Generative AI Agents with Large Language Model for Satellite Networks via a Mixture of Experts Transmission

Ruichen Zhang, Hongyang Du, Yinqiu Liu, Dusit Niyato, *Fellow, IEEE*, Jiawen Kang,
Zehui Xiong, Abbas Jamalipour, *Fellow, IEEE*, and Dong In Kim, *Fellow, IEEE*

Abstract—In response to the needs of 6G global communications, satellite communication networks have emerged as a key solution. However, the large-scale development of satellite communication networks is constrained by complex system models, whose modeling is challenging for massive users. Moreover, transmission interference between satellites and users seriously affects communication performance. To solve these problems, this paper develops generative artificial intelligence (AI) agents for model formulation and then applies a mixture of experts (MoE) approach to design transmission strategies. Specifically, we leverage large language models (LLMs) to build an interactive modeling paradigm and utilize retrieval-augmented generation (RAG) to extract satellite expert knowledge that supports mathematical modeling. Afterward, by integrating the expertise of multiple specialized components, we propose an MoE-proximal policy optimization (PPO) approach to solve the formulated problem. Each expert can optimize the optimization variables at which it excels through specialized training through its own network and then aggregate them through the gating network to perform joint optimization. The simulation results validate the accuracy and effectiveness of employing a generative agent for problem formulation. Furthermore, the superiority of the proposed MoE-ppo approach over other benchmarks is confirmed in solving the formulated problem. The adaptability of MoE-PPO to various customized modeling problems has also been demonstrated.

R. Zhang, Y. Liu, and D. Niyato are with the College of Computing and Data Science, Nanyang Technological University, Singapore (e-mail: ruichen.zhang@ntu.edu.sg, yinqiu001@e.ntu.edu.sg, dniyato@ntu.edu.sg).

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

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

Z. Xiong is with the Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore (e-mail: zehui_xiong@sutd.edu.sg).

A. Jamalipour is with The University of Sydney, Sydney NSW 2006, Australia (e-mail: a.jamalipour@ieee.org).

D. I. Kim is with the Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, South Korea (email:dongin@skku.edu).

I. INTRODUCTION

In the era of 6G, the demand for global communication continues to increase rapidly, spotlighting the significant role of satellite communications, particularly due to their unparalleled advantages in overcoming the constraints faced by ground communication systems [1]. Among these, low-earth orbit (LEO) satellite communication networks have emerged as a pivotal solution to the limitations of ground communication systems, offering high-speed, low-latency communication services with extensive coverage and enhanced security [2]. This makes LEO satellite systems critical for ensuring connectivity in remote areas such as oceans and mountainous regions that are difficult to reach with traditional communications systems [3]. Moreover, with an increasing number of satellite users, there is an urgent need for increased capacity in satellite communication networks [4]. This surge in demand underscores the importance of advancing and refining these systems to meet future communication requirements, presenting researchers with the challenge of scaling and optimizing satellite communications to address these evolving requirements [5]. However, to meet these demands, satellite communication networks face two main challenges as follows.

- **Challenge I:** The mathematical modeling of satellite communications is much more complex than that of ground communications systems [6]. It is necessary to consider the curvature of the earth, the impact of the atmosphere on the signal, and the non-uniformity of communications traffic [7]. This complexity makes satellite scenarios (i.e., homogeneous or heterogeneous systems), channel modeling (i.e., static or dynamic channels), access protocols (i.e., rate-splitting multiple access or space-division multiple access), optimization goals (i.e., energy efficiency (EE) and spectral efficiency (SE)) becomes extremely challenging, which not only sets a high threshold for newcomers but also poses a challenge to interdisciplinary researchers who need to understand the operating mechanism of satellite communications.
- **Challenge II:** Compared with ground communication systems, satellite communication networks face more complex challenges in resource allocation [8]. Due to the wide coverage of satellite communications and the large space

where users are distributed, coupled with the limited resources and fixed beam coverage design of satellites, it is challenging to meet the diverse and varying requirements of different kinds of users [9]. Moreover, transmission interference between satellites, between beams, and between users seriously affects communication performance [10]. Therefore, developing adaptive resource allocation schemes to mitigate interference and manage resources effectively is critical to improving the quality of service (QoS) of satellite communication networks.

To address Challenge I, an interactive generative artificial intelligent (AI) agent has been introduced [11]. Generative AI Agents can integrate the power of large language models (LLMs) such as ChatGPT and LLaMA with retrieval-augmented generation (RAG) technologies, aiming to generate solutions to specific problems through interactive sessions with human users [12]. Specifically, the LLM is able to respond by understanding the deep meaning and context of natural language and generating detailed answers to specific questions [13]. RAG is able to retrieve relevant information from massive documents, providing LLMs with real-time, rich background knowledge, allowing LLMs to provide more accurate and informative content when answering queries [14]. In the context of optimizing satellite communication networks, the interactive approach enables the generative AI agent to deeply understand and adapt to specific application scenarios and requirements [15]. For example, when faced with modeling challenges, including changes in satellite orbit parameters, signal propagation characteristics, or complex interference conditions, generative AI agents can refine a problem definition and collect necessary data to generate a more accurate and adaptive system model by interacting with human users and network environments.

To address Challenge II, the mixture of experts (MoE) model has been introduced. MoE integrates multiple specialized deep neural network (DNN) components, i.e., experts, to handle complex tasks [16]. Specifically, each expert is fine-tuned and specialized for a specific type of sub-task or sub-dataset, providing expertise in efficient problem-solving capabilities within their domain, which can be more effective than monolithic models [17]. MoE uses a gating network technique to divide larger tasks into smaller, more manageable units and dynamically select the most appropriate combination of experts to handle these tasks based on the input data characteristics. Unlike traditional neural network models, MoE selectively activates a subset of experts based on specific input features, thereby improving the model's processing efficiency and performance [18]. In satellite communication networks, MoE can provide customized solutions for various challenges based on different experts, such as spectrum allocation, signal interference reduction, and energy efficiency optimization. For example, under dynamically changing communication environments and diverse requirements of different users, MoE can achieve effective resource allocation and interference management by selecting a combination of experts that are most suitable for handling current signal conditions and user distribution [19].

Inspired by the above discussion, as shown in Fig. 1, this paper first studies a framework for modeling using generative AI agents and then optimizes the transmission strategy of satellite communication networks through an MoE model. For the generative AI agents, unlike traditional system modeling, which typically relies on the designer's knowledge and experience to manually formulate mathematical models, our generative AI agents leverage the vast knowledge embedded in LLMs and the precise contextual information provided by RAG to customize and adapt the modeling process to specific satellite communication needs. For the transmission strategy, instead of using a single comprehensive solution system to optimize all variables or employing a multi-agent system to address different aspects, our proposed MoE approach utilizes multiple specialized experts, each focusing on optimizing specific variables, thereby improving efficiency. *To the best of the authors' knowledge, this is the first work in the networking field to adopt generative AI agents for customized modeling and then solve the formulated problem by MoE.*[1] The contributions of this work are summarized as follows.

- We introduce a generative AI agent framework to customize system problem formulation. This approach effectively addresses the diverse modeling requirements within satellite communication networks via a two-layer semantic router and RAG. The framework leverages the rich knowledge of LLMs and the satellite expertise derived by RAG to provide precise modeling solutions for specific problems in satellite communication networks. (For Challenge I)
- Based on different satellite scenarios, we employ different access strategies, channel models, and optimization objectives for heterogeneous and homogeneous satellite networks through the generative AI agent framework to customize our system modeling. This step ensures that the models accurately reflect the complexity and specific requirements of the actual communication environment in both LEO and Geostationary Earth Orbit (GEO) satellite networks. By doing so, we enable a more efficient selection of modeling parameters based on designers' requirements, reducing the risk of errors and saving time. (For Challenge I)
- We propose an MoE-based proximal policy optimization (PPO) method to optimize the formulated problem generated by the generative AI agent. Specifically, each expert can optimize the optimization variables for which it has been trained through its network and then aggregate them through the gating network to perform joint optimization, which can maximize the efficiency of spectrum resource utilization while meeting communication quality requirements. (For Challenge II)

The rest of this paper is organized as follows. The related work is reviewed in Section II. Section III shows our proposed Generative AI Agent framework. An MoE-based PPO approach is proposed in Section IV. Section V conducts some simulations of the proposed generative AI agent framework and MoE-based PPO approach. Finally, we conclude the whole work in Section VI.

---

[1]https://github.com/RickyZang/GAI-agent-satellite

II. Related Work

In this section, we review literature across three domains, i.e., satellite communication networks, Generative AI Agents, and the MoE paradigm, highlighting the advancements and identifying gaps that our work aims to bridge.

A. Satellite Communication Networks

Recent advancements in satellite communication networks have garnered significant attention due to their potential benefits of global connectivity. For instance, Chen et. al. [3] proposed federated learning in LEO satellite communication networks to support massively interconnected devices and reduce communication overheads, showing through simulations that the method significantly reduces latency and improves satellite communication performance in 6G systems. Also, in [20], an optimization problem for maximizing the sum rate in massive MIMO LEO satellite networks, considering both imperfect successive interference cancellation (SIC) and channel state information (CSI). Similarly, Huang et. al. [21] addressed the sum rate maximization problem in LEO satellite-terrestrial networks utilizing RSMA and proposed a PPO-based approach to meet diverse QoS requirements. Khan et. al. [22] focused on sum rate maximization in heterogeneous satellite networks, introducing efficient subcarrier beam assignment and power allocation strategies for secondary LEO satellites while managing interference with primary GEO satellites. Besides spectral efficiency, in [23], the system EE was maximized in integrated satellite-terrestrial radio access networks by jointly optimizing the cache sharing vector, block placement vector, and power allocation vector, where the information EH QoS requirements of users and the power budget are considered as constraints. Despite these contributions, the complex models inherent in these efforts may create barriers to understanding for newcomers to the field, which requires a framework that makes entry into this field of research more accessible.

B. Generative AI Agent

The field of generative AI has seen significant interest, particularly in its application to interactive agents [24]. For instance, Wang et. al. [25] developed an interactive chatbot leveraging LLMs to mimic character-specific personality traits, achieving a high correlation (i.e., about 82.8%) with human perception. Also, The SWIFTSAGE framework, introduced by Lin et. al. [26], combines behavioral cloning with LLM prompting for enhanced action planning in complex tasks. Furthermore, generating agents are also used in the networking field. For instance, Du et. al. [13] demonstrated a generative AI agent empowered by LLMs to simulate user interactions, offering insights into real-time Quality of Experience (QoE) feedback, where experiments showed that using the proposed agent can improve performance by 15%. Also, Zhang et. al. [12] proposed an optimization framework including environment, action, brain, and perception by using an Generative AI Agent, where pluggable LLM and RAG modules are employed for decision-making knowledge base and contextual memory. However, these aforementioned works focus mainly
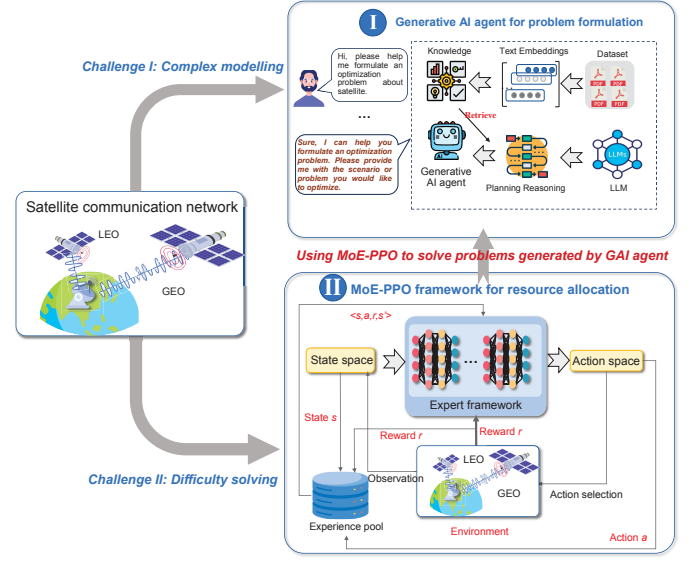

Fig. 1. The outline of the paper. Specifically, to address the challenges of complex modeling and difficulty in solving formulated problems within satellite communication networks, we introduce a generative AI agent and an MoE approach. The generative AI agent is tasked with formulating the optimization problems, while the MoE framework is employed to effectively solve the optimization problems formulated by the generative AI agent.

on direct policy generation and do not consider the potential for customized and selective system modeling and network configuration.

C. Mixture-of-Experts

The MoE model, initially introduced by Jacobs et. al. [27], represented a shift in supervised learning, facilitating parallel processing by multiple specialized networks. With the emergence of deep neural networks (DNN), Eigen et. al. [28] proposed an architecture that can extend multiple experts in parallel on each neural network layer. Each expert has its own weight matrix, and then the gating network distributes the weights to multiple experts, requiring all experts to perform calculations during reasoning. In the era of large models, Du et. al. [29] developed a large model combining LLM and MoE, where MoE can dynamically select only the most relevant experts for the current task to be calculated, thereby reducing the computation of floating-point operations. Thanks to these developments and advantages, MoE is also used in networking fields. For instance, Jaiswal et. al. [30] leveraged MoE for a data-driven transfer learning approach, enhancing radio map models through a combination of location-based and independent expertise. Also, Lopez et. al. [31] applied an MoE-based learning scheme to joint source-channel coding (JSCC), demonstrating the MoE models' versatility and robustness. Nonetheless, existing works focus mainly on MoE to enhance model architecture rather than exploiting its potential for resource optimization in networking. Therefore, this is the primary focus of this work to fill the gap.

III. Generative AI Agent Framework

As shown in Fig. 1, to address Challenge I, this section proposes a generative AI agent to achieve customized model formulation.

A. Dataset Construction

Due to the curvature of the earth, the influence of the atmosphere on signals of satellites, mathematical modeling of satellite communications is difficult and complex, where main four aspects (i.e., scenarios, access protocols, channel models, and optimization goals) should be considered. Thus, in the process of customizing a satellite communication model, the first step is to establish a specialized database due to the lack of ready-made databases. To support subsequent problem formulation, we need to build such a database based on the above four aspects so that the generated AI agent can perform generation.

1) Scenarios: There are significant differences in satellite communication models in different scenarios. Therefore, we consider two typical scenarios, i.e., the homogeneous satellite communication model and the heterogeneous satellite communication model [32].

For the homogeneous scenarios, we consider a downlink LEO satellite and terrestrial network, where an N_{T}-antenna LEO satellite serves K single-antenna LEO ground users (LGUs). Let $\mathbb{E}\{|s_k|^2\} = 1$ where $k \in \mathcal{K} = \{1, ..., K\}$. Thus, the transmit signal is given by

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{w}_k s_k, \tag{1}$$

where $\mathbf{w}_k \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$ denotes the beamforming vector associated with the stream s_k.

For the heterogeneous scenarios, we consider a downlink heterogeneous satellite network composed of one N_{M}-antenna GEO satellite and one N_{T}-antenna LEO satellite, where these satellites use the same spectrum. The GEO satellite serves M single-antenna GEO ground users (GGUs) and each LEO satellite serves K LGUs. For clarity, Let $\mathbb{E}\{|s_k|^2\} = 1$ and $\mathbb{E}\{|s_m|^2\} = 1$. We use 0 to denote the index of the GEO satellite, $m \in \mathcal{M} = \{1, 2, ..., M\}$ to denote the index of the m-th GGU and $k \in \mathcal{K} = \{1, 2, ..., K\}$ to denote the index of the k-th LGU. Thus, the transmit signals of GEO and LEO are, respectively, given by

$$\mathbf{x}_{\mathrm{G}} = \sum_{m=1}^{M} \mathbf{w}_m s_m, \tag{2}$$

and

$$\mathbf{x}_{\mathrm{L}} = \sum_{k=1}^{K} \mathbf{w}_k s_k. \tag{3}$$

2) Access Protocols: In different satellite communication scenarios, the selected access protocols, such as SDMA and RSMA, will also have different impacts on system modeling. For example, when we choose the heterogeneous scenario, for SDMA, it allows to use spatial separation technology to allow the system to serve multiple ground users simultaneously. Thus, the corresponding SINR at the m-th GGU and the k-th LGU are, respectively, given by

$$\Gamma_m\left(\{\mathbf{w}_k\}\right) = \frac{|\mathbf{h}_{\mathrm{G},m}^H \mathbf{w}_m|^2}{\sum\limits_{m' \neq m}^{M} |\mathbf{h}_{\mathrm{G},m}^H \mathbf{w}_{m'}|^2 + \sum\limits_{k=1}^{K} |\mathbf{h}_{\mathrm{L},m}^H \mathbf{w}_k|^2 + \sigma_a^2}, \tag{4}$$

and

$$\Gamma_k\left(\{\mathbf{w}_k\}\right) = \frac{|\mathbf{h}_{\mathrm{L},k}^H \mathbf{w}_k|^2}{\sum\limits_{k' \neq k}^{K} |\mathbf{h}_{\mathrm{L},k}^H \mathbf{w}_{k'}|^2 + \sum\limits_{m=1}^{M} |\mathbf{h}_{\mathrm{G},k}^H \mathbf{w}_m|^2 + \sigma_b^2}, \tag{5}$$

where $\mathbf{h}_{\mathrm{G},m}$ and $\mathbf{h}_{\mathrm{L},k}$ denote the direct channel vector between GEO and the m-th GGU as well as channel vector between LEO and the k-th LGU, respectively. σ_a^2 and σ_b^2 denote the corresponding Gaussian noise at m-th GGU and k-th LGU, respectively.

For RSMA, it decomposes signals into common and private parts to improve system performance. To mitigate the co-channel interference in heterogeneous scenarios, the 1-layer RSMA scheme is adopted at the LEO satellite [33]. Thus, the corresponding SINR at the m-th GGU is given by

$$\Gamma_m\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}\right)$$
$$= \frac{|\mathbf{h}_{\mathrm{G},m}^H \mathbf{w}_m|^2}{\sum\limits_{m' \neq m}^{M} |\mathbf{h}_{\mathrm{G},m}^H \mathbf{w}_{m'}|^2 + \sum\limits_{k=1}^{K} |\mathbf{h}_{\mathrm{L},m}^H \mathbf{w}_k^{\mathrm{p}}|^2 + |\mathbf{h}_{\mathrm{L},m}^H \mathbf{w}^{\mathrm{c}}|^2 + \sigma_a^2}, \tag{6}$$

where $\mathbf{w}_n^{\mathrm{c}} \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$ and $\mathbf{w}_k^{\mathrm{p}} \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$ denote the LEO satellite beamforming vectors associated with the common signal stream and private signal stream, respectively. Next, the corresponding common SINR and private SINR at the k-th LGU are, respectively, given by

$$\Gamma_k^{\mathrm{c}}\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}\right) = \frac{|\mathbf{h}_{\mathrm{L},k}^H \mathbf{w}^{\mathrm{c}}|^2}{\sum\limits_{k'=1}^{K} |\mathbf{h}_{\mathrm{L},k}^H \mathbf{w}_{n,k'}^{\mathrm{p}}|^2 + \sum\limits_{m=1}^{M} |\mathbf{h}_{\mathrm{G},k}^H \mathbf{w}_m|^2 + \sigma_b^2}, \tag{7}$$

and

$$\Gamma_k^{\mathrm{p}}\left(\{\mathbf{w}_k^{\mathrm{p}}\}\right) = \frac{|\mathbf{h}_{\mathrm{L},k}^H \mathbf{w}_k^{\mathrm{p}}|^2}{\sum\limits_{k' \neq k}^{K} |\mathbf{h}_{\mathrm{L},k}^H \mathbf{w}_{k'}^{\mathrm{p}}|^2 + \sum\limits_{m=1}^{M} |\mathbf{h}_{\mathrm{G},k}^H \mathbf{w}_m|^2 + \sigma_b^2}. \tag{8}$$

3) Channel Models: Channel modeling in satellite communication models is quite different, such as fixed and time-varying channels [34]. For the fixed channel model, it assumes that the channel statistical characteristics remain unchanged over a long time, where the corresponding channel model is expressed by

$$\mathbf{h} = \sqrt{G_s G_k \left(\frac{c}{4\pi f_c d_s}\right)^2} \mathbf{g}, \tag{9}$$

where G_s and G_k denote the satellite antenna gain and the user antenna gain, respectively. c is the light speed, f_c is the carrier frequency, and d_s is the distance between the corresponding satellite and the user. $\mathbf{g}$ is the small-scale fading vector with Rician distribution. For notational simplification, we omit the index of $\mathbf{h}_{\mathrm{L},k}$ and $\mathbf{h}_{\mathrm{G},m}$ here.

For the time-varying channel model, it is typically that channel statistical characteristics such as Doppler shift change rapidly with time. Following Jakes model [35], keeping the large-scale fading in (9) unchanged, the small-scale fading vector $\mathbf{g}$ is modeled as a first-order complex Gauss Markov process, i.e.,

$$\mathbf{g}(t) = \rho \mathbf{g}(t-1) + \sqrt{1-\rho^2} \mathbf{e}. \tag{10}$$

In (10), $\mathbf{e}$ denotes the additive complex Gaussian noise with the same distribution as $\mathbf{g}$ and the correlation coefficient ρ is determined by

$$\rho = J_0(2\pi f_d T_s), \tag{11}$$

where $J_0(\cdot)$, T_s, and f_d denote the first kind zero-order Bessel function, the time interval between adjacent instants, and the maximum Doppler frequency, respectively.

4) Optimization Goals: The optimization goals of satellite communications focus on improving various performance indicators, such as maximizing SE and EE [36].

For SE, the achievable information rate at the m-th GGU is given by

$$R_m\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}\right) = \log_2(1 + \Gamma_m(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\})). \tag{12}$$

Similarly, the achievable information rates of the common signal part and the private part at the k-th LGU are, respectively, given by

$$R_k^{\mathrm{c}}\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}\right) = \log_2(1 + \Gamma_k^{\mathrm{c}}\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}\right)), \tag{13}$$

and

$$R_k^{\mathrm{p}}\left(\{\mathbf{w}_k^{\mathrm{p}}\}\right) = \log_2(1 + \Gamma_k^{\mathrm{p}}\left(\{\mathbf{w}_k^{\mathrm{p}}\}\right)). \tag{14}$$

To guarantee that the common signal is successfully decoded at all LGUs, the rate of the common message should be chosen as $\min_k R_k^{\mathrm{c}}\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}\right)$. Denote c_k as the data rate for receiving the common message at the k-th LGU, which should satisfy that

$$\sum_{k=1}^{K} c_k \leq \min_k R_k^{\mathrm{c}}\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}\right). \tag{15}$$

Then, the sum achievable rate of the LEO area is expressed as

$$R\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}, c_k\}\right) = \sum_{k=1}^{K} \left(c_k + R_k^{\mathrm{p}}\left(\{\mathbf{w}_k^{\mathrm{p}}\}\right)\right). \tag{16}$$

For EE, it is a ratio of the sum achievable rate to the total power consumption and the system EE at the LEO area is expressed by

$$\mathrm{EE}\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}, c_k\}\right) = \frac{R\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}, c_k\}\right)}{P_{\mathrm{T}}\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}\right)}, \tag{17}$$

where

$$P_{\mathrm{T}}\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}\right) = \mu\left(\|\mathbf{w}^{\mathrm{c}}\|^2 + \sum_{k=1}^{K}\|\mathbf{w}_k^{\mathrm{p}}\|^2\right) + P_{\mathrm{C}}, \tag{18}$$

with $\mu \in [1, \infty)$ and P_{C} being the power amplifier efficiency factor and the constant power consumption by circuit modules, respectively.

Remark 1. *It can be observed that human users need to carefully configure the above four aspects to build a reasonable satellite model manually. Nonetheless, errors are prone to occur during model building due to the varying situations and multiple choices. For example, if the user neglects rate ordering when choosing RSMA as the access protocol or mistakenly adopts stable channels to communicate with rapidly moving satellites, the model correctness and the corresponding*

communication efficiency might be significantly affected.

To this end, a generative AI agent framework is constructed, as shown in Fig. 2. Specifically, we first utilize LLMs to establish a conversational modeling procedure. In each round of interaction, the semantic router [37] integrated into LLM extracts task-relevant semantics from users' natural language and invokes the corresponding function calls. Since LLMs are trained on general knowledge and lack satellite expertise, we build a RAG [14] system, which contains massive expert knowledge regarding satellite communications to support complicated mathematical modeling. The RAG functions called by the semantic router retrieve relevant knowledge from the dataset. Afterward, the agent leverages LLMs to analyze the retrieved knowledge and generate the configurations of each modeling aspect step-by-step, based on user descriptions and requirements. Next, we introduce the detailed design.

B. Semantic Router

As shown in Fig. 2, we segment the entire expertise dataset into four blocks, each corresponding to one aspect mentioned in Section III-A. Moreover, each block contains two sub-blocks, corresponding to the specific configurations in each aspect. In this case, to generate accurate system models, the prerequisite for our generative agent is to locate the correct sub-block for expertise retrieval. The semantic router acts as the nerves of agents, transforming users' descriptions in natural language to the calling for the specific sub-block. To do so, it employs a two-layer structure, in which layer-1 routing realizes the mapping between user description and block. Afterward, the layer-2 routing further determines which sub-block to call. The two routing selection layers are based on semantic similarity [37]. Specifically, denoting a user description/query as $q \in \mathbf{q}$, the semantic router will use an encoder $\mathbf{E}_q$ to encode it as text embeddings, i.e., $\mathbf{E}_q(q)$. Here, text embeddings play a crucial role in bridging humans and LLMs by encoding natural language into a numerical format that neural networks can process. In this paper, we leverage the pre-trained *text-embedding-ada-002* [38] model to generate embeddings of textual inputs. By training on large corpora, this model acquires outstanding contextual understanding capabilities, projecting each word/phrase into the continuous feature space with minimized conflicts. Moreover, we adopt the widely used *cosine distance* [37] to measure the semantic similarity between two embeddings.

The labels of blocks and sub-blocks are $\mathbf{b}_i$ and $(\mathbf{b}_i, \mathbf{s}_j^i)$, where $\mathbf{b}_i$ and $\mathbf{s}_j^i$ ($i \in \{1,2,3,4\}, j \in \{1,2\}$) refer to the aspect and configuration, respectively. For instance, $\mathbf{b}_3$ = "*channel model*", $\mathbf{s}_1^3$ = "*SDMA*", and $\mathbf{s}_2^3$ = "*RSMA*" (see Fig. 2). Using cosine similarity as the metric for similarity measurement, the layer-1 routing of query q can be defined as follows:

$$\min_i\left(1.0 - \frac{\sum_{k=1}^{\zeta}\left(\mathbf{E}_q(q)[k] \times \mathbf{E}_q(\mathbf{b}_i)[k]\right)}{\sqrt{\sum_{k=1}^{\zeta}\left(\mathbf{E}_q(q)[k]\right)^2}\sqrt{\sum_{k=1}^{\zeta}\left(\mathbf{E}_q(\mathbf{b}_i)[k]\right)^2}}\right), \tag{19}$$
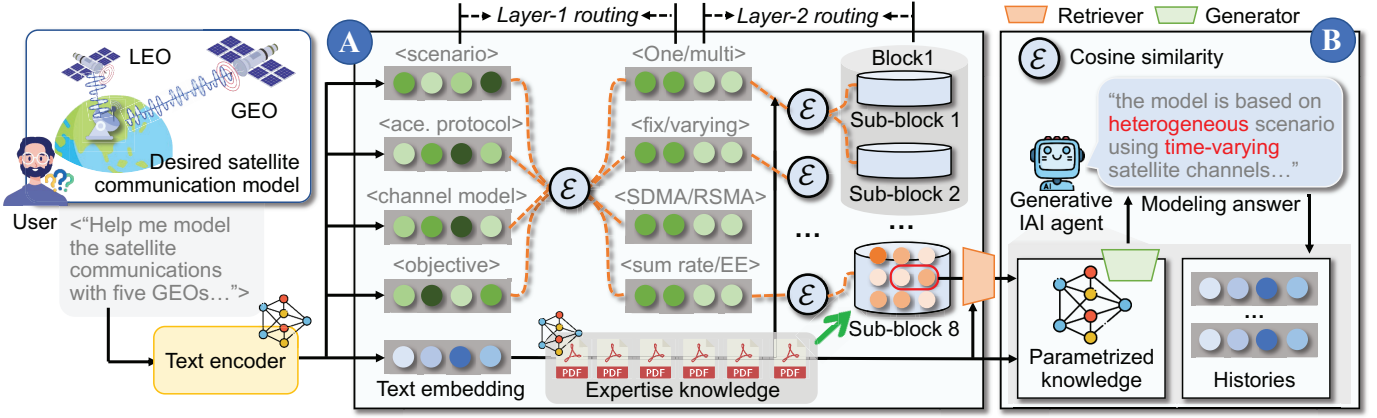
Fig. 2. The design of generative AI agent for satellite communication modeling. **A**: The two-layer semantic router. The expertise knowledge is organized into 4 blocks, each of which contains 2 sub-blocks. **B**: The answer generation for satellite communications modeling.

where ζ denotes the length of text embeddings. From this equation, we can observe that the user query will be routed to the block whose label is the most relevant. Similarly, layer-2 routing repeats Eq. (19) while aiming to further route q to the most relevant sub-block.

Remark 2. *The semantic router is presented to enhance the efficiency of RAG for satellite communications modeling. This is because the rate of correct retrieval decreases with the increasing knowledge volume. By organizing expertise knowledge into eight sub-blocks, the space for retrieval can be reduced by eight times, making our proposal suitable for complicated satellite communication modeling. Moreover, we note that the block and sub-block configuration is customizable. Users can include/remove blocks/sub-blocks according to the specific application scenarios. After routing the user query to the specific sub-block, the RAG-assisted modeling can be performed, which is discussed in the next part.*

C. Retrieval-Augmented Generation

RAG aims to retrieve relevant expertise from the knowledge sub-block to support satellite communications modeling. Such a process can be modeled as using input sequence q to fetch knowledge z, subsequently utilized as supplementary context for helping LLMs to generate the modeling answers m. Inspired by [14], we develop our RAG system atop *RAG-Sequence*, utilizing identical knowledge to forecast each token within the target sequence. As illustrated in Fig. 2, our system encompasses two pivotal elements: (i) a retrieval mechanism $p_\eta(z|q)$ characterized by parameters η, tasked with deriving a truncated distribution over knowledge embeddings in response to a query q, and (ii) a generative module $p_\theta(m_i|q, z, m_{1:i-1})$, parameterized by θ, responsible for the sequential generation of tokens, each contingent upon the antecedent $i-1$ tokens $m_{1:i-1}$, the initial query q, and retrieved knowledge z. Such a process encompasses the marginalization of the seq2seq probability $p_{\text{Seq}}(m|x)$ via a top-K approximation [14], formalized as follows:

$$
\begin{aligned}
p_{\text{Seq}}(m|q) &\approx \sum_{z \in \text{top-K}(p(\cdot|q))} p_\eta(z|q) p(m|z, q) \\
&= \sum_{z \in \text{top-K}(p(\cdot|q))} p_\eta(z|q) \prod_{i}^{L} p_\theta(m_i|q, z, m_{1:i-1}),
\end{aligned}
\tag{20}
$$

where L represents the length of the sequence. Note that top-K approximation is utilized to bring diversity to model output. Next, we introduce the retriever and generator design, as well as the training and inference paradigms.

Algorithm 1: Generative AI agent for satellite communications modeling

---

1 **Input** : User query, embedding models, LLM;
2 Initialize environment;
3 Listen to user query $\mathbf{q} = \{q_1, q_2, \ldots, q_n\}$
4 **for** *each q in $\mathbf{q}$* **do**
5   Encode q and acquire embeddings $\mathbf{E}_q(q)$;
6   **for** *each i in {1, 2, 3, 4}* **do**
7   Calculate cosine similarity between $\mathbf{E}_q(q)$ and $\mathbf{E}_q(\mathbf{b}_j)$ based on (19);
8   Route to the corresponding block i; *# layer-1 routing*
9   **for** *each j in {1, 2}* **do**
10   Calculate cosine similarity between $\mathbf{E}_q(q)$ and $\mathbf{E}_q(\mathbf{s}_j^i)$ based on (19);
11   Route to the corresponding sub-block j; *# layer-2 routing*
12   **for** *each chunk in the sub-block* **do**
13   Calculate cosine similarity between $\mathbf{E}_q(q)$ and the chunk based on (19);
14   Fetch the most prevalent knowledge chunks;
15   Call LLM to generate answers for q based on retrieved knowledge;
16 **Output** : The entire satellite communications model containing scenario, access protocol, channel, and optimization goal;

---

1) Retrieval Mechanism: The retriever function $p_\eta(z|q)$ is predicated on the dense passage retriever (DPR) model [14], embracing a bi-encoder architecture, i.e.,

$$p_\eta(z|q) \propto \exp\left(\mathbf{E}_{\mathrm{z}}(z)^\top \mathbf{E}_{\mathrm{q}}(q)\right), \quad (21)$$

where $\mathbf{E}_{\mathrm{z}}(z)$ is a dense representation of the knowledge produced by the aforementioned text encoder [39], and $\mathbf{E}_{\mathrm{q}}(q)$ denotes the query representation produced by a query encoder. Due to its strong representation ability, such encoders are now usually implemented on top of LLM, such as text-embedding-ada-002 [38], i.e.,

$$\mathbf{E}_{\mathrm{z}}(z) = \mathrm{LLM}_e(z) \ \& \ \mathbf{E}_{\mathrm{q}}(q) = \mathrm{LLM}_q(q). \quad (22)$$

The knowledge in the selected sub-block is split into multiple chunks. Based on Eq. (21), the most relevant chunks can be retrieved to support the generation of satellite communications modeling.

2) Generative Module: The generative module is pluggable and can be served by any mainstream LLM. Here, we adopt a generative pre-trained transformer (GPT) to realize $p_\theta(m_i|q, z, m_{1:i-1})$, organizing retrieved knowledge and generating coherent and context-aware answers about satellite communications modeling.

3) Training Approaches: The training scheme jointly optimizes the retriever and generator without explicit directives on document retrieval. The objective minimizes the negative marginal log-likelihood of the target sequences across a corpus of input/output pairs (q_j, m_j), employing stochastic gradient descent *Adam*.

4) Generation Strategies: During inference, *RAG-Sequence* models adopt distinct approximations to $p_{\mathrm{Seq}}(m|q)$. This paradigm requires searching on each document z, merging hypotheses across documents to form a candidate set. The likelihood $p_{\mathrm{Seq}}(m|q)$ is then deduced by aggregating generator probabilities $p_\theta(m_i|q, z, m_{1:i-1})$ across the document set, thereby generating rational and accurate answers about satellite communications modeling according to user descriptions/queries. In summary, Algorithm 1 illustrates the process of performing customized satellite communications modeling via the proposed generative AI agent.

D. Problem Formulation

Using such a framework, human users can effectively tailor optimization problems to meet their unique requirements and objectives. For instance, when selecting a heterogeneous scenario with RSMA in a time-varying channel and focusing on SE, the human users aim to enhance the sum rate in LEO satellite regions. This involves a complex joint optimization of transmit private beamforming vectors $\{\mathbf{w}_k^{\mathrm{p}}\}$, common beamforming vectors $\{\mathbf{w}^{\mathrm{c}}\}$, and common rates $\{c_k\}$, constrained by the achievable rate requirements at GGUs and LGUs. Note that the sum rate maximization in time-varying channels for heterogeneous satellite networks with RSMA has remained an open issue thus far, where generative AI agent can help construct the formulated problem, i.e.,

$$\max_{\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}, c_k\}} R\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}, c_k\}\right) \quad (23a)$$

$$\mathrm{s.t.} \quad \|\mathbf{w}^{\mathrm{c}}\|^2 + \sum_{k=1}^{K} \|\mathbf{w}_n^{\mathrm{p}}\|^2 \le P_{\max}, \quad (23b)$$

$$R_m\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}\right) \ge \xi_{\mathrm{GGU}}, \quad (23c)$$

$$R_k\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}, c_k\}\right) \ge \xi_{\mathrm{LGU}}, \quad (23d)$$

$$\sum_{k=1}^{K} c_k \le \min_k R_k^{\mathrm{c}}\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}\right) \quad (23e)$$

$$c_k \ge 0, \quad (23f)$$

$$\forall m \in \mathcal{M}, \forall n \in \mathcal{N}, \forall k \in \mathcal{K}. \quad (23g)$$

In constraint (23b), $P_{\max}$ denotes the maximum power budget at LEO satellite. In constraints (23c) and (23d), ξ_{GGU} and ξ_{LGU} denote the minimum required achievable information rate threshold of each GGU and LGU, respectively. Constraints (23e) and (23f) guarantee that the common message can be successfully decoded by each LGU.

Remark 3. *It is worth noting that Problem (23) is just one of many potential configurations that a generative AI agent can customize. The framework is universally adaptable and can incorporate various human user-defined needs based on the database from above four aspects, such as non-orthogonal multiple access (NOMA) in protocols and reconfigurable intelligent surface (RIS) in scenarios. This adaptability highlights the utility of generative AI agents in building complex models that can be time-consuming to model and highly error-prone if relying solely on human expertise.*

Proposition 1. *For any how customized modeling is adopted based on the above four aspects (i.e., scenarios, access protocols, channel models, and optimization goals), the formulated problem is an NP-hard problem.*

Proof. To prove the NP-hardness of formulated problem, it is equivalent to reducing the formulated problem to one of the proven NP-hard problems. Particularly, for any how customized modeling is adopted, with given some optimization variables such as $\{c_k\}$ or $\mathbf{w}^{\mathrm{c}}$, the formulated problem is reduced to a multi-ratio fractional programming problem, which has been reported to be an NP-hard problem in [40]. $\square$

IV. PROPOSED PPO WITH MoE APPROACH

As demonstrated in Proposition 1, due to the existence of heterogeneous variables and the complexity of the optimization environment, Problem (23) is an NP-hard problem, which poses a significant challenge in finding reasonable solutions. Therefore, in this section, we introduce an MoE-PPO approach. Note that generative AI agent can generate a variety of optimization problems, we focus on addressing one of the most challenging formulated problems, i.e., (23), constructed by the generative AI agent. For other simpler formulated problems that the proposed AI agent constructs, fewer experts and a more straightforward action space can be used.

A. Overview of MoE

1) Architecture of the MoE Model: MoE is the integration of multiple specialized sub-networks (i.e., "experts") under the

guidance of a central "gating network" [41]. This design consists of multiple expert networks $(E_1, E_2, ..., E_I)$, each with unique parameters and running simultaneously. The central gating network labeled G_σ generates a weight vector that enables the model to direct attention to the most relevant experts based on the current input.

In the MoE architecture, each expert is usually designed as a feed-forward neural network, processing inputs independently and producing outputs with the same dimensions, where the core function of the MoE model is described as

$$y = \sum_{i=1}^{I} G_\sigma(x)_i E_i(x), \tag{24}$$

In (24), $E_i(x)$ is the output from the i-th expert network for a given input. The gating network $G_\sigma(x)$ typically utilizes a Softmax function, i.e.,

$$G_\sigma(x) = \frac{1}{1 + \exp(-x \cdot W_g)}, \tag{25}$$

where W_g is a learnable weight matrix. This design ensures that the MOE model leverages the most pertinent experts for any specific input through $G_\sigma(x)$'s selective activation mechanism, thereby optimizing the model's performance.

2) MoE Model Training: For the training stage, MoE involves adjusting the parameters of the expert network and the gating network to minimize the loss function that reflects the difference between the predicted output and the real output, where the loss function based on mean-squared error (MSE) is defined as

$$\mathcal{L} = \frac{1}{B} \sum_{b=1}^{B} \left(y_b - \sum_{i=1}^{I} G_\sigma(x_b)_i E_i(x_b) \right)^2, \tag{26}$$

where B represents the total number of training samples, y_b denotes the actual output for the b-th sample, and x_b denotes the corresponding input. Next, we introduce an MOE-PPO framework, showing how to integrate task decomposition and policy optimization to further improve the learning performance of PPO.

B. Mixture-of-Experts with PPO (MoE-PPO) Framework

Integrating the MoE with the PPO approach, we enhance the policy's capacity to model complex behaviors in environments. Specifically, the MoE-PPO framework adopts an actor-critic structure, where the actor network, parameterized by $\boldsymbol{\theta}_A$, encapsulates a dynamic ensemble of expert policies for decision-making. Concurrently, the critic network, parameterized by $\boldsymbol{\theta}_C$, assesses the state-value function, $V_{\boldsymbol{\theta}_C}(\mathbf{s})$, guiding the actor's policy improvement.

1) Actor network with MoE: The actor network in the MoE-PPO framework is defined as

$$\pi_{\boldsymbol{\theta}_A}(a_t|\mathbf{s}_t) = \sum_{i=1}^{I} w_{\theta_i}(\mathbf{s}_t)\pi_{\psi_i}(a_t|\mathbf{s}_t), \tag{27}$$

where each π_{ψ_i} represents an expert policy within the mixture, and $w_{\theta_i}(\mathbf{s}_t)$ denotes the dynamic weighting of expert i for state $\mathbf{s}_t$, managed by the routing function parameterized by θ.

2) Gating Function with PPO: The gating function in the MOE model plays a crucial role by dynamically selecting and aggregating expert policies based on the current state $\mathbf{s}_t$. These selecting and aggregating processes, however, involves making decisions from a categorical distribution, which presents challenges for optimization within the PPO framework. To tackle this issue, a new approach to estimate the gradient of the actor network under gating function is introduced.

Given a state $\mathbf{s}_t$, the value of advantage $A_\phi(\mathbf{s}_t, a_t^i)$ decides the contribution of each expert policy π_{ψ_i} to final action decisions. For each available action a_t^i sampled from the respective expert policy $\pi_{\psi_i}(\cdot|\mathbf{s}_t)$, we compute the advantage $A_\phi(\mathbf{s}_t, a_t^i)$, reflecting the expected return of selecting a_t^i in state $\mathbf{s}_t$. The optimal expert policy for $\mathbf{s}_t$ is identified as the one yielding the highest advantage.

Definition 1 (Back propagation maximization). *For optimizing actor network, the gradient related to the expert i, denoted as $grad_{\theta_{A_i}}$, is estimated using the back propagation maximization approach, i.e.,*

$$grad_{\theta_{A_i}} = \delta_{\theta_{A_i}} \nabla_{\theta_{A_i}} \pi_{\theta_{A_i}}(\mathbf{s}_t), \tag{28}$$

where

$$\delta_{\theta_{A_i}} = \mathbb{1}_{\{i=\arg\max_j A_\phi(\mathbf{s}_t, a_t^j)\}} - \pi_{\theta_{A_i}}(\mathbf{s}_t), \tag{29}$$

with $\mathbb{1}_{\{i=\arg\max_j A_\phi(\mathbf{s}_t, a_t^j)\}}$ being an indicator function that equals 1 if expert i's action a_t^i maximizes the advantage $A_\phi(\mathbf{s}_t, a_t^j)$ over all experts, and 0 otherwise.

Based on the back propagation maximization, the MOE-PPO can be effectively optimized to refine the overall strategy $\pi_{\boldsymbol{\theta}_A}$ from multiple experts' policies.

3) Objective Function with MOE: To align the MOE framework within the PPO, the surrogate objective function is adapted to accommodate the MoE while maintaining the core principles of PPO [42], which is expressed by

$$J(\boldsymbol{\theta}_A) = \mathbb{E}_t \left[\rho_t(\boldsymbol{\theta}_A) A_t^{\pi_{\boldsymbol{\theta}_A^{old}}} \right], \tag{30}$$

where $\rho_t(\boldsymbol{\theta}_A) = \frac{\pi_{\boldsymbol{\theta}_A}(a_t|\mathbf{s}_t)}{\pi_{\boldsymbol{\theta}_A^{old}}(a_t|\mathbf{s}_t)}$ denotes the probability ratio of the chosen action under the current policy to the old policy, reflecting the essence of the importance sampling technique used in PPO to estimate the expected advantage of the policy update. A_t is the advantage function, which is used for highlighting the role of each expert in the collective policy through their weighted contributions to the estimated rewards, i.e.,

$$A_t^{\pi_{\boldsymbol{\theta}_A^{old}}} = r(\mathbf{s}_t, \mathbf{a}_t) + \gamma V_{\boldsymbol{\theta}_C^{old}}(\mathbf{s}_{t+1}) - V_{\boldsymbol{\theta}_C^{old}}(\mathbf{s}_t). \tag{31}$$

However, the monotonic improvement theory requires that $\pi_{\boldsymbol{\theta}_A}(\mathbf{a}_t|\mathbf{s}_t)$ and $\pi_{\boldsymbol{\theta}_A^{old}}(\mathbf{a}_t|\mathbf{s}_t)$ satisfy the following trust region constraint, i.e.,

$$\mathbb{E}_t \left[KL \left(\pi_{\boldsymbol{\theta}_A}(\mathbf{a}_t|\mathbf{s}_t) || \pi_{\boldsymbol{\theta}_A^{old}}(\mathbf{a}_t|\mathbf{s}_t) \right) \right] \leq \varepsilon, \tag{32}$$

where $KL(\cdot)$ represents the Kullback-Leibler divergence function, and ε is a positive hyperparameter.

To facilitate the computation, $J(\boldsymbol{\theta}_A)$ with $\boldsymbol{\theta}_A$ satisfying (32)

Fig. 3. The diagram of the proposed MoE-PPO approach.

can be approximated by

$$J^{\text{CLIP}}(\boldsymbol{\theta}_{\text{A}}) = \qquad (33)$$
$$\mathbb{E}\left[\min\left(\rho_t(\boldsymbol{\theta}_{\text{A}}) A_t^{\pi_{\boldsymbol{\theta}_{\text{A}}^{\text{old}}}}, \text{clip}\left(\rho_t(\boldsymbol{\theta}_{\text{A}}), 1-\varepsilon, 1+\varepsilon\right) A_t^{\pi_{\boldsymbol{\theta}_{\text{A}}^{\text{old}}}}\right)\right],$$

where $\text{clip}(\cdot)$ is the clip function to restrain $\rho_t(\boldsymbol{\theta}_{\text{A}})$ to lie in the range $[1-\varepsilon, 1+\varepsilon]$ and ε is a hyperparameter which is decayed during training stage.

4) Policy and Critic Update Mechanisms: The parameters of both the actor and critic networks are refined through mini-batch stochastic gradient descent (SGD) using experiences sampled from the environment, where their update rules are, respectively, formulated as

$$\boldsymbol{\theta}_{\text{A}} = \boldsymbol{\theta}_{\text{A}}^{\text{old}} - \alpha_{\text{A}}\frac{1}{B}\sum_{t=1}^{B}\left(\nabla_{\boldsymbol{\theta}_{\text{A}}}\tilde{J}_t^{\text{CLIP}}(\boldsymbol{\theta}_{\text{A}})\right), \qquad (34)$$

and

$$\boldsymbol{\theta}_{\text{C}} = \boldsymbol{\theta}_{\text{C}}^{\text{old}} - \alpha_{\text{C}}\frac{1}{B}\sum_{t=1}^{B}\nabla_{\boldsymbol{\theta}_{\text{C}}}\left(V_{\boldsymbol{\theta}_{\text{C}}}(\mathbf{s}_t) - V_{\text{tar}}(\mathbf{s}_t)\right)^2, \quad (35)$$

where $\tilde{J}_t^{\text{CLIP}}(\boldsymbol{\theta}_{\text{A}})$ reflects the realization of the CLIP-modified objective function, and $V_{\text{tar}}(\mathbf{s}_t)$ represents the target value for the state $\mathbf{s}_t$, which is computed by

$$V_{\text{tar}}(\mathbf{s}_t) = r(\mathbf{s}_t, \mathbf{a}_t) + \cdots + \gamma^{n+1}V_{\boldsymbol{\theta}_{\text{C}}^{\text{old}}}(\mathbf{s}_{t+n+1}), \qquad (36)$$

where $\gamma \in [0,1]$ denotes a discount factor. For clarity, the diagram of the proposed MoE-PPO approach is shown in Fig. 3.

Remark 4. *Note that, in our proposed MoE-PPO approach, we use a back-propagation maximization method to update the experts. This method uses the gradient of the most*

advantageous function value across all experts to update each expert, effectively allowing the best-performing expert to guide the improvement of the other experts. Additionally, the gating network dynamically adjusts weights, ensuring that the influence of poorly performing experts is mitigated. By doing so, the impact of the worst-performing experts on the whole network is significantly minimized. Such the back-propagation maximization method and gating network have been adopted in existing studies, e.g., [43] and [44].

C. MDP Formulation

In MDP, we define the framework through the components $\langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \gamma \rangle$. These components encapsulate the state space $\mathcal{S}$, action space $\mathcal{A}$, reward function $\mathcal{R}$, and a discount factor γ, where the detailed designs are shown as follows.

1) Action Space: In action space, each expert is assigned to generate specific actions in terms of $\{\mathbf{w}_k^{\text{P}}\}$, $\mathbf{w}^{\text{c}}$, and c_k. Considering the complexity of handling complex-valued beamforming vectors in neural networks and inspired by [45], a decomposition approach is adopted, where each beamforming vector $\{\mathbf{w}_v\}$ (including $\{\mathbf{w}_k^{\text{P}}\}$ and $\mathbf{w}^{\text{c}}$) is split into its magnitude and a unit-norm direction vector, i.e.,

$$\mathbf{w}_v = \|\mathbf{w}_v\| \overline{\mathbf{w}}_v, \forall v \in \{1, 2, ..., K+1\}. \qquad (37)$$

In (37), $\|\mathbf{w}_v\|$ denotes the magnitude of the corresponding transmit power and $\overline{\mathbf{w}}_v$ denotes the unit-norm beam direction vector.

For the magnitude of the corresponding transmit power $\|\mathbf{w}_v\|$, it is determined by the hyperbolic tangent function to ensure that the power constraints are met, which is expressed by

$$\|\mathbf{w}_v\| = \frac{\sqrt{P_{\text{max}}}}{2}\left(\underbrace{\frac{\exp(x_v^{\text{POW}}) - \exp(-x_v^{\text{POW}})}{\exp(x_v^{\text{POW}}) + \exp(-x_v^{\text{POW}})}}_{\text{hyperbolic tangent function}} + 1\right),$$
$$(38)$$

where $\frac{\exp(x_v^{\text{POW}}) - \exp(-x_v^{\text{POW}})}{\exp(x_v^{\text{POW}}) + \exp(-x_v^{\text{POW}})} \in [-1, +1]$ is employed as an activation function to guarantee that the outputs of the deep neural networks (DNNs), i.e., the selected transmit power, satisfy the corresponding constraints [46]. $x_v^{\text{POW}} = f(W \cdot x_v^{\text{POW}'} + b^l)$ denotes the output of the activation function with respect to the corresponding selected transmit power. $f(\cdot)$, W, $x_v^{\text{POW}'}$, and b^l denote the activation function of the current layer, the corresponding weight, the output in the previous layer and the corresponding bias, respectively.

For the unit-norm beam direction vector $\overline{\mathbf{w}}_v$, following [45], we leverage the maximum ratio transmission (MRT) and zero forcing (ZF) strategies, i.e.,

$$\overline{\mathbf{w}}_v = \begin{cases} \frac{\sum_{k=1}^{K}\tilde{\mathbf{h}}_{\text{L},k}^H}{\|\sum_{k=1}^{K}\tilde{\mathbf{h}}_{\text{L},k}^H\|}, & \text{if } v = K+1, \\ \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}, & \text{if } v \neq K+1, \end{cases} \qquad (39)$$

where $\tilde{\mathbf{h}}_{\text{L},k} = \mathbf{h}_{\text{L},k}^H$ and $\mathbf{v}_k$ is the k-th column of $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_K]$ where $\mathbf{V} = \mathbf{G}^H(\mathbf{G}\mathbf{G}^H)^{-1}$ and $\mathbf{G} = [\tilde{\mathbf{h}}_{\text{L},1}, ..., \tilde{\mathbf{h}}_{\text{L},K}]$.

For the achievable rate of the common message $\{c_k\}$, it is also determined by the hyperbolic tangent function to ensure

that the common rate constraints are met, which is expressed by

$$c_k = \frac{\min_k R_k^c\left(\{\mathbf{w}_k^p, \mathbf{w}^c\}\right)}{2}\left(\underbrace{\frac{\exp(x_k^{\text{COM}}) - \exp(-x_k^{\text{COM}})}{\exp(x_k^{\text{COM}}) + \exp(-x_k^{\text{COM}})}}_{\text{hyperbolic tangent function}} + 1\right),$$

(40)

where x_k^{COM} denotes the output of the activation function with respect to the corresponding selected achievable rate of the common message.

Therefore, the action space is defined as

$$\mathcal{A} = \left\{\{\mathbf{w}_k^p\}, \mathbf{w}^c, \{c_k\}\right\},$$

(41)

where the cardinal number of the action space $\mathcal{A}$ is $(2K+1)$.

2) State Space: The action space comprises the set of current information of all the users $\mathbf{u} \in \mathbb{R}^{(2K+M)\times 1}$, the selected action vector $\mathbf{a}$, and the instant reward r.

For the current information of all users $\mathbf{u}$, it is expressed by

$$\mathbf{u} = \left\{\left\{\Gamma_k^{(C)}\right\}, \left\{\Gamma_k^{(P)}\right\}, \{\Gamma_m\}\right\},$$

(42)

where $\Gamma_k^{(C)}$, $\Gamma_k^{(P)}$, and Γ_m represent the corresponding SINR of GGU and LGU in the last time step.

For the selected action vector $\mathbf{a}$, it is chosen from the set of possible actions $\mathcal{A}$, which reflects the strategy implemented in the last time step. The instant reward r is generated from the state and the chosen action, which indicates the efficacy with which the agent addresses the described problem. In the subsequent subsection, we provide a definition of instant reward r relevant to our proposed approach.

Therefore, the state space is defined as

$$\mathcal{S} = \{\mathbf{u}, \mathbf{a}, r\},$$

(43)

where the cardinal number of the state space $\mathcal{S}$ is $(4K+M+1)$.

3) Reward Function: The reward function considers both the objective function and the constraints of the Problem, which comprises two elements: the direct reward term, which reflects the unconstrained energy efficiency (EE), and the penalty term, which ensures compliance with the constraints. To harmonize the relationship between the direct reward and the penalty term, we define a penalty-based reward function that balances these two aspects, i.e.,

$$r = R\left(\{\mathbf{w}_k^p, \mathbf{w}^c, c_k\}\right)$$

(44)

$$\times \underbrace{(\Omega_{\text{Pow}} \times \Omega_{\text{Com}} \times \Omega_{\text{LEO}} \times \Omega_{\text{GEO}})}_{\text{Penalty term}},$$

where

$$\Omega_{\text{Pow}} = \begin{cases} 1, & \|\mathbf{w}^c\|^2 + \sum_{k=1}^K \|\mathbf{w}_k^p\|^2 \leq P_{\max}, \ \forall k \in \mathcal{K}, \\ 0, & \|\mathbf{w}^c\|^2 + \sum_{k=1}^K \|\mathbf{w}_k^p\|^2 > P_{\max}, \ \forall k \in \mathcal{K}, \end{cases}$$

(45)

$$\Omega_{\text{Com}} = \begin{cases} 1, & \sum_{k=1}^K c_k \leq \min_k R_k^c\left(\{\mathbf{w}_k^p, \mathbf{w}^c\}\right), \ \forall k \in \mathcal{K}, \\ 0, & \sum_{k=1}^K c_k > \min_k R_k^c\left(\{\mathbf{w}_k^p, \mathbf{w}^c\}\right), \ \forall k \in \mathcal{K}, \end{cases}$$

(46)

$$\Omega_{\text{LEO}} = \begin{cases} 1, & \min_k R_k\left(\{\mathbf{w}_k^p, \mathbf{w}^c, c_k\}\right) \geq \xi_{\text{LGU}}, \ \forall k \in \mathcal{K}, \\ 0, & \min_k R_k\left(\{\mathbf{w}_k^p, \mathbf{w}^c, c_k\}\right) < \xi_{\text{LGU}}, \ \forall k \in \mathcal{K}, \end{cases}$$

(47)

and

$$\Omega_{\text{GEO}} = \begin{cases} 1, & \min_m(R_m\left(\{\mathbf{w}_k^p, \mathbf{w}^c\}\right)) \geq \xi_{\text{GEO}}, \ \forall m \in \mathcal{M}, \\ 0, & \min_m(R_m\left(\{\mathbf{w}_k^p, \mathbf{w}^c\}\right)) < \xi_{\text{GEO}}, \ \forall m \in \mathcal{M}. \end{cases}$$

(48)

Ω_{Pow}, Ω_{Com}, Ω_{LEO}, and Ω_{GEO} respectively represent the penalties for actions without satisfying the power budget requirement (23b), the common message decoding requirement (23e), the LEO QoS requirement (23d), and the GEO QoS requirement (23c).

Remark 5. *The penalty terms have been widely used to guarantee a feasible solution, which, however, may induce sparse rewards and high variance. Hence, we only adopt the penalty for the complicated constraint, i.e., (23b)-(23e), while the simple constraint, i.e., (23f) is guaranteed via the activation function based on (40).*

Note that a non-zero reward is awarded only when all the constraints of Problem P_0 are met, which means that our objective can be achieved. Therefore, for the agent to receive positive rewards, it must optimize the sum rate and ensure that all constraints are fully satisfied. In summary, Algorithm 2 describes the training procedure for the proposed MoE-PPO approach.

V. Simulation Results

This section presents simulations to evaluate our proposed generative AI agent framework and our MoE-PPO approach.

A. Simulation Parameters and Setup

1) Scenario Settings: Since our customized system is based on heterogeneous scenarios, the simulated scenario is composed of one LEO and one GEO. The number of antennas at GGU and LGU are both set as 8. The number of GGU and LGU are both set to 2. The power budget of GEO and LEO are both set to 50 dBm. Following [21], the achievable information rate requirements of GGU and LGU are set as 1 b/s/Hz and 0.1 b/s/Hz, respectively. The noise power of each GGU and LGU are both set as -104 dBm/Hz. Moreover, the remaining hyper-parameters are summarized in Table I.

2) Generative AI Agent Settings: The generative AI agent is implemented as follows. First, the encoders for user description and expertise knowledge are realized by *OpenAIEmbeddings* [38]. The LLM for analyzing retrieved knowledge and generating satellite communication models is GPT-3.5 [47]. Finally, the semantic router, agent memory, and conversation chain are developed atop LangChain libraries [48].
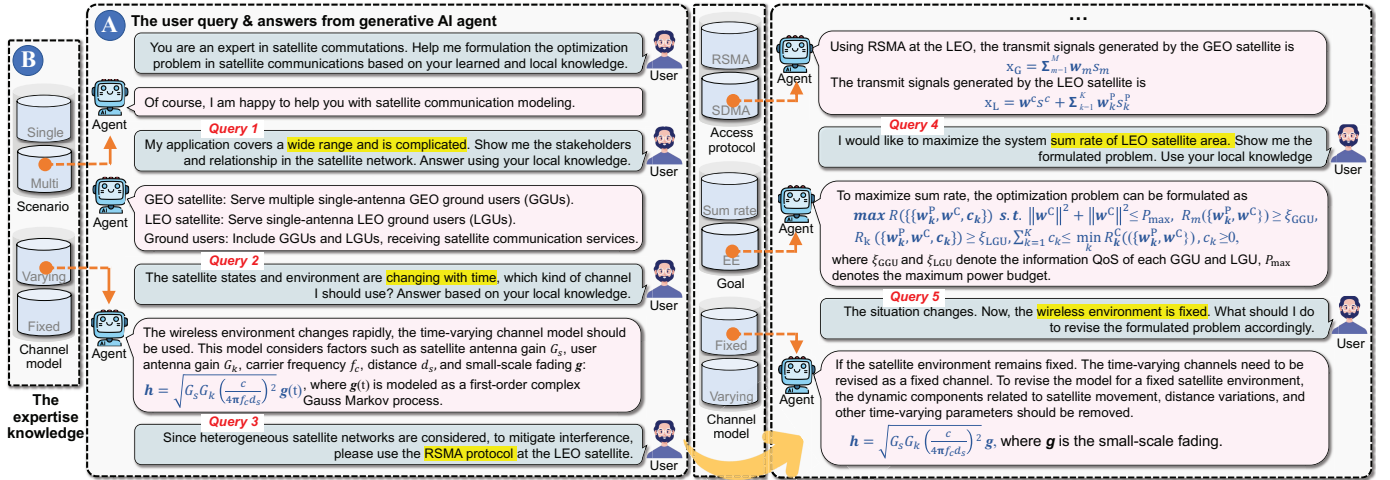
Fig. 4. The process of satellite communication modeling using proposed generative AI agent. **(A)**: The user description and answers of generative AI agent. The words with key semantics are highlighted in yellow. **(B)**: The retrieval from expertise knowledge.

Algorithm 2: The proposed MoE-PPO approach

1 Input : Episodes, exploration steps, update episodes, learning rate, channel status information, number of experts;

2 Initialize environment;

3 Observe initial state $\mathbf{s}_0$ **for** *each itertion* **do**

4   **for** *each exploration step t* **do**

5   Each expert makes its own action a_t' based on sub-policies;

6   Aggregate actions $a_t = \{\mathbf{w}_k^p, \mathbf{w}^c, c_k\}$ based on the current state s_t and gating mechanism;

7   Obtain the current reward r_t;

8   Observe the next state s_{t+1};

9   Store transition (s_t, a_t, r_t, s_{t+1}) in $\mathcal{D}$;

10   **for** *each update step t* **do**

11   Sample B mini-batch of transitions from $\mathcal{D}$ from all experts;

12   Update actor network θ_A and critic network θ_C via SGD according to (34) and (35) for half episodes of total update episodes;

13   Calculate advantage function $A(s, a)$ for all experts and obtain the maximization value $A'(s, a)$ through Back-propagation maximization approach;

14   Update state $s_t \leftarrow s_{t+1}$;

15 Output : The actions $\{\mathbf{w}_k^p, \mathbf{w}^c, c_k\}$;

---

3) MoE-PPO Approach Settings: For the proposed MoE-PPO approach, the whole structure is set as a four-layer feed-forward deep neural network, i.e., one input layer, one output layer, and two hidden layers. The total number of input ports and output ports are $(4K + M + 1)$ and $(2K + 1)$, respectively. Moreover, the remaining hyper-parameters are also summarized in Table I.

TABLE I
HYPER PARAMETERS SETTINGS

System Parameter	Value
Carrier frequency f_c	4 GHZ
Satellite antenna gain G_s	35 dBi
LEO altitude	300 KM
GEO altitude	4000 KM
Doppler frequency f_d	10Hz
Time interval T_s	2×10^{-3}s
Rician factor	4
The number of NN layers	4
Clipping parameter ϵ	0.2
DNN optimizer	*Adam*
Activation function	ReLU and Tanh
Hidden layer size	256
Learning rate α	3×10^{-4}
Discount factor γ	0
Clipped probability ratio μ	0.2
Memory size	40960
Time slots	10
Batch size	2048

B. The Effectiveness of Generative AI Agent

1) Satellite Communications Modeling: First, we evaluate the effectiveness of the proposed generative AI agent in customizing satellite communication models. As shown in Fig. 4, the users apply the *role assignment* strategy at the beginning, enabling the LLM to recall pre-trained satellite knowledge by asking it to act as satellite experts. Then, the users describe the requirements for the satellite communication model from four aspects. We can observe that our agent can precisely capture critical semantic keys from user natural language descriptions and call the corresponding RAG database. Then, the LLM can leverage the retrieved knowledge to generate coherent answers about satellite communications modeling according to user requests. Note that even though the user query is vague, the semantic router can still find the most relevant sub-block by semantic similarity. For instance, the agent precisely routes to the time-varying channel without explicit mention by users. In this case of Fig. 4, the user's requirement corresponds to
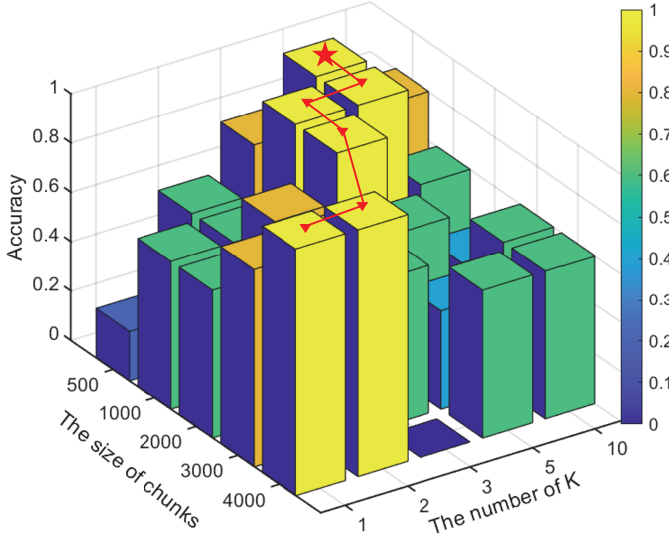
Fig. 5. The impact of chunk size and number on generation accuracy.

Fig. 7. The effectiveness of the different numbers of experts.
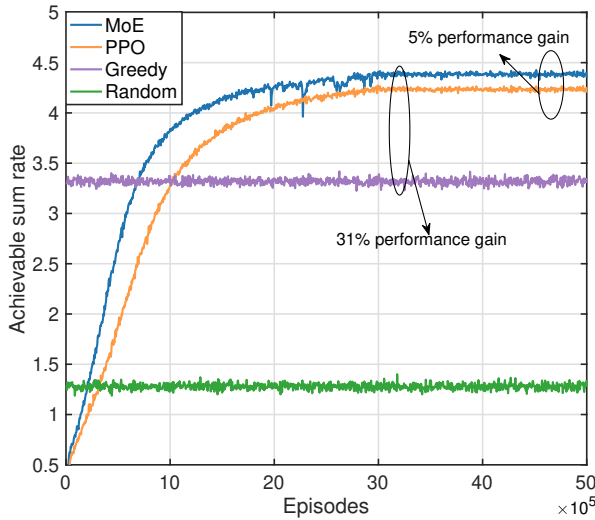

Fig. 6. The convergence of the proposed MoE-PPO approach.

a multiple-satellite network using RSMA protocol and time-varying channels, with the goal of maximizing the system EE. We can observe that the AI agent successfully constructs the entire model using 6 rounds of interactions, reaching the minimum for accomplishing five-step modeling (one extra round is for *role assignment*). Moreover, with the varying network conditions, the user requirement may change over time. Our agent can associate the requirement change to the parts of the models and revise such parts automatically, thus avoiding human-caused errors. As illustrated in Fig. 4, the agent precisely updates the channel model from Rayleigh fading to Rician when the network dynamics decrease.

2) Generation Accuracy: Next, we evaluate the accuracy of the generated model under different agent settings. For our proposed generative AI agent, it leverages RAG to help human users perform customized satellite communications modeling. Expert modeling knowledge is fed to the agent
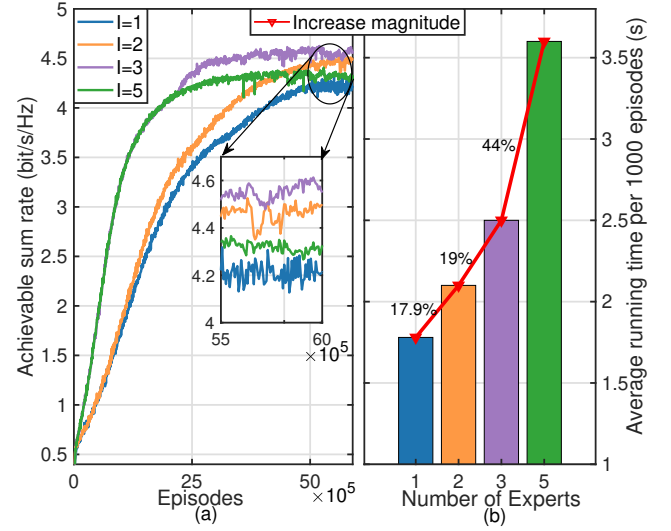
through chunked embeddings. If the correct knowledge chunk[2] that corresponds to the specific user-described scenario is retrieved, the generated answer can be authentic and rational. Therefore, to check the authenticity and rationality of the generated content, the metric named retrieval rate (RR) [49] can be utilized as accuracy. Specifically, RR indicates the ratio that the generative AI agent successfully retrieves the correct knowledge chunk and constructs the answer. Here, we adjust two hyperparameters in RAG, namely the chunk size and the number of chunks. The former means the number of tokens in each chunk; the latter means the number of chunks that can be retrieved at one time. Note that generation accuracy is defined as $n_c/5$, which means the number of correct answers among all five queries in Fig. 4. Fig. 5 demonstrates that low chunk size and number lead to poor generation accuracy. This is because the RAG system cannot fetch expertise to support answer generation. However, if retrieving too much redundant expertise, the LLM of the agent can hardly analyze it effectively, resulting in poor accuracy as well. Therefore, we set chunk size and number to be 500 and 5, respectively, which ensures that our agent can generate precise models with the minimum number of retrieved tokens (2500 in total).

C. The Effectiveness of PPO with MoE Approach

1) The Convergence of PPO with MoE Approach: Fig. 6 illustrates the convergence behavior of the proposed MoE-PPO approach in comparison to some different benchmarks, such as the traditional PPO, greedy, and random selection strategies. The figure reveals that both the MoE-PPO and the standard PPO approaches demonstrate convergence as the number of episodes increases. It shows that MoE-PPO consistently surpasses the performance of PPO (i.e., about 5%), achieving higher sum rates in each iteration. The performance enhancements of MoE-PPO can be attributed to

[2]Chunk is the unit of embedding segments. To increase the retrieval efficiency, in each sub-block, the agent splits the expert knowledge embedding into multiple chunks [14].
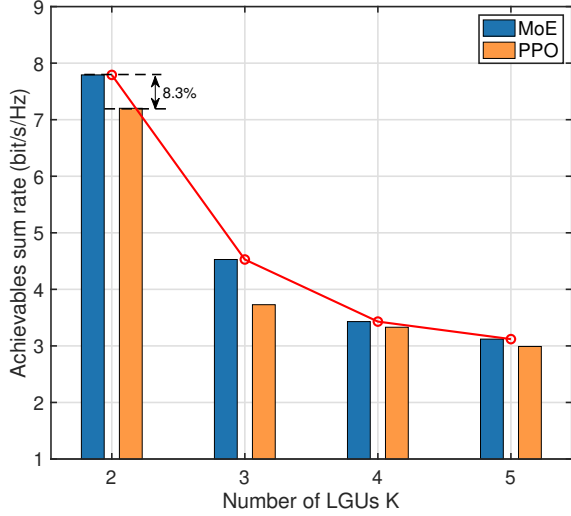
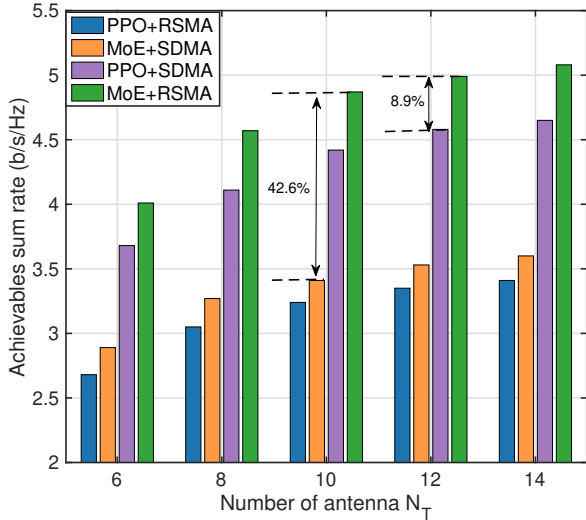Fig. 8. The effectiveness of MoE-PPO under different number of LGUs.

Fig. 9. SDMA design versus RSMA design by using different approaches.

- I=2: Two experts handle the tasks, with one expert optimizing the beamforming variables, i.e., $\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}$, and another expert focusing on the common rate variables $\{c_k\}$.
- I=3: Three experts handle the tasks, where the first expert is responsible for optimizing private beamforming variables, i.e., $\{\mathbf{w}_k^{\mathrm{p}}\}$, the second is responsible for optimizing common beamforming variables, i.e., $\mathbf{w}^{\mathrm{c}}$, and the third expert is responsible for optimizing common rate variables, i.e., $\{c_k\}$.
- I=5: Five experts are involved, with each expert specializing in the optimization of a specific variable.

Fig. 7 illustrates the impact of varying the number of experts within the MoE-PPO framework on system performance. As depicted in Fig. 7(a), the MoE-PPO approach converges across the different numbers of experts, with the achievable sum rate improving as the number of experts increases. Notably, when the number of experts is set to three, the system achieves the highest sum rate. This optimal performance is attributed to the allocation of experts to distinct variable categories, enabling each expert to specialize in and optimize variables within their expertise domain, followed by an effective gating process that aggregates the optimized variables. In contrast, a decrease in performance is observed when the number of experts is five. This counterintuitive result can be explained by over-segmentation of the optimization task; each expert optimizes a single variable, leading to errors in the gating process due to lack of visibility into other variables or their corresponding category variables. Furthermore, Fig. 7(b) presents the average running time with the different numbers of experts. An exponential increase in running time is observed with more experts, a consequence of the additional computational overhead required to calculate and aggregate the decisions from a greater number of experts. Therefore, to balance effectiveness and efficiency, the number of experts should be equal to the types of variables to ensure that each expert can fully utilize their expertise without incurring excessive computational costs or gating errors.

3) Different number of LGUs: Fig. 8 illustrates the achievable sum rate as a function of the number of LGUs. It is observed that the achievable sum rate decreases as the number of users increases for both the proposed MoE-PPO approach and the pure PPO approach. This decline occurs because, with increasing users, the approach must allocate more communication resources to users with poor channel quality to meet constraints, reducing the overall sum rate. Moreover, it is evident that the proposed MoE-PPO method consistently outperforms the PPO method, where the MoE-PPO approach achieves an 8.3% higher sum rate compared to the PPO approach when there are 2 LGUs. This superiority is due to the MoE's ability to leverage multiple specialized experts to handle different aspects of the optimization problem more effectively. Specifically, MoE-PPO can dynamically select the most relevant experts, optimizing resource allocation more efficiently than a single comprehensive network.
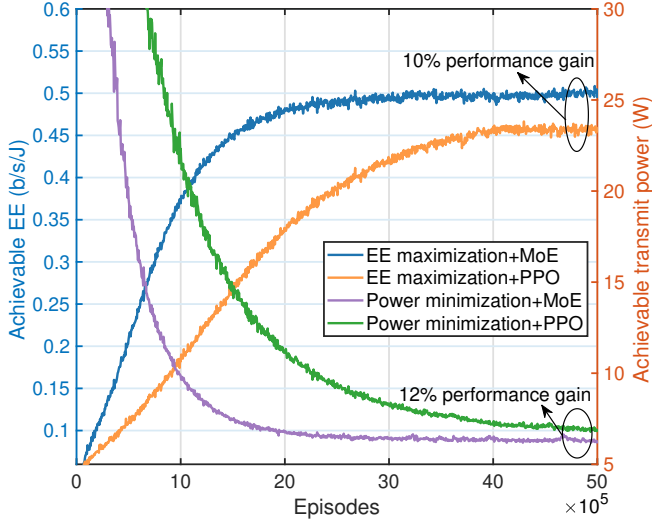
the integration of the MoE framework, which utilizes expert policy specialization and adaptive policy weights to more effectively navigate the policy space. To further validate the effectiveness of MoE-PPO, its comparison with the greedy and random baselines after convergence underscores a significant improvement (i.e., about 220% and 38%). This improvement is facilitated by MoE-PPO's enhanced collaborative capabilities and more efficient utilization of the state-action space using gating network.

To validate the adaptability of the proposed MoE-PPO approach, we conducted studies under varying protocols (i.e., considering SDMA) and optimization goals (i.e., considering EE or power minimization).

2) Different Number of Experts: To illustrate the impact of the number of experts on optimization outcomes, we consider the scenarios with different numbers of experts:

- I=1: One expert optimizes all variables, i.e., $\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}, c_k\}$.

Fig. 10. The effectiveness of MoE-PPO under different optimization goals with $\mu = 1$ and $P_C = 1$.

D. The Adaptability of the generative AI agent and MoE-PPO approach

1) MoE-PPO approach for different access protocols from generative AI agent construction: Fig. 9 presents the achievable sum rate obtained by RSMA and SDMA versus the number of transmit antennas N_t. It shows that the achievable sum rate increases with the number of N_t for all considered protocols and approaches. This increment is attributable to the additional spatial degrees of freedom provided by more antennas, which help mitigate interference and enhance system capacity. Moreover, it is observed that the system performance under RSMA outperforms that under SDMA (about 42.6% gain). This superiority of RSMA can be ascribed to the employment of successive interference cancellation (SIC) techniques, which effectively remove a part of the interference. Additionally, under each antenna configuration, our MoE-PPO approach always surpasses the PPO approach, which is consistent with the previous simulations.

2) MoE-PPO approach for different optimization goals from generative AI agent construction: Furthermore, Fig. 10 illustrates the different system performance of the proposed MoE-PPO approach under a time-varying mode with different optimization goals, i.e., EE maximization and transmit power minimization. In this context, the corresponding reward functions for MoE-PPO approach are respectively designed as follows.

- For EE maximization:

$$
\begin{aligned}
r_{\mathrm{EE}} = & \mathrm{EE}\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}, c_k\}\right) \\
& \times \underbrace{(\Omega_{\mathrm{Pow}} \times \Omega_{\mathrm{Com}} \times \Omega_{\mathrm{LEO}} \times \Omega_{\mathrm{GEO}})}_{\text{Penalty term}},
\end{aligned}
\tag{49}
$$

- For power minimization:

$$
r_{\mathrm{P}} = \frac{\overbrace{(\Omega_{\mathrm{Pow}} \times \Omega_{\mathrm{Com}} \times \Omega_{\mathrm{LEO}} \times \Omega_{\mathrm{GEO}})}^{\text{Penalty term}}}{P_{\mathrm{T}}\left(\{\mathbf{w}_k^{\mathrm{p}}, \mathbf{w}^{\mathrm{c}}\}\right)}.
\tag{50}
$$

It is observed that whether maximizing EE or minimizing power consumption, our proposed MoE-PPO approach always outperforms the traditional PPO (about 10% and 12%) with the increment of episodes. The reason is that the MoE structure is able to efficiently allocate computing resources because each expert can operate in parallel to solve different variables in which they specialize. This parallelism facilitates deep exploration of the state space, resulting in better policy updates at each stage. Next, it also demonstrates the adaptability of the proposed MoE-PPO approach, confirming its ability to meet different channel conditions and optimization goals.

VI. CONCLUSION

This paper proposed a generative AI agent framework and an MoE-PPO method for network modeling and transmission strategy design in satellite communications networks, respectively. Specifically, the proposed framework utilized LLM and RAG for adaptive system modeling and configuration automation, while MoE-PPO optimized resource allocation and interference management by integrating expert knowledges. Simulation results confirmed the effectiveness of the generative AI agent framework and MoE-PPO approach in satellite communication networks.

REFERENCES

[1] J. Liu, Y. Shi, and et. al., "Space-air-ground integrated network: A survey," *IEEE Commun. Surv. Tut.*, vol. 20, no. 4, pp. 2714–2741, 2018.

[2] S. Liu, Z. Gao, and et. al., "LEO satellite constellations for 5G and beyond: How will they reshape vertical domains?" *IEEE Commun. Mag.*, vol. 59, no. 7, pp. 30–36, 2021.

[3] H. Chen, M. Xiao, and Z. Pang, "Satellite-based computing networks with federated learning," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 78–84, 2022.

[4] O. Kodheli and et. al., "Satellite communications in the new space era: A survey and future challenges," *IEEE Commun. Surv. Tut.*, vol. 23, no. 1, pp. 70–109, 2021.

[5] L. Qian, P. Yang, Y. L. Guan, and et. al., "Multi-dimensional polarized modulation for land mobile satellite communications," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 383–397, 2021.

[6] P. Yue, J. An, J. Zhang, and et. al., "Low earth orbit satellite security and reliability: Issues, solutions, and the road ahead," *IEEE commun. Surv. Tut.*, vol. 25, no. 3, pp. 1604–1652, 2023.

[7] A. Jamalipour and et. al., "Performance of an integrated voice/data system in nonuniform traffic low earth-orbit satellite communication systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 2, pp. 1–1, 1995.

[8] B. Li and et. al., "A novel frequency reuse model for co-existing LEO and GEO satellites," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2024.

[9] R. Deng, B. Di, H. Zhang, and et. al., "Ultra-dense LEO satellite constellations: How many LEO satellites do we need?" *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 4843–4857, 2021.

[10] Z. Xiao, J. Yang, T. Mao, and et. al., "LEO satellite access network (LEO-SAN) towards 6G: Challenges and approaches," *IEEE Wireless Commun.*, pp. 1–8, 2022.

[11] J. S. Park and et. al., "Generative agents: Interactive simulacra of human behavior," in *Proc. ACM USIT*, 2023, pp. 1–22.

[12] R. Zhang, H. Du, Y. Liu, D. Niyato, and et. al., "Interactive AI with retrieval-augmented generation for next generation networking," *IEEE Network*, 2024.

[13] H. Du, R. Zhang, D. Niyato, and et. al., "User-centric interactive AI for distributed diffusion model-based AI-generated content," *arXiv preprint arXiv:2311.11094*, 2023.

[14] P. Lewis and et. al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," *NeurIPS*, vol. 33, pp. 9459–9474, 2020.

[15] R. Zhang, H. Du, and et. al., "Generative AI for space-air-ground integrated networks (SAGIN)," *arXiv preprint arXiv:2311.06523*, 2023.

[16] S. E. Yuksel and et. al., "Twenty years of mixture of experts," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 23, no. 8, pp. 1177–1193, 2012.

[17] S. Masoudnia and R. Ebrahimpour, "Mixture of experts: a literature survey," *Artif. Intell. Rev.*, vol. 42, pp. 275–293, 2014.

[18] Z. Chen, Y. Shen, M. Ding, and et. al., "Mod-Squad: Designing mixtures of experts as modular multi-task learners," in *Proc. IEEE/CVF CVPR*, 2023, pp. 11 828–11 837.

[19] J. Zhang, Z. Tang, M. Li, and et. al., "Crosssense: Towards cross-site and large-scale wifi sensing," in *Proc. ICMCN*, 2018, pp. 305–320.

[20] Z. Gao, A. Liu, C. Han, and X. Liang, "Sum rate maximization of massive MIMO NOMA in LEO satellite communication system," *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1667–1671, 2021.

[21] J. Huang, Y. Yang, J. Lee, and et. al., "Deep reinforcement learning based resource allocation for RSMA in LEO satellite-terrestrial networks," *IEEE Trans. Commun.*, 2023.

[22] W. U. Khan, Z. Ali, E. Lagunas, and et. al., "Rate splitting multiple access for next generation cognitive radio enabled leo satellite networks," *IEEE Trans. Wireless Commun.*, 2023.

[23] J. Li, K. Xue, and et. al., "Energy efficiency and traffic offloading optimization in integrated satellite/terrestrial radio access networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2367–2381, 2020.

[24] H. Du, R. Zhang, Y. Liu, J. Wang, Y. Lin, Z. Li, D. Niyato, J. Kang, Z. Xiong, S. Cui *et al.*, "Enhancing deep reinforcement learning: A tutorial on generative diffusion models in network optimization," *IEEE Commun. Surv. Tut.*, 2024.

[25] X. Wang, Y. Fei, Z. Leng, and C. Li, "Does role-playing chatbots capture the character personalities? assessing personality traits for role-playing chatbots," *arXiv preprint arXiv:2310.17976*, 2023.

[26] B. Y. Lin and et. al., "Swiftsage: A generative agent with fast and slow thinking for complex interactive tasks," *arXiv preprint arXiv:2305.17390*, 2023.

[27] R. A. Jacobs and et. al., "Adaptive mixtures of local experts," *Neural computation*, vol. 3, no. 1, pp. 79–87, 1991.

[28] D. Eigen and et. al., "Learning factored representations in a deep mixture of experts," *arXiv preprint arXiv:1312.4314*, 2013.

[29] N. Du and et. al., "Glam: Efficient scaling of language models with mixture-of-experts," in *Proc. ICML*, 2022, pp. 5547–5569.

[30] R. K. Jaiswal, M. Elnourani, S. Deshmukh, and et. al., "Leveraging transfer learning for radio map estimation via mixture of experts," *Authorea Preprints*, 2023.

[31] L. M. Lopez-Ramos and et. al., "Channel gain cartography via mixture of experts," in *Proc. IEEE Globecom*, 2020, pp. 1–7.

[32] B. Deng, C. Jiang, H. Yao, S. Guo, and S. Zhao, "The next generation heterogeneous satellite communication networks: Integration of resource management and deep reinforcement learning," *IEEE Wireless Comm.*, vol. 27, no. 2, pp. 105–111, 2019.

[33] B. Clerckx, Y. Mao, E. A. Jorswieck, and et. al., "A primer on rate-splitting multiple access: Tutorial, myths, and frequently asked questions," *IEEE J. Sel. Area Comm.*, 2023.

[34] R. Zhang, K. Xiong, Y. Lu, B. Gao, P. Fan, and K. B. Letaief, "Joint coordinated beamforming and power splitting ratio optimization in MU-MISO SWIPT-enabled HetNets: A multi-agent ddqn-based approach," *IEEE J. Sel. Area. Comm.*, vol. 40, no. 2, pp. 677–693, 2022.

[35] Y. S. Nasir and D. Guo, "Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks," *IEEE J. Sel. Area Comm.*, vol. 37, no. 10, pp. 2239–2250, 2019.

[36] K. Ardah and et. al., "Hybrid analog-digital beamforming design for SE and EE maximization in massive MIMO networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 377–389, 2019.

[37] Semantic router. 2024. [Online]. Available: https://python.langchain.com/docs/expression_language/cookbook/embedding_router

[38] Open ai textual embedding model. 2024. [Online]. Available: https://openai.com/blog/new-and-improved-embedding-model

[39] E. Kasneci, K. Seßler, and et. al., "Chatgpt for good? On opportunities and challenges of large language models for education," *Learning and individual differences*, vol. 103, p. 102274, 2023.

[40] K. Shen and W. Yu, "Fractional programming for communication systems-Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, 2018.

[41] H. Ye and D. Xu, "Taskexpert: Dynamically assembling multi-task representations with memorial mixture-of-experts," in *Proc. IEEE/CVF ICCV*, 2023, pp. 21 828–21 837.

[42] R. Zhang, K. Xiong, Y. Lu, P. Fan, D. W. K. Ng, and K. B. Letaief, "Energy efficiency maximization in RIS-assisted SWIPT networks with rsma: A ppo-based approach," *IEEE J. Sel. Area. Comm.*, vol. 41, no. 5, pp. 1413–1430, 2023.

[43] J. Ren, Y. Li, Z. Ding, W. Pan, and H. Dong, "Probabilistic mixture-of-experts for efficient deep reinforcement learning," *arXiv preprint arXiv:2104.09122*, 2021.

[44] J. Obando-Ceron, G. Sokar, T. Willi, C. Lyle, J. Farebrother, J. Foerster, G. K. Dziugaite, D. Precup, and P. S. Castro, "Mixtures of experts unlock parameter scaling for deep rl," *arXiv preprint arXiv:2402.08609*, 2024.

[45] B. Clerckx and et. al., "Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: A simple two-user rate analysis," *IEEE Wireless Comm. Lett.*, vol. 9, no. 3, pp. 349–353, 2019.

[46] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE J. Sel. Area Comm.*, vol. 38, no. 8, pp. 1839–1850, 2020.

[47] The gpt-3.5 model. 2024. [Online]. Available: https://openai.com/blog/gpt-3-5-turbo-fine-tuning-and-api-updates

[48] The langchain library. 2024. [Online]. Available: https://www.langchain.com/

[49] J. Chen, H. Lin, X. Han, and L. Sun, "Benchmarking large language models in retrieval-augmented generation," in *Proc. AAAI*, vol. 38, no. 16, 2024, pp. 17 754–17 762.

Ruichen Zhang is currently working as a Post-Doctoral Research Fellow with the College of Computing and Data Science, Nanyang Technological University (NTU), Singapore. He received the B.E. degree from Henan University (HENU), China, in 2018, and the Ph.D. degree from Beijing Jiaotong University (BJTU), China, in 2023. In 2024, he was a Visiting Scholar with the College of Information and Communication Engineering, Sungkyunkwan University, Suwon, South Korea. His research interests include LLM-empowered networking, reinforcement learning-enabled wireless communication, generative AI models, and heterogeneous networks.

Hongyang Du is an assistant professor at the Department of Electrical and Electronic Engineering, The University of Hong Kong. He received the BEng degree from the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, in 2021, and the PhD degree from the Interdisciplinary Graduate Program at the College of Computing and Data Science, Energy Research Institute @ NTU, Nanyang Technological University, Singapore, in 2024. He is the Editor-in-Chief assistant of IEEE Communications Surveys & Tutorials (2022-2024). He is the recipient of the IEEE Daniel E. Noble Fellowship Award from the IEEE Vehicular Technology Society in 2022, the IEEE Signal Processing Society Scholarship from the IEEE Signal Processing Society in 2023, the Chinese Government Award for Outstanding Students Abroad in 2023, and the Singapore Data Science Consortium (SDSC) Dissertation Research Fellowship in 2023. He was recognized as an exemplary reviewer of the IEEE Transactions on Communications and IEEE Communications Letters in 2021. His research interests include edge intelligence, generative AI, semantic communications, and network management.

Yinqiu Liu received B.Eng. degree from Nanjing University of Posts and Telecommunications, China in 2020 and M.Sc. degree from the University of California, Los Angeles in 2022. He is currently a Ph.D. student at the College of Computing and Data Science, Nanyang Technological University, Singapore. He won the Honorary Mention award in the ComSoc Student Competition from IEEE Communications Society in 2023, and the First Prize in the 2024 ComSoc Social Network Technical Committee (SNTC) Student Competition. His current research interests include wireless communications, mobile AIGC, and generative AI.

Dusit Niyato (M'09-SM'15-F'17) is a professor in the College of Computing and Data Science, at Nanyang Technological University, Singapore. He received B.Eng. from King Mongkuts Institute of Technology Ladkrabang (KMITL), Thailand and Ph.D. in Electrical and Computer Engineering from the University of Manitoba, Canada. His research interests are in the areas of mobile generative AI, edge intelligence, decentralized machine learning, and incentive mechanism design.

Abbas Jamalipour (S'86-M'91-SM'00-F'07) received the Ph.D. degree in Electrical Engineering from Nagoya University, Nagoya, Japan in 1996. He holds the positions of Professor of Ubiquitous Mobile Networking with The University of Sydney and since January 2022, the Editor-in-Chief of the IEEE Transactions on Vehicular Technology. Prof. Jamalipour is a recipient of the number of prestigious awards, such as the 2019 IEEE ComSoc Distinguished Technical Achievement Award in Green Communications, the 2016 IEEE ComSoc Distinguished Technical Achievement Award in Communications Switching and Routing, the 2010 IEEE ComSoc Harold Sobol Award, the 2006 IEEE ComSoc Best Tutorial Paper Award, as well as over 15 Best Paper Awards. He was the President of the IEEE Vehicular Technology Society (2020-2021). Previously, he held the positions of the Executive Vice-President and the Editor-in-Chief of VTS Mobile World and has been an elected member of the Board of Governors of the IEEE Vehicular Technology Society since 2014. He was the Editor-in-Chief IEEE WIRELESS COMMUNICATIONS, the Vice President-Conferences, and a member of Board of Governors of the IEEE Communications Society. He sits on the Editorial Board of the IEEE ACCESS and several other journals and is a member of Advisory Board of IEEE Internet of Things Journal. He has been the General Chair or Technical Program Chair for several prestigious conferences, including IEEE ICC, GLOBECOM, WCNC, and PIMRC. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), the Institute of Electrical, Information, and Communication Engineers (IEICE), and the Institution of Engineers Australia, an ACM Professional Member, and an IEEE Distinguished Speaker

Jiawen Kang (M'18-SM'23) received the Ph.D. degree from the Guangdong University of Technology, China in 2018. He was a postdoc at Nanyang Technological University, Singapore from 2018 to 2021. He currently is a full professor at Guangdong University of Technology, China. His research interests mainly focus on blockchain, security, and privacy protection in wireless communications and networking. He has published more than 170 research papers in leading journals and flagship conferences including 12 ESI highly-cited papers and 3 ESI hot papers. He has won IEEE VTS Best PaperAward, IEEE Communications Society CSIM Technical Committee Best Journal Paper Award, IEEE Best Land Transportation Paper Award, IEEE HITC Award for Excellence in Hyper-Intelligence Systems (Early Career Researcher award), IEEE Computer Society Smart Computing Special Technical Community Early-Career Award, and 14 best paper awards of international conferences.

Dong In Kim (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1990. He was a Tenured Professor with the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada. He is currently a Distinguished Professor with the College of Information and Communication Engineering, Sungkyunkwan University, Suwon, South Korea. He was the first recipient of the NRF of Korea Engineering Research Center in Wireless Communications for RF Energy Harvesting from 2014 to 2021. He received several research awards, including the 2023 IEEE ComSoc Best Survey Paper Award and the 2022 IEEE Best Land Transportation Paper Award. He was selected the 2019 recipient of the IEEE ComSoc Joseph LoCicero Award for Exemplary Service to Publications. He has been listed as a Highly Cited Researcher by Clarivate Analytics in 2020 and 2022. He was the General Chair of the IEEE ICC 2022, Seoul. Since 2001, he has been serving as an Editor, an Editor at Large, and an Area Editor for Wireless Communications I for IEEE Transactions on Communications. From 2002 to 2011, he served as an Editor and a Founding Area Editor for Cross-Layer Design and Optimization for IEEE Transactions on Wireless Communications. From 2008 to 2011, he served as the Co-Editor-in-Chief for the IEEE/KICS Journal of Communications and Networks. He served as the Founding Editor-in-Chief for the IEEE Wireless Communications Letters from 2012 to 2015. He is a Fellow of the Korean Academy of Science and Technology and a member of the National Academy of Engineering of Korea.

Zehui Xiong (Senior Member, IEEE) is currently an Assistant Professor at Singapore University of Technology and Design, and also an Honorary Adjunct Senior Research Scientist with Alibaba-NTU Singapore Joint Research Institute, Singapore. He received the PhD degree in Nanyang Technological University (NTU), Singapore. He was the visiting scholar at Princeton University and University of Waterloo. His research interests include wireless communications, Internet of Things, blockchain, edge intelligence, and Metaverse. He has won over 10 Best Paper Awards in international flagship conferences. He is now serving as the editor or guest editor for many leading journals including IEEE Journal on Selected Areas in Communications, IEEE Transactions on Vehicular Technology, IEEE Internet of Things Journal, IEEE Transactions on Cognitive Communications and Networking, and IEEE Transactions on Network Science and Engineering. He is the recipient of many prestigious awards including Forbes Asia 30u30, IEEE Early Career Researcher Award for Excellence in Scalable Computing, IEEE Technical Committee on Blockchain and Distributed Ledger Technologies Early Career Award, IEEE Internet Technical Committee Early Achievement Award, IEEE TCSVC Rising Star Award, IEEE TCI Rising Star Award, IEEE TCCLD Rising Star Award, IEEE ComSoc Outstanding Paper Award, IEEE Best Land Transportation Paper Award, IEEE CSIM Technical Committee Best Journal Paper Award, IEEE SPCC Technical Committee Best Paper Award, and IEEE VTS Singapore Best Paper Award.