [2].

A. Background

Within the realm of GAI, AI-Generated Content (AIGC), i.e., digital content including text, image, audio and video is generated by machine learning (ML) algorithms automatically, stands as a notable application in information technology field. Recently, many AIGC products with high efficiency and knowledgeability meeting the huge demand from people on data acquisition, have attracted much attention. One of the most significant reasons is that AIGC services are capable to deal with large-scale database in a short time due to the advanced computing ability. For instance, Claude, delivered by Anthropic's GAI, could process 100,000 tokens of text (equal to about 75,000 words in a minute) by May 2023 [3]. Especially, some advanced AIGC services are realized by sophisticated multimodal GAI models which can cope with more than one data formats. A renowned example, ChatGPT-4 [4], allows users to share images and engage in voice conversations. It significantly enriches the user experience, offering a more dynamic and interactive form of communication compared to ChatGPT-3.5's primarily text-based interactions.

In addition to enjoying benefits offered by GAI, it is anticipated that new challenges will come into the landscape of wireless communication networks [5]. To accommodate novel communication scenarios with GAI and AIGC services, the proliferation in traffic demands along with more stringent latency and reliability requirements are inevitable. Whereas, current wireless communication systems continue to operate under the framework of traditional Shannon's theory, that is, regardless of data format, all content is encoded into binary bits via predefined codebooks for transmission [6]. They concentrate solely on bits, neglecting the message's meaning, thus leading to a low bandwidth utilization. Furthermore, these traditional systems focus on improving network performance in terms of data rate, latency, throughput, reliability, etc., which are tightly dependent on the amount of bits rather than content meaning. This rigid scheme is unable to exploit contextual information, customize content and make intelligent decisions based on the knowledge of the communication pairs, thus failing to adapt the knowledge-driven AIGC services [7].

Fortunately, semantic communication (SemCom), has been regarded as a groundbreaking paradigm shift. Compared with conventional communication systems, SemCom focuses on exchanging the meaning of information rather than reproducing the source data. Typically, a transmitter in SemCom begins by extracting the hidden semantics from source data and then adapting the encoding bits to the wireless channel conditions [8], [9]. The information is then transmitted through a wireless channel, with the receiver working to recover the source's meaning aiming to minimize semantic ambiguity. In this process, SemCom only conveys essential semantics and filters out irrelevant information, which leads to significant savings in wireless resource consumption. Moreover, the semantic encoder can personalize content creation based on the individual's background knowledge. Further, aided by the mutual knowledge between the transmitter and receiver, the syntactic errors are corrected via context-based reasoning while recovering received semantics at the receiver. Thereby, SemCom can sustain good performance even under harsh channel conditions.

B. Motivations

Considering the superiority of SemCom, it should be expected as a promising paradigm for AIGC transmission. It is observed that the structure and logic of the AIGC are inherently tied to GAI models, making it possible for meaning inference according to immediate context. This is highly compatible with the framework of SemCom. Meanwhile, SemCom systems enable to handle high-volume data and diverse content types with sustainable resource consumption, sufficing the needs of intricate AIGC services while alleviating Internet strain. By leveraging the knowledge collected from user's history and sensing data, SemCom systems allow more intelligent personalized services.

On the other hand, GAI brings numerous benefits to SemCom design. At its core, the SemCom encoder and decoder are augmented by GAI. Through the generation of context-sensitive, adaptable, and semantically dense content, GAI greatly bolsters SemCom's ability to effectively transmit content. Moreover, GAI can be employed to continuously refine the knowledge bases (KBs) and learning models, guaranteeing that the SemCom system is able to observe network environments and adapts to evolving dynamic network conditions.

Nevertheless, the synthesis of SemCom and GAI in intelligent wireless communication networks inevitably encounters many challenges, including:

- *Challenge 1: How to construct the SemCom systems fusing GAI to process any data format?* The associated semantics are produced and interpreted by GAI through semantic encoder/decoder in SemCom. However, basic data processing falls short of meeting the demands posed by data-intensive AIGC services, necessitating the use of multimodal algorithms to handle diverse data types. Additionally, the computational time and power required for training must be factored in. During transmission, channel encoders and decoders should adaptively compress semantic information based on varying channel conditions.
- *Challenge 2: How to measure the effectiveness of information generated by GAI in SemCom-based networks?* As SemCom emphasizes the conveyed message's meaning rather than transmitted bits, conventional performance indicators derived from Shannon's framework, are not suitable for evaluating SemCom networks [10]. To offer enhanced services, information effectiveness measurement is tied to the achievement of specific objectives and time. Additionally, interactions now incorporate both human-to-machine and machine-to-machine, moving beyond just human-to-human communication. Thus, determining appropriate metrics considering different goals and scenarios, poses another challenge.
- *Challenge 3: How to manage SemCom-based networks with GAI technologies?* The rising ubiquity of ML tools across all network nodes necessitates coordinated management of resources for computation, communication, and control. Significantly, SemCom heavily relies on background knowledge for semantic representation and interpretation. In this context, knowledge can be considered a resource that requires storage and bandwidth costs for construction and sharing. However, devices with limited storage capacity may need to reduce the size of their KBs. Furthermore, finding the right balance between knowledge freshness and update cost is crucial, as updating the KBs too frequently incurs high costs, while updating it too infrequently may result in outdated or stale knowledge. Consequently, the development and implementation of efficient knowledge management strategies are essential in SemCom-based networks, presenting the third challenge.

C. Related Surveys, Contributions and Organization

Several brilliant works are conducted recently as listed in Table I. In terms of SemCom, [12] presents a detailed

TABLE I
SUMMARY OF RELATED SURVEYS VERSUS OUR WORK

References	Contributions	GAI-assisted Information Creation	SemCom-enabled Information Transmission	Information Effectiveness	Resource Allocation	Knowledge Management
[11]	Presents a detailed survey on advancements in SemCom intelligent wireless networks including AI/ML schemes, SemCom architecture, and networking.	×	√	×	×	×
[12]	Provides a comprehensive survey on the implementation of SemCom in 6G and discusses 6G applications in potential SemCom-empowered network architecture.	×	√	√	√	×
[13]	Provides a comprehensive review of recent challenges and results in the field of GAI with applications to mobile telecommunications networks.	√	×	×	×	×
[14]	Provides a comprehensive survey of AIGC that summarizes GAI in terms of techniques and applications.	√	×	×	×	×
[15]	Provides an extensive overview of AIGC, covering its definition, essential conditions, cutting-edge capabilities, and advanced features.	√	×	×	×	×
[16]	Provides a comprehensive survey on the definition, lifecycle, models, and evaluation metrics of AIGC within mobile edge networks through the combined efforts of mobile-edge-cloud communication, computing, and storage infrastructure.	√	√	×	√	×
[17]	Introduces a framework of GAI-assisted SemCom networks that integrates global and local GAI with semantic coding models in a collaborative cloud-edge-mobile design.	√	√	×	×	×
[18]	Introduces a GAI-aided SemCom framework without necessitating joint training, aiming at consuming fewer computational resources and less energy compared to conventional SemCom systems.	√	√	×	√	×
This paper	Proposes a novel framework for GAI-driven SemCom networks and discusses the cooperation of GAI and SemCom in terms of information creation, AIGC transmission, information effectiveness, and network management.	√	√	√	√	√

survey on the recent technological trends in regard to SemCom for intelligent wireless networks. Reference [11] provides a survey on the implementation of SemCom in 6G and discusses potential applications of 6G in SemCom-empowered network architectures. As for GAI and AIGC, [13] provides a review of recent challenges and results in the field of GAI with application to mobile telecommunications networks. References [14], [15] provide surveys of techniques, applications and challenges of AIGC as the exploration of GAI. Furthermore, [16] presents a comprehensive survey on the definition, lifecycle, models, and evaluation metrics of AIGC within mobile edge networks through the combined efforts of mobile-edge-cloud communication, computing, and storage infrastructures. It also bridges GAI technology with SemCom in the discussion on AIGC transmission. Considering the collaboration between GAI and SemCom, the authors in [17] propose a framework of GAI-assisted SemCom network that integrates global and local GAI with semantic coding models in a collaborative cloud-edge-mobile design. Moreover, the authors in [18] propose a GAI-aided SemCom framework without necessitating joint training with a reduction in both computational complexity and energy cost compared to conventional SemCom methods. These two works delve into the detailed frameworks of SemCom networks assisted by GAI, but without extensive discussions on information effectiveness and knowledge management in wireless networks. In this survey, compared with the relevant works in Table I, we present a comprehensive survey and focus specially on the interplay between GAI and SemCom in wireless communication networks involving GAI-assisted information creation, SemCom-enabled information transmission, information effectiveness, resource allocation and knowledge management.

The organization and our main contributions in this paper can be summarized as:

- Section II- We primarily present a GAI-driven SemCom framework for AIGC delivery. It is the first paper to delve into the framework architecture, components, KPIs, and network management approaches of the synthesis of SemCom systems and GAI algorithms.

- Section III- We provide an extensive review of GAI models from the perspectives of unimodal and multimodal, then classify them into text-to-text, vision-to-vision, audio-to-audio, text-to-X (Text2X), X-to-text (X2Text) and voice bots perspectively.
- Section IV- We conduct a comprehensive exploration into the benefits that SemCom can offer for the delivery of AIGC with the introduction of SemCom's mathematical theory and transceiver design.
- Section V- We conduct an investigation on AIGC's information effectiveness for three perspectives: task-oriented systems, age of information (AoI), value of information (VoI) and causal control.
- Section VI- We introduce new architecture and related algorithms for optimizing communication and computing resource allocation and knowledge management, including knowledge construction, update and sharing, to operate and maintain GAI-driven SemCom networks.
- Section VII- We discuss several use cases and explore the benefits that the envisioned systems offer in various scenarios, such as autonomous driving, smart cities, and the Metaverse.

II. Generative AI-Driven Semantic Communication Network

In this section, we first present the basics of GAI and SemCom. In this context, we envision the GAI-driven SemCom networks from three perspectives: data plane, physical infrastructure and network control plane.

A. Basics of GAI and AIGC

GAI employs specific ML algorithms such as variational autoencoder (VAE) and generative adversarial network (GAN) to generate new content that resembles training data [19]. Unlike common ML algorithms that focus on analyzing data to find patterns and make accurate predictions, these generative algorithms learn the underlying patterns and distributions of the training data to generate new samples. Examples of GAI applications include image generation, text generation, music composition, and speech synthesis.

In turn, AIGC leverages the power of GAI models to automatically produce, manipulate, and modify valuable and diverse content [14], [15], [16], [20]. With GAI's capability to process large volumes of data and generate content at near real-time speed, AIGC significantly presents a marked advantage in the era of big data and enhances operations in the environment where immediate responses are crucial, e.g., live chat interactions, real-time gaming and instantaneous content recommendations. Additionally, AIGC services are redefining the boundaries of creativity by synthesizing entirely new content unbounded by the limitations of traditional human cognition. This creative capacity finds significant applications in areas ranging from digital art and music composition to advertising and product design. Furthermore, AIGC services improve user experiences across multiple fields, including healthcare, education, entertainment, and customer service, by generating content tailored to individual preferences, past interactions, and specific contexts.

B. Basics of SemCom

SemCom diverges from conventional Shannon communication by integrating human-like "comprehension" and "reasoning" into the data encoding and decoding processes, rather than striving for precise bit reproduction [21]. It typically prioritizes transmitting the most significant and relevant information extracted from transmitter to receivers given constrained channel bandwidth, instead of focusing on maximizing bit throughput [22]. To be concrete, in semantic encoding, the original messages are first represented into semantic information carrying background knowledge and context-related information [23]. The KB of semantic encoder needs to be shared with that in semantic decoder before transmission to ensure that the receiver can understand received message. Under equivalent background knowledge, the semantic decoding serves as the reverse operation of encoding to interpret received semantic information.

C. GAI-Driven SemCom Network Architecture

Given the basics and features of GAI and SemCom, next, we present a synergistic interaction between GAI algorithms and SemCom networks. Thus, in this section, we present the vision of GAI-driven SemCom network architecture, as depicted in Fig. 1. We illustrate this architecture from the perspectives of physical infrastructure, data plane and network control plane.

1) Physical Infrastructure: Similar to conventional communication networks, the physical infrastructure in GAI-driven SemCom networks consists of multiple wireless terminal devices (TDs), access points (APs), base stations (BSs), edge servers, and central cloud servers [24]. Besides performing conventional functions in communication systems, these entities are armed with additional intelligent techniques to support novel AIGC services. To be specific, TDs, such as smartphones, tablets, and laptops, are equipped with KBs and well-trained GAI models including encoder and decoder modules in SemCom system. Before transmission, TDs upload sensing data, as well as download knowledge and well-trained models through APs and BSs, thus integrating knowledge and updating KBs.

In GAI-driven SemCom networks, the edge nodes, including mobile edge computing (MEC) servers and BSs, enable to pre-train and fine-tune GAI models with the knowledge from themselves, connected TDs and central cloud servers. Then, edge nodes will offload the well-trained models to TDs corresponding to their tasks and environments. Additionally, the edge servers account for managing knowledge sharing and update with optimization of resource consumption (energy, bandwidth, etc.).

Due to the large storage and computing resource of central cloud servers, the large-scale GAI models can be employed and pre-trained. Virtually, most global AIGC services (e.g., ChatGPT) are trained in cloud utilizing the data from many

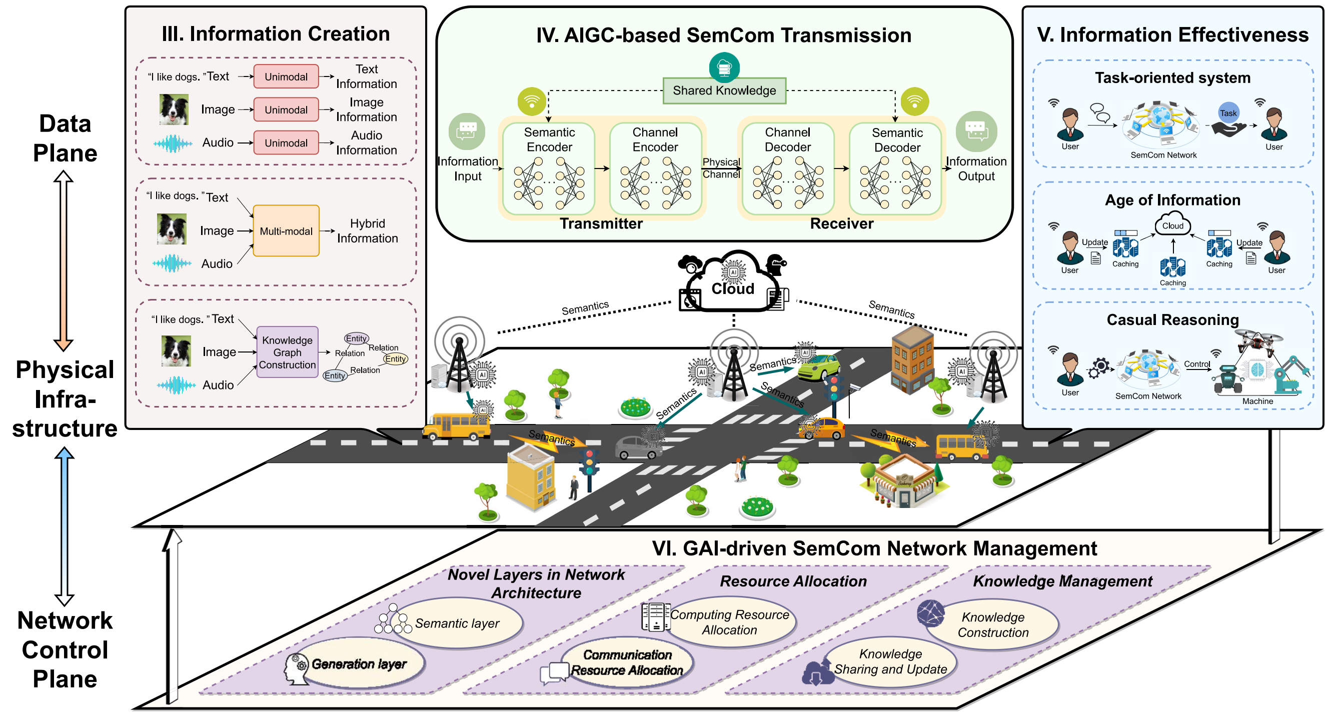


Fig. 1. The architecture of the GAI-driven SemCom networks involving three perspectives: data plane, physical infrastructure and network control plane. The data plane layer includes information creation (Section III), AIGC-based SemCom transmission (Section IV) and information effectiveness (Section V). The network control plane layer includes GAI-driven SemCom network management (Section VI).

data suppliers. Meanwhile, the centralized model will be updated, absorbing new knowledge to refresh models and adjusting resource allocation strategies.

2) Data Plane: The AIGC data is generated, transmitted and evaluated on the data plane of this network. First, the AIGC information is created through GAI models, including unimodal and multimodal models, which will be discussed in Section III. Then, AIGC data is transmitted through a wireless channel in the approach of SemCom. To be concrete, the source messages are fed into semantic encoder and channel encoder at the transmitter to extract and compress their semantic information. Subsequently, the compressed semantic information passes through a wireless channel. At the other end, the distorted data are recovered by the channel decoder and semantic decoder based on the shared knowledge beforehand. Through this approach, SemCom could enhance the efficiency of AIGC transmission and resource utilization by transmitting only essential semantic information of AIGC data are transmitted.

Another function achieved by the data plane is to measure the AIGC information effectiveness from the perspectives of task completion, data freshness and relevance, as well as causal reasoning. First, some performance metrics on evaluating the task implementation for task-oriented systems are delivered [25]. Next, the AoI [26] is regarded as an important metric to measure how fresh the information is, which is significant in real-time supervision system and update system. If the information is expired, it may reduce the accuracy and reliability of system decision. Moreover, the VoI focusing on the importance and relevance of the information being transmitted is also an practicable metric for information effectiveness measurement in SemCom [27]. Finally, due to the dynamics in wireless communication environments, new measurements related to causal reasoning are envisioned considering the state of SemCom networks.

3) Network Control Plane: Unlike conventional communication network, in GAI-driven SemCom networks, the network management should be more intelligent, knowledgeable and adaptive to GAI. Consequently, the network control plane encompasses network architecture, knowledge management, and resource allocation. First, the novel layers in the proposed networks are discussed including semantic level and generation level. Next, the knowledge management features the utilization of KB which are essential in the processes of training both GAI and SemCom models containing public and private knowledge, especially for the personalized function. In this network, the key procedures consist of KB construction, sharing and update. To create a KB, the raw data, such as users' history and channel status, are collected, classified, and encoded. In turn, KBs are continuously monitored by GAI automatically, updated based on new knowledge and users' feedback, ensuring knowledge remains dynamic and reliable over time. Also, KBs in transceivers need to be aligned since the inconsistent KBs would lead to content misunderstanding.

Additionally, since the limited resource restricts the implementation of AIGC services with extensive data, new resource allocation methods for GAI-driven SemCom networks are urgently required. Beyond the conventionally utilized communication resources such as energy and bandwidth, some unprecedented issues are explored for SemCom networks, e.g., the matching degree of physical channel and KB. Furthermore, GAI acts as an add-on module to boost network performance.

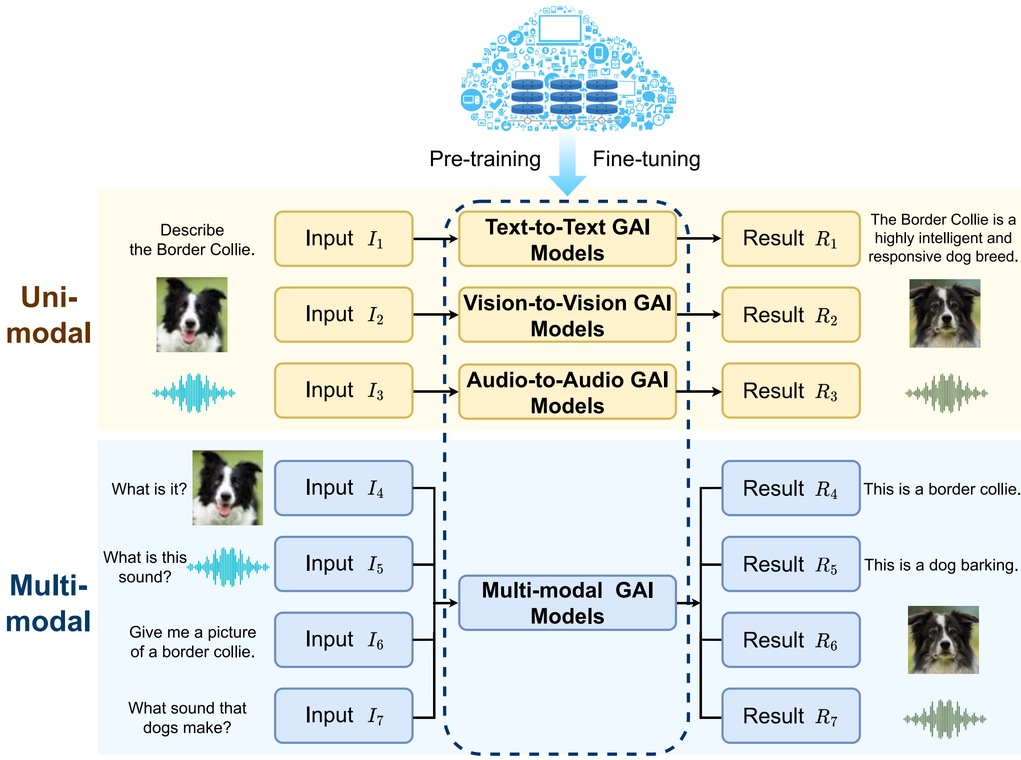


Fig. 2. Two types of GAI models for information creation: unimodal and multimodal. Unimodal GAI models specialize in processing a single type of data, while multimodal GAI models integrate and interpret multiple data types.

The strategies for resource allocation are decided by GAI automatically and they can be adjusted dynamically according to new network status.

III. Information Creation via Generative AI

This section introduces the overview of GAI models for information creation in AIGC with two categories: unimodal and multimodal as depicted in Fig. 2. Unimodal generative models focus on learning the distribution within a single modality and generating new samples that resemble the training data. In contrast, multimodal generative models learn the joint distribution across multiple modalities simultaneously, capturing the relationships and dependencies between them to generate samples that exhibit coherence and alignment among the modalities.

A. Unimodal Models

As shown in Table II, we classify unimodal models based on the type of input data they work with, including text, vision (image and video) and audio.

1) Text-to-Text: Text-to-text GAI models are particularly effective in tasks like text generation, machine translation, summarization and question-answering (QA). These models can be divided into four categories: autoregressive models, VAE-based models, GAN-based models and diffusion-based models.

Autoregressive models, such as Seq2Seq models [28], [29], [30], [31] and Transformer-based models [32], [33], [34], [35], process text sequences and predict the next textual element based on the previously generated text. Particularly, Seq2Seq models are designed to handle variable-length input and output sequences with the architectures such as long short-term memory (LSTM) [36], recurrent neural networks (RNN) [37] and gated recurrent unit (GRU) [38]. They are straightforward to train but tend to generate generic text. Transformer-based models are stemmed from Transformers [39], have become the backbone of many state-of-the-art GAI models (e.g., GPT 2-3 [32], [33], bidirectional encoder representations from Transformers (BERT) [34], and T5 [35]) with their self-attention mechanism, excelling at handling long-range dependencies in text. Although autoregressive models can generate high-quality and coherent outputs, their sequential generation process often results in slower inference times compared to other types of generative models.

VAE-based models [40], [41], [42], [43] function by encoding an input sequence into a fixed-dimensional representation and subsequently decoding it. They provide a probabilistic framework for learning latent representations and can be used for controlled text generation by manipulating the latent space. However, they struggle to capture long-range dependencies in text sequences due to their use of fixed-size latent representations. GAN-based models [44], [45], [46], [47], [48] are featured by GAN which consists of two neural networks, a generator and a discriminator, that compete with each other utilize a game-theoretic approach. Most of these GAN-based models and VAE-based models mentioned before employ RNNs in their generator and discriminator/encoder and decoder to generate text. Though GAN-based models can

TABLE II
OVERVIEW OF STATE-OF-THE-ART AIGC APPLICATIONS AND MODELS

<table>
<tr><th colspan="3">Model Types</th><th>AIGC Applications</th><th>GAI Models</th><th>Model Architectures</th></tr>
<tr><td rowspan="3">Uni-
modal</td><td colspan="2">Text-to-Text</td><td>ChatGPT-3.5 [51],
Bing AI [52],
Megatron-Turing NLG [53],
Claude 3 [54]</td><td>[28]–[31], GPT-2,3 [32], [33],
BERT [34], T5 [35],
[40]–[43], SeqGAN [44], [45], [46],
VGAN [47], TranGAN [48],
Diffusion-LM [49], DiffuSeq [50]</td><td>Autoregressive models,
VAE, GAN,
Diffusion models</td></tr>
<tr><td colspan="2">Vision-to-Vision
(Image-to-Image,
Video-to-Video)</td><td>PaintMe.AI [55],
Vizcom [56],
Steve.AI [57]</td><td>PixelRNN [58], PixelCNN [59],
IntroVAE [60], VQ-VAE-2 [61],
CycleGAN [62], StyleGAN [63],
CVAE-GAN [64], Zero-VAE-GAN [65],
Glow [66], Real NVP [67], DDPM [68],
DDIM [69], MoCoGAN [70], [71]</td><td>Autoregressive models,
VAE, GAN,
Flow-based models,
Diffusion models</td></tr>
<tr><td colspan="2">Audio-to-Audio</td><td>Murf.AI [72],
Resemble.AI [73],
MetaVoice [74]</td><td>WaveNet [75], SampleRNN [76],
GANSynth [77], WaveGAN [78],
SpecGAN [79], VAE-VC [80],
ArchiSound [81]</td><td>Autoregressive models,
VAE, GAN,
Diffusion models</td></tr>
<tr><td rowspan="7">Multi-
modal</td><td rowspan="3">Text2X</td><td>Text-to-Image</td><td>DALL-E 2 [82],
NightCafe [83],
Dream Studio [84]</td><td>DALL-E [85], DALL-E 2 [86],
Magic3D [87], DreamFusion [88],
Imagic [89], Uni-ControlNet [90]</td><td rowspan="3">Autoregressive models,
GAN, VAE,
Flow-based models,
Diffusion models</td></tr>
<tr><td>Text-to-Video</td><td>Synthesia [91],
Pictory [92],
Make-A-Video [93]</td><td>Phenaki [94],
CogVideo [95],
Tune-A-Video [96]</td></tr>
<tr><td>Text-to-Audio</td><td>Murf AI [97],
PlayHT [98]</td><td>Tacotron [99], AudioLM [100]</td></tr>
<tr><td rowspan="3">X2Text</td><td>Image-to-Text</td><td>Transkribus [101]</td><td>VisualGPT [102], ViT [103]</td><td rowspan="3">Transformer, VAE, GAN,
CNN-RNN models</td></tr>
<tr><td>Video-to-Text</td><td>Google Cloud
Video Intelligence API</td><td>UniVL [104], VideoCLIP [105]</td></tr>
<tr><td>Audio-to-Text</td><td>Speak AI [106]</td><td>DeepSpeech [107], wav2vec 2.0 [108]</td></tr>
<tr><td>Voice
Bots</td><td>Speech-to-Text
and
Text-to-Speech</td><td>Siri [109], XiaoIce [110],
Google Assistant [111],
Amazon Alexa [112]</td><td>Pipelines involving ASR,
NLU and NLG models [113]</td><td>Autoregressive models,
GAN, VAE, RL</td></tr>
</table>

generate more diverse and generative content, they may be more challenging to train.

Diffusion-based models, originally designed for image generation, have recently been adapted for text generation tasks [49], [50]. These models learn to generate text by iteratively denoising a sequence of randomly corrupted text samples. At each step, these models learn to recover the original data distribution by estimating the noise that was added and removing it from the corrupted input. However, these models are computationally expensive and have a slower sampling process compared to other generative models.

2) Vision-to-Vision: Vision-to-vision GAI models, which consist of image-to-image and video-to-video models, are utilized in tasks like photo/video editing, medical imaging, and altering facial expressions. Most of these models are built on convolutional neural networks (CNNs) to capture spatial hierarchies and local patterns in images and videos.

In particular, image-to-image models can be broadly classified into five categories: autoregressive models, VAE-based models, GAN-based models, flow-based models and diffusion-based models. Autoregressive models, such as PixelRNN [58] and PixelCNN [59], treat an image as a sequence of pixels, capture the local dependencies and patterns, and predict the value of each pixel based on the previously generated pixels with a lower speed. VAE-based models, such as IntroVAEs [60] and VQ-VAE-2 [61], aim to make the latent vectors of image encoding follow a Gaussian distribution. They allow for parallel image generation, while the images they generated may be blurry and lack sharp details. GAN-based models, like CycleGAN [62] and StyleGAN [63], make the distribution of the generated images increasingly similar to that of the real images. While they have demonstrated the ability to generate clear and realistic images, they still face challenges, including a lack of diversity in the generated outputs and potential instability during the training process. Therefore, some works like CVAE-GAN [64] and Zero-VAE-GAN [65] combine VAE and GAN models for improved generation quality and stable training.

Flow-based models, such as Glow [66] and real NVP [67], generate images by learning an invertible transformation between the data distribution and a known distribution, typically a Gaussian distribution. These models have demonstrated improved quantitative performance as measured by logarithmic likelihood, while they may have limited expressiveness and a reduced ability to capture long-range dependencies in the data. Diffusion-based models, such as denoising diffusion probabilistic models (DDPM) [68] and denoising diffusion implicit models (DDIM) [69], learn to generate a clean image by reversing a gradual noising process. Despite their impressive results in image generation tasks, diffusion-based models have some disadvantages, including high computational complexity, slow inference speed, lack of explicit latent representation and significant memory requirements.

Video-to-video GAI models are designed to tackle various video processing tasks, such as video super-resolution, video inpainting and video denoising. Similar to image-to-image models, video-to-video models can employ autoregressive models, VAE-based, GAN-based, flow-based and diffusion-based models as well. MoCoGAN [70] and the video diffusion model proposed in [71] are two notable examples. MoCoGAN, a GAN-based model, generates videos by decomposing the video generation process into separate motion and content components, whereas the video diffusion model, a diffusion-based approach, produces realistic and diverse video clips from random noise through an iterative denoising process.

3) Audio-to-Audio: Audio-to-audio GAI models can generate new sounds based on input audio for the tasks of music generation, speech synthesis, audio editing, etc. Some of these models work directly with raw audio waveforms or other audio representations. Examples include autoregressive models, such as WaveNet [75] and SampleRNN [76], as well as GAN-based models, like GANSynth [77] and WaveGAN [78]. On the other hand, some audio-to-audio GAI models first transform the audio data into a visual representation, such as a spectrogram or mel-spectrogram. They allow the audio data to be treated as an image, enabling the use of well-established image-based generative models, including GAN, VAEs, and diffusion models. SpecGAN [79], VAE-VC [80], and ArchiSound [81] are examples of GAN, VAE, and diffusion models respectively. By leveraging the success of these image-based models, researchers can generate, manipulate, and analyze audio data in the visual domain before converting the results back into the audio domain.

B. Multimodal Models

The majority of current popular AIGC services like DALL-E 2 are empowered by multimodal GAI models. Compared with unimodal GAI models, multimodal ones are more complex and versatile to process multiple types of input and output. Text data, being simple and interpretable, are often used in multimodal GAI models as textual labels or descriptions to provide supervision for training image and audio models. By leveraging the associations between text and other modalities, multimodal models can learn to generate or manipulate data across different modalities. Thus, we categorize multimodal GAI models into three main categories: text input (text2X), text output (X2text), and voice conversation (voice bots) as shown in Table II.

1) Text-to-X: Text-to-X GAI models transform textual input into diverse output formats, such as images, videos, and audio. Text-to-image models interpret the semantic content of the input text and generate corresponding visual representations. To achieve the complex task of translating text into images, most text-to-image models [85], [86], [89], [90] employ a two-stage architecture that integrates a text understanding model and an image generation model. Particularly, DALL-E [85] and DALL-E 2 [86] leverage Transformer-based architecture to process and understand the input text. However, DALL-E uses a discrete VAE and DALL-E 2 uses a hierarchical vector quantized (VQ)-VAE to generate high-resolution images with fine-grained details. DALL-E 2 also incorporates a diffusion model which is conditioned on the input text using a Contrastive Language-Image Pre-training (CLIP)-like encoder. Besides, Magic3D [87] can create high quality 3D mesh models based on text prompts by improving the design of diffusion models in DreamFusion [88].

Text-to-video models generate dynamic multi-frame videos that include motion and temporal coherence. A popular approach is to generate videos from text sequence-to-sequence, like Phenaki [94], where video tokens are predicted from the paired text embeddings using a bidirectional Transformer architecture. While this approach can generate high-quality video, it requires substantial computing resources. Another approach is to generate sequential images using text-to-image models and connect or interpolate them to generate a video, such as CogVideo [95] and Tune-A-Video [96].

Text-to-audio models learn the mapping between text and audio, allowing them to synthesize natural-sounding speech for any given text input. A famous example is Tacotron [99], which is a Seq2Seq model that maps character embeddings to mel-scale spectrograms, which are then converted to audio using a vocoder like WaveNet or Griffin-Lim. Another example is AudioLM [100], which can produce speech extensions that are both syntactically and semantically coherent through a multi-stage Transformer-based language model.

2) X-to-Text: In contrast to the text-to-X models, X-to-text GAI models enable accurate interpretation and generation of descriptive textual content from diverse input modalities, such as images, videos and audio.

Image-to-text and video-to-text GAI models are trained on the relationship between visual content and its corresponding textual representation. For example, VisualGPT [102] and Vision Transformer (ViT) [103] leverage Transformer-based architecture to generate textual descriptions from image features. Moreover, video-to-text models can capture temporal information from the sequential frames of a video.

UniVL [104] and VideoCLIP [105] are two examples, both of which utilize Transformer-based architectures. UniVL leverages the self-attention mechanism to capture the relationships between video frames and text tokens, while VideoCLIP extends the contrastive learning approach of the CLIP model to the video domain.

Audio-to-text models, also known as speech recognition models, focus on transcribing spoken language into written text. These models can learn the relationship between acoustic features extracted from audio signal and textual description, e.g., DeepSpeech [107] and wav2vec 2.0 [108]. To recognize speech, DeepSpeech employs an RNN-based system that directly maps input audio spectrograms to textual transcriptions, while wav2vec 2.0 utilizes a Transformer-based architecture that learns powerful representations from masked speech input in the latent space.

3) Voice Bots: Voice bots, also known as voice-based chatbots or voice assistants, enable voice conversation in human-computer interaction [114]. Several products have been widely used, like Amazon's Alexa [112], Apple's Siri [109], and Google Assistant [111]. Basically, voice bots interpret user's spoken input through automatic speech recognition (ASR) algorithms, and convert the audio into text. This text is then processed using natural language understanding (NLU) techniques to understand the intent and context behind the user's query. Once understood, voice bots generate a response which converts the text back into human-like speech through natural language generation (NLG) algorithms [113].

IV. Semantic Communication for AIGC Transmission

With the goal of going into SemCom for AIGC delivery, we first present a literature review on the advanced information theory used for SemCom. Subsequently, relevant works on SemCom transceiver design are discussed as per four categories (text, image, audio and video delivery). Finally, we explore the intricacies of GAI and SemCom's interaction and potential implications. Also, through numerical results, we demonstrate the robustness and effectiveness of the proposed system compared to a classical SemCom system and a traditional wireless communication system.

A. Information Theory for SemCom

Information theory plays a crucial role in understanding and modeling communication systems. For SemCom systems, semantic information theory (SIT) provides a mathematical framework to quantify the amount of information being exchanged, assess the capacity of communication channels, and determine the optimal coding schemes to minimize errors and maximize efficiency. The goal of SIT is to quantify, measure, and optimize the semantic content of messages, taking into account the context, purpose, and impact of the communicated information. Recently, SIT has evolved over offering a broader range of perspectives on the core nature of semantic information. The authors in [115] introduce a unique theory on semantic information, emphasizing its distinct position within the information trinity. From a physics standpoint, the work of [116] characterizes semantic information as the syntactic data between a system and its surroundings, which causally aids the system's ongoing operation. The work of [116] later provides a layered interpretation of semantic information across various strata of communication systems, and employs semantic entropy used in [117] for its assessment.

Particularly, central to this theory are the concepts of semantic entropy and semantic-aware channel capacity. Derived from information entropy developed by Shannon, semantic entropy quantifies the amount of semantic information contained in a message, considering factors such as the semantic similarity between symbols or the relevance of the information to the intended receiver [9]. The authors in [118] primarily measure the semantic information using the degree of confirmation. Then, the authors in [119] explain the theory of SemCom and the entropy-based quantification of semantic information. Semantic-aware channel capacity, on the other hand, characterizes the maximum rate at which semantic information can be reliably transmitted over a communication channel, taking into account the semantic noise and distortions that may affect the interpretation of the received message. Also in [119], the authors present the calculation of semantic channel capacity of a discrete memoryless channel.

B. Transceiver Design in SemCom

Transceivers within a SemCom system can be significantly enhanced by GAI models on optimizing the understanding, transmission, and management of information. Essentially, the purpose of semantic transceivers is to extract semantics at the sender's end and restore it at the receiver's end with minimum semantic errors over different channel conditions. Learned from KB, the semantic features are first distilled by semantic encoder, then compressed by channel encoder. After passing through a physical channel, these distorted semantic features are restored by channel decoder and semantic decoder. In recent years, the prevalent architecture of SemCom has undergone significant enhancements and refinements through numerous groundbreaking works that tackle a wide range of tasks. This section categorizes SemCom's transceiver designs based on the type of source data they handle, including text, image, speech, and video.

1) Text Delivery: By leveraging advanced NLP techniques and deep learning (DL) models, SemCom systems can extract and transmit the essential semantic information from the textual tokens. Transformer, a widely adopted architecture in SemCom systems for text delivery, has gained significant popularity due to its ability to effectively capture contextual relationships through the attention mechanism. A notable milestone in the development of DL-based transceiver design for SemCom is called DeepSC [120], which is built on Transformer architecture. This groundbreaking work has inspired numerous variations and advancements and leaded to a proliferation of innovative approaches. The authors in [121] integrate semantic encoding with Reed-Solomon coding, a hybrid automatic repeat request mechanism and a similarity detection network to enhance the reliability of transmitting textual semantics. The authors in [122] also propose a

Transformer-based SemCom system with a new loss function to quantify the impact of semantic distortion, allowing for a dynamic balance between semantic compression loss and semantic accuracy. In order to adapt the trained model for Internet-of-Things (IoT) devices with limited capability, the authors in [123] come up with a lite version of DeepSC (L-DeepSC) through pruning and quantizing the fully trained DeepSC models to achieve as large as 40$\times$ compression ratio without performance degradation.

Different from that KBs in the aforementioned research merely acting as corpora with unprocessed text, the KGs consisting of interconnected entities and their relations, enhance reasoning ability and improve personalization of SemCom. The KG-based SemCom systems are capable of predicting words based on the relationships delineated by KG, rather than solely relying on context, hence enhancing the accuracy of prediction. For example, [124] introduces a KG-driven SemCom system and utilizes Text2KG and KG2Text networks in semantic encoder and semantic decoder. Additionally, [125] delivers a more reliable SemCom system by integrating extracted semantics and KG. Specifically, it aggregates context in KG extraction and semantic restoration, which shows great robustness especially when the channel quality is poor.

Furthermore, recent research has focused on semantic generative communication systems that integrate GAI techniques into SemCom systems for text delivery. In contrast to classical SemCom systems, these integrated approaches leverage pre-trained GAI models and employ prompt processing techniques to achieve superior performance in terms of accuracy and latency. For example, the authors in [126] propose a semantic importance-aware communication scheme using pre-trained language models (e.g., ChatGPT, BERT, etc.). In addition, [127] focuses on utilizing GAI techniques to assist knowledge construction in SemCom.

2) Image Delivery: SemCom systems for image delivery can capture the semantic information from images by learning from the relevance between adjacent pixels. Transformer, CNNs, GANs are usually used for transceiver design to process high-dimensional image data. For example, the authors in [128], [129], [130] leverage GANs and ViT in SemCom systems to enhance the efficiency and quality of image transmission. Moreover, researchers harness adaptive and task-oriented solutions, such as reinforcement learning (RL)-based adaptive semantic coding [131] and unified transmission-classification systems [132]. These approaches aim to optimize the transmission and processing of images based on their semantic content and specific vision-related tasks. The application of SemCom in unmanned aerial vehicle (UAV) scenarios has also gained attention, with studies focusing on task-oriented scene classification [133] and personalized semantic encoding for energy-efficient transmission [134].

Furthermore, cooperating with multimodal GAI models is a promising approach for image SemCom systems. In [135], the authors address a GAN-based SemCom framework to reduce communication overhead and maintain the QoS of emerging applications. The authors in [17] propose a GAI-assisted SemCom network framework in a cloud-edge-mobile design with a case study built on Stable Diffusion for image transmission service. The author in [136] also introduce a diffusion-based SemCom system with multimodal prompts for accurate content decoding.

3) Video Delivery: Video delivery in SemCom systems presents unique challenges compared to image delivery, as it requires maintaining temporal consistency between sequential frames to account for the time dimension. Furthermore, video content allows for reasoning based on action/trajectory logic and behavior patterns, enabling a deeper understanding of the video semantics. Researchers have proposed various approaches to address these challenges and optimize video transmission in SemCom systems. In [137], the authors design a joint source-channel coding strategy that optimizes the trade-off between transmission rate and distortion for over-the-air video transmission. In the context of video conferencing, [138] introduces a novel method that employs a semantic error detector and utilizes a still photo of the speaker as prior information. By leveraging this prior information, the system can reconstruct the speaker's facial expressions more effectively. Another notable contribution is the VISTA framework proposed in [139]. This framework segments each video frame into environment and action parts, extracting and transmitting their semantics separately. By doing so, it reduces redundancy in repeated environmental images and unnecessary action images.

Moreover, researchers have explored novel approaches that go beyond traditional video transmission techniques, such as text-based video editing and transmission. In [140], the authors introduce a method for editing talking-head videos through text modification, allowing for flexible and efficient video manipulation. Similarly, the authors in [141] propose transmitting only text instead of video to substantially relieve network traffic, leveraging the power of NLP to convey video content.

4) Audio Delivery: Utilizing cutting-edge NLP techniques, spoken words can be efficiently converted to text, which can then be channeled into SemCom. However, compared to text, which is purely composed of characters, the intricacies of speech signals make them more challenging to handle. This complexity arises from factors beyond just the fidelity and volume of the speech, encompassing its frequency and tone as well. The authors in [142] introduce a speech-focused variant of DeepSC, called DeepSC-S. The authors in [143] extend DeepSC-S to DeepSC-ST which leverages RNNs to extract and transmit textual semantic content from speech signals. Also, the authors in [144] extend DeepSC-S to accommodate multiple users, deploying federated learning (FL) to collaboratively train a CNN-based encoder and decoder across various local devices and a central server. Additionally, in [145], the authors propose a novel audio SemCom system based on a diffusion model which can simultaneously restore the received information from multiple degradations, including corruption noise and missing parts caused by transmission over the noisy channel.

C. GAI's Role in SemCom

By learning from the aforementioned review, we can observe that GAI technologies can significantly enhance SemCom systems, as illustrated in Fig. 3. Herein, classical

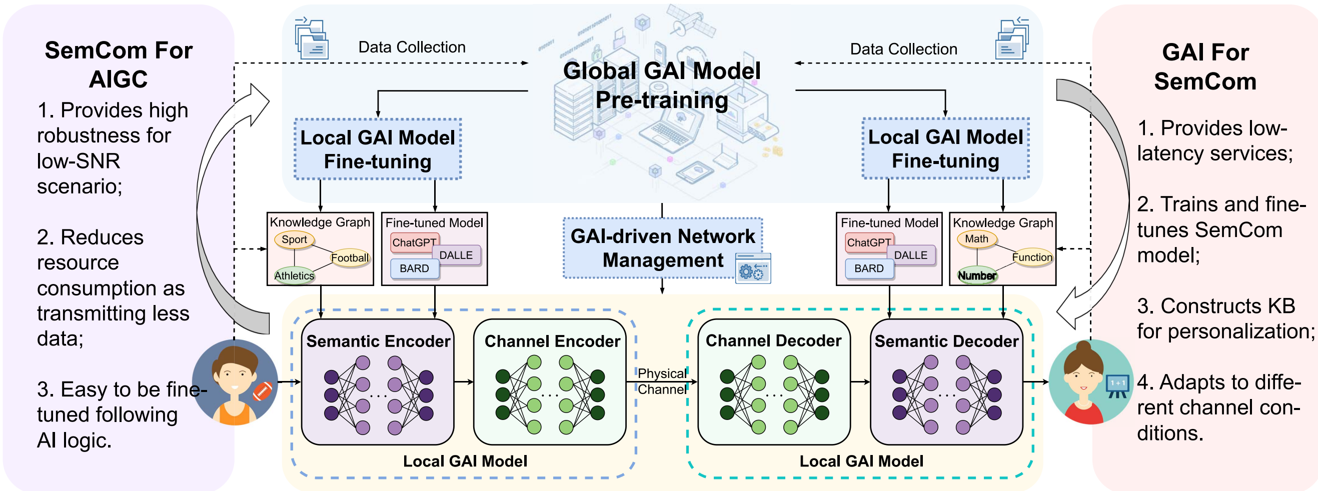


Fig. 3. The synergistic relationship between SemCom and GAI in SemCom-empowered AIGC transmission. The central part of the diagram depicts the framework for AIGC transmission, outlining the process from the cloud server to user devices. The left part highlights the contributions of SemCom to AIGC, and the right part details the advantages GAI offers to SemCom.

SemCom models encompass three important components: semantic content generation and reconstruction module, transceiver's KBs and semantic-aware transmission module. GAI can assist in these components as follow.

- *Semantic Content Generation and Reconstruction:* GAI can significantly benefit the semantic content generation and reconstruction modules due to its ability to generate diverse and personalized content. In semantic content generation, GAI can help train models to better understand and interpret diverse data inputs, converting them into meaningful semantic information. SemCom systems can leverage well-trained GAI models that learn from global knowledge, such as GPT-3 and diffusion models, to generate semantic content. Although this approach may not be well-suited for all tasks, it can greatly reduce the computational resources and time required for pre-training. Moreover, GAI algorithms can be pre-trained within SemCom models, enabling them to learn from vast amounts of global and private data. This approach can generate more accurate semantic data and allow researchers to design specific details in GAI algorithms for diverse tasks, while requiring more computing time. Furthermore, GAI can be used to fine-tune pre-trained SemCom models to specific domains, tasks, or user preferences, enhancing their adaptability and performance.
  The robust learning abilities of GAI models demonstrate that semantic decoders can reconstruct source messages using only a limited amount of semantic information, such as prompts, without requiring joint training with the semantic encoder [136]. Also, GAI enables accurate and personalized content reconstruction to better cater to the unique needs and interests of the receiver by learning from the receiver's personal knowledge and adapting to individual contexts. This approach can enhance user experience, especially for personalized AIGC services.
- *KB Construction:* Since SemCom and GAI models are trained based on background knowledge, the construction of KBs is crucial. Various input data types and sources may utilize diverse data forms, including data corpora, trees, ontologies, and graphs. GAI can structure and extract information from these diverse data sources into a unified form, which is particularly beneficial for multimodal SemCom systems. Additionally, GAI can generate new knowledge by reasoning from existing knowledge, further expanding the available information.
  Moreover, SemCom models require a great amount of data for pre-training. However, some mobile devices may not have the computational resources and storage capacity to learn and store large KBs. In this regard, GAI can enhance knowledge storage methods by optimizing the representation and organization of information, deleting unnecessary knowledge, and highlighting personal knowledge that is most relevant to the user's needs and preferences.
- *Semantic-aware Transmission:* GAI enables the prioritization of the transmission of semantically critical information, adapts transmission strategies based on the semantic content, or routes data based on semantic relevance to the recipients. In addition, GAI can significantly bolster the adaptability of communication channels by enabling dynamic adjustments based on content, context, and user behavior. GAI models analyze real-time data streams to optimize channel performance, adjusting parameters like bandwidth and signal modulation to maintain high-quality transmission under varying conditions. They can predict and preemptively mitigate potential disruptions, ensuring seamless communication. Furthermore, GAI's predictive capabilities allow for anticipatory resource allocation, reducing latency and enhancing the overall user experience.

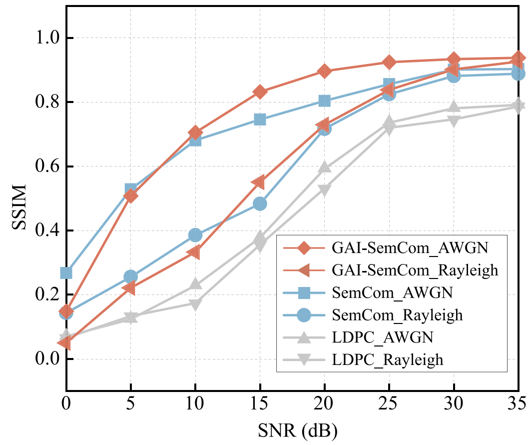


Fig. 4. SSIM performance of images reconstructed by a GAI-driven SemCom system, a classical SemCom system and a traditional wireless communication system versus varying SNRs.

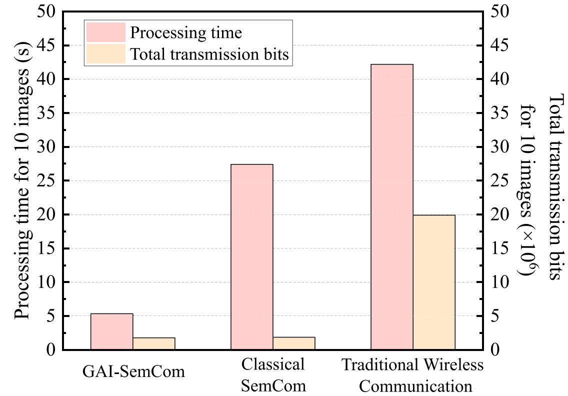


Fig. 5. Processing time and total transmission bits for 10 images of a GAI-driven SemCom system, a classical SemCom system and a traditional wireless communication system.

D. Experiments and Results

We conduct numerical evaluations to demonstrate the performance of GAI-driven SemCom systems, and we implement all subsequent simulations in a computer with an Intel Core i9 CPU and NVIDIA Geforce RTX 3090 Ti GPU processors where the main software environment is Python 3.9. The architecture of this system is learned from the work in [136] which delivers images with multi-model prompts for accurate content decoding. Particularly, this system utilizes multi-modal prompts which incorporate visual prompts to restore images' structural fidelity, and textual prompts to capture the semantic information of images. Transmitting these prompts within the confines of an open wireless environment necessitates robust protective measures. Moreover, this system employs well-trained diffusion model and does not require joint training in pre-training process, which offers a reduction in both computational complexity and energy cost compared to conventional SemCom methods.

As shown in Fig. 4, the GAI-driven SemCom system shows great performance of structural similarity index measure (SSIM) over a wide range of Signal-to-Noise Ratios (SNRs) spanning from 0 to 35 dB, comparing with classical SemCom system [128] and traditional wireless image transmission system using LDPC codes [146]. The channel types are additive white Gaussian noise (AWGN) channel and Rayleigh fading channel, and the channel coding approach adopted is binary phase-shift keying (BPSK). We can observe that the superior performance of the GAI-driven SemCom system is particularly pronounced under varying SNRs from 10 to 35 dB, where the SSIM value rapidly improves, showcasing its exceptional capabilities in favorable conditions. In lower SNR conditions, the generation of prompts is adversely impacted by the presence of noise, leading to a degradation in the quality and accuracy of the reconstructed images. As illustrated in Fig. 5, the GAI-driven SemCom system demonstrates a significant reduction in processing time compared to traditional wireless communication systems. This improvement can be attributed to the utilization of a well-trained diffusion model, which efficiently processes and generates the transmitted data. Moreover, the GAI-driven SemCom system requires the transmission of fewer bits compared to traditional wireless communication systems, highlighting its enhanced efficiency in data compression and transmission. However, this system may transmit slightly more bits compared to classical SemCom systems due to its capability to handle multimodal data transmission, which involves the integration and processing of various data types.

V. INFORMATION EFFECTIVENESS FOR AIGC

Clearly, the conventional network performance metrics, such as throughput, latency/delay, packet loss, and bit error rate (BER), are no longer adaptive for the brand-new GAI-driven SemCom networks which are content oriented and sensitive to data freshness. Thus, how to evaluate information effectiveness for such networks is imperative. In this section, we shed light on new views considering the communication goal, data freshness, and causal reasoning for three varieties of AIGC regimes, i.e., task-oriented systems, AoI, VoI and causal control systems.

1) Task-Oriented Systems: A task-oriented SemCom system is designed to focus not just on transmitting data but ensuring that the communicated data are efficiently used to fulfill a specific task or objective. This approach goes beyond merely delivering information to guarantee that the communicated information is of maximum utility in achieving the recipient's goals. For instance, in a manufacturing setting, a task-oriented SemCom system might be used to transmit machine sensing data. Rather than merely sending raw sensor readings, the system could be designed to only transmit when readings indicate a potential issue, such as a sudden spike in temperature or vibration, that could signify a problem requiring intervention.

Some related works propose various methods for measuring information effectiveness in task-oriented SemCom systems. Focusing on transmitting single and multiple modality data, task-oriented multi-user SemComs with image and text source inputs are explored in [147]. Three intelligent tasks are chosen as representative case studies: image retrieval and machine translation for single-modality data transmission, and the more complex visual QA task for multimodal data transmission.

Moreover, the authors in [134] present an energy-efficient SemCom-based framework for UAVs, incorporating a personalized semantic encoder to selectively transmit images that align with the user's specific interests. The task involves optimizing objectives to enhance resource utilization in the proposed multi-user resource allocation scheme, which is grounded in game theory.

2) Age of Information (AoI) and Value of Information (VoI): AoI is a concept in communication systems that quantifies the freshness or timeliness of information received at the destination as proposed in [26]. AoI can be utilized in some applications that deeply require recipients to receive the most recent status updates to ensure the correctness of the actions taken, such as IoT networks [148], wireless sensor networks [149], cloud gaming [150], etc.

Therefore, AoI can be a useful criteria in SemCom system optimization, considering its emphasis on time sensitivity [12], [151]. The authors in [152] regard AoI as a semantic measure for an age-aware SemCom system from data significance perspective. Furthermore, a new age-related metric named Age of Incorrect Information (AoII) is presented in [153]. In contrast to AoI, AoII represents an advancement by incorporating an information-penalty aspect and a time-related function, thereby enhancing SemCom systems.

Besides, VoI refers to the importance and relevance of the information being transmitted, as proposed in [27], as opposed to merely focusing on the quantity of data. Its emphasis is on ensuring that the information shared is meaningful and useful for the intended purpose or recipient, which is a suitable metric for SemCom [154]. VoI can also be regarded as a nonlinear relationship between the value and freshness denoted by $f(AoI(t))$ at a given time t, where $f(\cdot)$ represent the AoI penalty function describing how VoI evolves as AoI grows [155]. For instance, in [156], the authors utilize VoI to formulate a rate-regulation tradeoff between the packet rate and the regulation cost with an event trigger and a controller in networked control systems.

3) Causal Reasoning: Causal reasoning techniques have gained popularity by providing a structured framework to understand the mechanism of cause-and-effect for reasoning and inference. Central to this approach is causal inference, also known as counterfactual inference, which aims at addressing questions like "What if I had acted differently?". Although causality stands out for its interpretability and capacity to extrapolate data, it is challenging for traditional causal models to deal with high-dimensional, unstructured data. Fortunately, GAI models can approximate complex and high-dimensional data with relative ease, therefore complementing the strengths of causal models. Moreover, SemCom systems ensure that the causal analysis conducted is not just data-driven but meaning-driven with low latency, leading to more accurate and real-time inferences. In addition, the knowledge update and sharing in SemCom networks can to be monitored dynamically through casual reasoning.

Hence, the wireless network state and dynamics are promising to measure information effectiveness for SemCom systems. For instance, the authors in [157] propose an emergent SemCom framework composed of a signaling game for emergent language

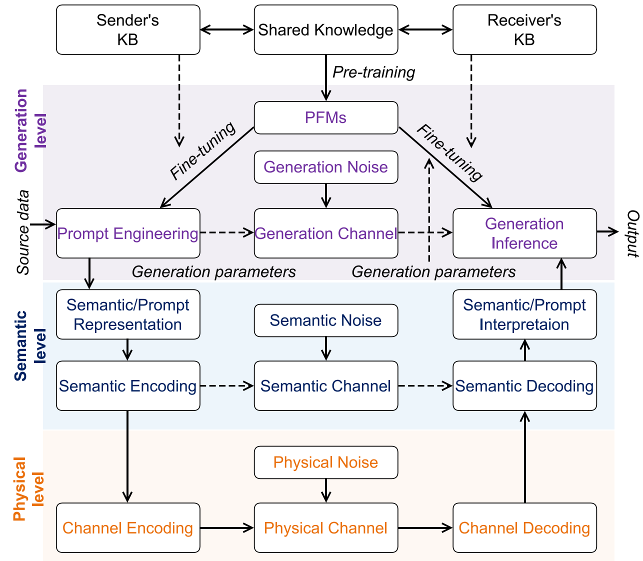


Fig. 6. GAI-driven SemCom networks with physical, semantic, and generation levels [159].

design and a neuro-symbolic AI approach for causal reasoning. They utilize metrics on causal influence within this emergent SemCom framework to capture semantic effectiveness, rather than relying on classical mutual information metrics. In the work of [158], a new information measure for the learned structural causal models at the imitator is introduced in a causal SemCom framework. Also, this work presents a new semantic state abstraction concept based on the dynamic network states, which utilizes the intrinsic information concept from integrated information theory. Furthermore, the GAI architecture and components for "Network State Model" in the proposed framework are discussed.

VI. Generative AI-Driven Semantic Communication Network Management

This section delves into the management of GAI-driven SemCom networks. We first illustrate the novel layers introduced in the network to manage resources from an architecture perspective. Then, we discuss the knowledge management, including knowledge construction and knowledge sharing and update. Finally, we investigate the computing and communication resource allocation strategies.

A. Novel Layers in GAI-Driven SemCom Architecture

To manage the GAI-driven SemCom networks, some research works propose novel layers for dedicatedly dealing with semantic message. A novel GAI-assisted SemCom network framework in a cloud-edge-mobile design is proposed in [17], which enables multimodal semantic content provisioning, semantic-level joint-source-channel coding, and AIGC acquisition. The authors in [12] come up with a two-tier architecture primarily including physical and semantic levels in the semantic-aware network management and communications realm.

Drawing inspiration from previous research, the proposed network architecture, as depicted in Fig. 6, comprises three distinct layers: the *physical layer*, *the semantic layer*, and the newly introduced *generation layer*. The *physical layer* handles the actual data transmission, encompassing the encoding and decoding of signals. The *semantic layer* focuses on the meaning or context of the transmitted information. The innovative *generation layer* utilizes semantic information and algorithmic parameters to guide GAI models, producing content that aligns with specific communication goals. Crucially, knowledge is shared between the sender and receiver beforehand to pre-train the GAI foundation models. The sender employs prompt engineering, based on the sender's KB, to create generation parameters from source data. Once transmitted through a channel, these signals are processed into generation inferences at the receiver end, guided by the receiver's KB, and then interpreted into the final output. This architecture represents a cohesive integration of physical transmission, semantic understanding, and content generation, tailored to enhance communication efficacy for AIGC services.

Moreover, the pre-trained foundation models (PFMs) can be fine-tuned for the specific tasks such as feature extraction and parameter optimization, to provide personalized services and meet the unique demands of various applications [159]. The introduced novel layers also restrict the exposure of sensitive information as only semantic instructions and prompts are transmitted. Hence, the integration of GAI into SemCom models is envisioned to herald a new era of unparalleled personalization, adaptability, and security.

B. Knowledge Management in GAI-Driven SemCom Networks

Knowledge, regarded as the foundation of GAI, comprises two categories as follow:

- *Background Knowledge:* Task-specific parameters at the transmitter end and required expertise for model fine-tuning at the receiver end form the components of this tier.
- *Common Knowledge:* A shared database enables both the transmitter and receiver to pull relevant data, facilitating the use of PFMs for further refinements.

In this sense, knowledge management is significant in the proposed networks since GAI relies on accumulative knowledge for learning, while the network hinges on constant knowledge sharing and updating for seamless operation [160], [161]. To be concrete, knowledge is first constructed from raw data, then shared and updated between each communication entity.

1) Knowledge Construction: The repository of knowledge, KB, contains a variety of data, e.g., model parameters, sensing data, knowledge graph (KG), as well as comprehensive corpuses of text, image, audio, and video. Especially, the KG sophisticates the data framework by weaving in entity and relation triples, thereby enhancing the structure of the stored information. Prominent examples of KGs, such as Freebase [162], DBpedia [163], Wikidata [164], and Google's Knowledge Graph [165], have bolstered GAI's deductive prowess requiring a high degree of personalization, like recommendation engines and QA systems [166], [167], [168].

In terms of knowledge construction in GAI-driven SemCom networks, KBs are compiled from an amalgamation of public and proprietary data sources, processed through GAI algorithms. These sources are diverse, ranging from crowdsourced content to data marketplaces, from the input of IoT sensors to passive collections, as well as encompassing user histories and records [169]. This expansive data assimilation is vital for the effective operation of the proposed networks.

Besides, the KG creation is more complex, which has three fundamental processes: knowledge extraction, knowledge representation learning (KRL), and knowledge graph completion (KGC) [168]. The initial stage in the knowledge management process involves leveraging algorithms for named entity recognition (NER) and relation extraction to distill valuable entities and their connections from unstructured data, forming a network of triples. These triples consist of head entities, relations, and tail entities, denoted as (h, r, t) and organized by the resource description framework (RDF) [170]. GAI algorithms then employ KRL to convert these triples into compact, low-dimensional vectors, rendering complex knowledge into a machine-interpretable format. Finally, KGC algorithms are responsible for inferring and inserting the missing pieces within these triples via triple and relation based reasoning to ensure data integrity and completeness [171].

2) Knowledge Sharing and Update: After collecting the knowledge in various communication nodes, knowledge sharing and update are crucial for maintaining high accuracy and relevance of knowledge in SemCom systems. The processes of knowledge sharing and updating ensure that GAI's decisions are created from the latest data, fostering efficiency and innovation. Regularly refreshing KBs is vital to enable quick adaptation to new market trends, technologies, and user demands. Especially in customer-centric services, it enhances personalization and dedicated user experience.

- *Knowledge Sharing:* Sharing among edge nodes enables collective learning and cooperative knowledge creation, often facilitated by methods like FL [172], [173], [174]. Additionally, a specific application for knowledge sharing in the context of the Industrial Internet of Things is presented via edge GAI platforms [172].
- *Knowledge Update:* To sustain the accuracy and relevance of the KB community, periodic audits are employed to identify and excise outdated or incorrect data, while concurrently integrating new research and insights [175], [176], [177]. Tracking KB versions periodically is advisable to streamline the management of these updates. Such a system allows for the archiving of significant updates as separate versions, providing users the flexibility to compare changes and revert to prior versions if needed [172].

Through these multifaceted strategies, the KB community maintains high integrity, adaptability, and utility, thereby serving as a robust asset in GAI-driven SemCom networks.

C. Resource Allocation in GAI-Driven SemCom Networks

Resource allocation strategies in GAI-driven SemCom networks prioritize knowledge-related metrics, focusing on the intelligent use of computing and communication resources such as bandwidth, spectrum, and energy. This section reviews intelligent and strategic approaches to managing network resources for these networks.

1) Computing Resource Allocation: In GAI-driven SemCom networks, the efficient allocation of computing resources is crucial for ensuring optimal performance and scalability. SemCom and GAI models often require significant computational power for training and inference. To accommodate these resource demands, advanced strategies for computing resource allocation, such as distributed computing and edge computing, must be employed [178], [179]. By leveraging distributed computing architectures, SemCom and GAI models' workloads can be parallelized and processed across multiple nodes, enabling faster training and inference times. Edge computing approaches bring computation closer to the data sources, reducing latency and enabling real-time processing of SemCom tasks. The authors in [21] discuss how the integration of edge intelligence techniques in SemCom networks can lead to significant improvements in the efficient allocation and utilization of computing resources. Moreover, the authors in [180] propose a semantic-aware joint communication and computation resource allocation framework for MEC systems. In particular, the computing resource allocation problem is constructed as an optimization framework aiming to acquire the optimal local computing rate and uploading power, remote computing capacity, and semantic extraction factors.

2) Communication Resource Allocation: In GAI-driven SemCom networks, the communication resources including bandwidth and power are limited and must be effectively managed to support the transmission of AIGC. However, broadcasting large SemCom and GAI models and knowledge sharing can lead to high communication overhead and potential congestion due to the vast number of parameters involved. Therefore, these network require more intelligent and semantic-aware communication resource allocation strategies to solve this issue.

The first feasible approach is employing semantic-aware scheduling and dynamic modulation, which can prioritize the transmission of semantically critical information, ensuring that the most relevant and time-sensitive data are transmitted with minimal delay. For example, the authors in [181] propose an optimal allocation scheme that outperforms both the average allocation scheme and the confidence-based allocation scheme. Particularly, the confidence is generated by the object detector to quantify the importance of each piece of semantic information. Furthermore, some works utilize a knowledge-matching degree as a new metric to evaluate the relevance of transceiver's KBs in SemCom systems. In [10], the authors develop a bit-to-message transformation function based on the specific knowledge-matching degree between transceivers for the first time to optimize resource management in SemCom-enabled cellular networks. Besides, the authors in [169] propose an efficient and low-latency semantic service provisioning solution in SemCom-enabled vehicular networks for the knowledge-matching problem between pairing vehicles.

Moreover, dynamic modulation techniques enable the dynamic adjustment of transmission parameters based on the current channel conditions and the semantic requirements of the data. Particularly, scheduling the broadcasting of model parameters and knowledge during off-peak hours when network traffic is lower is a promising approach to reduce network congestion. Some previous works have explored feasible communication allocation strategies for GAI-driven SemCom systems. For example, the authors in [182], [183] introduce a generative diffusion model (GDM)-based resource allocation scheme for SemCom-aided AIGC services. They construct the bandwidth allocation problem considering the time and data rate of the generation and transmission of semantic information, AIGC content, and rendering results. Also, the authors in [184] construct a near-optimal resource allocation strategy based on the diffusion model for performing semantic extraction, content generation, and graphic rendering in the Metaverse. Additionally, the authors in [185] propose a dynamic resource allocation scheme for the task-oriented SemCom network based on DRL.

VII. CASE STUDY

In this section, we conceive important cases for GAI-driven SemCom networks, including autonomous driving, smart cities, and the Metaverse.

A. Autonomous Driving

In the realm of autonomous driving, autonomous vehicles (AVs) need to actively gather sensing data and swiftly analyze the data to form a perception of their surrounding environment. However, the data collection and transmission processes for AVs are often cumbersome and expensive [186], [187], [188]. To this end, several prominent studies [189], [190] have been conducted on SemCom systems for autonomous driving. Reference [189] focuses on knowledge sharing strategy to improve driving decisions in autonomous vehicle systems. Reference [190] presents a high altitude platform (HAP)-supported fully connected AV network where the traffic infrastructure (TI) transmits its semantic information to the macro BS whenever it observes a connected AV.

Looking ahead, GAI-driven SemCom systems have the potential to revolutionize autonomous driving by enabling AVs to exchange semantic information with other nodes swiftly and efficiently. By generating semantic data between communication nodes, the latency of data transmission can be significantly reduced, enabling real-time decision-making in dynamic driving environments. However, the limited communication bandwidth and on-board processing capacity in connected and autonomous vehicles (CAVs) pose a critical challenge in terms of resource management and competition. To address this issue, advanced AI algorithms can be employed to intelligently allocate resources based on the specific requirements of each AV and the overall traffic situation. By considering factors such as the criticality and time-sensitivity of the semantic data being exchanged, these networks can ensure that safety-critical information, such as collision warnings or sudden changes in road conditions, is given the highest priority and allocated the necessary resources to guarantee minimal latency and maximum reliability in transmission. One potential solution to tackle the resource management challenge is the development of a collaborative multi-objective optimization framework that takes into account various performance metrics related to driving safety, vehicle string stability, and road traffic throughput.

B. Smart City

Smart cities represent intricate socio-technical networks made up of various interrelated components like IoT devices, mobile phones, other portable devices, physical infrastructures, services, applications, and the data shared among these elements [191]. The high complexity of smart city networks comes from dealing with numerous data, diverse content types, distributed control systems, and the intricate interconnections between various urban subsystems spanning physical, digital, organizational, and societal spheres [192]. Some existing works [193], [194] utilize ML algorithms and semantic models to handle smart city issues. The authors in [194] focus on the application of semantic technologies that can enhance interoperability among Internet-of-Everything components in smart cities. In [193], a smart city digital twin architecture is introduced, which facilitates the representation and reasoning of semantic knowledge.

Compared with these works, GAI-driven SemCom networks can develop smart cities by embedding multimodal sensing data into semantic space, connecting them with the semantics. The scalability of these networks allows them to be applied to a wide range of tasks, benefiting multiple domains within the smart city ecosystem. These network can reduce the cost of training different systems while meeting the diverse requirements for different tasks. For instance, in the domain of transportation, GAI-driven SemCom networks can significantly enhance efficiency by optimizing traffic flow, reducing congestion, and improving public transit services. By analyzing real-time traffic data, these networks can dynamically adjust traffic signal timings, recommend optimal routes to drivers, and predict and mitigate potential bottlenecks. Energy conservation is another domain where GAI-driven SemCom networks can be employed. By leveraging semantic data from smart meters, weather sensors, and building management systems, these networks can optimize energy consumption at both the individual building and city-wide levels.

C. Metaverse

Metaverse is a collective virtual shared space, emerging from the fusion of enhanced virtual depictions of our physical world and persistent digital realms. Advancements in GAI technologies have led to a significant increase in Metaverse applications, notably in augmented reality (AR), virtual reality (VR), and extended reality (XR) [195], [196], [197], [198]. Both academia and industry are exploring the Metaverse to create immersive, dynamic virtual landscapes that can adapt in real-time, reflecting user interactions and inclinations. However, to authentically mirror our physical world within these virtual domains, vast amounts of data, spanning text, images, and videos are essential. Recent studies [184], [199], [200] have started to explore potential solutions to this problem through GAI-driven SemCom systems. These frameworks utilize diffusion models [184], Magic3D [199], and GANs [200] to generate digital content, render graphics, and exchange semantic information between transceivers' local semantic multiverses. Additionally, a trustworthy SemCom system using FL and intelligent radio is conceived for privacy protection [200].

Furthermore, user experience quality is of utmost importance in Metaverse applications. Future research can focus on designing and employing GAI-driven SemCom systems to generate and transmit personalized content, enabling tailored services for individual users. By analyzing user preferences, behavior patterns, and contextual information, these networks can infer the content that users are most interested in and adapt their models accordingly to meet users' specific requirements.

VIII. Conclusion

In this survey, we have investigated the integration of GAI with SemCom, leading to GAI-driven SemCom networks. We have detailed the fundamentals of GAI and SemCom, emphasizing their synergy and its impact on AIGC services. We have also delved into network management, covering aspects of novel layers, knowledge management, and resource allocation. The comprehensive exploration in this work not only underscores the revolutionary potential of GAI-driven SemCom networks in various domains but also highlights the emerging trends and future directions in wireless communications. The insights gained from this survey could pave the way for innovative solutions in enhancing network efficiency and user experience. We further explored practical applications in fields of autonomous driving, smart cities, and the Metaverse, showcasing the real-world potential of these technologies in GAI-driven SemCom networks.

References

[1] M. Jovanovic and M. Campbell, "Generative artificial intelligence: Trends and prospects," *Computer*, vol. 55, no. 10, pp. 107–112, Oct. 2022.
[2] M. Chui et al., "The economic potential of generative AI: The next productivity frontier." Jun. 2023. [Online]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-AI-the-next-productivity-frontier#key-insights
[3] Anthropic PBC, "Video AI," Accessed: Jul. 19, 2023. [Online]. Available: https://www.promptengineering.org/claudes-100k-context-window-how-this-will-transform-business-education-and-research/
[4] "ChatGPT can now see, hear, and speak." OpenAI. Accessed: Dec. 9, 2023. [Online]. Available: https://openai.com/blog/chatgpt-can-now-see-hear-and-speak
[5] H. Du et al., "Enabling AI-generated content (AIGC) services in wireless edge networks," 2023, *arXiv:2301.03220*.
[6] D. Gündüz et al., "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 5–41, Jan. 2023.
[7] R. Cheng, Y. Sun, D. Niyato, L. Zhang, L. Zhang, and M. A. Imran, "A wireless AI-generated content (AIGC) provisioning framework empowered by semantic communication," 2023, *arXiv:2310.17705*.
[8] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 210–219, Feb. 2022.
[9] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic communications: Principles and challenges," 2021, *arXiv:2201.01389*.

[10] L. Xia, Y. Sun, D. Niyato, X. Li, and M. A. Imran, "Joint user association and bandwidth allocation in semantic communication networks," *IEEE Trans. Veh. Technol.*, vol. 73, no. 2, pp. 2699–2711, Feb. 2024.
[11] S. Iyer et al., "A survey on semantic communications for intelligent wireless networks," *Wireless Pers. Commun.*, vol. 129, no. 1, pp. 569–611, 2023.
[12] 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.
[13] A. Karapantelakis, P. Alizadeh, A. Alabassi, K. Dey, and A. Nikou, "Generative AI in mobile networks: A survey," *Ann. Telecommun.*, vol. 79, pp. 1–19, Aug. 2023.
[14] 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*.
[15] J. Wu, W. Gan, Z. Chen, S. Wan, and H. Lin, "AI-generated content (AIGC): A survey," 2023, *arXiv:2304.06632*.
[16] M. Xu et al., "Unleashing the power of edge-cloud generative AI in mobile networks: A survey of AIGC services," 2023, *arXiv:2303.16129*.
[17] L. Xia, Y. Sun, C. Liang, L. Zhang, M. A. Imran, and D. Niyato, "Generative AI for semantic communication: Architecture, challenges, and outlook," 2023, *arXiv:2308.15483*.
[18] H. Du et al., "Generative AI-aided joint training-free secure semantic communications via multi-modal prompts," 2023, *arXiv:2309.02616*.
[19] G. Harshvardhan, M. K. Gourisaria, M. Pandey, and S. S. Rautaray, "A comprehensive survey and analysis of generative models in machine learning," *Comput. Sci. Rev.*, vol. 38, 2020, Art. no. 100285.
[20] Y. Li, L. Feng, J. Xu, T. Zhang, Y. Liao, and J. Li, "Full-reference and no-reference quality assessment for compressed user-generated content videos," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, 2021, pp. 1–6.
[21] W. Yang et al., "Semantic communication meets edge intelligence," *IEEE Wireless Commun.*, vol. 29, no. 5, pp. 28–35, Oct. 2022.
[22] P. Basu, J. Bao, M. Dean, and J. Hendler, "Preserving quality of information by using semantic relationships," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, 2012, pp. 58–63.
[23] F. Zhao, Y. Sun, R. Cheng, and M. A. Imran, "Background knowledge aware semantic coding model selection," in *Proc. IEEE 22nd Int. Conf. Communication Technol. (ICCT)*, 2022, pp. 84–89.
[24] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.
[25] C. Cai, X. Yuan, and Y.-J. A. Zhang, "Multi-device task-oriented communication via maximal coding rate reduction," 2023, *arXiv:2309.02888*.
[26] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE INFOCOM*, 2012, pp. 2731–2735.
[27] R. A. Howard, "Information value theory," *IEEE Trans. Syst. Sci. Cybern.*, vol. 2, no. 1, pp. 22–26, Aug. 1966.
[28] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.
[29] A. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," 2017, *arXiv:1704.04368*.
[30] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.
[31] O. Vinyals and Q. Le, "A neural conversational model," 2015, *arXiv:1506.05869*.
[32] A. Radford et al., "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, no. 8, p. 9, 2019.
[33] T. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1877–1901.
[34] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
[35] C. Raffel et al., "Exploring the limits of transfer learning with a unified text-to-text transformer," *J. Mach. Learn. Res.*, vol. 21, no. 1, pp. 5485–5551, 2020.
[36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[37] T. Mikolov, M. Karafiát, L. Burget, J. Cernockỳ, and S. Khudanpur, "Recurrent neural network based language model," in *Proc. Interspeech*, vol. 2, 2010, pp. 1045–1048.
[38] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, 2017, pp. 1597–1600.
[39] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–15.
[40] L. Fang, C. Li, J. Gao, W. Dong, and C. Chen, "Implicit deep latent variable models for text generation," 2019, *arXiv:1908.11527*,
[41] W. Wang et al., "Topic-guided variational autoencoders for text generation," 2019, *arXiv:1903.07137*.
[42] X. Zhang, Y. Yang, S. Yuan, D. Shen, and L. Carin, "Syntax-infused variational autoencoder for text generation," 2019, *arXiv:1906.02181*.
[43] S. Dai, Z. Gan, Y. Cheng, C. Tao, L. Carin, and J. Liu, "APo-VAE: Text generation in hyperbolic space," 2020, *arXiv:2005.00054*.
[44] L. Yu, W. Zhang, J. Wang, and Y. Yu, "SeqGAN: Sequence generative adversarial nets with policy gradient," in *Proc. AAAI Conf. Artif. Intell.*, vol. 31, 2017, pp. 1–11.
[45] Y. Zhang et al., "Adversarial feature matching for text generation," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 4006–4015.
[46] J. Guo, S. Lu, H. Cai, W. Zhang, Y. Yu, and J. Wang, "Long text generation via adversarial training with leaked information," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, 2018, pp. 1–14.
[47] H. Wang, Z. Qin, and T. Wan, "Text generation based on generative adversarial nets with latent variable," in *Proc. Adv. Knowl. Disc. Data Min. 22nd Pacific–Asia Conf.*, 2018, pp. 92–103.
[48] C. Zhang, C. Xiong, and L. Wang, "A research on generative adversarial networks applied to text generation," in *Proc. 14th Int. Conf. Comput. Sci. Educ. (ICCSE)*, 2019, pp. 913–917.
[49] X. Li, J. Thickstun, I. Gulrajani, P. S. Liang, and T. B. Hashimoto, "Diffusion-LM improves controllable text generation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 4328–4343.
[50] S. Gong, M. Li, J. Feng, Z. Wu, and L. Kong, "DiffuSeq: Sequence to sequence text generation with diffusion models," 2022, *arXiv:2210.08933*.
[51] "ChatGPT: Optimizing language models for dialogue." OpenAI. Accessed: Jul. 13, 2023. [Online]. Available: https://openai.com/blog/chatgpt/
[52] (Microsoft, Redmond WA, USA). *Reinventing Search With a New AI-Powered Microsoft Bing and Edge, Your Copilot for the Web*. Accessed: Jul. 13, 2023. [Online]. Available: https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/
[53] S. Smith et al., "Using DeepSpeed and megatron to train megatron-turing NLG 530B, a large-scale generative language model," 2022, *arXiv:2201.11990*.
[54] "Introducing the next generation of claude." Anthropic. Accessed: Mar. 4, 2024. [Online]. Available: https://www.anthropic.com/news/claude-3-family
[55] "PaintMe AI: Home." PaintMe. Accessed: Jul. 19, 2023. [Online]. Available: https://www.paintme.ai/index.php
[56] "The next generation of product visualization." Vizcom. Accessed: Jul. 19, 2023. [Online]. Available: https://www.vizcom.ai/
[57] "Convert photos to videos instantly using steve.AI." Steve.AI. Accessed: Jul. 19, 2023. [Online]. Available: https://www.steve.ai/photo-video-maker
[58] A. Van Den Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1747–1756.
[59] A. Van den Oord, N. Kalchbrenner, L. Espeholt, O. Vinyals, A. Graves, and K. Kavukcuoglu, "Conditional image generation with PixelCNN decoders," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–13.
[60] H. Huang, R. He, Z. Sun, T. Tan, and Z. Li, "IntroVAE: Introspective variational Autoencoders for photographic image synthesis," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–12.
[61] A. Razavi, A. Van den Oord, and O. Vinyals, "Generating diverse high-fidelity images with VQ-VAE-2," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–15.
[62] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.
[63] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4401–4410.
[64] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, "CVAE-GAN: Fine-grained image generation through asymmetric training," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2745–2754.
[65] R. Gao et al., "Zero-VAE-GAN: Generating unseen features for generalized and transductive zero-shot learning," *IEEE Trans. Image Process.*, vol. 29, pp. 3665–3680, 2020.

[66] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1x1 convolutions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–15.
[67] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real NVP," 2016, *arXiv:1605.08803*.
[68] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.
[69] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," 2020, *arXiv:2010.02502*.
[70] S. Tulyakov, M.-Y. Liu, X. Yang, and J. Kautz, "MoCoGAN: Decomposing motion and content for video generation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1526–1535.
[71] J. Ho, T. Salimans, A. Gritsenko, W. Chan, M. Norouzi, and D. J. Fleet, "Video diffusion models," 2022, *arXiv:2204.03458*.
[72] "Transform your Voiceover from a home recording to a professional AI voice." Murf.AI. Accessed: Jul. 19, 2023. [Online]. Available: https://murf.ai/voice-changer
[73] "Real-time speech-to-speech voice conversion." Resemble.AI. Accessed: Jul. 19, 2023. [Online]. Available: https://www.resemble.ai/speech-to-speech/
[74] "MetaVoice." Accessed: Jul. 19, 2023. [Online]. Available: https://themetavoice.xyz/
[75] A. v. D. Oord et al., "WaveNet: A generative model for raw audio," 2016, *arXiv:1609.03499*.
[76] S. Mehri et al., "SampleRNN: An unconditional end-to-end neural audio generation model," 2016, *arXiv:1612.07837*.
[77] J. Engel, K. K. Agrawal, S. Chen, I. Gulrajani, C. Donahue, and A. Roberts, "GANSynth: Adversarial neural audio synthesis," 2019, *arXiv:1902.08710*.
[78] C. Donahue, J. McAuley, and M. Puckette, "Adversarial audio synthesis," 2018, *arXiv:1802.04208*.
[79] Y. Meng, W. Li, S. Lei, Z. Zou, and Z. Shi, "Large-factor super-resolution of remote sensing images with spectra-guided generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022.
[80] K. Akuzawa, K. Onishi, K. Takiguchi, K. Mametani, and K. Mori, "Conditional deep hierarchical variational autoencoder for voice conversion," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, 2021, pp. 808–813.
[81] F. Schneider, "ArchiSound: Audio generation with diffusion," 2023, *arXiv:2301.13267*.
[82] "DALL·E 2: Exploring the limits of data-driven image generation with transformers." OpenAI. Accessed: Jul. 19, 2023. [Online]. Available: https://openai.com/dall-e-2
[83] "NightCafe studio—AI art generator." NightCafe Studio. Accessed: Jul. 19, 2023. [Online]. Available: https://creator.nightcafe.studio/
[84] "DreamStudio—Image generator." DreamStudio. Accessed: Jul. 19, 2023. [Online]. Available: https://dreamstudio.ai/generate
[85] A. Ramesh et al., "Zero-shot text-to-image generation," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 8821–8831.
[86] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with CLIP latents," 2022, *arXiv:2204.06125*.
[87] C.-H. Lin et al., "Magic3D: High-resolution text-to-3D content creation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 300–309.
[88] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, "DreamFusion: Text-to-3D using 2D diffusion," 2022, *arXiv:2209.14988*.
[89] B. Kawar et al., "Imagic: Text-based real image editing with diffusion models," 2023.
[90] S. Zhao et al., "Uni-ControlNet: All-in-one control to text-to-image diffusion models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2024, pp. 1–23.
[91] "Create professional videos without Mics, cameras, or actors." Synthesia. Accessed: Jul. 19, 2023. [Online]. Available: https://www.synthesia.io/
[92] "Script to video creation in minutes." Pictory. Accessed: Jul. 19, 2023. [Online]. Available: https://pictory.ai/pictory-features/script-to-video
[93] "Meta AI," Accessed: Jul. 19, 2023. [Online]. Available: https://makeavideo.studio/
[94] R. Villegas et al., "Phenaki: Variable length video generation from open domain textual description," 2022, *arXiv:2210.02399*.
[95] W. Hong, M. Ding, W. Zheng, X. Liu, and J. Tang, "CogVideo: Large-scale pretraining for text-to-video generation via transformers," 2022, *arXiv:2205.15868*.
[96] J. Z. Wu et al., "Tune-a-video: One-shot tuning of image diffusion models for text-to-video generation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 7623–7633.
[97] "Go from text to speech with a versatile AI voice generator," Murf.AI. Accessed: Jul. 19, 2023. [Online]. Available: https://murf.ai/
[98] "AI powered text to voice generator." PlayHT. Accessed: Jul. 19, 2023. [Online]. Available: https://play.ht/blog/ai-text-to-speech-voice-cloning-technology-overview/
[99] Y. Wang et al., "Tacotron: Towards end-to-end speech synthesis," 2017, *arXiv:1703.10135*.
[100] Z. Borsos et al., "AudioLM: A language modeling approach to audio generation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 2523–2533, 2023.
[101] "Unlock historical documents with AI." READ-COOP. Accessed: Jul. 19, 2023. [Online]. Available: https://readcoop.eu/transkribus/
[102] J. Chen, H. Guo, K. Yi, B. Li, and M. Elhoseiny, "VisualGPT: Data-efficient adaptation of pretrained language models for image captioning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 18030–18040.
[103] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
[104] H. Luo et al., "UniVL: A unified video and language pre-training model for multimodal understanding and generation," 2020, *arXiv:2002.06353*.
[105] H. Xu et al., "VideoCLIP: Contrastive pre-training for zero-shot video-text understanding," 2021, *arXiv:2109.14084*.
[106] "Turn your language data into insights, fast and with no code," SpeakAI. Accessed: Jul. 19, 2023. [Online]. Available: https://speakai.co/
[107] A. Hannun et al., "Deep speech: Scaling up end-to-end speech recognition," 2014, *arXiv:1412.5567*.
[108] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "Wav2vec 2.0: A framework for self-supervised learning of speech representations," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 12449–12460.
[109] "Siri." Apple. Accessed: Jul. 19, 2023. [Online]. Available: http://www.apple.com/ios/siri/
[110] "XiaoIce." Accessed: Jul. 19, 2023. [Online]. Available: https://www.xiaoice.com/
[111] "Google: Google Assistant." Accessed: Jul. 19, 2023. [Online]. Available: https://assistant.google.com/
[112] "Alexa." Amazon. Accessed: Jul. 19, 2023. [Online]. Available: https://developer.amazon.com/public/solutions/alexa/
[113] L. Zhou, J. Gao, D. Li, and H.-Y. Shum, "The design and implementation of XiaoIce, an empathetic social Chatbot," *Comput. Linguist.*, vol. 46, no. 1, pp. 53–93, 2020.
[114] R. Batish, *Voicebot and Chatbot Design: Flexible Conversational Interfaces with Amazon Alexa, Google Home, and Facebook Messenger*. Birmingham, U.K.: Packt Publ. Ltd, 2018.
[115] Y. Zhong, "A theory of semantic information," *China Commun.*, vol. 14, no. 1, pp. 1–17, 2017.
[116] A. Kolchinsky and D. H. Wolpert, "Semantic information, autonomous agency and non-equilibrium statistical physics," *Interface Focus*, vol. 8, no. 6, 2018, Art. no. 20180041.
[117] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Stat. Probab. Contrib. Theory Stat.*, vol. 4, 1961, pp. 547–562.
[118] R. Carnap and Y. Bar-Hillel, "An outline of a theory of semantic information," *J. Symbolic Logic*, vol. 19, no. 3, pp. 230–232, 1954.
[119] J. Bao et al., "Towards a theory of semantic communication," in *Proc. IEEE Netw. Sci. Workshop*, 2011, pp. 110–117.
[120] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, 2021.
[121] P. Jiang, C.-K. Wen, S. Jin, and G. Y. Li, "Deep source-channel coding for sentence semantic transmission with HARQ," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5225–5240, Aug. 2022.
[122] M. Sana and E. C. Strinati, "Learning semantics: An opportunity for effective 6G communications," in *Proc. IEEE 19th Annu. Consum. Commun. Netw. Conf. (CCNC)*, 2022, pp. 631–636.
[123] H. Xie and Z. Qin, "A lite distributed semantic communication system for Internet of Things," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 142–153, Jan. 2021.
[124] F. Zhou, Y. Li, X. Zhang, Q. Wu, X. Lei, and R. Q. Hu, "Cognitive semantic communication systems driven by knowledge graph," in *Proc. IEEE Int. Conf. Commun.*, 2022, pp. 4860–4865.
[125] S. Jiang et al., "Reliable semantic communication system enabled by knowledge graph," *Entropy*, vol. 24, no. 6, p. 846, 2022.

[126] S. Guo, Y. Wang, S. Li, and N. Saeed, "Semantic importance-aware communications using pre-trained language models," *IEEE Commun. Lett.*, vol. 27, no. 9, pp. 2328–2332, Sep. 2023.
[127] F. Zhao, Y. Sun, L. Feng, L. Zhang, and D. Zhao, "Enhancing reasoning ability in semantic communication through generative AI-assisted knowledge construction," *IEEE Commun. Lett.*, vol. 28, no. 4, pp. 832–836, Apr. 2024.
[128] H. Yoo, T. Jung, L. Dai, S. Kim, and C.-B. Chae, "Real-time semantic communications with a vision transformer," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2022, pp. 1–2.
[129] D. Huang, X. Tao, F. Gao, and J. Lu, "Deep learning-based image semantic coding for semantic communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2021, pp. 1–6.
[130] M. U. Lokumarambage, V. S. S. Gowrisetty, H. Rezaei, T. Sivalingam, N. Rajatheva, and A. Fernando, "Wireless end-to-end image transmission system using semantic communications," *IEEE Access*, vol. 11, pp. 37149–37163, 2023.
[131] D. Huang, F. Gao, X. Tao, Q. Du, and J. Lu, "Toward semantic communications: Deep learning-based image semantic coding," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 55–71, Jan. 2023.
[132] C.-H. Lee, J.-W. Lin, P.-H. Chen, and Y.-C. Chang, "Deep learning-constructed joint transmission-recognition for Internet of Things," *IEEE Access*, vol. 7, pp. 76547–76561, 2019.
[133] X. Kang, B. Song, J. Guo, Z. Qin, and F. R. Yu, "Task-oriented image transmission for scene classification in unmanned aerial systems," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5181–5192, Aug. 2022.
[134] J. Kang et al., "Personalized saliency in task-oriented semantic communications: Image transmission and performance analysis," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 186–201, Jan. 2023.
[135] A. D. Raha, M. S. Munir, A. Adhikary, Y. Qiao, and C. S. Hong, "Generative AI-driven semantic communication framework for NextG wireless network," 2023, *arXiv:2310.09021*.
[136] H. Du et al., "Generative AI-aided joint training-free secure semantic communications via multi-modal prompts," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2024, pp. 12896–12900.
[137] S. Wang et al., "Wireless deep video semantic transmission," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 214–229, Jan. 2023.
[138] P. Jiang, C.-K. Wen, S. Jin, and G. Y. Li, "Wireless semantic communications for video conferencing," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 230–244, Jan. 2023.
[139] C. Liang et al., "VISTA: Video transmission over a semantic communication approach," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2023, pp. 1777–1782.
[140] O. Fried et al., "Text-based editing of talking-head video," *ACM Trans. Graph.*, vol. 38, no. 4, pp. 1–14, 2019.
[141] P. Tandon et al., "Txt2Vid: Ultra-low bitrate compression of talking-head videos via text," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 107–118, Jan. 2023.
[142] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2434–2444, Aug. 2021.
[143] Z. Weng, Z. Qin, X. Tao, C. Pan, G. Liu, and G. Y. Li, "Deep learning enabled semantic communications with speech recognition and synthesis," *IEEE Trans. Wireless Commun.*, vol. 22, no. 9, pp. 6227–6240, Sep. 2023.
[144] H. Tong et al., "Federated learning for audio semantic communication," *Front. Commun. Netw.*, vol. 2, Sep. 2021, Art. no. 734402.
[145] E. Grassucci, C. Marinoni, A. Rodriguez, and D. Comminiello, "Diffusion models for audio semantic communication," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2024, pp. 13136–13140.
[146] C. Zhong and J. Havlicek, "LDPC codes for robust transmission of images over wireless channels," in *Proc. Conf. Rec. 35th Asilomar Conf. Signals, Syst. Comput.*, vol. 1, 2001, pp. 797–800.
[147] H. Xie, Z. Qin, X. Tao, and K. B. Letaief, "Task-oriented multi-user semantic communications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2584–2597, 2022.
[148] J. Hribar, M. Costa, N. Kaminski, and L. A. DaSilva, "Updating strategies in the internet of things by taking advantage of correlated sources," in *Proc. IEEE Global Commun. Conf.*, 2017, pp. 1–6.
[149] Q. Abbas, S. A. Hassan, H. K. Qureshi, K. Dev, and H. Jung, "A comprehensive survey on age of information in massive IoT networks," *Comput. Commun.*, vol. 197, pp. 199–213, Jan. 2023.
[150] R. D. Yates, M. Tavan, Y. Hu, and D. Raychaudhuri, "Timely cloud gaming," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2017, pp. 1–9.
[151] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, May 2021.
[152] E. Uysal et al., "Semantic communications in networked systems: A data significance perspective," 2021, *arXiv:2103.05391*.
[153] A. Maatouk, M. Assaad, and A. Ephremides, "The age of incorrect information: An enabler of semantics-empowered communication," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2621–2635, Apr. 2023.
[154] E. Uysal et al., "Semantic communications in networked systems: A data significance perspective," *IEEE Netw.*, vol. 36, no. 4, pp. 233–240, Jul./Aug. 2022.
[155] A. Li, S. Wu, S. Meng, and Q. Zhang, "Towards goal-oriented semantic communications: New metrics, open challenges, and future research directions," 2023, *arXiv:2304.00848*,
[156] T. Soleymani, J. S. Baras, and S. Hirche, "Value of information in feedback control: Quantification," *IEEE Trans. Autom. Control*, vol. 67, no. 7, pp. 3730–3737, Jul. 2022.
[157] C. K. Thomas and W. Saad, "Neuro-symbolic causal reasoning meets signaling game for emergent semantic communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4546–4563, May 2024.
[158] C. K. Thomas, W. Saad, and Y. Xiao, "Causal semantic communication for digital twins: A generalizable imitation learning approach," 2023, *arXiv:2304.12502*.
[159] G. Liu et al., "Semantic communications for artificial intelligence generated content (AIGC) toward effective content creation," 2023, *arXiv:2308.04942*.
[160] Q. Hu, G. Zhang, Z. Qin, Y. Cai, G. Yu, and G. Y. Li, "Robust semantic communications with masked VQ-VAE enabled codebook," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 8707–8722, Dec. 2023.
[161] G. Shi, Y. Xiao, Y. Li, and X. Xie, "From semantic communication to semantic-aware networking: Model, architecture, and open problems," *IEEE Commun. Mag.*, vol. 59, no. 8, pp. 44–50, Aug. 2021.
[162] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "FreeBase: A collaboratively created graph database for structuring human knowledge," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2008, pp. 1247–1250.
[163] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBpedia: A nucleus for a Web of open data," in *Proc. 6th Int. Semantic Web 2nd Asian Conf. Asian Semantic Web Conf.*, 2007, pp. 722–735.
[164] D. Vrandečić and M. Krötzsch, "Wikidata: A free collaborative knowledgebase," *Commun. ACM*, vol. 57, no. 10, pp. 78–85, 2014.
[165] "Introducing the knowledge graph: Things, not strings." Google. Accessed: Jul. 19, 2023. [Online]. Available: https://blog.google/products/search/introducing-knowledge-graph-things-not/
[166] X. Wang, D. Wang, C. Xu, X. He, Y. Cao, and T.-S. Chua, "Explainable reasoning over knowledge graphs for recommendation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 5329–5336.
[167] W. Zheng, L. Yin, X. Chen, Z. Ma, S. Liu, and B. Yang, "Knowledge base graph embedding module design for visual question answering model," *Pattern Recognit.*, vol. 120, p. 108153, 2021.
[168] X. Chen, S. Jia, and Y. Xiang, "A review: Knowledge reasoning over knowledge graph," *Expert Syst. Appl.*, vol. 141, Mar. 2020, Art. no.112948.
[169] L. Xia, Y. Sun, D. Niyato, D. Feng, L. Feng, and M. A. Imran, "xURLLC-aware service provisioning in vehicular networks: A semantic communication perspective," 2023, *arXiv:2302.11993*.
[170] S. Ji, S. Pan, E. Cambria, P. Marttinen, and S. Y. Philip, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 2, pp. 494–514, Feb. 2022.
[171] D. Q. Nguyen, T. Vu, T. D. Nguyen, D. Q. Nguyen, and D. Phung, "A capsule network-based embedding model for knowledge graph completion and search personalization," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguist. Human Lang. Technol.*, vol. 1, 2019, pp. 2180–2189.
[172] Y. Lin, X. Wang, H. Ma, L. Wang, F. Hao, and Z. Cai, "An efficient approach to sharing edge knowledge in 5G-enabled Industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 930–939, Jan. 2023.
[173] H. Chai, S. Leng, Y. Chen, and K. Zhang, "A hierarchical blockchain-enabled federated learning algorithm for knowledge sharing in Internet of Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 3975–3986, Jul. 2021.

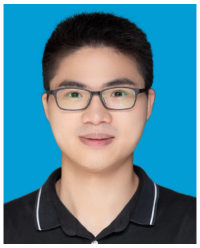
[174] Z. Cai and X. Zheng, “A private and efficient mechanism for data uploading in smart cyber-physical systems,” *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 766–775, Apr.–Jun. 2020.
[175] R. Mullins and M. T. Barros, “Cognitive network management for 5G,” White Paper, 5GPPP Working Group on Network Management and QoS, 5GPPP, Pisa, Italy, 2017.
[176] T. Maksymyuk, S. Dumych, M. Brych, D. Satria, and M. Jo, “An IoT based monitoring framework for software defined 5G mobile networks,” in *Proc. 11th Int. Conf. Ubiquitous Inf. Manage. Commun.*, 2017, pp. 1–4.
[177] H. Du, J. Wang, D. Niyato, J. Kang, Z. Xiong, and D. I. Kim, “AI-generated incentive mechanism and full-duplex semantic communications for information sharing,” 2023, *arXiv:2303.01896*.
[178] C. Hu, W. Bao, D. Wang, and F. Liu, “Dynamic adaptive DNN surgery for inference acceleration on the edge,” in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2019, pp. 1423–1431.
[179] Z. Lin, S. Bi, and Y.-J. A. Zhang, “Optimizing AI service placement and resource allocation in mobile edge intelligence systems,” *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7257–7271, Nov. 2021.
[180] Y. Cang et al., “Online resource allocation for semantic-aware edge computing systems,” *IEEE Internet Things J.*, early access, Oct. 17, 2023, doi: 10.1109/JIOT.2023.3325320.
[181] B. Du et al., “YOLO-based semantic communication with generative AI-aided resource allocation for digital twins construction,” 2023, *arXiv:2306.14138*.
[182] H. Du et al., “Rethinking wireless communication security in semantic Internet of Things,” *IEEE Wireless Commun.*, vol. 30, no. 3, pp. 36–43, Jun. 2023.
[183] H. Du et al., “Beyond deep reinforcement learning: A tutorial on generative diffusion models in network optimization,” 2023, *arXiv:2308.05384*.
[184] Y. Lin et al., “A unified framework for integrating semantic communication and AI-generated content in metaverse,” 2023, *arXiv:2305.11911*.
[185] H. Zhang, H. Wang, Y. Li, K. Long, and A. Nallanathan, “DRL-driven dynamic resource allocation for task-oriented semantic communication,” *IEEE Trans. Commun.*, vol. 71, no. 7, pp. 3992–4004, Jul. 2023.
[186] Q. Ye, W. Shi, K. Qu, H. He, W. Zhuang, and X. Shen, “Joint RAN slicing and computation offloading for autonomous vehicular networks: A learning-assisted hierarchical approach,” *IEEE Open J. Veh. Technol.*, vol. 2, pp. 272–288, 2021.
[187] K. Qu, W. Zhuang, Q. Ye, W. Wu, and X. Shen, “Model-assisted learning for adaptive cooperative perception of connected autonomous vehicles,” *IEEE Trans. Wireless Commun.*, early access, Jan. 24, 2024, doi: 10.1109/TWC.2024.3354507.
[188] W. Zhuang, Q. Ye, F. Lyu, N. Cheng, and J. Ren, “SDN/NFV-empowered future IoV with enhanced communication, computing, and caching,” *Proc. IEEE*, vol. 108, no. 2, pp. 274–291, Feb. 2020.
[189] S. Zhang, I.-L. Yen, F. Bastani, H. Moeini, and D. Moore, “A semantic model for information sharing in autonomous vehicle systems,” in *Proc. IEEE 11th Int. Conf. Semantic Comput. (ICSC)*, 2017, pp. 32–39.
[190] A. Deb Raha, M. S. Munir, A. Adhikary, Y. Qiao, S.-B. Park, and C. Seon Hong, “An artificial intelligent-driven semantic communication framework for connected autonomous vehicular network,” in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, 2023, pp. 352–357.
[191] L. Carvalho, “Smart cities from scratch? A socio-technical perspective,” *Cambridge J. Regions, Econ. Soc.*, vol. 8, no. 1, pp. 43–60, 2015.
[192] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, “Identifying, understanding, and Analyzing critical infrastructure interdependencies,” *IEEE Control Syst. Mag.*, vol. 21, no. 6, pp. 11–25, Dec. 2001.
[193] M. Austin, P. Delgoshaei, M. Coelho, and M. Heidarinejad, “Architecting smart city digital twins: Combined semantic model and machine learning approach,” *J. Manage. Eng.*, vol. 36, no. 4, 2020, Art. no. 4020026.
[194] P. Antonios, K. Konstantinos, and G. Christos, “A systematic review on semantic interoperability in the IoE-enabled smart cities,” *Internet Things*, vol. 22, Jul. 2023, Art. no. 100754.
[195] L. Xia et al., “WiserVR: Semantic communication enabled wireless virtual reality delivery,” *IEEE Wireless Commun.*, vol. 30, no. 2, pp. 32–39, Apr. 2023.
[196] C. Wang et al., “Adaptive semantic-bit communication for extended reality interactions,” *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 5, pp. 1080–1092, Sep. 2023.
[197] B. Zhang, Z. Qin, Y. Guo, and G. Y. Li, “Semantic sensing and communications for ultimate extended reality,” 2022, *arXiv:2212.08533*.
[198] B. Zhang, Z. Qin, and G. Y. Li, “Semantic communications with variable-length coding for extended reality,” 2023, *arXiv:2302.08645*.
[199] J. Park, J. Choi, S.-L. Kim, and M. Bennis, “Enabling the wireless metaverse via semantic multiverse communication,” in *Proc. 20th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, 2023, pp. 85–90.
[200] J. Chen, J. Wang, C. Jiang, Y. Ren, and L. Hanzo, “Trust-worthy semantic communications for the metaverse relying on federated learning,” 2023, *arXiv:2305.09255*.

Chengsi Liang (Student Member, IEEE) is currently pursuing the Ph.D. degree with the James Watt School of Engineering, University of Glasgow, U.K. Her research interests include semantic communications and networking.

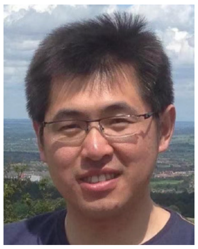
Hongyang Du (Graduate Student Member, IEEE) received the Ph.D. degree from the Interdisciplinary Graduate Program, College of Computing and Data Science, Energy Research Institute @ NTU, Nanyang Technological University, Singapore, in 2024. His research interests include edge intelligence, generative AI, semantic communications, and network management. He is the recipient of the IEEE Daniel E. Noble Fellowship Award from the IEEE Vehicular Technology Society in 2022, the IEEE Signal Processing Society Scholarship from the IEEE Signal Processing Society in 2023, the Chinese Government Award for Outstanding Students Abroad in 2023, and the Singapore Data Science Consortium Dissertation Research Fellowship in 2023.

Yao Sun (Senior Member, IEEE) is currently a Lecturer with the James Watt School of Engineering, University of Glasgow, Glasgow, U.K. His research interests include intelligent wireless networking, semantic communications, blockchain system, and resource management in next-generation mobile networks. He has won the IEEE Communication Society of TAOS Best Paper Award in 2019 ICC, the IEEE IoT Journal Best Paper Award in 2022, and the Best Paper Award in 22nd ICCT.

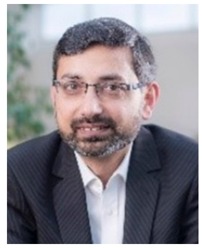
Dusit Niyato (Fellow, IEEE) received the B.Eng. degree from King Mongkut’s Institute of Technology Ladkrabang, Thailand, and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Canada. He is a Professor with the College of Computing and Data Science, Nanyang Technological University, Singapore. His research interests are in the areas of mobile generative AI, edge intelligence, decentralized machine learning, and incentive mechanism design.



Jiawen Kang (Senior Member, IEEE) received the Ph.D. degree from the Guangdong University of Technology, China, in 2018. He was a Postdoctoral Fellow with Nanyang Technological University, Singapore, from 2018 to 2021. He is currently a Full Professor with the Guangdong University of Technology. His research interests mainly focus on blockchain, security, and privacy protection in wireless communications and networking.



Dezong Zhao (Senior Member, IEEE) received the B.Eng. and M.S. degrees in control science and engineering from Shandong University, Jinan, China, in 2003 and 2006, respectively, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2010. He is a Reader of Autonomous Systems with the James Watt School of Engineering, University of Glasgow and a Turing Fellow with the Alan Turing Institute. His research interests include autonomous vehicles, robotics, machine intelligence, and control engineering. He was awarded a Royal Society-Newton Advanced Fellow in 2020 and an EPSRC Innovation Fellow in 2018.



Muhammad Ali Imran (Fellow, IEEE) received the M.Sc. (Distinction) and Ph.D. degrees from Imperial College London, U.K., in 2002 and 2007, respectively. He is a Professor of Communication Systems with the University of Glasgow. He has a global collaborative research network spanning both academia and key industrial players in the field of wireless communications. He has supervised 50+ successful Ph.D. graduates and published over 600 peer-reviewed research papers including more than 100 IEEE TRANSACTION papers.