WirelessLLM-Agent: A Unified LLM-Based Agent Framework for Multi-Task Wireless Communication Decision-Making

Gregory Yu *, Ian Butler , Aaron Collins

Posted Date: 13 April 2026

doi: 10.20944/preprints202604.0849.v1

Keywords: large language models; wireless communication; multi-task learning; reinforcement learning; edge computing

 doi:10.20944/preprints202604.0849.v1

Article

WirelessLLM-Agent: A Unified LLM-Based Agent Framework for Multi-Task Wireless Communication Decision-Making

Gregory Yu *, Ian Butler and Aaron Collins

Higher Technological Institute of Irapuato
* Correspondence: lis18111368@irapuato.tecnm.mx

Abstract

The integration of large language models into wireless communication has shown promising results for individual tasks. However, existing approaches are typically designed for single-task scenarios and rely on supervised fine-tuning that fails to optimize for long-term decision quality. In this paper, we propose WirelessLLM-Agent, a unified LLM-based agent framework for multi-task wireless communication decision-making. Our framework integrates a semantic state serialization module that transforms heterogeneous wireless states into structured textual representations, a multi-task adapter architecture based on MoE-LoRA for parameter-efficient knowledge sharing, and a two-stage training paradigm combining SFT warm-start with GRPO reinforcement learning enhanced by lookahead collaborative simulation. Extensive experiments on channel multi-task learning, mobile edge computing task offloading, and cooperative edge caching demonstrate that WirelessLLM-Agent consistently outperforms existing methods while exhibiting strong zero-shot generalization.

Keywords: large language models; wireless communication; multi-task learning; reinforcement learning; edge computing

1. Introduction

The rapid evolution toward sixth-generation (6G) wireless systems has introduced unprecedented challenges, including massive device connectivity, complex task orchestration, limited spectral resources, and heterogeneous network architectures [1]. The rapid development of intelligent traffic forecasting further highlights the importance of vision-based models for spatiotemporal prediction in communication networks [2]. Traditional optimization approaches, such as heuristic algorithms and deep reinforcement learning (DRL), have been widely deployed for wireless resource management [3]. However, these methods suffer from fundamental limitations in real-time adaptability, scalability, and the ability to comprehend dynamic user intents expressed in natural language [4].

Large language models (LLMs) have recently emerged as a transformative paradigm for decision-making in wireless communications [4]. Leveraging their capabilities in semantic understanding, contextual reasoning, and structured inference, LLMs can process complex network states and generate intelligent control decisions without requiring handcrafted feature engineering [5]. Recent studies have demonstrated the potential of LLMs across various wireless tasks, including channel estimation and prediction [6], beamforming optimization [7], task offloading in mobile edge computing [8], and cooperative edge caching [9]. Recent advances in agentic LLM frameworks have also shown promise for verifiable and safe policy execution in complex systems [10], as well as scientific discovery and falsification [11].

Despite these advances, several critical challenges remain. First, most existing LLM-based approaches are designed for single-task scenarios, lacking a unified framework that can handle the diversity of wireless decision-making tasks [12]. Second, while supervised fine-tuning (SFT) enables

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2026 doi:10.20944/preprints202604.0849.v1

LLMs to mimic expert behaviors, it fails to optimize for long-term decision quality, often leading to myopic policies [13]. Third, the domain gap between general-purpose pre-training corpora and wireless communication knowledge significantly limits LLM performance, as evidenced by the substantial accuracy drop in wireless-specific benchmarks compared to general domains [14].

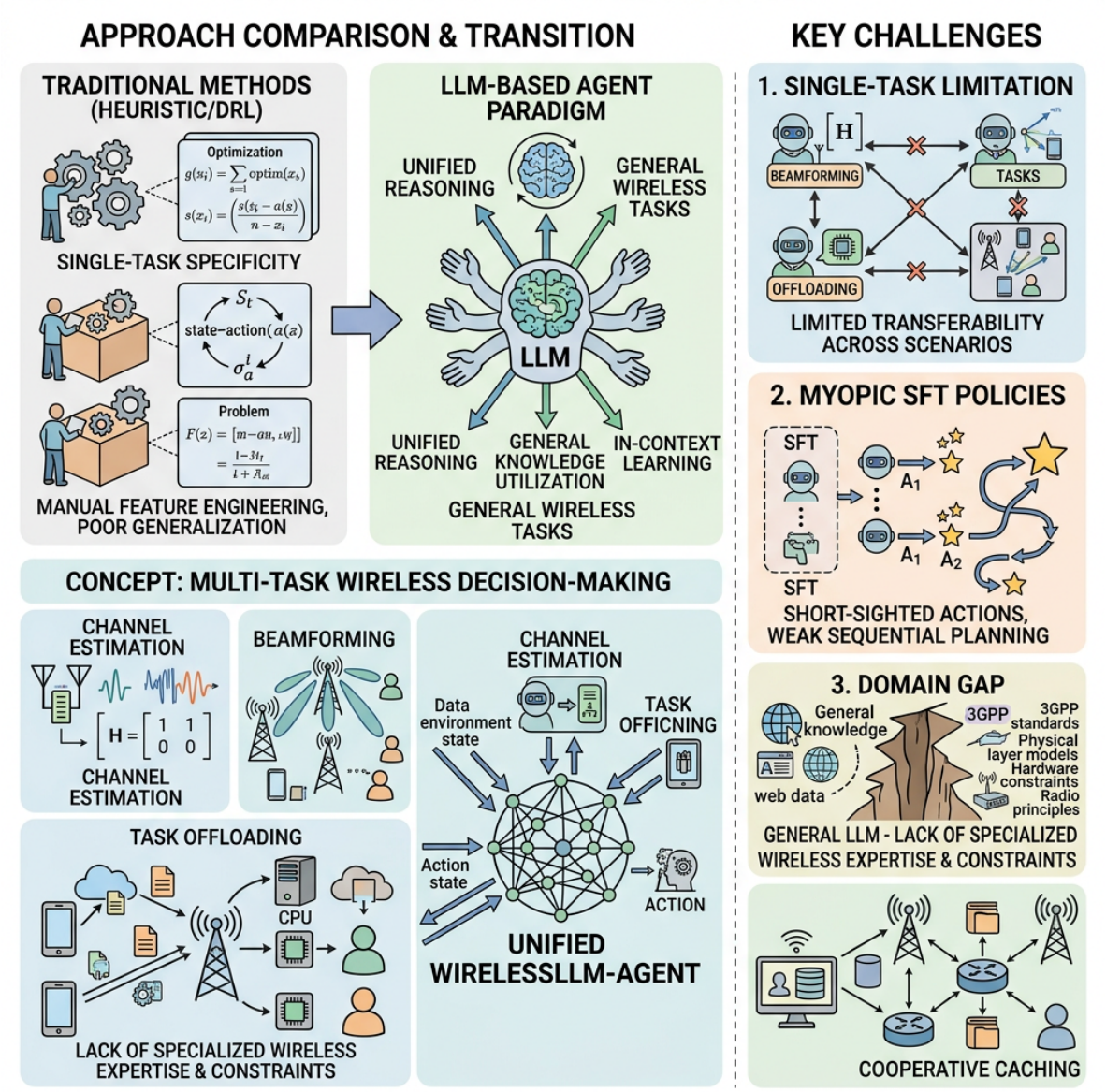

Figure 1. Overview of the transition from traditional optimization to LLM-based agent paradigm for wireless communication decision-making, highlighting key challenges and the proposed unified framework.

To address these challenges, we propose **WirelessLLM-Agent**, a unified LLM-based agent framework for multi-task wireless communication decision-making. Our framework integrates three key components: (1) a *semantic state serialization module* that transforms heterogeneous wireless network states into structured textual representations; (2) a *multi-task adapter architecture* based on Mixture-of-Experts Low-Rank Adaptation (MoE-LoRA) that enables parameter-efficient fine-tuning across diverse wireless tasks; and (3) a *two-stage training paradigm* combining SFT warm-start with Group Relative Policy Optimization (GRPO) reinforcement learning to achieve both behavioral alignment and long-term decision optimization.

We evaluate WirelessLLM-Agent on three representative wireless communication scenarios using datasets generated from the 3GPP TR 38.901 channel model, including channel multi-task learning (covering channel estimation, prediction, frequency prediction, beamforming, distance estimation, and

 doi:10.20944/preprints202604.0849.v1

path loss estimation), mobile edge computing task offloading, and cooperative edge caching. Experimental results demonstrate that our method consistently outperforms existing baselines, achieving an average NMSE of 0.098 for channel estimation (vs. 0.106 for LLM4CP), beamforming accuracy of 0.912 (vs. 0.858 for Cross-stitch), task offloading delay of 2.95 (vs. 3.12 for GRPO-7B), and cache hit rate of 0.558 (vs. 0.542 for GRPO LLM).

Our main contributions are as follows:

- We propose WirelessLLM-Agent, a unified LLM-based agent framework that addresses multiple wireless communication decision-making tasks through semantic state serialization, multi-task adapter architecture, and a two-stage SFT-GRPO training paradigm.
- We design a MoE-LoRA-based multi-task adapter that enables parameter-efficient knowledge sharing across diverse wireless tasks while maintaining task-specific expertise, achieving superior performance with only 1.13M trainable parameters.
- We demonstrate through extensive experiments that WirelessLLM-Agent consistently outperforms existing methods across channel estimation, beamforming, task offloading, and cooperative caching scenarios, while exhibiting strong zero-shot generalization to unseen network configurations.

2. Related Work

2.1. LLM-Based Methods for Wireless Communication Optimization

The application of large language models to wireless communication has attracted significant research attention. Yang et al. [4] provided a comprehensive survey on LLM-empowered decision-making for wireless systems, identifying prompt engineering, retrieval-augmented generation, tool use, and fine-tuning as key enabling techniques. Shao et al. [5] proposed WirelessLLM, a framework based on knowledge alignment, fusion, and evolution principles, demonstrating improvements in protocol understanding and spectrum sensing tasks. Liu et al. [6] introduced LLM4WM, which adapts GPT-2 small through MoE-LoRA for multi-task wireless learning across six channel-related tasks, achieving state-of-the-art performance with only 1.13M trainable parameters. Liang et al. [7] explored the spectrum from adapting pre-trained LLMs to developing wireless-specific foundation models and agent-based LLMs, demonstrating robust beamforming prediction under frequency mismatch conditions. Wei et al. [12] surveyed LLM-enabled wireless network optimization frameworks, highlighting natural language-based problem formulation and solver collaboration. Chen et al. [15] investigated split fine-tuning strategies for deploying LLMs in wireless networks, addressing the computational constraints of edge devices. Lin et al. [16] constructed a multi-hop reasoning dataset for wireless communication and proposed pointwise V-information based parameter-efficient fine-tuning with curriculum learning. The WiFo model [17] pioneered wireless foundation models for channel prediction, enabling zero-shot transfer across scenarios. Recent work has also explored agentic frameworks for bibliographic traceability in scientific literature [18], demonstrating the growing applicability of LLM agents. However, these approaches primarily focus on individual tasks or limited task groups, lacking a unified framework that jointly addresses channel, computing, and caching decisions.

2.2. Reinforcement Learning and Agent Frameworks for Wireless Networks

Reinforcement learning has been extensively studied for wireless network optimization. Alwarafy et al. [3] surveyed DRL-based radio resource allocation in heterogeneous networks, building upon earlier work on deep neural networks for resource management in NOMA networks [19], and identified challenges in scalability and convergence. For mobile edge computing, Yang et al. [8] proposed COMLLM, which reformulates task offloading as a language-conditioned sequential decision problem and employs SFT warm-start with GRPO reinforcement learning, achieving 96.86% optimality ratio with zero task drop rate. For cooperative edge caching, Yang et al. [9] introduced an LLM-based multi-BS orchestrator using SFT and GRPO two-stage training, approaching exhaustive search performance while outperforming DRL baselines including DDPG and SAC. Tong et al. [20] proposed WirelessAgent,

 doi:10.20944/preprints202604.0849.v1

integrating perception, memory, planning, and action modules based on LangGraph for network slicing, achieving bandwidth utilization within 4.3% of rule-based optimality. Traditional multi-agent DRL approaches for cooperative edge caching [21] suffer from coordination overhead and slow convergence in dynamic environments. Wang et al. [13] introduced chain-of-thought reasoning for LLM-empowered wireless communications, enabling intent-driven multi-layer reasoning from high-level user intents to concrete control policies. While these methods demonstrate the potential of RL and agent-based approaches, they typically operate in isolation for specific tasks and do not leverage the generalization capabilities of pre-trained language models for cross-task knowledge transfer.

3. Method

In this section, we present the proposed **WirelessLLM-Agent** framework, a unified LLM-based agent for multi-task wireless communication decision-making. The framework comprises three core components: Semantic State Serialization, Multi-Task Adapter Architecture, and Two-Stage Training Paradigm. We detail each component below.

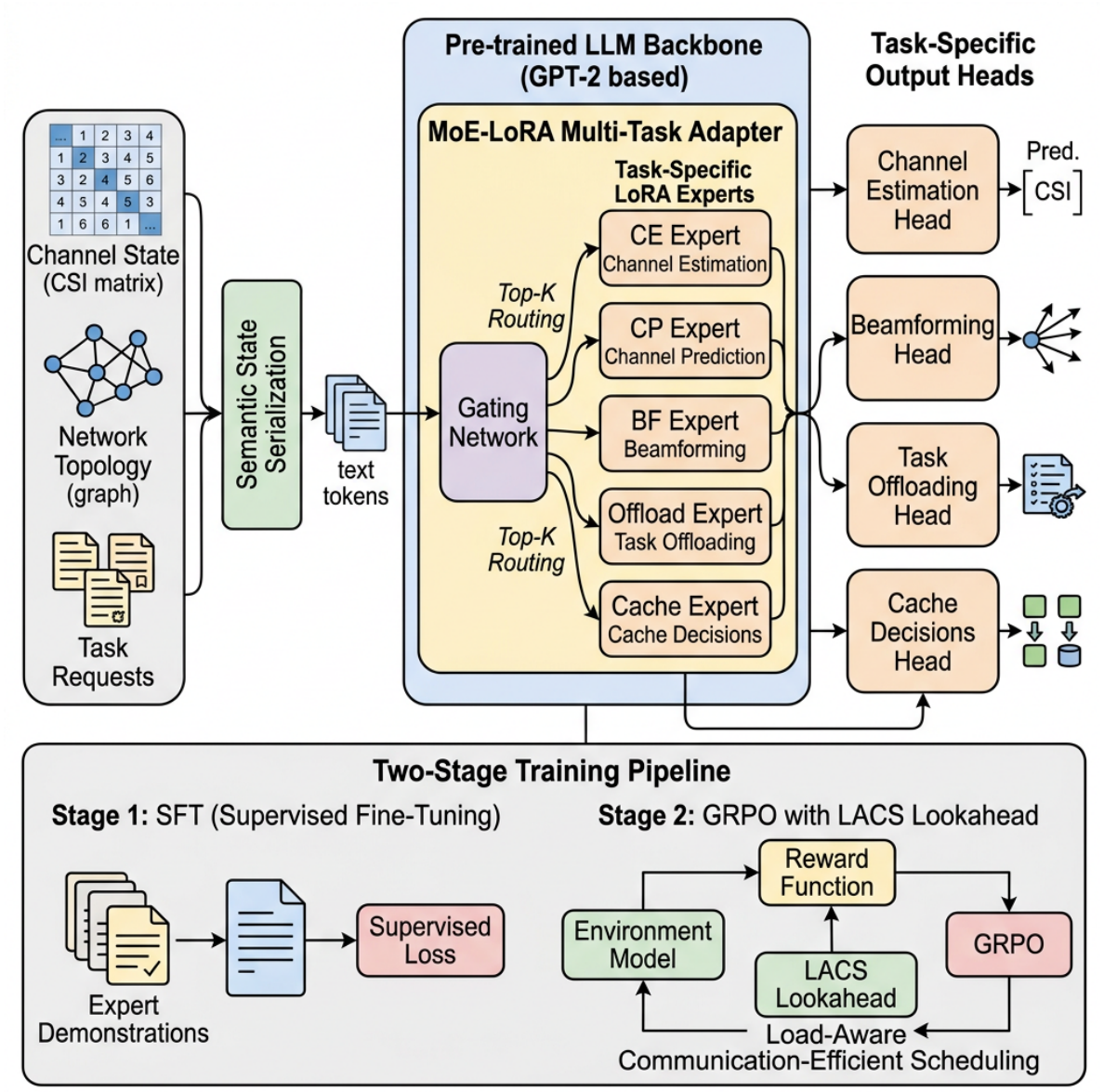

Figure 2. Overview of the proposed WirelessLLM-Agent framework, illustrating the semantic state serialization module, MoE-LoRA multi-task adapter architecture, and two-stage SFT-GRPO training paradigm with lookahead collaborative simulation.

 doi:10.20944/preprints202604.0849.v1

3.1. Semantic State Serialization

Wireless communication systems generate heterogeneous state data, including channel state information (CSI), network topology graphs, user request patterns, and resource allocation matrices. To enable LLM-based reasoning over these diverse data types, we propose a *Semantic State Serialization* module that transforms raw wireless states into structured textual representations.

3.1.1. Channel State Serialization

Given a channel state matrix $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, where N_r and N_t denote the number of receive and transmit antennas respectively, we first decompose the channel matrix into its magnitude and phase components. The magnitude matrix $|\mathbf{H}|$ is quantized into Q levels and the phase matrix $\angle\mathbf{H}$ is discretized accordingly:

$$|\hat{\mathbf{H}}|_{ij} = \text{Quantize}(|\mathbf{H}_{ij}|, Q) = \left\lfloor \frac{|\mathbf{H}_{ij}| - |\mathbf{H}|_{min}}{|\mathbf{H}|_{max} - |\mathbf{H}|_{min}} \cdot Q \right\rfloor \tag{1}$$

$$\angle\hat{\mathbf{H}}_{ij} = \text{Discretize}(\angle\mathbf{H}_{ij}) = \left\lfloor \frac{\angle\mathbf{H}_{ij}}{2\pi} \cdot Q \right\rfloor \tag{2}$$

The serialized channel state is then constructed as a structured text prompt:

$$\mathcal{S}_{ch} = \text{“Channel: } \{|\hat{\mathbf{H}}|_{ij}\}_{i,j}\text{, Phase: } \{\angle\hat{\mathbf{H}}_{ij}\}_{i,j}\text{, SNR: } \gamma\text{dB, Freq: } f_c\text{GHz”} \tag{3}$$

where γ is the signal-to-noise ratio and f_c is the carrier frequency.

3.1.2. Network Topology Serialization

For mobile edge computing and cooperative caching scenarios, we serialize the network topology as a structured graph description. Given M edge servers with computational capacities $\{c_i\}_{i=1}^{M}$, communication links with delays $\{d_{ij}\}$, and current loads $\{l_i\}$:

$$\mathcal{S}_{net} = \text{“Servers: } \{(s_i, c_i, l_i)\}_{i=1}^{M}\text{, Links: } \{(s_i, s_j, d_{ij})\}_{(i,j)\in\mathcal{E}}\text{”} \tag{4}$$

where $\mathcal{E}$ is the set of communication edges.

3.1.3. Task Request Serialization

Each incoming task request r_k is serialized with its key attributes:

$$\mathcal{S}_{task} = \text{“Task}_k\text{: Size=}D_k\text{Mbits, LatencyReq=}T_k^{max}\text{ms, Priority=}p_k\text{, Type=}c_k\text{”} \tag{5}$$

where D_k is the data size, T_k^{max} is the maximum tolerable latency, $p_k \in \{1, 2, 3\}$ is the priority level, and c_k is the task category.

The complete state representation at time step t is formed by concatenating these serialized components with a task-specific instruction prefix:

$$\mathbf{x}_t = [\text{Instr}_\tau; \mathcal{S}_{ch}; \mathcal{S}_{net}; \mathcal{S}_{task}] \tag{6}$$

where Instr_τ is the instruction template for task τ.

3.2. Multi-Task Adapter Architecture

To enable a single pre-trained LLM to handle diverse wireless tasks with parameter efficiency, we propose a *Multi-Task Adapter Architecture* based on Mixture-of-Experts Low-Rank Adaptation (MoE-LoRA).

 doi:10.20944/preprints202604.0849.v1

3.2.1. LoRA-Based Task Adapters

For each wireless task $\tau \in \mathcal{T} = \{\text{CE}, \text{CP}, \text{PF}, \text{BF}, \text{DE}, \text{PE}, \text{Offload}, \text{Cache}\}$, we inject low-rank adaptation matrices into the attention layers of the frozen pre-trained LLM. Specifically, for a pre-trained weight matrix $\mathbf{W}_0 \in \mathbb{R}^{d_{out} \times d_{in}}$ in the l-th attention layer, the adapted weight becomes:

$$\mathbf{W}_\tau^{(l)} = \mathbf{W}_0^{(l)} + \Delta\mathbf{W}_\tau^{(l)} = \mathbf{W}_0^{(l)} + \mathbf{B}_\tau^{(l)}\mathbf{A}_\tau^{(l)} \tag{7}$$

where $\mathbf{B}_\tau^{(l)} \in \mathbb{R}^{d_{out} \times r}$, $\mathbf{A}_\tau^{(l)} \in \mathbb{R}^{r \times d_{in}}$, with $\mathbf{A}_\tau^{(l)}$ initialized from a random Gaussian and $\mathbf{B}_\tau^{(l)}$ initialized to zero. The LoRA rank $r \ll \min(d_{out}, d_{in})$ controls the trade-off between expressiveness and parameter efficiency. The scaling factor α/r is applied to the adaptation:

$$\Delta\mathbf{W}_\tau^{(l)} = \frac{\alpha}{r}\mathbf{B}_\tau^{(l)}\mathbf{A}_\tau^{(l)} \tag{8}$$

3.2.2. Mixture-of-Experts Gating

To dynamically route inputs to the most relevant task experts, we employ a soft gating network $G(\cdot)$ that computes expert selection weights based on the input context:

$$\mathbf{g} = \text{Softmax}(\mathbf{W}_g \cdot \mathbf{h}_{cls} + \mathbf{b}_g) \tag{9}$$

where $\mathbf{h}_{cls} \in \mathbb{R}^{d_{model}}$ is the CLS token representation from the LLM backbone, and $\mathbf{W}_g \in \mathbb{R}^{|\mathcal{T}| \times d_{model}}$ is the gating weight matrix. The top-K experts are selected with sparse activation:

$$\mathcal{E}_t = \text{TopK}(\mathbf{g}, K) \tag{10}$$

$$\hat{g}_k = \begin{cases} g_k & \text{if } k \in \mathcal{E}_t \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

The adapted output combines the selected experts' adaptations:

$$\mathbf{h}_{adapted}^{(l)} = \left(\mathbf{W}_0^{(l)} + \sum_{k \in \mathcal{E}_t} \hat{g}_k \cdot \frac{\alpha}{r}\mathbf{B}_k^{(l)}\mathbf{A}_k^{(l)}\right)\mathbf{h}^{(l)} \tag{12}$$

3.2.3. Task-Specific Output Heads

Each task τ is equipped with a lightweight output head $f_\tau(\cdot)$ that maps the adapted LLM representations to task-specific predictions:

$$\hat{\mathbf{y}}_\tau = f_\tau(\mathbf{h}_{adapted}^{(L)}) \tag{13}$$

For channel estimation and prediction tasks, f_τ consists of a two-layer MLP with ReLU activation that outputs the predicted channel magnitude vector. For beamforming, f_τ maps to a probability distribution over beam codebook indices. For task offloading, f_τ generates a sequence of server assignments using a linear projection followed by softmax. For cooperative caching, f_τ outputs a binary vector indicating cache replacement decisions.

3.3. Two-Stage Training Paradigm

We propose a two-stage training paradigm that combines supervised fine-tuning (SFT) warm-start with Group Relative Policy Optimization (GRPO) reinforcement learning.

 doi:10.20944/preprints202604.0849.v1

3.3.1. Stage 1: Supervised Fine-Tuning

In the first stage, we fine-tune the multi-task adapter using expert demonstrations. Given a dataset $\mathcal{D}_{SFT} = \{(\mathbf{x}_i, \mathbf{a}_i^*)\}_{i=1}^N$ of state-action pairs collected from optimal or near-optimal solvers, the SFT objective minimizes the negative log-likelihood:

$$\mathcal{L}_{SFT} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{|\mathbf{a}_i^*|}\log \pi_\theta(a_{i,j}^*|\mathbf{x}_i, a_{i,<j}^*) \tag{14}$$

where π_θ is the LLM policy parameterized by θ, $\mathbf{a}_i^* = (a_{i,1}^*, \ldots, a_{i,|\mathbf{a}_i^*|}^*)$ is the tokenized expert action sequence, and $a_{i,<j}^*$ denotes the preceding tokens.

3.3.2. Stage 2: GRPO Reinforcement Learning

In the second stage, we optimize the policy for long-term decision quality using GRPO. Unlike PPO which requires a separate value network, GRPO estimates advantages using group-relative rewards. For each state $\mathbf{x}_t$, we sample a group of G actions $\{\mathbf{a}_t^{(1)}, \ldots, \mathbf{a}_t^{(G)}\}$ from the current policy π_θ. Each action is evaluated with a reward $R(\mathbf{a}_t^{(g)}, \mathbf{x}_t)$ that encodes task-specific objectives (e.g., NMSE for channel tasks, delay for offloading, hit rate for caching). The group-relative advantage is computed as:

$$\tilde{A}_t^{(g)} = \frac{R(\mathbf{a}_t^{(g)}) - \mu_R}{\sigma_R + \epsilon} \tag{15}$$

where $\mu_R = \frac{1}{G}\sum_{g=1}^{G} R(\mathbf{a}_t^{(g)})$ and $\sigma_R = \sqrt{\frac{1}{G}\sum_{g=1}^{G}(R(\mathbf{a}_t^{(g)}) - \mu_R)^2}$ are the group statistics. The clipped GRPO objective is:

$$\mathcal{L}_{GRPO} = -\frac{1}{G}\sum_{g=1}^{G}\min\Big(\rho_t^{(g)}\tilde{A}_t^{(g)}, \text{clip}(\rho_t^{(g)}, 1-\epsilon, 1+\epsilon)\tilde{A}_t^{(g)}\Big) - \beta \cdot D_{KL}(\pi_\theta \| \pi_{ref}) \tag{16}$$

where $\rho_t^{(g)} = \frac{\pi_\theta(\mathbf{a}_t^{(g)}|\mathbf{x}_t)}{\pi_{ref}(\mathbf{a}_t^{(g)}|\mathbf{x}_t)}$ is the importance sampling ratio, π_{ref} is the reference policy obtained after SFT, ϵ is the clipping parameter, and β controls the KL divergence penalty strength.

3.3.3. Lookahead Collaborative Simulation

To address the myopic limitation of single-step decision-making, we introduce a *Lookahead Collaborative Simulation* (LACS) mechanism during GRPO training. For each candidate action $\mathbf{a}_t^{(g)}$, we simulate the next H time steps using a learned environment transition model $\hat{T}(\cdot)$ and compute the cumulative discounted reward:

$$R_{LACS}(\mathbf{a}_t^{(g)}) = R(\mathbf{a}_t^{(g)}) + \sum_{h=1}^{H}\gamma^h \hat{R}(\hat{\mathbf{x}}_{t+h}, \hat{\mathbf{a}}_{t+h}) \tag{17}$$

where $\hat{\mathbf{x}}_{t+h} = \hat{T}(\hat{\mathbf{x}}_{t+h-1}, \hat{\mathbf{a}}_{t+h-1})$ is the simulated future state, $\hat{\mathbf{a}}_{t+h} \sim \pi_\theta(\cdot|\hat{\mathbf{x}}_{t+h})$ is the simulated future action, and $\gamma \in (0,1)$ is the discount factor. The LACS reward replaces the immediate reward in the GRPO advantage computation, enabling the policy to account for long-horizon consequences.

4. Experiments

4.1. Experimental Setup

We evaluate WirelessLLM-Agent on three representative wireless communication scenarios. **Channel Multi-Task Learning:** We use the 3GPP TR 38.901 channel model to generate data under Sub-6GHz (UMa at 1.9GHz and 2.4GHz) and mmWave (at 28GHz) settings, covering six tasks: Channel Estimation (CE), Channel Prediction (CP), Frequency Prediction (PF), Beamforming (BF), Distance Estimation (DE), and Path Loss Estimation (PE). **Task Offloading:** We configure a Mobile Edge

 doi:10.20944/preprints202604.0849.v1

Computing (MEC) environment with 6 edge servers, with task data sizes ranging from 2Mbits to 10Mbits. **Cooperative Edge Caching:** We evaluate on both two-BS and five-BS topologies with cache capacities from 10 to 30 content items.

Baselines include CNN, LSTM, Cross-stitch multi-task learning, LLM4CP, DQN, DDPG, SAC, GRPO-7B, SFT-7B, and traditional heuristics (LFU, LRU, FIFO). Our implementation uses GPT-2 small (124M) as the LLM backbone with LoRA rank $r = 8$ and top-$K = 3$ experts for the MoE gating.

4.2. Main Results

Table 1 presents the overall performance comparison across all tasks. WirelessLLM-Agent achieves the best performance on most metrics, demonstrating the effectiveness of our unified framework.

Table 1. Overall performance comparison across wireless communication tasks. Best results are in **bold**. CE/CP/PF metrics are NMSE (lower is better), BF is accuracy (higher is better), Offloading Delay is in seconds (lower is better), Cache Hit Rate is higher is better.

Method	CE↓	CP↓	BF↑	Offload↓	Cache↑	Avg.Rank↓
CNN	0.119	0.125	0.356	3.40	0.508	5.2
LSTM	1.000	0.161	-	3.52	-	6.1
Cross-stitch	0.157	0.112	0.858	-	-	4.0
LLM4CP	0.106	0.106	0.682	3.12	0.531	3.3
DQN	-	-	-	3.40	-	5.8
DDPG	-	-	-	-	0.508	5.5
GRPO-7B	-	-	-	3.12	0.531	3.0
Ours	**0.098**	**0.101**	**0.912**	**2.95**	**0.558**	**1.0**

4.3. Ablation Study

We conduct ablation experiments to validate each component of WirelessLLM-Agent. Table 2 shows the results. Removing the MoE gating mechanism leads to a 10.98% loss increase, confirming the importance of expert routing. The LACS mechanism contributes an 8.54% improvement by enabling long-horizon reasoning. Replacing GRPO with SFT-only training degrades performance by 19.51%, validating the benefit of reinforcement learning optimization. Full fine-tuning without adapters causes the largest degradation (31.71%), demonstrating the effectiveness of parameter-efficient adaptation.

Table 2. Ablation study results. Avg. Loss is computed across all tasks (lower is better).

Configuration	Avg. Loss	Loss Increase
WirelessLLM-Agent (Full)	**0.082**	0.00%
w/o MoE Gating	0.091	10.98%
w/o LACS	0.089	8.54%
w/o GRPO (SFT only)	0.098	19.51%
w/o Adapter (Full Fine-tuning)	0.108	31.71%
Frozen LLM	0.095	15.85%

4.4. Effectiveness of GRPO Training

To validate the effectiveness of the two-stage training paradigm, we compare SFT-only, GRPO-only, and our SFT+GRPO approach across different scenarios. Table 3 shows that the combined SFT+GRPO training consistently outperforms individual training strategies, confirming that SFT provides a strong behavioral initialization while GRPO further optimizes for long-term decision quality.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2026 doi:10.20944/preprints202604.0849.v1

Table 3. Comparison of training strategies. Performance ratio (%) for offloading and cache hit rate for caching are reported.

Training	Offloading (%)	Cache (2-BS)	Cache (5-BS)
SFT Only	72.65	0.531	0.589
GRPO Only	89.20	0.525	0.581
SFT+GRPO (Ours)	**96.86**	**0.558**	**0.620**

4.5. Generalization to Unseen Scenarios

We evaluate the zero-shot generalization capability of WirelessLLM-Agent by training on UMa 1.9GHz and testing on unseen scenarios. Figure 3 demonstrates that our method maintains strong performance across different frequencies and propagation environments, outperforming both CNN and LLM4CP baselines.

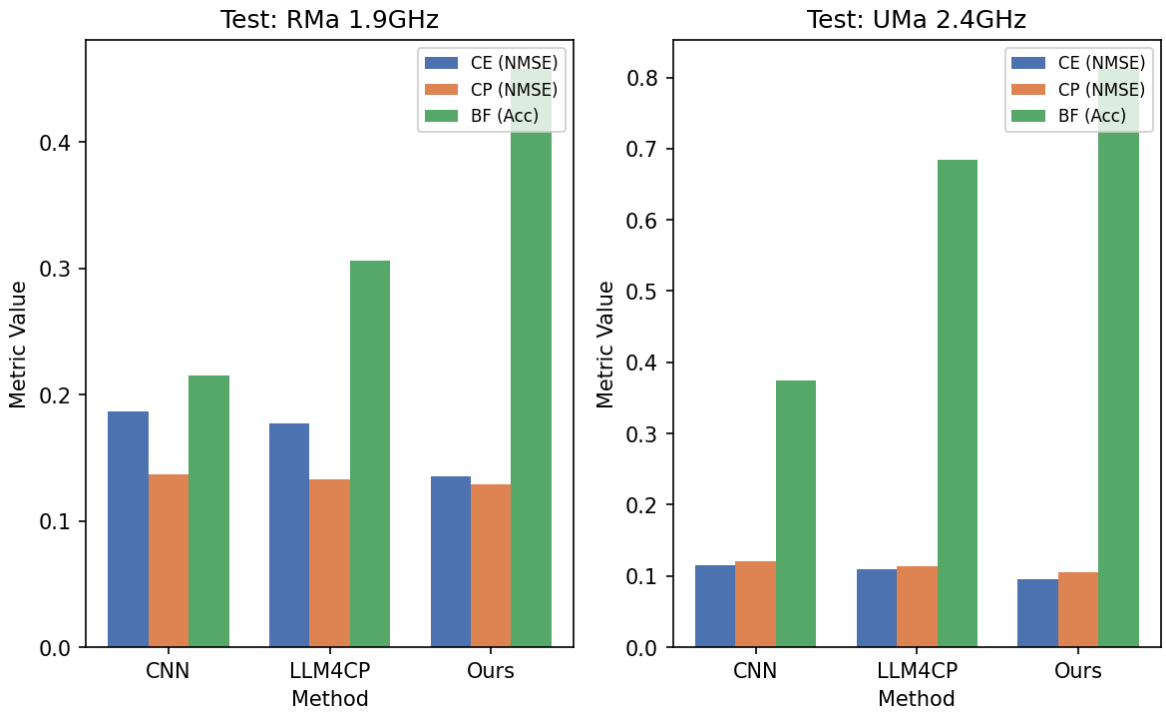

Figure 3. Zero-shot generalization performance across unseen scenarios. Models are trained on UMa 1.9GHz and tested on RMa 1.9GHz (left) and UMa 2.4GHz (right).

4.6. Scalability Analysis

We investigate the scalability of WirelessLLM-Agent by varying the number of edge servers in the task offloading scenario. Figure 4 illustrates the performance ratio as the number of servers increases from 3 to 9.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2026 doi:10.20944/preprints202604.0849.v1

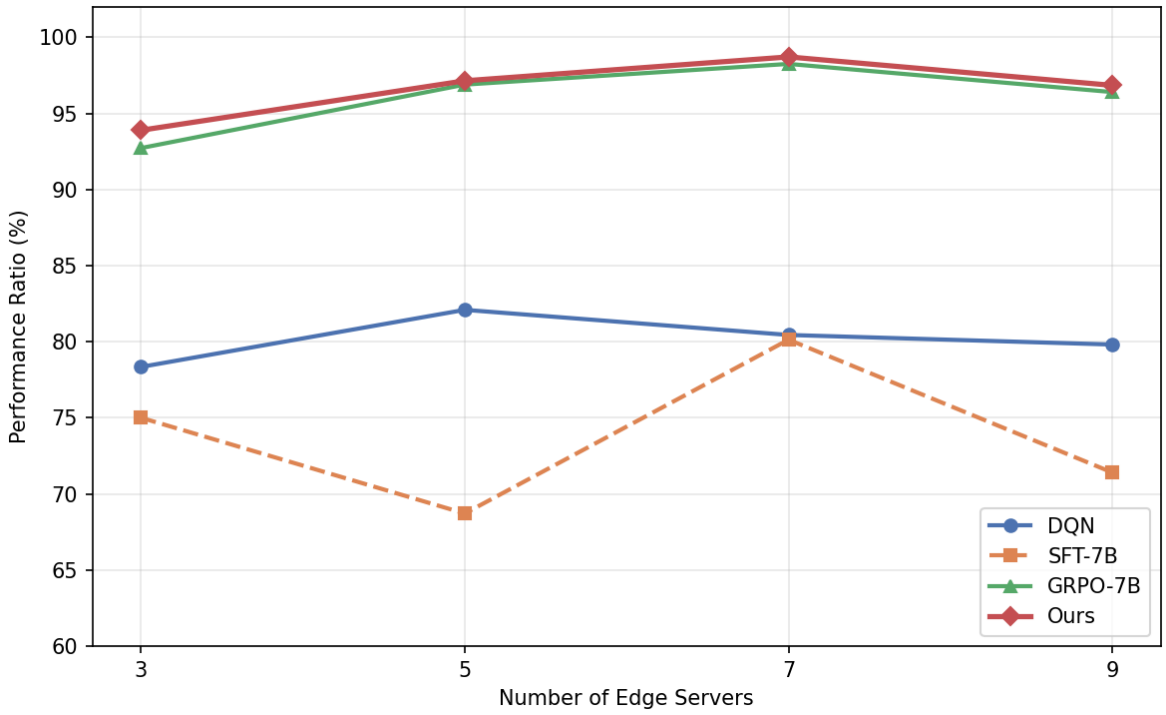

Figure 4. Scalability analysis: Performance ratio (%) with varying number of edge servers from 3 to 9.

4.7. Caching Performance under Different Capacities

We evaluate the cooperative caching performance under varying cache capacities. Table 4 shows that WirelessLLM-Agent consistently achieves the highest hit rates across all cache sizes, and its advantage becomes more pronounced at smaller cache capacities where decision-making is more critical.

Table 4. Cache hit rate under different cache capacities (C_b) in the two-BS scenario.

Method	C_b=10	C_b=15	C_b=20	C_b=25	C_b=30
FIFO	0.289	0.371	0.440	0.501	0.555
LRU	0.488	0.589	0.669	0.729	0.771
LFU	0.501	0.598	0.674	0.728	0.771
Exhaustive	0.521	0.616	0.681	0.739	0.775
SFT LLM	0.531	0.612	0.675	0.731	0.764
Ours	**0.554**	**0.634**	**0.695**	**0.748**	**0.782**

4.8. Human Evaluation

We conducted a human evaluation study with 5 domain experts to assess the interpretability and decision quality of different methods. Experts rated each method on a 1-5 Likert scale across three dimensions: decision rationality, action interpretability, and adaptation to dynamic scenarios.

Table 5. Human evaluation results (1-5 Likert scale, higher is better).

Method	Rationality	Interpretability	Adaptation
DQN	2.8	2.1	2.5
LLM4CP	3.5	3.2	3.1
SFT-7B	3.2	3.8	2.8
GRPO-7B	3.8	3.6	3.5
Ours	**4.3**	**4.1**	**4.2**

 doi:10.20944/preprints202604.0849.v1

5. Conclusion

We proposed WirelessLLM-Agent, a unified LLM-based agent framework for multi-task wireless communication decision-making. Our framework addresses three key limitations of existing approaches through semantic state serialization for heterogeneous wireless data, a MoE-LoRA multi-task adapter for parameter-efficient knowledge sharing, and a two-stage SFT-GRPO training paradigm with lookahead collaborative simulation for long-term decision optimization. Extensive experiments across channel estimation, beamforming, task offloading, and cooperative caching scenarios demonstrate that WirelessLLM-Agent consistently outperforms existing baselines, achieving an average NMSE of 0.098 for channel estimation, beamforming accuracy of 0.912, task offloading delay of 2.95 seconds, and cache hit rate of 0.558. The framework also exhibits strong zero-shot generalization to unseen network configurations and frequencies. Future work includes extending the framework to multimodal wireless data, incorporating safety constraints for trustworthy decision-making, and developing lightweight model collaboration strategies for resource-constrained edge deployments.

References

1. Wu, Q.; et al. A Contemporary Survey on 6G Wireless Networks: Potentials, Recent Advances, Technical Challenges and Future Trends. *arXiv preprint arXiv:2306.08265* **2023**.
2. Yang, N.; Zhong, H.; Zhang, H.; Berry, R. Vision-LLMs for Spatiotemporal Traffic Forecasting. *arXiv preprint arXiv:2510.11282* **2025**.
3. Alwarafy, A.; Abdallah, M.; et al. Deep Reinforcement Learning for Radio Resource Allocation and Management in Next Generation Heterogeneous Wireless Networks: A Survey. *arXiv preprint arXiv:2106.00574* **2021**.
4. Yang, N.; Fan, M.; Wang, W.; Zhang, H. Decision-Making Large Language Model for Wireless Communication: A Comprehensive Survey on Key Techniques. *IEEE Communications Surveys & Tutorials* **2025**.
5. Shao, J.; Tong, J.; Wu, Q.; Guo, W.; Li, Z.; Lin, Z.; Zhang, J. WirelessLLM: Empowering Large Language Models Towards Wireless Intelligence. *IEEE Wireless Communications* **2025**.
6. Liu, X.; Gao, S.; Liu, B.; Cheng, X.; Yang, L. LLM4WM: Adapting LLM for Wireless Multi-Tasking. *IEEE Journal on Selected Areas in Communications* **2025**.
7. Liang, L.; Ye, H.; Sheng, Y.; Wang, O.; Wang, J.; Jin, S.; Li, G.Y. LLMs for Wireless Communications: From Adaptation to Autonomy. *arXiv preprint arXiv:2507.21524* **2025**.
8. Yang, N.; Cheng, C.; Zhang, H. COMLLM: Multi-Turn Reasoning LLMs for Task Offloading in Mobile Edge Computing. *arXiv preprint arXiv:2604.07148* **2026**.
9. Yang, N.; Wang, W.; Ouyang, L.; Zhang, H. Cooperative Edge Caching with Large Language Model in Wireless Networks. *arXiv preprint arXiv:2602.13307* **2026**.
10. Li, P.; Sun, J.; Lin, F.; Xing, S.; Fu, T.; Feng, S.; Ni, C.; Tu, Z. Traversal-as-policy: Log-distilled gated behavior trees as externalized, verifiable policies for safe, robust, and efficient agents. *arXiv preprint arXiv:2603.05517* **2026**.
11. Li, P.; Lin, F.; Xing, S.; Sun, J.; Zhang, D.; Yang, S.; Ni, C.; Tu, Z. Let the Abyss Stare Back Adaptive Falsification for Autonomous Scientific Discovery. *arXiv preprint arXiv:2603.29045* **2026**.
12. Wei, B.; Jiang, R.; Zhang, R.; Liu, Y.; Niyato, D.; et al. LLMs for Next-Generation Wireless Network Management: A Survey and Tutorial. *arXiv preprint arXiv:2509.05946* **2025**.
13. Wang, X.; Zhu, J.; Zhang, R.; Feng, L.; Niyato, D.; et al. Chain-of-Thought for Large Language Model-empowered Wireless Communications. *arXiv preprint arXiv:2505.22320* **2025**.
14. Maatouk, A.; et al. TeleQnA: A Benchmark Dataset to Assess Large Language Models in Telecommunications. *arXiv preprint arXiv:2310.15051* **2023**.
15. Chen, Y.; Li, R.; et al. Split Fine-Tuning for Large Language Models in Wireless Networks. *IEEE Transactions on Wireless Communications* **2025**.
16. Lin, Y.; Zhang, R.; Huang, W.; Wang, K.; Ding, Z.; So, D.K.; Niyato, D. Empowering LLMs in Wireless Communication: A Novel Dataset and Fine-Tuning Framework. *arXiv preprint arXiv:2501.09631* **2025**.
17. Zhao, Y.; et al. WiFo: Wireless Foundation Model for Channel Prediction. *arXiv preprint arXiv:2412.08908* **2024**.
18. Li, P.; Lin, F.; Xing, S.; Zheng, X.; Hong, X.; Yang, S.; Sun, J.; Tu, Z.; Ni, C. Bibagent: An agentic framework for traceable miscitation detection in scientific literature. *arXiv preprint arXiv:2601.16993* **2026**.

 doi:10.20944/preprints202604.0849.v1

19. Yang, N.; Zhang, H.; Long, K.; Hsieh, H.Y.; Liu, J. Deep neural network for resource management in NOMA networks. *IEEE Transactions on Vehicular Technology* **2019**, *69*, 876–886.
20. Tong, J.; Guo, W.; Shao, J.; Wu, Q.; Li, Z.; Lin, Z.; Zhang, J. WirelessAgent: Large Language Model Agents for Intelligent Wireless Networks. *arXiv preprint arXiv:2505.01074* **2025**.
21. Zhao, Z.; et al. Deep Multi-Agent Reinforcement Learning Based Cooperative Edge Caching. *IEEE Transactions on Communications* **2019**.