GenSC-6G: A Prototype Testbed for Integrated Generative AI, Quantum, and Semantic Communication

Brian E. Arfeto, Shehbaz Tariq, Uman Khalid, Trung Q. Duong, *Fellow, IEEE*, and Hyundong Shin, *Fellow, IEEE*

***Abstract*—We introduce a prototyping testbed—GenSC-6G—developed to generate a comprehensive dataset that supports the integration of generative artificial intelligence (AI), quantum computing, and semantic communication for emerging sixth-generation (6G) applications. The GenSC-6G dataset is designed with noise-augmented synthetic data optimized for semantic decoding, classification, and localization tasks, significantly enhancing flexibility for diverse AI-driven communication applications. This adaptable prototype supports seamless modifications across baseline models, communication modules, and goal-oriented decoders. Case studies demonstrate its application in lightweight classification, semantic upsampling, and edge language inference under noise conditions. The GenSC-6G dataset serves as a scalable and robust resource for developing goal-oriented communication systems tailored to the growing demands of 6G networks.**

I. Introduction

The integration of generative artificial intelligence (AI) with semantic communication (SC) marks a transformative paradigm shift in communications, transitioning from basic data transmission to goal-oriented, context-aware information exchange [1], [2]. By leveraging advanced AI and foundation models [3], [4], these systems enhance efficiency and adaptability, tailoring transmissions to align with specific communicative goals. At its core, the SC employs a knowledge base (KB) with semantic encoders and decoders, prioritizing context and intent over raw data [5]. This innovative approach enables ultra-efficient compression, making it ideal for applications such as Internet of Things (IoT), cloud services, autonomous systems, and other cutting-edge sixth-generation (6G) use cases, as illustrated in Fig. 1.

Recent advancements have demonstrated the successful integration of advanced AI and SC systems, paving the way for more adaptive and intelligent communication networks [6]. Building on this progress, numerous studies have explored how generative AI can enhance data generation and transmission quality in 6G networks [7], [8]. For example, generative AI has been utilized at the network edge to improve visual data transmission quality by leveraging multimodal data inputs [9]. Similarly, integrating AI into communication devices has facilitated support for diverse data formats and translation tasks [10]. Moreover, advanced AI techniques have been embedded in decoders to allow the generation of new information on the receiver end, further enhancing the potential of AI-generated content (AIGC) [11]. Generative models play a crucial role in creating and updating the KB within SC systems by dynamically generating refined knowledge representations [12]. This KB serves as a repository for learned semantics, enabling efficient encoding and decoding while minimizing raw data transmission. Constructing a robust KB, however, requires extensive labeled datasets for effective model training.

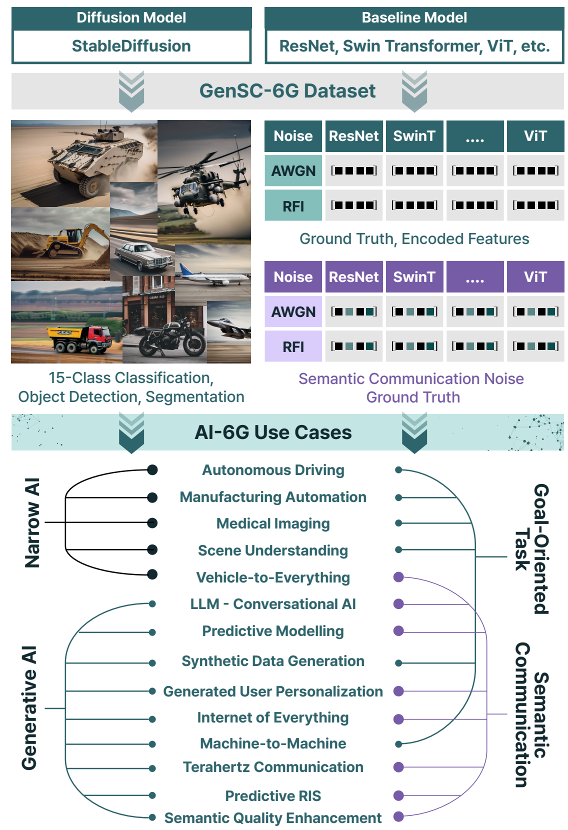

Fig. 1. GenSC-6G dataset structures. The dataset consists of the ground-truth data, encoded features, and additive noise. The AI-6G use cases span narrow AI and generative AI applications over SC and goal-oriented tasks in multiple fields. ResNet, ViT, SwinT, AWGN, RFI, LLM, and RIS stand for residual network, vision transformer, swin transformer, additive white Gaussian noise, radio frequency interference, large language model, and reconfigurable intelligent surface, respectively.

While testbeds, such as DeepSense 6G [13], offer realistic environments for sub-terahertz (THz) beam prediction, their high cost and limited adaptability pose scalability challenges. Generative AI addresses these limitations by producing realistic synthetic datasets, reducing the manual data collection burden, and enhancing scalability across domains. To support the further development of 6G technologies and semantic-native communication, generative semantic communication for 6G (GenSC-6G) is designed as a flexible and sustainable testbed that bridges multiple stacks and integration. In addition to scalability, sustainability is critical in communication

B. E. Arfeto, S. Tariq, U. Khalid, and H. Shin (corresponding author) are with Kyung Hee University, Korea; T. Q. Duong is with Memorial University, Canada, and Queen's University Belfast, UK.

systems, especially for 6G applications on edge devices. Hybrid quantum-classical (HQC) processing and model pruning contribute to this goal by offloading computationally intensive tasks and enabling faster inference [14]. Furthermore, recent work in quantum SC has highlighted robustness and security advantages of quantum integration, further broadening the potential of sustainable SC frameworks [15]. To address these challenges, this paper offers the following contributions:

- **Adaptable SC Framework**: A flexible prototype that enables customization of baseline models, communication modules, and decoders for SC.
- **Generative AI-Driven SC**: The integration of generative AI for synthetic data generation and enriching the KB and leveraging large language model (LLM) capabilities for enhanced semantic tasks.
- **Noise-Augmented Dataset**: A labeled dataset with injected noise, specifically optimized for semantic tasks such as target recognition, localization, and recovery.
- **Case Study on Semantic Tasks**: A detailed case study that evaluates baseline models across various semantic tasks, assessing performance and adaptability.

II. GenSC-6G Dataset Structures

The GenSC-6G dataset is meticulously organized to support machine learning (ML) tasks, including classification, segmentation, object detection, and edge LLM tasks. Each ML or semantic task is associated with a standalone ground-truth data collection or a combination of multiple collections in a generic format. These collections achieve dual objectives: i) they ensure that models can be effectively trained and evaluated across different and interconnected tasks; and ii) the generic format of the dataset provides scalability and flexibility for SC tasks, enabling easy modification of wireless methods, parameters, or noise levels. This adaptability allows it to serve goal-oriented tasks in any environment. The dataset is available at `https://github.com/CQILAB-Official/GenSC-6G`. Each collection of semantic tasks is structured as follows.

1) Ground-Truth Data: The ground truth data collection is precisely annotated for each of the 15 defined classes, with corresponding class and segmented labels. This dataset includes a total of $4,829$ instances for training and $1,320$ instances for testing. The context of the dataset pertains to common vehicle types in both military and civilian sectors. The input data collection is designed for the semantic upsampling task.

2) Base-Model Features: The dataset includes features extracted from base models in the form of matrices, as illustrated in Fig. 1. These extracted features serve as data representations, enabling the processing across multiple decoding components within the large AI framework.

3) Additive Noise Features: In this component, the dataset includes features with various levels additive white Gaussian noise (AWGN). These noise features allow for testing the resilience and adaptability of the base models while also enabling in-depth analysis of noise effects and modulation.

4) Framework Source Code: A codebase is provided including the transmission configuration, base model, and metrics, as illustrated in Fig. 2.

III. Prototyping a Large-AI SC Testbed

In this section, we present a modular framework integrating large AI models and HQC computing with SC to enhance data generation, transmission efficiency, and adaptability within the joint source-channel coding (JSCC) framework.

A. Generative AI Auto Dataset

We first introduce the concept of dataset auto-creation by integrating diffusion models and automated mechanisms to streamline data generation, labeling, and training processes.

1) Diffusion-Driven Data Generator: The GenSC-6G testbed leverages text-to-image diffusion models to automate the generation of diverse and trainable images for the base model. This framework uses the latent diffusion model (LDM), which transforms high-dimensional data into low-dimensional latent spaces, optimizing computational efficiency, output quality, and diversity. The LDM operates in two main stages: forward diffusion, in which an image is progressively transformed into a noisy representation by adding Gaussian noise over several steps, and reverse diffusion, where this noisy latent representation is iteratively denoised back into a coherent image guided by text prompts.

2) Auto Inference Mechanism: With the GenSC-6G dataset pipeline, an auto-generation mechanism is employed in real time to continuously generate new data. The workflow involves feeding the trained model dynamically with input data through a diffusion model, producing diverse data instances. These instances are then labeled automatically and passed to the training pipeline, with the model saved for subsequent training sections. A masked autoencoder model with a vision transformer (ViT) such as the segment anything model (SAM) or you only look once (YOLO) can be potentially employed for auto-localization tasks (instance segmentation and object detection). This approach allows for consecutive model training without the need for manual data collection.

B. Base-Model Encoder

We now outline the framework components responsible for encoding, transmission, and task-aware decoding.

1) Backbone Encoder Selection: To ensure optimal performance across semantic decoding tasks, we employ a spectrum of backbone encoders—convolutional neural networks (CNNs) for robust local feature extraction, ViTs for capturing global context, and lightweight neural networks for real-time edge inference—chosen based on device capabilities and operating conditions, and leveraging pretrained weights for accelerated convergence or training from scratch when required. Contingent upon quantum processing unit (QPU) availability, we integrate HQC, harnessing quantum superposition and entanglement to generate exponentially enriched high-dimensional embeddings with inherent noise resilience. This flexible design allows substituting any backbone encoder, ensuring compatibility with current and future model architectures.

2) Mobile Deep Learning: To accommodate limited computational resources available on edge devices, the backbone framework is alterable in an optional way to use lightweight neural networks such as EfficientNet and MobileNet. These models feature significantly fewer parameters compared to

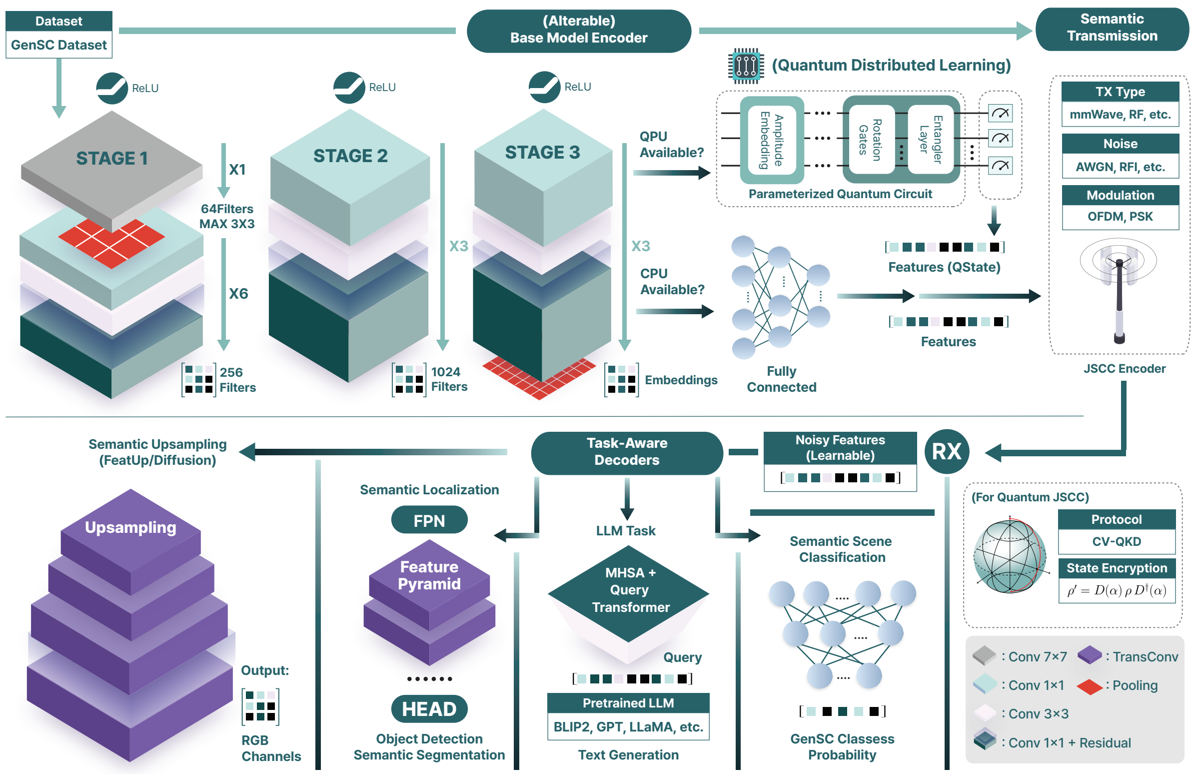

Fig. 2. A GenSC-6G testbed prototype. The large-AI SC testbed framework prototypes a flexible architecture in which the backbone encoder and communication modules are alterable to fit any backend, and the semantic decoders can be adapted for various downstream goal-oriented tasks. The overall pipeline is as follows. Inputs from the GenSC-6G dataset pass through a configurable backbone encoder to extract semantic features. These embeddings are either enhanced via quantum distributed learning on QPUs or processed classically on CPUs. The features are then transmitted over JSCC-enabled channels (classical or CV-QKD quantum links). At the receiver, task-aware decoders conditionally perform semantic upsampling, localization, classification, and LLM-based text generation. ReLU, TX, RX, mmWave, OFDM, PSK, CV-QKD, MHSA, BLIP, GPT, LLaMA, and FPN stand for the rectified linear unit, transmitter, receiver, millimeter wave, orthogonal frequency division multiplexing, phase-shift keying, continuous-variable quantum key distribution, multi-head self-attention, bootstrapping language-image pretraining, generative pretrained transformer, LLM Meta AI, and feature pyramid network, respectively.

traditional deep learning (DL) models, effectively reducing the parameter number and computational complexity at each layer. This reduction translates into faster inference and load times, which are crucial for real-time applications (see Table I). EfficientNet and MobileNet achieve this by optimizing network layers, employing depthwise separable convolutions, and reducing the overall model size, all while striving to maintain high accuracy with lower computational demands.

3) Feature Extraction: Feature extraction transforms raw input data into compact, informative representations that preserve essential semantic and structural information for downstream SC tasks. In the GenSC-6G, this is performed by an interchangeable backbone encoder—either a CNN baseline or ViT—chosen depending on the target application. CNNs apply stacked convolution and pooling layers to hierarchically extract local features, while ViTs split the input into patches and use a self-attention mechanism to model global context. As shown in Fig. 2, the backbone encoder begins by passing the input data through several convolutional layers, progressively reducing data dimensionality while retaining key semantic components throughout the patching and filters.

4) Conditional Quantum Embeddings: This process maps the semantic features extracted from the backbone encoder into quantum states. The features are first passed through a fully connected layer for dimensionality reduction, preparing them for quantum encoding. The prototype can then utilize amplitude-embedding and angle-embedding techniques to encode these features into qubits. In amplitude embedding, the normalized feature vector is directly used to define the amplitudes of the quantum state, allowing for efficient representation of high-dimensional data in a quantum system. This method maps the classical information into the amplitudes of a superposed quantum state. In angle embedding, the features are encoded into the rotation angles of quantum gates, such as parameterized rotation gates on the bloch sphere, which manipulate the state of the qubits accordingly. These embeddings translate classical data into quantum formats suitable for quantum processing and transmission. Additionally, an entanglement layer can be incorporated into the quantum encoder to exploit the entanglement properties. This layer uses quantum gates, such as the controlled-NOT or controlled-phase gates, to establish entanglement among qubits, thereby enabling the quantum system to capture complex feature interactions and dependencies in larger dimensions within the semantic data.

5) Quantum Parallel Processing: HQC computation in the large-AI SC testbed framework combines QPUs with traditional graphics processing units (GPUs) to execute DL tasks more efficiently. QPUs work alongside GPUs to handle tasks such as HQC optimization, state preparation, and sampling, which complement GPU operations on matrix multiplication and dense processing. The testbed framework employs skip connections and resource splitting, enabling QPUs to manage quantum-specific computations while GPUs process standard neural network layers. Specifically, to implement this, the model architecture is enhanced with quantum layers, consisting of a basic entanglement layer and amplitude embedding. The skip connection features are conditionally encoded into these quantum layers for QPUs. In general, this task division helps manage computational loads on backbone feature extraction, decoding, and inference. Hence, the framework incorporates quantum kernels, which introduce quantum properties into the learning process to enhance model performance. These quantum kernels, in the form of embeddings within QPUs, map data into high-dimensional quantum Hilbert spaces, providing more powerful data representations that can improve classification accuracy and decision-making processes. Additionally, the framework supports quantum distributed learning, partitioning model training and gradient aggregation across multiple QPU nodes to accelerate convergence. This hybrid approach allows the system to dynamically offload complex tasks to cloud-based QPUs when on-premises hardware reaches its computational limits.

6) Semantic Compression: The semantic compression process is performed by extracting and prioritizing only the most relevant features from the input data. A critical step of this compression process is quantization. After feature extraction, the data is quantized to map the continuous feature space into a discrete set of values, reducing data precision in a controlled manner. The backbone maintains a balance between compression efficiency and semantic relevance preservation.

7) Fine-Tuning, Retraining, and Output: Once the encoder is initialized with a pretrained network, the model can be fine-tuned or retrained. Fine-tuning involves adjusting the pre-trained weights with a lower learning rate to adapt to the new task while preserving useful representations—freezing some layers. Retraining involves filling in missing weights, such as those in the decoder and fully connected layers, or resetting the weights entirely—allowing the model to be trained from scratch. After the feature extraction step, the encoder produces a compressed representation of the input data. This compact form reduces data dimensionality while retaining the most significant semantic information, which is then prepared for transmission. The output from the bottleneck can be reused for various goal-oriented tasks, such as classification, object detection, or segmentation.

C. Semantic Transmission

This task entails encoding extracted semantic features into latent representations and conveying them over classical or quantum channels using bandwidth-efficient JSCC schemes.

1) Bandwidth-Efficient SC: Efficient compression of semantic data is critical to transmitting relevant information without exhausting resources. The model compression mechanism, such as entropy coding, represents the most frequent semantic components with fewer bits, maximizing compression efficiency. After passing through stages of feature extraction and quantization, where the complexity and number of filters increase at each stage, the data is compressed. This compression is achieved through techniques such as pooling and fully connected layers, which map the extracted features into a compact representation. The compressed data is then encoded into a format suitable for transmission, reducing the burden on available bandwidth by compressing them into latent features.

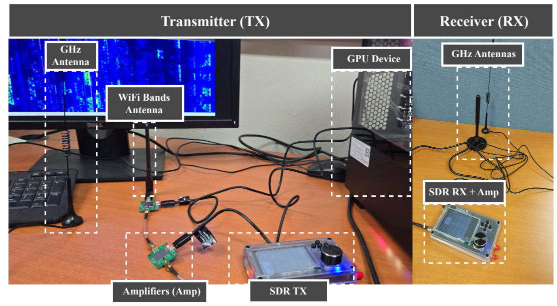

Fig. 3. A transceiver setup to capture noise features as part of the testbed. The testbed leverages the Wi-Fi 7 (802.11be) OFDM communication with file streaming. On the transmitter side, a programmable SDR setup with gigahertz (GHz) antennas and amplifiers sends high-frequency signals to the receiver.

2) Classical and Quantum JSCC Modules: The GenSC-6G JSCC module is designed to evaluate the quality degradation and performance of data transmission under both classical and quantum channel conditions.

Classical Channel: The classical channel in the GenSC-6G prototype is configured to emulate realistic communication scenarios, particularly focusing on noise conditions, modulation schemes, and transmission protocols representative of current and next-generation communication standards.

- **Channel Noise:** Noise is implemented within the testbed to replicate real-world communication scenarios where data transmission is disturbed. The system model simulates noise conditions by incorporating AWGN with a specific signal-to-noise ratio (SNR) and random transmitter noise from a programmable software-defined radio (SDR) to dynamically adjust noise characteristics. An example setup is shown in Fig. 3. By injecting controlled and random noise levels during transmission, the system can assess the impact on semantic data, helping to refine error correction and noise mitigation within the network decoder. The DL model adapts to varying noise conditions by optimizing its parameters to minimize the loss function, which measures the discrepancy between the predicted noisy outputs and the ground-truth data.
- **JSCC Transmission:** The prototype utilizes JSCC to efficiently transmit semantic information over bandwidth-limited channels. By combining source compression and channel coding within a single DL framework, the system learns an end-to-end mapping from input data to transmitted signals and from received signals to reconstructed outputs. In this process, the semantic features are directly mapped into channel symbols, bypassing traditional separate source and channel coding schemes. The testbed transmission employs orthogonal frequency-division multiplexing (OFDM) with Wi-Fi 7 (802.11be) (see Fig. 3). The system can be adapted to various frequency bands, including THz and sub-THz bands, which

TABLE I
PERFORMANCE OF LARGE-AI BASE MODELS TRAINED ON THE GENSC-6G DATASET FOR CLASSIFICATION, UPSAMPLING, AND EDGE LLM TASKS.

Device	Backbone Encoder	Decoder Classifier (Upsampler)	Parameters (Upsamplers)	Processing Unit	AWGN (SNR = 10 dB) Accuracy (LPIPS)	AWGN (SNR = 10 dB) F1 (CLIP-S*)	AWGN (SNR = 10 dB) Recall (CLIP-S†)	AWGN (SNR = 30 dB) Accuracy (LPIPS)	AWGN (SNR = 30 dB) F1 (CLIP-S*)	AWGN (SNR = 30 dB) Recall (CLIP-S†)
Classical	ViT-L-32	3xFC	306.79	GPU	0.8477	0.8514	0.8512	0.8485	0.8522	0.8516
	ViT-L-32	(FeatUp)	(30.75)	GPU	(0.4211)	(28.0641)	(29.9747)	(0.4038)	(27.5533)	(30.0994)
	ResNet-50	3xFC	25.81	GPU	0.8447	0.8468	0.8462	0.8485	0.8507	0.8510
	ResNet-50	(FeatUp)	(30.75)	GPU	(0.4163)	(27.7008)	(29.9923)	(0.4027)	(27.9054)	(30.0679)
	VGG-16	3xFC	138.61	GPU	0.8144	0.8163	0.8158	0.8167	0.8158	0.8161
	Inception-V3	3xFC	27.42	GPU	0.8561	0.8569	0.8553	0.8644	0.8650	0.8641
	DINO-V2	(FeatUp)	(30.75)	GPU	(0.4210)	(27.6070)	(29.8617)	(0.4153)	(27.7631)	(30.0411)
Mobile	EfficientNet-B1	2xFC	7.79	CPU	0.8689	0.8702	0.8700	0.8705	0.8720	0.8735
	MobileNet-V3	2xFC	5.48	CPU	0.7871	0.7889	0.7896	0.8197	0.8365	0.8192
HQC	ViT-L-32	QNN	306.79	GPU/QPU	0.8303	0.8250	0.8232	0.8485	0.8522	0.8516
	ResNet-50	QNN	25.81	GPU/QPU	0.8144	0.8181	0.8185	0.8356	0.8383	0.8381

Note: * LLaMA-3; † BLIP-2

provide extensive bandwidth for semantic applications.

Quantum Channel: In the quantum communication scenario, the JSCC module is adapted to encode semantic information into quantum states for transmission over quantum channels that exploit quantum properties. The typical integration in quantum transmission can be described as follows.

- **Quantum Protocols:** Quantum communication protocols leverage fundamental quantum properties to enable advanced tasks like quantum teleportation, distributed sensing, and secure key distribution. In continuous-variable quantum key distribution (CV-QKD), the transmitter encodes information onto Gaussian-modulated coherent states using optical quadratures, which are transmitted over an optical channel. At the receiver, homodyne or heterodyne measurements extract correlated classical data, forming the basis for generating secure cryptographic keys protected by quantum mechanics.
- **Bit-Flip and Phase-Flip Noise:** Qubits are susceptible to various types of noise, including bit-flip and phase-flip errors. Bit-flip noise alters the state of a qubit from $|0\rangle$ to $|1\rangle$ or vice versa, while phase-flip noise changes the phase between states without affecting their probabilities. These noise types are typically modeled using Pauli operators, specifically the Pauli-X (bit-flip) and Pauli-Z (phase-flip) gates. Quantum error correction codes, such as Shor or Steane codes, mitigate these errors by leveraging entanglement and redundancy.

D. Task-Aware Decoder: Case Study

We now present case studies demonstrating the application of the GenSC-6G prototype to various semantic decoding tasks. Each task utilizes the compressed semantic features transmitted over the communication channel and focuses on different aspects of semantic understanding. We evaluate the performance using relevant metrics and provide benchmarks for comparison.

1) Lightweight Classification: The semantic classification task involves categorizing images into predefined classes based on their content. Using the GenSC-6G dataset, we train several baseline models, including ViT-L-32, residual network (ResNet)-50, visual geometry group (VGG)-16, and Inception-V3, on single or combined processing units, as well as lightweight models like EfficientNet-B1 and MobileNet-V3 suited for edge devices. All baseline models are trained under various SNR conditions (e.g., at 10 dB and 30 dB) to evaluate their robustness to noise in the communication channel. The decoder architecture is defined as a lightweight module with three fully connected layers designed to downsample the features extracted by the encoder. Fig. 4 shows the confusion matrix for the ResNet-50 model under AWGN with the SNR of 10 dB (lower-left plot), illustrating the balance between true and false positive rates. The model achieves an accuracy of 84.47% (see Table I), demonstrating robust—but improvable—baseline performance in noisy conditions. Notably, the highest baseline accuracy is achieved by EfficientNet-B1 with a score of 86.89% while maintaining a mobile-friendly architecture. Additional metrics, including F1 score and recall, are also provided in Table I. By focusing on transmitting essential semantic features and utilizing lightweight decoders, the models demonstrate that high classification accuracy is achievable even in the presence of significant channel noise. These compressed features are reusable across tasks, making them efficient for diverse downstream AI tasks.

2) Semantic Localization: We first utilize the YOLO model for object detection of vehicles within the images from the GenSC-6G dataset. After detecting the vehicles with YOLO, we perform semantic segmentation to achieve pixel-level localization. The ground truth provided by the GenSC-6G segments is used to train and validate the segmentation models. By leveraging these detailed annotations, the models learned to accurately delineate vehicle boundaries, enhancing the localization precision. As shown in Fig. 4 (upper plot), we use localization metrics such as the intersection over union (IoU) and mean pixel accuracy (MPA). The higher IoU for Image 1 at the SNR of 10 dB indicates better overlap between the predicted segmentation and the ground truth, suggesting accurate vehicle localization by the model in that instance. The MPA values, while relatively low, provide insight into pixel-level classification accuracy across the entire image, including both the object and the background. The close MPA

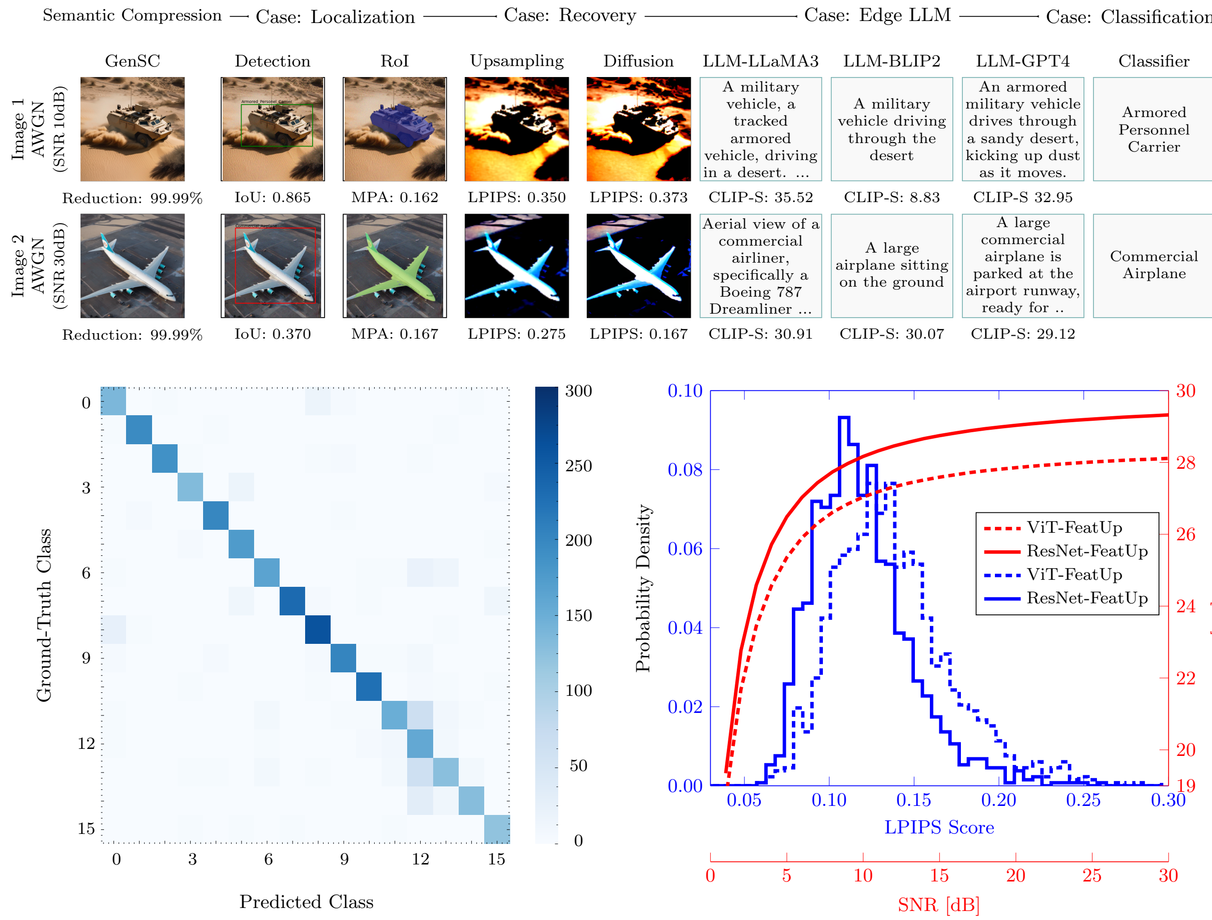

Fig. 4. Overview of case studies demonstrating downstream goal-oriented tasks from encoded features and their common metrics, including semantic compression, object localization, recovery through upsampling and diffusion, and post-processing with LLM, illustrating the adaptability of encoded features. Herein, the average compression rate of the dataset reduces to 99.993%, which is compressed to nearly zero. IoU and MPA assess object detection and segmentation performance, while CLIP-S measures the alignment between generated text and visual content. The confusion matrix (bottom left) illustrates the classification performance of the ResNet-50 model at the SNR of 30 dB. The LPIPS probability distribution and the PSNR (bottom right) contrast perceptual similarity scores between different models and reflect image quality maintenance across varying noise levels, showing the adaptability and effectiveness of these models. IoU, MPA, and CLIP-S stand for the intersection-over-union, mean pixel accuracy, and contrastive language-image pretraining score, respectively.

values between the two images indicate consistent pixel-level performance, though there is room for improvement, especially in challenging conditions. These results demonstrate that the SC framework effectively supports semantic localization tasks by preserving essential spatial features necessary for accurate object detection and segmentation.

3) Semantic Upsampling Recovery: The semantic upsampling recovery task focuses on enhancing low-resolution images received over noisy channels by reconstructing high-resolution outputs. Two distinct approaches are evaluated for upsampling combined with feature upsampling (FeatUp) for two baseline models (ResNet and ViT). FeatUp enhances deep features by restoring lost spatial information through high-resolution signal guidance or implicit modeling to improve performance in dense prediction tasks. To ascertain upsampling recovery performance, we evaluate the learned perceptual image patch similarity (LPIPS) and the peak SNR (PSNR) for the GenSC-6G dataset in Fig. 4 (lower-right plot). As depicted, there is an inverse relationship between the probability of accurate reconstruction and the LPIPS score. Lower LPIPS values indicate that reconstructed images are perceptually closer to the ground truth. The empirical distribution demonstrates that the ResNet with FeatUp has a higher probability density in the lower LPIPS score range (from 0.05 to 0.15), indicating more accurate reconstructions, while the ViT baseline with FeatUp remains notable performance. The PSNR performance is also depicted as a function of SNR. Here, ResNet-FeatUp outperforms ViT-FeatUp again, especially at high SNR values, demonstrating considerable noise resilience and image fidelity during upsampling recovery.

4) Edge LLM: The edge LLM task involves integrating advanced models, such as the bootstrapping language-image pretraining (BLIP), generative pretrained transformer (GPT), and LLM Meta AI (LLaMA or large language model Meta AI), which run as a post-processing service on edge nodes, ingesting semantic embeddings to generate text from visual semantic features and enriching output semantics. As depicted in Fig. 2, the prototype architecture employs a feature encoder combined with a querying transformer. These queries are then input into pretrained LLMs that specialize in transforming visual-semantic representations into meaningful text outputs. The visual-text encoder architecture plays a pivotal role in this setup, combining visual feature extraction with a text generation pipeline. The contrastive language-image pretraining score (CLIP-S) is used to measure the alignment between the generated text and the visual context (see Fig. 4 upper plot). The CLIP-S evaluates how well the descriptions or captions match the visual input, reflecting the effectiveness of generated text in comprehending the scene. For example, in GenSC-6G image captioning, LLaMA-3 reaches a CLIP-S of 35.52 for Image 1, demonstrating its level of contextual understanding and the richness of the text generated from the image.

IV. Open Challenges

Despite the results demonstrated by the GenSC-6G framework, several challenges remain in fully realizing its potential. Sustainability is a significant concern, particularly in maintaining energy-efficient operations for large models and continuous data processing at the edges. This is critical for ensuring that AI-driven SC systems align with the sustainability goals of future networks. Model robustness in the face of varying noise conditions and unpredictable environments also remains challenging. Furthermore, quantum distributed learning across multiple QPU nodes faces challenges in synchronization and error accumulation, which can impede reliable convergence. Deploying these models on mobile devices poses another challenge, as many of the state-of-the-art AI models, including LLM, are computationally intensive and difficult to scale down to the limited resources of edge devices. Solutions such as model pruning, quantization, and efficient architecture designs require further refinement to enable real-time on-device processing. Generalization to more downstream AI tasks, including complex multimodal tasks, is also crucial to expand the SC utility. Finally, quantum communication and computing face challenges in ensuring noise resilience, enhancing scalability, and managing the complexity of parallel processing. Looking forward, the GenSC-6G framework is inherently compatible with the envisioned architecture of 6G SC stacks. Its modularity enables seamless integration across the physical, semantic, and application layers, supporting core functionalities. Although further integration efforts are required, the SC framework is well positioned to interface with emerging 6G technologies such as terahertz communication, open radio access networks, and intelligent reflecting surfaces. These synergies establish GenSC-6G as a versatile platform for advancing semantic-native system design, robust model training, and validation in real-world 6G scenarios.

V. Conclusion

This paper has introduced the GenSC-6G framework, which integrates large AI, HQC optimization, and SC, tailored to optimize 6G networks through scalable and alterable communications models. The testbed offers a flexible prototype that enables the modification of baseline models, communication modules, and goal-oriented decoders, supporting a variety of downstream tasks. This modular framework leverages generative AI to enhance the KB by generating realistic synthetic data, thus improving model diversity and adaptability in real-world scenarios. Synergistically, these rich semantic inputs are conditionally transformed by quantum processing into ultra-high-dimensional embeddings and parallelized across training to enable a scalable testbed through synthetic data generation. Through the detailed case studies, we have demonstrated the effectiveness of our approach in various semantic decoding tasks, including lightweight classification, semantic localization, and upsampling recovery. Evaluations across different communication conditions, as seen in downstream tasks such as semantic classification and edge LLM, highlight the practical adaptability of the GenSC-6G framework in a wide range of semantic tasks.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) under RS-2025-00556064 and by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2025-2021-0-02046) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation). The work of T. Q. Duong was supported in part by the Canada Excellence Research Chair (CERC) Program CERC-2022-00109 and in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant Program RGPIN-2025-04941.

References

[1] C. Chaccour *et al.*, "Less data, more knowledge: Building next generation semantic communication networks," *IEEE Commun. Surveys Tuts.*, Jun. 2024 (Early Access), doi:10.1109/COMST.2024.3412852.
[2] Z. Lu *et al.*, "Semantics-empowered communications: A tutorial-cum-survey," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 1, pp. 41–79, 2024.
[3] N. Fei *et al.*, "Towards artificial general intelligence via a multimodal foundation model," *Nat. Commun.*, vol. 13, no. 1, p. 3094, Jun. 2022.
[4] P. Xu *et al.*, "Multimodal learning with transformers: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, Oct. 2023.
[5] 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.
[6] F. Jiang *et al.*, "Large AI model empowered multimodal semantic communications," *IEEE Commun. Mag.*, Sep. 2024 (Early Access).
[7] F. Jiang *et al.*, "Large language model enhanced multi-agent systems for 6G communications," *IEEE Wireless Commun.*, Aug. 2024 (Early Access).
[8] S. Tariq *et al.*, "Integrating sustainable big AI: Quantum anonymous semantic broadcast," *IEEE Wireless Commun.*, vol. 31, no. 3, pp. 86–99, Jun. 2024.
[9] Y. Zhao *et al.*, "LaMoSC: Large language model-driven semantic communication system for visual transmission," *IEEE Trans. on Cogn. Commun. Netw.*, May 2024 (Early Access).
[10] C. Liang *et al.*, "Generative AI-driven semantic communication networks: Architecture, technologies and applications," *IEEE Trans. on Cogn. Commun. Netw.*, Jul. 2024 (Early Access).
[11] W. Yang *et al.*, "Rethinking generative semantic communication for multi-user systems with multi-modal LLM," *arXiv:2408.08765*, 2024.
[12] J. Ren *et al.*, "Knowledge base enabled semantic communication: A generative perspective," *IEEE Wireless Commun.*, vol. 31, no. 4, pp. 14–22, Aug. 2024.
[13] A. Alkhateeb *et al.*, "DeepSense 6G: A large-scale real-world multi-modal sensing and communication dataset," *IEEE Commun. Mag.*, vol. 61, no. 9, pp. 122–128, Jun. 2023.
[14] S. Tariq *et al.*, "Deep quantum-transformer networks for multimodal beam prediction in ISAC systems," *IEEE Internet Things J.*, vol. 11, no. 18, pp. 29 387–29 401, Sep. 2024.
[15] U. Khalid *et al.*, "Quantum semantic communications for Metaverse: Principles and challenges," *IEEE Wireless Commun.*, vol. 30, no. 4, pp. 26–36, Aug. 2023.

Brian E. Arfeto is a M.S. student at Kyung Hee University, Korea.

Shehbaz Tariq is a Ph.D. student at Kyung Hee University, Korea.

Uman Khalid is a postdoctoral fellow at Kyung Hee University, Korea.

Trung Q. Duong [F] is a Canada Excellence Research Chair at Memorial University, Canada and adjunct professor at Queen's University Belfast, UK.

Hyundong Shin [F] is a professor at Kyung Hee University, Korea.