Integrating Sustainable Big AI: Quantum Anonymous Semantic Broadcast

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

***Abstract*—Semantic communication (SC) with native artificial intelligence (AI) is a context-centric framework that intelligently extracts task-specific semantics from source data and efficiently regenerates the intended meaning at the destination. Hence, this computing-intensive methodology enables goal-oriented communication by maintaining a high semantic quality of service with a low requirement for data transfer. Recently, the emergence of big-AI foundation models such as the generative pre-trained transformer and diffusion models—with zero-shot task generalization and native cross-modal learning capabilities—has brought a paradigm shift in designing AI-native frameworks for wireless networks. However, deploying big AI in wireless networks involves inherent challenges such as large training parameters and computing requirements. To address these challenges, we use sustainability techniques such as pruning and fine-tuning to create sustainable (lightweight) models from big AI, which can reduce the resource consumption and environmental impact in computation-heavy SC systems while preserving or enhancing the task performance. Moreover, classical communication networks lack quantum-safe communication security and data privacy. In this article, we prototype a sustainable big AI-native quantum anonymous SC system. In this framework, we leverage big-AI models for semantic retrieval processing, i.e., semantic extraction and recovery, and employ quantum anonymous communication protocols to broadcast semantics. We detail the underlying functionalities, sustainable practices, and potential challenges of integrating big AI into a quantum anonymous semantic broadcast system. We also formulate case studies demonstrating the sustainability and reliability of the envisioned framework. This work provides a sustainable and quantum-safe semantic communication framework by integrating big AI and quantum anonymous communication.**

***Index Terms*—Big artificial intelligence (AI), semantic communication (SC), quantum anonymous broadcast.**

I. Introduction

UPON the era of integrating sensing, computing, and communication, enormous volumes of data are perpetually generated and transmitted across a variety of cutting-edge services such as digital twins, autonomous mobility, telehealth, and Metaverse [1]. However, designing communication systems that cater to seamless wireless transmission of large data volumes while maintaining a high-quality user experience is ultimately challenging. SC is a nascent communication paradigm that invokes interpretive abilities of AI to extract intended meanings from data into compact latent representations, enabling efficient and task-oriented information exchanges [2]–[4].

Shehbaz Tariq, Uman Khalid, Brian E. Arfeto, and Hyundong Shin (corresponding author) are with Kyung Hee University, Republic of Korea; Trung Q. Duong is with Memorial University of Newfoundland, Canada and Queen's University Belfast, UK; Shehbaz Tariq and Uman Khalid contributed equally to this paper.

The use of *narrow* AI in the SC framework is fraught with limitations. Primarily, these systems are unimodal, designed to excel within a specific data modality (such as text, images, or audio). While beneficial in certain scenarios, this narrowness restricts their adaptability across diverse data types [5], [6]. Moreover, task-oriented designs restrict flexibility to pivot to varied tasks without extensive retraining. This rigidity can lead to inefficiencies, especially in dynamic environments where communication paradigms are ever-evolving. Additionally, the data-driven nature of narrow AI demands a vast amount of labeled data for effective training, a requirement that becomes resource-intensive and somewhat infeasible [1].

Contrariwise, *big* AI emerges as a more scalable and sustainable solution featuring its inherent multimodal capability [7]–[12]. Big AI is adept at processing, understanding, and generating information across a spectrum of modalities, from texts and visuals to audio and beyond. This ability to seamlessly integrate and interpret diverse data sources offers a richer and more comprehensive understanding, invaluable in multifaceted SC tasks. The big-AI potential to leverage transfer learning further enhances its efficiency, enabling the application of knowledge from one domain to another. In short, narrow-AI models offer specialized solutions within confined boundaries, while big AI promises a holistic, adaptable, and encompassing approach to computation-heavy SC.

Wireless communication systems grapple with inherent security vulnerabilities in lieu of the susceptibility to undetected eavesdropping and the incapacity of classical encryption under quantum computing potentials. Quantum communications herald a paradigm shift, offering a secure framework grounded in the immutable principles of quantum mechanics [13]. A cornerstone of this quantum-safe security is the no-cloning law, which posits the impossibility of creating an exact replica of arbitrary unknown quantum states, thus rendering the intrusion detectable. Furthermore, quantum anonymous protocols ensure the communication anonymity that conceals the identities of communicating parties, providing a quantum layer of untraceable privacy with no classical counterpart (see [14] and references therein). Fig. 1 illustrates the big AI and quantum applications for wireless innovations along with pivotal milestones in the AI evolution. This figure highlights the transformative potentials of integrating big AI and quantum advantages in the realms of computing, networking, sensing, security, and privacy for wireless networks.

Owing to the potential benefits of big AI in the semantic extraction and recovery processes and quantum anonymous communication in ensuring security and privacy, we integrate them to develop a hybrid quantum-classical (HQC) SC system.

The paper is organized as follows.

- Firstly, we briefly summarize the paradigm of computing-intensive communications driven by narrow- and big-AI models along with their potential challenges.
- Subsequently, we delve into techniques of big-AI sustainability. We use pruning techniques and apply style transfers, with a focus on reducing the number of parameters, thereby making AI models more sustainably lightweight.
- Then, we prototype a big AI-native quantum anonymous SC framework. Herein, we detail the underlying essential functionalities, including quantum anonymous broadcast (QAB), and provide sustainable practices. We formulate a case study to demonstrate the envisioned framework.
- Finally, we highlight the central challenges of integrating big-AI models into quantum anonymous SC systems, including future prospects.

II. Computing-Intensive Communications

The scales of AI models are reshaping the paradigm of computing-intensive communication.

A. Narrow-AI SC

Narrow AI-driven SC encompasses computing-empowered communication systems that leverage both classical and noisy intermediate-scale quantum (NISQ) machine learning (ML).

1) ML Semantic Learning: Classical ML methods employed in semantic extraction and recovery processes are classified as follows.

- **Deep Semantic Learning**: Deep learning architectures and methodologies have been developed for SC systems by casting the communication task as an end-to-end information bottleneck problem [3]. The standard approach utilizes autoencoders for joint source-channel coding that intelligently maps image pixel values to channel inputs by capturing semantic patterns from the knowledge base (KB). These autoencoders are trained using cross-entropy and mutual information, which ensure both accuracy and preservation of semantic content. Notably, these architectures are based on convolutional neural networks (CNNs) and transformers. The CNN is mainly used to capture the local context, making it adept at recognizing patterns in short data sequences, while the transformer captures the global context in data and maintains semantic coherence within long data sequences.
- **Reinforcement Semantic Learning**: Reinforcement semantic learning empowers SC systems in challenging scenarios such as the indiscernibility of semantic similarity metrics and the unpredictability of noisy channels [15]. This reinforcement approach allows efficient and stable learning based on user-defined semantic measurements. In contrast to backpropagation, it tackles the indiscernible semantic channel optimization problem using self-critic stochastic iterative updating training on a decoupled semantic transceiver. However, the inability of such frameworks to generalize on diverse scenarios owes to the sparse rewards, non-stationary environments, and sample inefficiency.

2) NISQ Semantic Learning: Integrating quantum computing with semantics-centric communication has led to the development of quantum SC. In the current NISQ age, semantic extraction is performed by employing variational quantum computing or HQC computing, detailed as follows.

- **Quantum Semantic Learning**: Quantum machine learning (QML or quantum ML)—which combines quantum computing with classical ML to exhibit potential quantum advantages—is utilized for semantics learning. The raw data is encoded into variational quantum circuits implementing quantum artificial neural networks (ANNs). The variational circuit is trained over data to learn parameters through classical optimization. Currently, NISQ processing limitations render this method well-suited for only classification tasks involving small amounts of data.
- **HQC Semantic Learning**: The semantic feature learning is facilitated by both QML and classical ML methods. The hybrid models employ quantum CNN layers to formulate a quantum feature map from raw data. This map is then input to classical CNN layers for further processing. Despite NISQ limitations, this hybrid approach exhibits practical advantage over classical counterparts in some detection tasks while handling large amounts of data.

B. Big-AI SC

Big AI embodies a transformative leap in artificial general intelligence, aiming to emulate human-like cognitive capabilities across a vast spectrum of tasks. The evolution is catalyzed by the emergence of foundational models such as the bidirectional encoder representations from transformers (BERT) and the generative pre-trained transformer (GPT). Such models have pioneered the realm of self-supervised learning with multimodal data, enabling functionalities including zero-shot task generalization without specific training. Furthermore, the big AI models—classified into large language models (LLMs), vision foundation models (VFMs), and vision-language pre-training models (VLPMs)—can significantly enhance the semantic extraction and recovery processes, listed as follows.

1) Language Semantic Learning: The LLM is mainly used for the semantic extraction and recovery of textual data. The fine-tuned LLM enables tasks related to context understanding, sentiment analysis, text classification, text completion, and text summarization. The most notable semantic learning techniques with LLMs are as follows.

- **Autoregression**: The autoregression process handles sequences token by token, making predictions for the subsequent token by learning contextual information from the preceding ones. In the context of semantic extraction and recovery, autoregression is pivotal in developing a profound understanding of language semantics. These semantic models are trained over diverse and extensive corpora, enabling them to infer the meaning, context, and relationships within the input textual data. For instance, such capabilities allow ChatGPT to extract semantic information from ambiguous and incomplete sentences and generate coherent and contextually relevant responses.

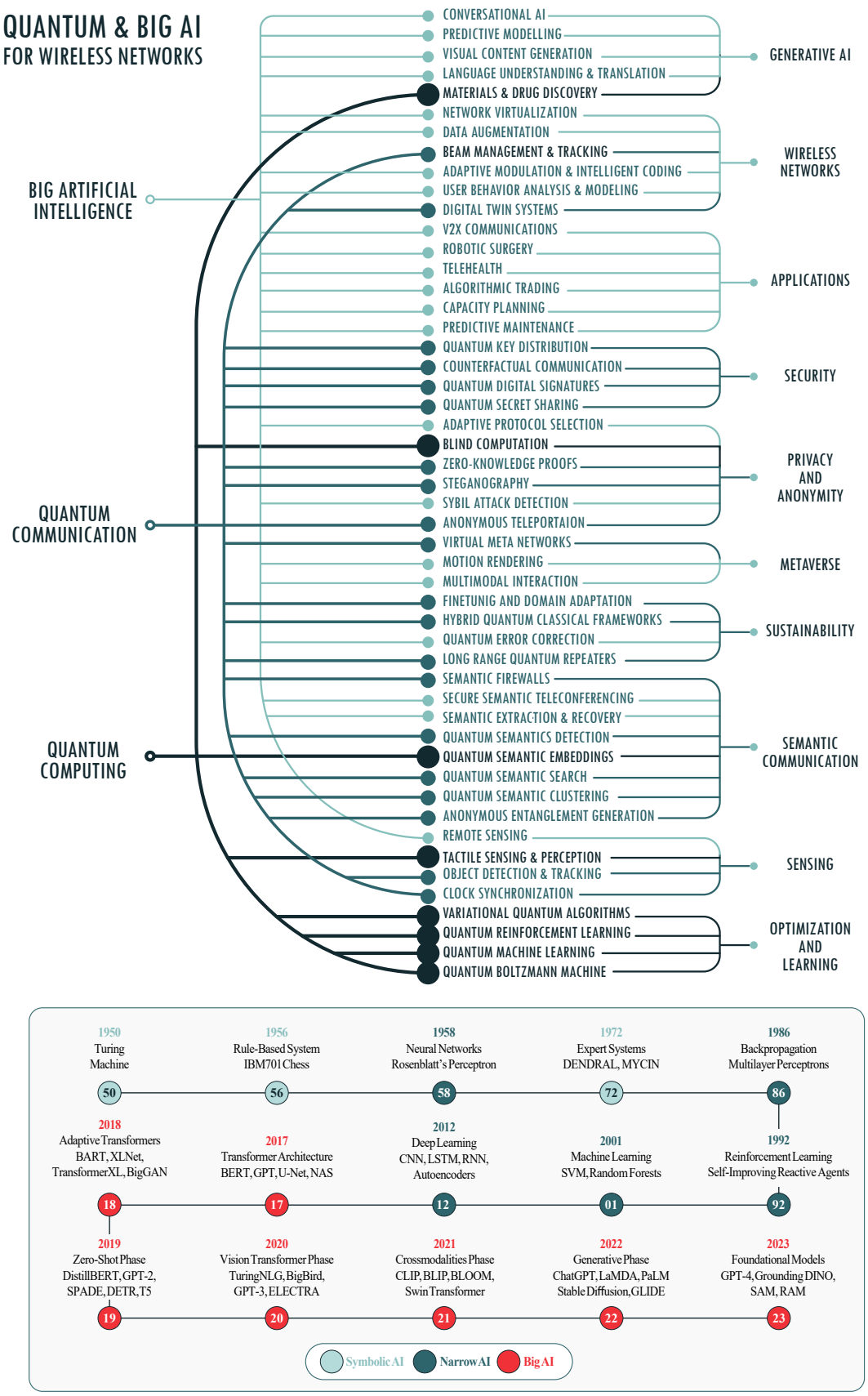


Fig. 1. A technology chart connecting big AI, quantum communication, and quantum computing with features, applications, use cases, and visions of wireless networks. This map underscores the potential integration advantages of big AI and quantum innovations in computing, networking, sensing, securing, and privacy-preserving for wireless solutions. In addition, the timeline traces key milestones in the evolution of AI, delineating the progression of symbolic AI origins, transitioning through narrow AI, and culminating in the current era of big AI (see Table I for all abbreviations).

TABLE I
A LIST OF ABBREVIATIONS

Abbreviation	Definition
AI	Artificial Intelligence
ANN	Artificial Neural Network
AWGN	Additive White Gaussian Noise
BART	Bidirectional and Auto-Regressive Transformers
BEP	Bit Error Probability
BERT	Bidirectional Encoder Representations from Transformers
BigGAN	Big Generative Adversarial Network
BLIP	Bootstrapping Language-Image Pre-Training
BLOOM	BigScience Large Open-science Open-access Multilingual Language Model
CLIP	Contrastive Language-Image Pre-Training
CNN	Convolutional Neural Network
DENDRAL	Dendritic Algorithm
DETR	Detection Transformer
DIV2K	Diverse 2K Resolution
DNN	Deep Neural Network
ELECTRA	Efficiently Learning an Encoder that Classifies Token Replacements Accurately
FID	Fréchet Inception Distance
GAN	Generative Adversarial Network
GHZ	Greenberger–Horne–Zeilinger
GPT	Generative Pre-trained Transformer
GPU	Graphics Processing Unit
HQC	Hybrid Quantum-Classical
KB	Knowledge Base
KPI	Key Performance Indicator
LaMDA	Language Model for Dialogue Applications
LLM	Large Language Model
LoRA	Low-Rank Adaptation
LPIPS	Learned Perceptual Image Patch Similarity
LRTD	Low-Rank Tensor Decomposition
LSTM	Long Short-Term Memory
METEOR	Metric for Evaluation of Translation with Explicit Ordering
ML	Machine Learning
MOS	Model Output Statistics
NAS	Neural Architecture Search
NISQ	Noisy Intermediate-Scale Quantum
PISQ	Perfect Intermediate-Scale Quantum
PaLM	Pathways Language Model
QAB	Quantum Anonymous Broadcast
QML	Quantum Machine Learning
QPU	Quantum Processing Unit
SAM	Segment Anything Model
SC	Semantic Communication
SNR	Signal-to-Noise Ratio
SPADE	Spatially-Adaptive Denormalization
SVM	Support Vector Machine
T5	Text-to-Text Transfer Transformer
TransformerXL	Extra Long Transformer
TuringNLG	Turing Natural Language Generation
U-Net	U-Shaped Network
VFM	Vision Foundation Model
VLPM	Vision-Language Pre-training Model
XLNet	Extra Long Network
YOLO	You Only Look Once

This makes it a powerful tool for various natural language understanding and applications.

- **Masked Autoencoding**: The masked autoencoding process operates by masking specific segments of the input textual data. It then predicts these masked words by learning the surrounding bidirectional context to enhance its understanding of their semantics. This parallel processing approach enables the model to develop a rich and nuanced understanding of word relationships and contextual meanings. The semantic learning with masked autoencoding employs big-AI models (e.g., BERT) to discern implicit meanings and relationships in given texts, making it adept at tasks such as entity recognition, relation extraction, and semantic role labeling.

2) Vision Semantic Learning: The VFM is mainly used for extracting semantics from visual (e.g., image and video) data, enabling it to perform tasks such as image recognition, object detection, image segmentation, visual understanding, and even generating images. The key vision semantic learning tasks are as follows [10], [11].

- **Localization and Recognition**: Localization focuses on spatially identifying specific regions or entities within

visual data using their distinctive attributes, such as color, intensity, shape, or texture. Recognition aims to understand, interpret, and categorize visual data. Their primary objective is determining the boundary of prominent object features and classifying objects within images. For instance, the segment anything model (SAM)—a promptable image segmentation model for zero-shot segmentation tasks—is utilized in semantic extraction to partition an image into coherent segments, each representing a distinct region of interest (RoI). This model allows for the segmentation of image data with just a simple prompt, thereby eliminating the need for labor-intensive image labeling and domain expertise.

- **Synthesis**: Image synthesis refers to generating new and unseen images, typically by leveraging the VFM to create visually coherent and contextually relevant images [12]. In the VFM, a user provides a prompt, such as a textual description or a set of attributes, and the model translates this prompt into a coherent visual representation. This model involves generative adversarial networks (GANs) in semantic recovery to interpret the semantic content of the prompt and synthesizes an image that visually represents the described scene, ensuring that the generated image aligns with the contextual and visual elements detailed in the prompt.

3) Vision-Language Semantic Learning: The VLPM is a fusion of learning with both the LLM and VFM. The process can be designed to understand and generate semantics for both textual and visual data. The VLPM is trained on large datasets containing paired images and text information and deals with semantic labeling, image captioning, visual question answering, and output textual semantics for images.

- **Contrastive Learning**: It involves comparing a positive pair (similar or related items) against a negative pair (dissimilar or unrelated items) and optimizing the model to bring similar items closer in the embedding space while pushing dissimilar items apart. By learning to associate visual features with linguistic features effectively, the contrastive learning model gains a deeper understanding of the semantic content of visual data. This can be particularly useful in tasks such as image captioning, visual question answering, and multimodal translation, where understanding the semantics is a primary concern.
- **Multimodal Learning**: It improves the semantic learning process by fusing semantic features extracted from different input data types to create a unified representation that captures the shared and complementary information across different modalities. This learning enables models to leverage the rich, high-level semantic information from text and the detailed, perceptual information from images to understand the intricate interplay between vision and language. The most notable models capable of understanding both images and text in the context include the bootstrapping language-image pre-training (BLIP), making it suitable for tasks such as image-text retrievals and textual descriptions for images.

C. Potential Challenges

The potential challenges in computing-intensive communication systems are summarized as follows.

1) Security and Privacy: Classical SC systems face challenges in ensuring the security and privacy of semantic data. These systems are vulnerable to privacy leakages and sophisticated threats, from man-in-the-middle to replay attacks. Moreover, classical encryption methods are increasingly vulnerable due to the lack of adequate defense mechanisms to counter the potential of emerging quantum computing. In contrast, quantum protocols yield quantum leaps in security and privacy. For example, counterfactual quantum communication facilitates information transfer without transmitting physical particles, ensuring semantic data remains safeguarded from potential semantic attacks. In addition to the confidentiality of semantic data, quantum anonymous protocols introduce a quantum layer of privacy, enabling the communicating parties to remain anonymous and untraceable [14].

2) Zero-Shot KBs: A KB is a repository shared by the SC transceiver that contains essential information and rules, offering context for semantic interpretation and efficient processing. Most notably, KBs exist as knowledge graphs, databases, and trained parametric or non-parametric models. The formation of the KB is a resource-intensive process that requires domain expertise for semantic extraction and hardcoded labeling from the underlying source data. In a dynamic environment where AI models perform various semantic tasks simultaneously, employing narrow AI becomes challenging, as it requires frequent knowledge updates and has limited knowledge representation due to parameter restriction. Contrarily, big-AI models—with billions of parameters—are trained on a large amount of multimodal data from diverse knowledge domains, forming a universal KB. Hence, the big-AI models natively possess zero-shot learning that enables them to perform various tasks and extract semantics without a predefined KB. This ability makes it suitable for embedding in resource-constrained edge devices. Moreover, data sources in distributed systems with diverse devices and sensors are heterogeneous, noisy, and complex. A multimodal zero-shot KB thus offers a sustainable and efficient solution to streamline semantic tasks in these systems by exploiting correlated representations and data isomerism among different modalities.

3) Adaptability: Adaptability refers to the capability of AI models for task specialization and domain adaptation. Task specialization deals with foundation models that perform some tasks within specific modalities, whereas domain adaptation involves the model's ability to apply knowledge learned from a specific data distribution to another distribution (style or dataset). This capability is particularly crucial for big-AI models in broadening their versatility far beyond the original training scope. Herein, a domain-shift process (such as multi-task and transfer learning) is employed by leveraging a KB gained from a source domain to improve the model performance in a target domain with different data distributions. On the other hand, narrow-AI models are pre-destined to learn from the data by optimizing the underlying parameters based on the learning pattern. Therefore, the narrow-AI models are inherently trained to handle singular and limited distributions.

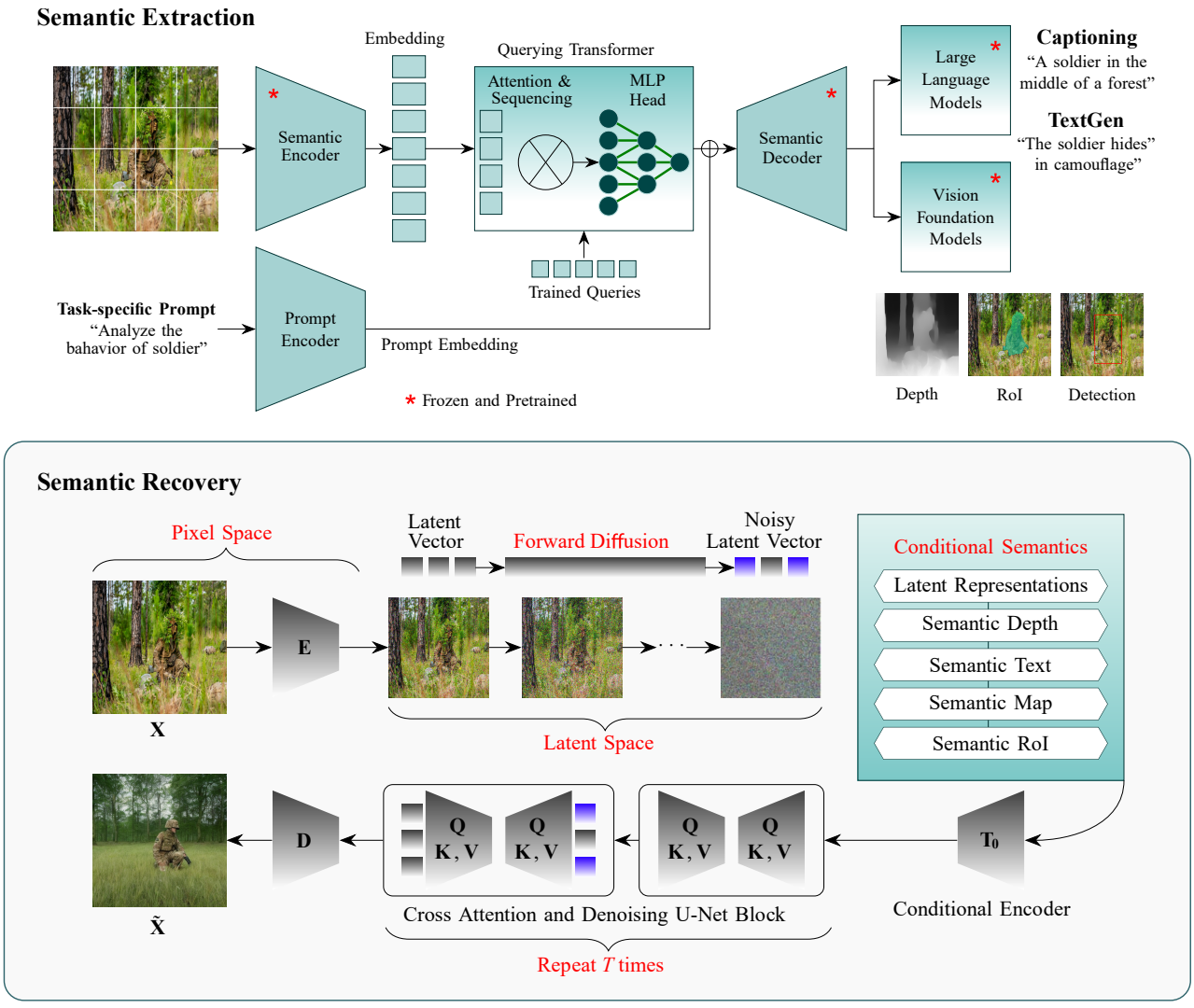
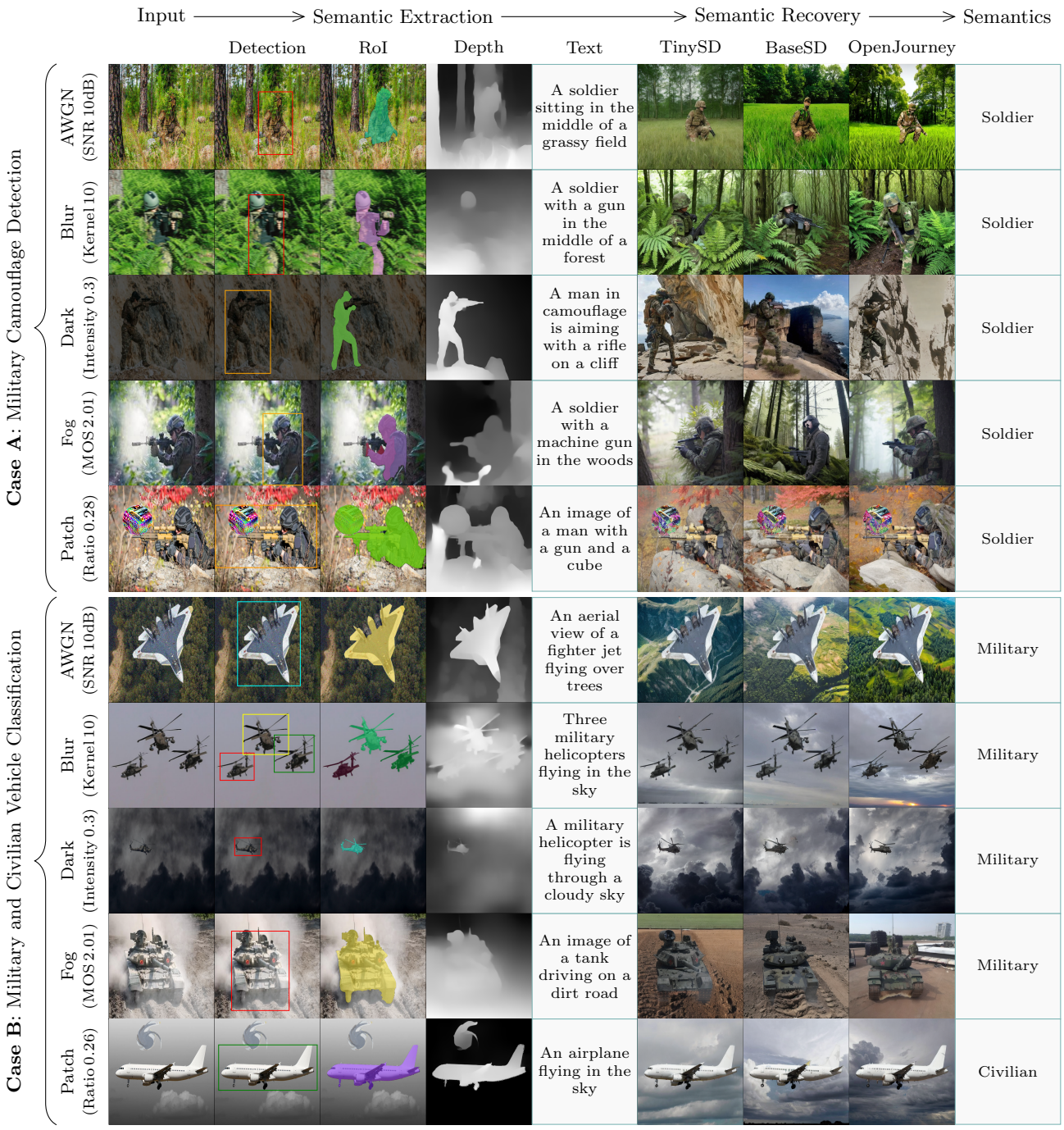


Fig. 2. A framework for big AI-native quantum anonymous semantic broadcast. For the first phase, semantic extraction consists of pre-trained big-AI models of a semantic encoder, a querying transformer, and a semantic decoder. Note that the prompt encoder is optional and only used in some specific cases (e.g., visual question answering) along with predetermined tasks. The caption and extracted semantics are then fed forward and used as diffusion-guided inputs in the next semantic recovery process. For the second recovery phase, an image is encoded into latent vectors, and then diffusion noise is gradually added within the interval time following a Markov chain. Conditional semantics represent the target output of forward diffusion, depending on the semantic recovery task. At the end, a switch is employed to determine the pipeline interval of U-Net on the reverse diffusion process. In the last phase, the recovered semantic information is broadcast to network parties using the quantum anonymous broadcast protocol. This quantum protocol ensures that a broadcasting party (Alice) remains anonymous and untraceable by virtue of the preshared multipartite entanglement and broadcast modulation.



(a) Military camouflage detection (Case A) and military-civilian vehicle classification (Case B)

Fig. 3. Case Study I: Semantic extraction and recovery using big-AI models. All simulations are performed with NVIDIA GeForce RTX 3090 Ti GPU with 24 gigabytes (GB) memory. (a) For two exemplary tasks (military camouflage detection and military-civilian vehicle classification), we test the robustness of our envisioned framework for extracting semantics from corrupted images subjected to various types of occlusions—e.g., AWGN, motion blur, dark brightness, synthesized fog, and adversarial patch. Textual semantics are obtained by BLIP 2, while we employ SAM (ViT-h) and YOLO v8 concurrently for the automated RoI (segmentation) and detection. YOLO detects RoIs in source images, which are given as a prompt to SAM for segmentation. Moreover, the image depth is extracted by a dense prediction transformer. These semantics are then processed by semantic recovery models (TinySD, BaseSD, and OpenJourney) for comparative analysis. TinySD is proposed for a sustainable model by incorporating the LoRA and pruning in the standard stable diffusion model (BaseSD) for semantic recovery. The background is created using text-only input, while the foreground is reconstructed by recovery models conditioned on the extracted segments. This approach uses text for background generation to minimize communication overhead and leverages the generative capabilities of large-scale AI at the receiver's end for semantically accurate information recovery. We use the Aiming soldiers image dataset and the Military and Civilian Vehicles classification dataset for Cases A and B, respectively.

These models encounter limitations, e.g., the lack of universality in the KB and the risk of overfitting to the specific training data. Moreover, the domain shifting in narrow AI requires a large amount of labeled data from both the source and target

domains, which may not be readily available. This ultimately leads to poor generalization and adaptation for the narrow-AI models. With regards to semantic learning, the big-AI models offer a broader understanding of semantic tasks and can easily adapt to crossmodalities. In addition, the domain shifting on big AI requires less data and computational resources since the underlying models feature a universal KB. For example, a big-AI model trained on a large neural network of multimodal data can be fine-tuned on a smaller corpus of medical texts to perform medical diagnosis. The efficient adaptability of big-AI models renders them versatile and robust semantic learners for a wide range of applications.

4) Sustainability: Training big-AI models for semantic learning is resource-intensive and costly. With their vast number of parameters, these models can be slow, inefficient, and environmentally taxing, especially when online semantic learning is required. In this case, pruning offers a solution to this challenge. By eliminating the less important weights or neurons from a neural network, the pruning can substantially reduce the model's size without compromising its performance significantly. The pruned models, having fewer parameters, demand decreased computational resources, resulting in faster inference times and reduced energy consumption. On the other hand, fine-tuning offers a sustainable practice by adjusting pre-trained models to ensure that the pruned model remains effective, addressing specific semantic challenges with precision. The prune-fine-tuned models—tailored for specific semantic challenges—can process information faster, more accurately, and with reduced environmental impacts. Hence, these models not only optimize computation but also serve sustainability, making big AI increasingly important for semantic learning.

III. Big AI-Native Quantum SC

In this section, we prototype a big AI-empowered quantum anonymous SC system (see Fig. 2), where the semantics of multimodal data are retrieved by big-AI models and broadcast by the QAB protocol.

A. Anonymous SC

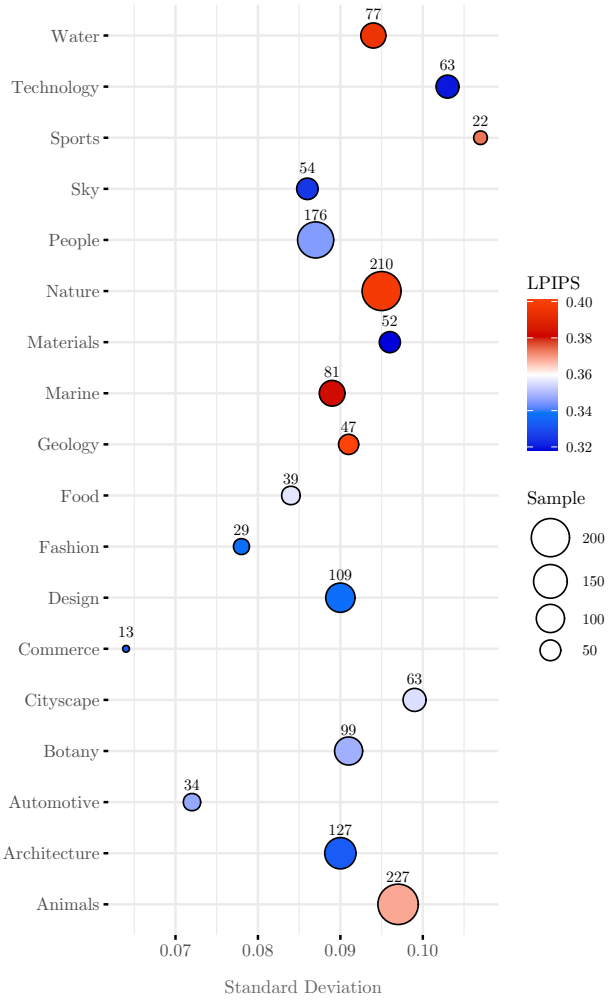
Now, we detail the key functional elements of an anonymous SC transceiver.

1) Semantic Extraction: The semantics of multimodal data are embedded, queried, and decoded using big-AI models as follows.

- **Semantic Representation Embedding**: Raw multimodal input data is first embedded into a sequence of vectors in a suitable format for subsequent processing. This representation embedding is primarily classified into patch and prompt embedding. In the patch embedding, raw data is divided into fixed-size patches. These patches are then linearly embedded into vectors, which serve as the initial input tokens for subsequent layers. It enables the model to handle data in manageable chunks, allowing for efficient parallel processing. On the other hand, the prompt embedding additionally utilizes predefined vectors to assist the model with semantic context and ensure that subsequent

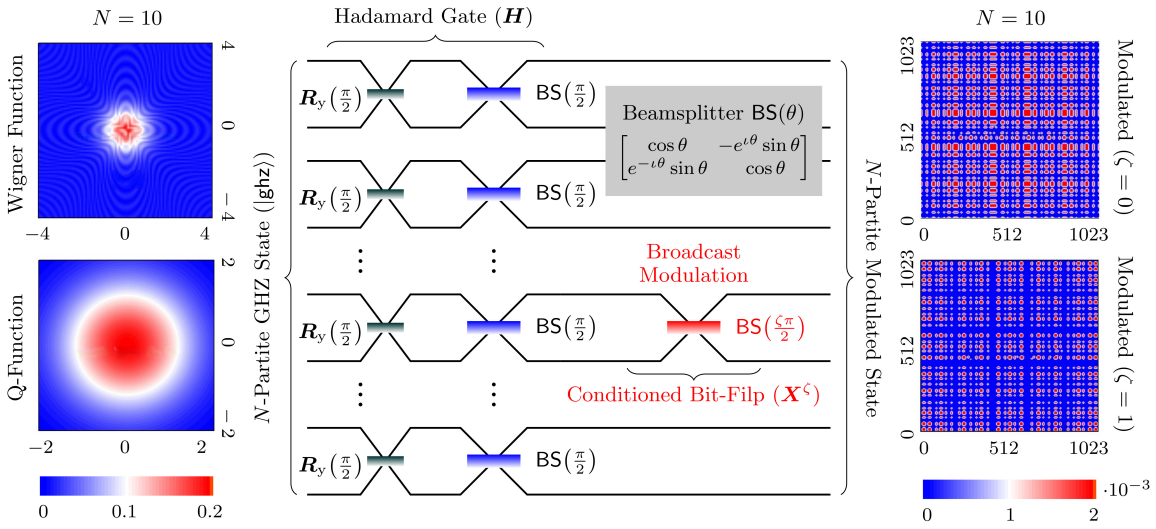
Model	KPI	TinySD	BaseSD	OpenJourney
Sustainability	Parameters [10^6]	**906**	1,920	1,060
	Memory [GB]	**0.337**	1.791	1.784
	Time [s]	**1.627**	1.644	1.660
	Power [J]	**0.588**	0.999	1.034
	Storage [GB]	**3.972**	7.174	6.021
Case A: Military Camouflage Detection	LPIPS	**0.299**	0.300	0.305
	FID	**21.301**	22.369	24.519
	METEOR	0.533	0.537	**0.545**
	BERT-S	0.958	0.957	**0.959**
Case B: Military and Civilian Vehicle Classification	LPIPS	0.261	**0.257**	0.264
	FID	11.264	10.874	**9.847**
	METEOR	**0.600**	0.596	0.580
	BERT-S	**0.966**	0.965	0.963

(b) Sustainability and similarity KPIs of semantic recovery

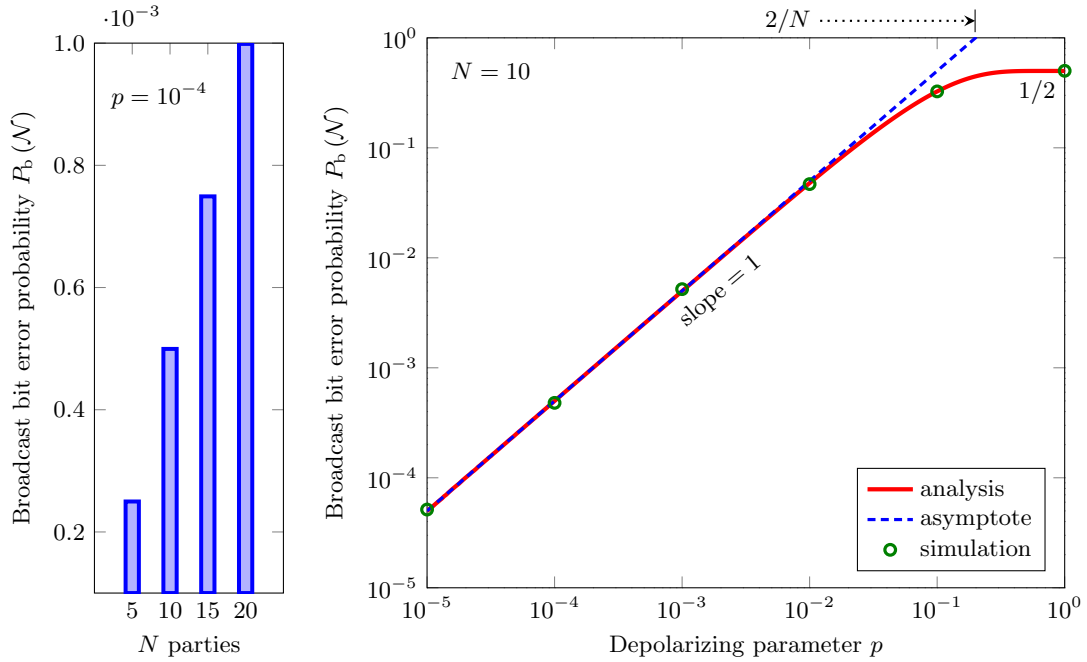


(c) Zero-shot semantic recovery for TinySD

Fig. 3. *(Continued.)* Case Study I: Semantic extraction and recovery using big-AI models. All simulations are performed with NVIDIA GeForce RTX 3090 Ti GPU with 24 GB memory. (b) Sustainability and similarity KPIs of semantic recovery are tabulated for TinySD, BaseSD, and OpenJourney models on the datasets for Cases A and B under AWGN occlusions at the SNR of 10 dB. To benchmark the precision metrics, such as LPIPS, FID, METEOR, and BERT similarity (BERT-S), we test the models on both image-image and text-image semantic recovery tasks. (c) Zero-shot semantic recovery is depicted for TinySD on the unseen DIV2K dataset under AWGN at the SNR of 10 dB. DIV2K dataset contains 1000 high-resolution images with multiple categories. The images in the dataset are initially captioned and categorized by BLIP 2, and then the recovery similarity is tested in terms of LPIPS.



(a) Broadcast modulation



(b) Broadcast BEP under depolarizing quantum noise $\mathcal{N}$

Fig. 4. Case Study II: Semantic anonymous broadcast using a preshared N-qubit GHZ state $|\mathsf{ghz}\rangle$. (a) The broadcast modulation for semantic information ζ in a classical bit is illustratively implemented using linear optical elements. For $N = 10$, we plot the Wigner and Husimi Q-functions for the 1024×1024 dimensional density operator $\boldsymbol{\rho} = |\mathsf{ghz}\rangle\langle\mathsf{ghz}|$ of the GHZ state as a function of the phase-space parameters using the QuTiP (quantum toolbox in Python) package (left). The tomography of the corresponding modulated state is also depicted for both $\zeta = 0$ and 1 (right). (b) The broadcast BEP $P_{\mathrm{b}}(\mathcal{N})$ for the QAB protocol is evaluated under the isotropic depolarizing noise $\mathcal{N}(\boldsymbol{\rho}) = q\boldsymbol{\rho} + \boldsymbol{I}p/2$ with the noise parameter p where $q = 1 - p$. Since the quantum state $\boldsymbol{\rho}$ is depolarized with probability p, we have $2P_{\mathrm{b}}(\mathcal{N}) = 1 - q^N$ and its asymptote $2P_{\mathrm{b}}(\mathcal{N}) = pN + o(p)$ as $p \to 0$. The broadcast BEP $P_{\mathrm{b}}(\mathcal{N})$ is plotted as a function of the number N of network parties when $p = 10^{-4}$ (left) and the depolarizing parameter p when $N = 10$ (right). For simulations, we use the NetSquid (a discrete-event simulator designed for quantum networks).

layers process data in a manner aligned with the desired outcome.

- **Semantic Querying Transformer**: Semantic transformers utilize multi-head attention, sequencing, and feed-forward fully-connected neural network layers to further process the tokenized embedded vectors. The multi-head attention mechanism evaluates the intricate correlations among the embedded patches. The sequencing preserves the spatial context of the embedded data by introducing a special token at the start of a sequence, thereby including positional embedding in addition to patch embedding. Finally, the fully-connected layers further refine the semantic features by applying nonlinear transformations such that the features tend to be more representative of

the distinctive patterns. These transformations are carried out under the pretext of pre-defined tasks. Such tasks are monitored by the pre-trained querying vectors corresponding to image captioning, semantic segmentation, depth estimation, and object detection.

- **Semantic Attribute Decoding**: Semantic attribute decoding is task-specific. Herein, embeddings are transformed into the desired semantic attributes. For example, in the context of images, the decoder extracts masks and depth information; marks the RoI; and generates a contextually relevant descriptive text. The decoding process leverages rich semantic information encapsulated in the embeddings to produce vectors representing text tokens for language-based decoders and pixel values for vision-based decoders. The choice of the decoder architecture depends on the extraction model employed at the source. For example, in the case of BLIP, which is a language-based extraction model that uses an LLM to generate embeddings from natural language, the semantic decoder is also an LLM that can generate natural language from embeddings. In contrast, in the SAM case, a vision-based extraction model that uses a VFM to generate embeddings from images, the semantic decoder is a mask decoder that can generate masks from image embeddings. Therefore, the semantic decoder should match the modality and format of the source and destination data.

2) Semantic Recovery: The extracted semantics are subjected to extensive regenerative processes to refine semantic fidelity. Notable semantic recovery modules rely mainly on diffusion models, variational autoencoders, and GANs. Diffusion models are a class of latent variable models that learn to generate data by reversing a Markov chain that gradually adds noise to the data. The Markov chain is defined by a score function that measures the likelihood of the data given the noise level. The diffusion models are trained beforehand to learn the reverse process and thus produce high-fidelity image data from semantic text descriptions by iterating backward over this chain. Variational autoencoders are ANNs that combine elements of both autoencoders and variational inference to recover the desired content. Herein, the received semantics are mapped to a probabilistic distribution in the latent space, enabling the generation of new data samples while allowing for smooth interpolation between different data points. These data samples are mapped back to pixel values in regenerating diverse images and variationally minimizing the reconstruction loss in terms of pixel-wise mean squared errors to restore high-fidelity images. Meanwhile, GANs consist of two ANNs, e.g., the generator and discriminator, which are trained together through a competitive process. During training, the generator improves its ability to generate realistic data by trying to deceive the discriminator, creating high-fidelity semantic data. The semantic recovery process with these modules is classified as follows.

- **Language Semantics to Image Recovery**: In diffusion models, the recovery process starts with a noisy image and iteratively refines it through the reverse Markov chain process guided by the textual semantics-conditioned score function. This processing outputs an image that not only appears closely related to training data but also aligns closely with the textual semantics, thereby restoring and synthesizing images from textual descriptions through forward inference. In the case of variational autoencoders, the encoder probabilistically maps the textual semantic data to the latent space by employing textual embeddings such as BERT. The decoder is trained to reconstruct the output images from a sample drawn from the latent distribution. Therefore, by conditioning the decoder on the textual input, the variational autoencoders learn to produce images that are highly consistent with the textual semantic description. In GANs, the generator network is conditioned on textual semantic features embedded by a character-level CNN. The CNN transforms the textual descriptions into rich high-level encoding, which guides the image generation process. The discriminator evaluates the authenticity of generated images in relation to both their appearance and semantic relevance to the textual descriptions.
- **Vision Semantics to Image Recovery**: In utilizing diffusion models to restore images from vision semantics, the score function employs these semantics as conditioning information and guides the reverse Markov chain. This processing restores high-fidelity images that align closely with visual semantics. In the variational autoencoders, the encoder maps the vision semantics to a probabilistic distribution in the latent space while the decoder reconstructs the output image from a sample drawn from this latent distribution. Herein, the loss function comprises two terms, i.e., a reconstruction loss that measures the pixel-wise difference between the input and output images and a Kullback-Leibler divergence loss that measures the difference between the latent distribution and the *a priori* standard distribution. In GANs, visual semantics are input into the generator that transforms them into the target image style. The discriminator attempts to distinguish between real and fake images from the target domain and the generator, respectively. This adversarial setup drives the GAN model towards generating increasingly realistic and high-quality images that closely resemble the input.

3) Quantum Anonymous Broadcast: We consider that the network consists of N parties, including Alice (see Fig. 2). The QAB protocol allows Alice (or any network party) to anonymously broadcast her recovered semantics in a classical bit to all other parties in the network without revealing her identity. This anonymous and untraceable broadcast is also crucial for anonymous teleportation. Specifically, using a preshared N-partite maximally entangled state, the QAB protocol takes the Hadamard-basis measurement, bit-flip operation, classical announcement, and modulo 2 sum calculation for anonymous broadcast of semantic information.

- **Preparation**: All the N parties in the network, including Alice, initially share an N-partite entangled GHZ state. By applying the Hadamard gate $\boldsymbol{H}$ to the first qubit and then sequentially performing controlled-NOT, e.g., controlled Pauli $\boldsymbol{X}$ (bit-flip) gates between the first qubit

(control) and all successive qubits (target), the N-qubit system is prepared in this N-partite maximally entangled state. These entangled qubits can be distributed across the network parties using quantum teleportation and quantum repeaters.

- **Broadcast Modulation**: All the network parties (including Alice) start the protocol by performing the Hadamard operation $\boldsymbol{H}$ on their respective qubits. Alice (broadcaster) performs $\boldsymbol{X}^{\zeta}$ on her qubit to modulate the broadcast semantics ζ in a classical bit, i.e., the broadcaster applies the bit-flip Pauli operator $\boldsymbol{X}$ if $\zeta = 1$ and leaves the state as it is, otherwise. The other network parties apply the identity operator $\boldsymbol{I}$ on their qubits, i.e., leave the qubit states as they are. Due to the Hadamard and bit-flip operations, this modulated state is in even superposition of all 2^{N-1} N-qubit states whose exclusive OR (XOR) or modulo 2 sum is equal to the broadcast semantics ζ.
- **Broadcast Detection**: All the N network parties measure their qubits in the computational basis and get their binary outcomes $\mu_1, \mu_2, \ldots, \mu_N$. The XOR (modulo 2 sum) of all these measurement outcomes is equal to the broadcast information ζ—due to the symmetry of the modulated state from Alice's bit-flip operation. These N-tuple binary outcomes appear randomly with an equal probability of $1/2^{N-1}$ due to the basis change from the Hadamard operations, even for the entangled state between the N network parties. This randomness completely conceals the fact that Alice has broadcast the semantics bit ζ by bit-flipping her qubit state. Now, all the parties utilize classical communication to announce their measurement outcomes. Finally, any recipient party calculates the modulo 2 sum of all announced measurement outcomes to recover the broadcast bit (semantics) without revealing the broadcaster's identity, i.e., Alice—thus preserving anonymity and untraceability in the broadcast process.

B. Sustainable Practices

Various sustainable practices can be incorporated into our proposed big AI-native quantum anonymous SC system.

1) Low-Rank Adaptation: The low-rank adaptation (LoRA) is a sustainable practice to reduce memory consumption by fine-tuning big-AI models for semantic retrievals. Fine-tuning refers to adjusting the parameters of a pre-trained model using domain-specific raw data. This tuning enables the model to adapt to domain-specific tasks while retaining general knowledge from initial training. The LoRA utilizes a low-rank approximation technique to reduce matrix complexity and dimensionality. In practice, the LoRA freezes original weights during fine-tuning and trains only a much smaller number of adaptable parameters obtained by the low-rank approximation, thereby improving training efficiency without increasing inference latency.

2) Pruning: Pruning is a sustainable practice that reduces the computational complexity of big-AI models by limiting over-parameterization in model training without affecting its semantic retrieval performance. It is categorized into structured and unstructured pruning. The structured pruning removes entire channels, filters, and neurons, while the unstructured pruning eliminates individual weights. However, aggressive pruning can degrade the model performance, which can be counterbalanced by strategic fine-tuning.

3) Knowledge Distillation: By incorporating the knowledge distillation in the big-AI framework, a larger pre-trained model (teacher module) is utilized to train a smaller model (student module) for semantic extraction. Herein, the student module learns by minimizing the loss function in terms of both the ground truths and predictions made by the teacher module. This process leverages the class probability distribution and the softmax function from the teacher module.

4) Low-Rank Tensor Decomposition: The low-rank tensor decomposition (LRTD) assists in reducing memory usage in textual semantics to image restoration without compromising the semantic recovery performance. It approximates multiple CNN layers with fewer components by finding an approximation of original high-dimensional tensors with fewer components, thus decreasing the memory requirement while maintaining system efficiency.

5) Quantization: The quantization is utilized in reducing the computational and memory requirements while extracting vision semantics from image data without impacting the semantic extraction performance. Herein, by decreasing the bit-width representation of the numbers for weights and biases in the network, the number of distinct pixel intensity levels is reduced. This results in lower precision calculations, which are computationally inexpensive and require less memory for storage. However, an excessive reduction in precision can degrade the semantic extraction performance significantly.

C. Case Study

We present a case study demonstrating the sustainable big AI-driven quantum anonymous SC framework.

1) Semantic Extraction and Recovery: We show an illustrative example of semantic extraction and recovery using big-AI models on the Aiming soldiers image dataset (Case A), the Military and Civilian Vehicles classification dataset (Case B), and the diverse 2K resolution (DIV2K) dataset in Fig. 3. To extract semantics from image data, we employ a combination of robust big-AI models, e.g., BLIP, SAM, and you only look once (YOLO), as shown in Fig. 3(a). We propose *TinySD* by incorporating sustainable practices, e.g., LoRA and pruning, in the standard stable diffusion model (BaseSD) to recover the image from the extracted semantics. As the big-AI models are computing-intensive, it is imperative to make them lightweight for sustainable deployment in wireless networks. Herein, we use the TinySD model (a lightweight variant of StableDiffusion) obtained through fine-tuning followed by structural pruning. In the fine-tuning step, we apply the LoRA that unfolds in three distinct phases. In the first phase, it employs the prior preservation of class images and sparse tokens to regularize the training process, thus preserving the model generalization capability while achieving high semantic fidelity. Then, it uses the inversion technique to instantiate new tokens, which learns the token embedding with the gradient descent. Lastly, the token embeddings are coupled with the prior preservation to

fine-tune the model. In the pruning step, we compress the U-shaped network (U-Net)—which is the most computation-heavy component in the architecture—of the StableDiffusion model to reduce the parameters. The U-Net performs multiple denoising steps on the latent representations conditioned by semantics and time-step embeddings. At each stage, it produces the noise residual to compute the next latent representation. We reduce this computation by using block-level elimination and feature pruning. Although the pruning degrades the recovery similarity, the LoRA applies style transfer and maintains high semantic fidelity at the output.

Fig. 3(a) shows the semantic recovery tests for the TinySD, BaseSD, and OpenJourney models with corrupted images in the two exemplary military-related tasks (detection and classification). The key performance indicators (KPIs) of sustainability and similarity are also tabulated in Fig. 3(b) for these semantic recovery models on the datasets for Cases A and B under additive white Gaussian noise (AWGN) occlusions. The learned perceptual image patch similarity (LPIPS) calculates the perceptual similarity between two images while the Fréchet inception distance (FID) compares the distribution between two data sets. For textual semantics, the metric for evaluation of translation with explicit ordering (METEOR) represents the alignment between the generated and referenced text, while the BERT similarity shows the cosine similarity of embedding vectors. Fig. 3(c) depicts the LPIPS for the zero-shot semantic recovery of the TinySD model on the unseen DIV2K dataset under AWGN occlusions.

2) Semantic QAB: Fig. 4 demonstrates the broadcast modulation and bit error probability (BEP) for the semantic QAB using a preshared GHZ state involving N entangled qubits. The semantic information retrieved by big-AI models is modulated using the Hadamard and bit-flip operations and broadcast anonymously in the QAB protocol. For example, the semantic recovery of soldier intrusion or a military vehicle for Case A or B is alerted by the broadcast bit $\zeta = 1$ to all network parties without revealing the identity of a broadcasting party.

Fig. 4(a) depicts an illustrative implementation of the broadcast modulation for the semantic information ζ using linear optical elements (mirrors and beamsplitters). To modulate the semantics ζ in the preshared GHZ state, the Hadamard optical gate $\boldsymbol{H}$ and the conditioned bit-flip Pauli $\boldsymbol{X}^{\zeta}$ operation are implemented using a combination of the Pauli-y rotation mirror $\boldsymbol{R}_{\mathrm{y}}(\pi/2)$ and the symmetric beamsplitter $\mathsf{BS}(\pi/2)$, and the symmetric beamsplitter $\mathsf{BS}(\zeta\pi/2)$, respectively. Specifically, we have $\mathsf{BS}(\pi/2)$ (bit-flip Pauli $\boldsymbol{X}$ optical gate) for $\zeta = 1$, whereas $\mathsf{BS}(0)$ (identity $\boldsymbol{I}$ optical gate) for $\zeta = 0$. In optical quantum setups, the density matrix of a quantum state can be represented in the phase-space formalism using the Wigner and Husimi Q-functions (quasi-probability distributions). For $N = 10$, we plot the Wigner and Husimi Q-functions for the N-qubit GHZ state and the tomography of the corresponding modulated state for the semantic information $\zeta = 0$ and 1. Fig. 4(b) shows the broadcast BEP for the QAB protocol under isotropic depolarizing noise in the multipartite entanglement distribution across the network. The depolarizing noise is a completely positive trace-preserving map that transforms a quantum state into a linear combination of itself and a completely mixed state. The depolarizing noise parameter p denotes a probability that a quantum state is *depolarized*, i.e., completely lost and evolves into the completely mixed state—while left untouched (noiseless) with probability $1-p$. We plot the broadcast BEP as a function of the number N of network parties when $p = 10^{-4}$ and the depolarizing parameter p when $N = 10$. The broadcast error performance degrades with increasing both the quantum noise degree and the broadcast network scale.

IV. Challenges and Opportunities

In this section, we outline some key challenges in integrating big AI into a quantum anonymous communication system. We also highlight the prospects of employing such a HQC SC framework.

A. NISQ Limitations

In the NISQ era, quantum computing and communications systems face the following challenges.

1) Computation: NISQ computing systems still lack in exhibiting practical advantage in the semantic learning process due to reasons, for example.

- **Quantum Fidelity**: The fidelity is an essential measure of quantum computing accuracy. NISQ computers generally exhibit the quantum gate fidelity around 99.9 %, while perfect intermediate-scale quantum (PISQ) computers are expected to surpass 99.999 %. The lower gate fidelity in NISQ devices can cause errors in quantum computations, making it challenging to achieve reliable computations for intricate problems. The higher gate fidelity is crucial to expand the practicality of quantum computing across various domains, as quantum errors accumulate in complex quantum algorithms, limiting their utility.
- **Quantum Volume**: A quantum volume is another crucial metric for assessing the computational power of quantum computing machines. The NISQ computers typically have a quantum volume in the range from 10^3 to 10^4, whereas the PISQ computers are projected to achieve a quantum volume of more than 10^5. The relatively low quantum volume of NISQ devices limits their ability to perform sophisticated computations efficiently. This limitation signifies that many practical applications requiring substantial quantum computational power remain inaccessible until PISQ devices are deployed.

2) Communication: NISQ communication is relatively developed and practically useful as compared to NISQ computing. However, some essential limitations are listed as follows.

- **Coherence Time**: The coherence time is the duration a qubit maintains its coherent superposition state before the information is lost to the environment, causing decoherence. The coherence time of NISQ communication bits is limited due to its sensitivity to quantum noise. This limitation ultimately reduces the efficiency, security, and privacy of NISQ communication protocols. Perpetual developments in quantum modalities exhibiting longer qubit coherence times suggest the prospects of secure and privacy-protecting PISQ networks.

- **Entanglement Range**: The reliability of NISQ communication is limited by distance. This distance limitation is caused by quantum channel noise, resulting in entanglement degradation during its distribution in an entangled quantum network. This degradation limits the entanglement range of NISQ networks. However, the ongoing significant improvements in quantum memory, quantum repeaters, and quantum error correction would noticeably advance the entanglement range of PISQ networks.

B. Big-AI Limitations

Despite being more practical than NISQ computing for semantic retrievals, big-AI models face some challenges that cause sustainability issues.

1) Computation: Big AI-native systems are computationally constrained due to reasons, for instance.

- **Processing Requirement**: Big AI often relies on graphics processing units (GPUs) to perform complex and intensive computations, such as matrix manipulations, convolution, and gradient descent. The GPUs are operationally expensive, requiring extensive energy and cooling mechanisms. However, future big-AI models are expected to be deployed on lightweight devices, such as smartphones, tablets, or wearables, with limited computational power. This integration can potentially benefit from knowledge distillation and fine-tuning, which aim at reducing the number of parameters.
- **Training Inefficiency**: Big-AI systems encounter burden training, which implies that they require a vast amount of data, longer training times, and computational resources to learn and enhance their performance. For example, the big-AI GPT model has been trained over a million floating point operations per second. Hence, the scalability of big AI raises serious sustainability concerns. However, using pre-trained models and fine-tuning under specific domains can reduce training inefficiency.

2) Privacy: Utilizing big AI in the semantic extraction and recovery stages can induce some privacy concerns.

- **Generative Risk**: The privacy risks associated with the diffusion models in the semantic recovery process are that such generative tools can create realistic images, memorize specific images from their training data, and reproduce them during generation. This capability poses a significant privacy threat, particularly in the context of privacy-sensitive data.
- **Ethical Risk**: Big-AI systems often use private and confidential data to infer and predict attributes and behaviors. These predictions can influence individual opportunities or outcomes in privacy-sensitive domains, such as education, employment, health, justice, etc. This risk leads to ethical issues such as discrimination, bias, and unfairness.

C. Big-AI Semantic QAB Limitations

The integration between big AI and quantum anonymous SC demands additional steps to leverage the potential of both domains as follows.

1) Communication: he effective and privacy-preserving transmissions in big-AI semantic QAB systems face challenges due to the following concerns.

- **Privacy**: The use of GHZ states preserves anonymity in semantic information encoding and broadcasting. However, these states are intrinsically vulnerable to environmental decoherence, presenting substantial hurdles in preserving high fidelity in big AI-native quantum SC setups. As depicted in Fig. 4, GHZ states under quantum noise lose their entanglement properties, thus compromising anonymity and broadcast error performance. Therefore, ensuring privacy in big AI-native quantum environments becomes more challenging amidst quantum noise and semantic attacks.
- **Scalability**: As the quantum network scales, the performance of anonymous SC across network parties deteriorates. These scalability issues are inherent in quantum networks that employ GHZ states for anonymous communication. In big-AI semantic QAB systems, the broadcast error increases with the number of network parties involved in GHZ states, as evident from Fig. 4. Herein, the semantic QAB becomes more complicated and resource-intensive as the network grows to include more parties. Such scaling issues severely impact the privacy, reliability, and efficiency of large-scale quantum anonymous semantic networking, urging innovative solutions in quantum information engineering.

2) Computation: Critical tasks of encoding classical semantic data into quantum communication setups (qubit modalities) and integrating GPUs with quantum processing units (QPUs) present several issues.

- **Quantum Embedding**: The integration of big AI with quantum systems requires transforming classical semantic data into a quantum-compatible format through quantum kernel methods. The classical data is mapped into quantum embeddings using parameterized quantum circuits. The transformation is computationally demanding, as it requires mapping high-dimensional data into a quantum state, which is currently underdeveloped in the NISQ era. This is countered by enabling big AI for downsampling tasks regarding semantic learning to minimize complexity and computational overheads.
- **Sustainability**: To harness the combined power of classical and quantum processing units (GPU-QPU), an effective and parallel integrating strategy is required. Achieving seamless coordination between GPU-QPU poses challenges in terms of data synchronization, load balancing, and minimizing latency. Efficiently distributing computational tasks between classical and quantum units is essential in sustainability issues for maximizing the overall processing speed and optimizing resource utilization in big-AI semantic QAB systems.

V. CONCLUSION

Integrating computing-intensive communication and quantum information technologies is crucial for developing a secure and effective communication framework as we advance into

next-generation networks. The application of big AI is essential for semantic retrieval processing (extraction and recovery), and quantum communication is preeminent for establishing a secure and privacy-protecting SC system that enables untraceable and anonymous communication of semantic information. However, it is imperative to delve deep into essential aspects for sustainable deployment. For instance, integrating quantum mechanics and advanced AI models entails a prudent approach to ensure enhanced security, privacy, and effective SC. Hence, we must design efficient quantum protocols and refine learning models to seamlessly consolidate these innovative solutions within evolving networks. Furthermore, developing effective sustainability strategies, e.g., model pruning and fine-tuning, is also crucial to address the challenges of extensive training parameters and high computational demands inherent in implementing big AI for communication networks. These strategies help evolve sustainable and adaptable models that align with the needs of next-generation networks. The envisioned hybrid SC framework serves as a stepping stone in introducing a HQC computing-intensive communication paradigm—by integrating big AI and anonymous quantum networking.

REFERENCES

[1] Z. Chen, Z. Zhang, and Z. Yang, "Big AI models for 6G wireless networks: Opportunities, challenges, and research directions," *arXiv:2308.06250*, Aug. 2023.

[2] U. Khalid, M. S. Ulum, A. Farooq, T. Q. Duong, O. A. Dobre, and H. Shin, "Quantum semantic communications for Metaverse: Principles and challenges," *IEEE Wireless Commun.*, vol. 30, no. 4, pp. 26–36, Aug. 2023.

[3] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 5–41, Nov. 2023.

[4] H. Du, J. Wang, D. Niyato, J. Kang, Z. Xiong, M. Guizani, and D. I. Kim, "Rethinking wireless communication security in semantic Internet of things," *IEEE Wireless Commun.*, vol. 30, no. 3, pp. 36–43, Jun. 2023.

[5] L. Ericsson, H. Gouk, C. C. Loy, and T. M. Hospedales, "Self-supervised representation learning: Introduction, advances, and challenges," *IEEE Signal Process. Mag.*, vol. 39, no. 3, pp. 42–62, May 2022.

[6] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu *et al.*, "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Feb. 2022.

[7] P. Xu, X. Zhu, and D. A. Clifton, "Multimodal learning with transformers: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 12 113–12 132, May 2023.

[8] N. Fei, Z. Lu, Y. Gao, G. Yang, Y. Huo, J. Wen, H. Lu, R. Song, X. Gao, T. Xiang *et al.*, "Towards artificial general intelligence via a multimodal foundation model," *Nat. Commun*, vol. 13, no. 1, p. 3094, Jun. 2022.

[9] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10 850–10 869, Mar. 2023.

[10] J. Zhang, J. Huang, S. Jin, and S. Lu, "Vision-language models for vision tasks: A survey," *arXiv:2304.00685*, Apr. 2023.

[11] M. Awais, M. Naseer, S. Khan, R. M. Anwer, H. Cholakkal, M. Shah, M.-H. Yang, and F. S. Khan, "Foundational models defining a new era in vision: A survey and outlook," *arXiv:2307.13721*, Jul. 2023.

[12] F. Pourpanah, M. Abdar, Y. Luo, X. Zhou, R. Wang, C. P. Lim, X.-Z. Wang, and Q. M. J. Wu, "A review of generalized zero-shot learning methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4051–4070, Jul. 2023.

[13] Z. Yang, M. Zolanvari, and R. Jain, "A survey of important issues in quantum computing and communications," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 1059–1094, Mar. Secondquarter 2023.

[14] F. Zaman, S. N. Paing, A. Farooq, H. Shin, and M. Z. Win, "Concealed quantum telecomputation for anonymous 6G URLLC networks," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 7, pp. 2278–2296, May 2023.

[15] K. Lu, R. Li, X. Chen, Z. Zhao, and H. Zhang, "Reinforcement learning-powered semantic communication via semantic similarity," *arXiv:2108.12121*, Apr. 2021.

Shehbaz Tariq received his B.S. degree in Electrical Engineering, University of Engineering and Technology (UET), Peshawar, Pakistan in 2020. He is currently pursuing the Ph.D. degree with the Department of Electronics and Information Convergence Engineering, Kyung Hee University, South Korea. His research interests include quantum machine learning, quantum information science, and artificial intelligence for 6G and beyond.

Uman Khalid received his B.S. degree in electronics engineering from the Ghulam Ishaq Khan (GIK) Institute, Topi, Pakistan, in 2015 and his Ph.D. in electronics engineering from Kyung Hee University, South Korea, in Feb. 2023. Since Mar. 2023, he has been a Post-Doctoral Fellow with the Department of Electronics and Information Convergence Engineering, Kyung Hee University. His research interests include quantum information science, quantum metrology, and quantum networks.

Brian Estadimas Arfeto received his B.S. degree in computer science from Universitas Indonesia, Depok, Indonesia, in 2020. He is currently pursuing the Ph.D. degree with the Department of Electronics and Information Convergence Engineering, Kyung Hee University, South Korea. His research interests include artificial intelligence, quantum information science, and wireless communication.

Trung Q. Duong [F] is a Canada Excellence Research Chair and Full Professor at Memorial University of Newfoundland, Canada. He is also an adjunct Professor in Telecommunications at Queen's University Belfast, UK and a Research Chair of the Royal Academy of Engineering. He was a Distinguished Advisory Professor at Inje University, South Korea (2017-2019). He is a Visiting Professor (under Eminent Scholar program) at Kyung Hee University, South Korea (2023-2024). His current research interests include quantum communications, wireless communications, signal processing, machine learning, and realtime optimisation.

Hyundong Shin [F] is a Professor at Kyung Hee University, Korea. His research interests include quantum information science, wireless communication, and machine intelligence. He received the IEEE Communications Society's Guglielmo Marconi Prize Paper Award and William R. Bennett Prize Paper Award. He served as a Publicity Co-Chair for IEEE PIMRC and a Technical Program Co-Chair for IEEE WCNC and IEEE GLOBECOM. He was an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE COMMUNICATIONS LETTERS.