
Hierarchical and Collaborative LLM-Based Control for Multi-UAV Motion and Communication in Integrated Terrestrial and Non-Terrestrial Networks

Zijiang Yan^{*1,2} Hao Zhou^{*3} Jianhua Pei⁴ Hina Tabassum¹

Abstract

Unmanned aerial vehicles (UAVs) have been widely adopted in various real-world applications. However, the control and optimization of multi-UAV systems remain a significant challenge, particularly in dynamic and constrained environments. This work explores the joint motion and communication control of multiple UAVs operating within integrated terrestrial and non-terrestrial networks that include high-altitude platform stations (HAPS). Specifically, we consider an aerial highway scenario in which UAVs must accelerate, decelerate, and change lanes to avoid collisions and maintain overall traffic flow. Different from existing studies, we propose a novel hierarchical and collaborative method based on large language models (LLMs). In our approach, an LLM deployed on the HAPS performs UAV access control, while another LLM onboard each UAV handles motion planning and control. This LLM-based framework leverages the rich knowledge embedded in pre-trained models to enable both high-level strategic planning and low-level tactical decisions. This knowledge-driven paradigm holds great potential for the development of next-generation 3D aerial highway systems. Experimental results demonstrate that our proposed collaborative LLM-based method achieves higher system rewards, lower operational costs, and significantly reduced UAV collision rates compared to baseline approaches.

1. Introduction

Unmanned aerial vehicles (UAVs) have seen widespread adoption across numerous application areas due to the ease of deployment, steadily reduced production costs, and the availability of line-of-sight (LOS) channels (Yu et al., 2022). Depending on the mission, a UAV may act as user equipment requiring cellular access or operate as an aerial base station (BS) to extend network coverage. Nonetheless, for UAVs flying beyond visual line-of-sight, robust command-and-control links are indispensable to guarantee both the operation reliability and security. Meanwhile, High Altitude Platform Stations (HAPS) has recently emerged as a crucial part of 6G Non-Terrestrial Networks (NTNs). HAPS, known for its extensive propagation coverage, long-distance reach, and utilization of eco-friendly energy sources like solar and wind power, is progressively becoming a fundamental component of 6G networks (Kurt et al., 2021).

Compared to the control of an individual UAV, the joint coordination of multiple UAVs inherently incurs substantially greater complexity, since it must address not only the increased state-space dimensionality but also the inter-UAV synchronization required to maintain formation and collaborative mission objectives (Cherif et al., 2021; Kotarski et al., 2020). In addition, when these UAV swarms are integrated with a HAPS within a NTN, additional challenges may arise from the heterogeneity of air-to-air and air-to-ground links. Resource allocation, network topology management, and control-plane algorithms will jointly contribute to the overall complexity across multi-UAV control, further complicating system design, real-time decision making, and end-to-end quality-of-service (QoS) guarantees.

Existing research primarily focuses on optimizing cellular link availability and QoS using reinforcement learning (RL) algorithms, without considering multi-UAV aerial traffic flow and motion dynamics of UAVs. In (Cherif, 2022), the authors proposed a RL algorithm that considers disconnectivity, handovers, and energy consumption for trajectory planning and cell association in cargo UAVs. Nesrine *et al.* leverage RL and propose an algorithm for joint trajectory planning and cell association in (Cherif et al., 2023), to minimize energy consumption and number of handoff events with QoS constraints. The authors in (Chen et al., 2020)

^{*}Equal contribution ¹Department of Electrical Engineering and Computer Science, York University, Toronto, Canada ²Bell Media Inc, Toronto, Canada ³Department of Computer Science, McGill University, Montréal, Canada ⁴School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan, China. Correspondence to: Hao Zhou <hao.zhou4@mail.mcgill.ca>.

present strategies based on deep learning to predict handovers in mmWave communications and optimize handover (HO) rates and radio link quality for known UAV trajectories. However, most existing algorithm's actions only consider the direction of motion with no speed and lane considerations (Li et al., 2020). These works have not considered the motion dynamics factors, such as acceleration, deceleration, and lane changes of multi-UAVs (Zhang et al., 2022). For instance, collision avoidance is crucial for guaranteeing the performance for multi-UAV traffic flows. Specifically, in densely populated aerial corridors, each UAV must continuously monitor the positions and velocities of its neighbors and execute timely trajectory adjustments to preserve safe separation distances. Failure to do so can lead to emergency maneuvers that disrupt coordinated flight patterns, introduce significant latency in UAV task execution.

Different from existing studies, this work considers a more challenging scenario of a 3D aerial highway, in which multiple UAVs have to jointly share the highway lanes by acceleration/deceleration, lane changes, etc. Meanwhile, from the communication perspective, the UAVs are expected to achieve higher data rates and minimal handover losses. The transportation decisions are closely coupled with the wireless communication system performance. For instance, increasing speed can increase transportation traffic flow, but it will also lead to frequent handovers, which can negatively impact the UAV communications.

To better handle such complexity, this work employs a hierarchical and collaborative control method based on large language models (LLMs). Leveraging the intrinsic strengths of LLMs, our approach harnesses their extensive pre-trained knowledge base to support both high-level planning and fine-grained decision-making. Through in-context learning, the LLM dynamically interprets real-time traffic patterns, channel state information, and mission objectives, allowing it to infer optimal acceleration, deceleration, and lane-change policies without requiring exhaustive task-specific retraining. In addition, the instruction-following capabilities further enable seamless incorporation of safety constraints, such as minimum inter-UAV spacing and maximum allowable handover rates, directly into the decision pipeline, ensuring both transportation and communication objectives are satisfied.

Existing studies have explored various application scenarios of LLM-enabled 6G networks, including grounding and alignment (Xu et al., 2024), prompt engineering (Zhou et al., 2025), network design (Qiu et al., 2024), power control (Zhou et al., 2024), etc. However, this work is different by including hierarchical and collaborative multiple LLM interactions. In particular, we consider one LLM deployed in the HAPS to manage the UAV-HAPS connections, and another LLM for UAV action control such as acceleration, de-

celeration, and lane change. By unifying transportation and communication control within an LLM-driven architecture, our method can significantly reduce the design complexity associated with traditional model-based controllers. It also accelerates adaptation to novel operational scenarios, such as varying UAV densities or spectrum availability, through prompt and instruction-guided updates. This end-to-end, knowledge-rich paradigm holds significant promise for next-generation 3D aerial highways, where the seamless interplay of mobility and connectivity is paramount. Finally, the simulations show that our proposed technique can achieve higher rewards, lower transportation and communication costs, and lower collision rates for multi-UAV scenarios.

2. System Model

Fig. 1 shows our considered aerial highway scenario, where N_{UAV} UAVs are flying along the highway lanes, and UAVs can connect HAPS or terrestrial BSs. The HAPS and terrestrial BSs form a set $\mathcal{B} = \{b_1, \dots, b_{N_R}, b_{\text{HAPS}}\}$, where each BS is denoted by i for $\forall i \in \mathcal{B}$. To simulate the UAVs' movements on a given aerial highway, we consider the continuous intelligent driver model that models acceleration as in (Treiber & Kesting, 2013).

2.1. G2A Channel Model

Firstly, the ground-to-air (G2A) channel model mainly depends on the BS's antenna gain, and the experienced path loss, and line-of-sight (LoS) probability (3GPP, 2018).

In cellular-connected aerial systems, UAVs depend on the sidelobe emissions of BS antennas to establish links with terrestrial infrastructure. Consequently, an accurate representation of BS 3D radiation characteristics is essential for reliable UAV connectivity. We adopt the 3GPP antenna pattern specification (3GPP, 2018), which closely approximates real-world radiation behavior. Conforming to this model, each BS is partitioned into three sectors, each armed with cross-polarized elements arranged in a uniform linear array (ULA). Every antenna element delivers a peak gain of $B_{\text{max}} = 8$ dBi along its principal lobe (3GPP, 2018), while sidelobe gains vary according to the UAV's azimuth and elevation angles relative to the array (Cherif, 2022):

$$B_{\text{az}}(\phi_t^{ij}) = \min \left\{ 12 \left(\phi_t^{ij} / \phi_{3\text{dB}} \right), B_m \right\}, \quad (1)$$

$$B_{\text{el}}(\zeta_t^{ij}) = \min \left\{ 12 \left(\zeta_t^{ij} / \zeta_{3\text{dB}} \right), \text{SLA} \right\}, \quad (2)$$

where ϕ_t^{ij} and ζ_t^{ij} are the azimuth and elevation angles between BS i and UAV j , and $\phi_{3\text{dB}} = \zeta_{3\text{dB}} = \frac{65\pi}{180}$ at 3dB bandwidths. B_m and SLA represent the antenna nulls thresholds, respectively. The, the gain of antenna element is

$$B(\zeta_t^{ij}, \phi_t^{ij}) = B_{\text{max}} - \min \{ -(B_{\text{az}}(\phi_t^{ij}) + B_{\text{el}}(\zeta_t^{ij})), B_m \}. \quad (3)$$

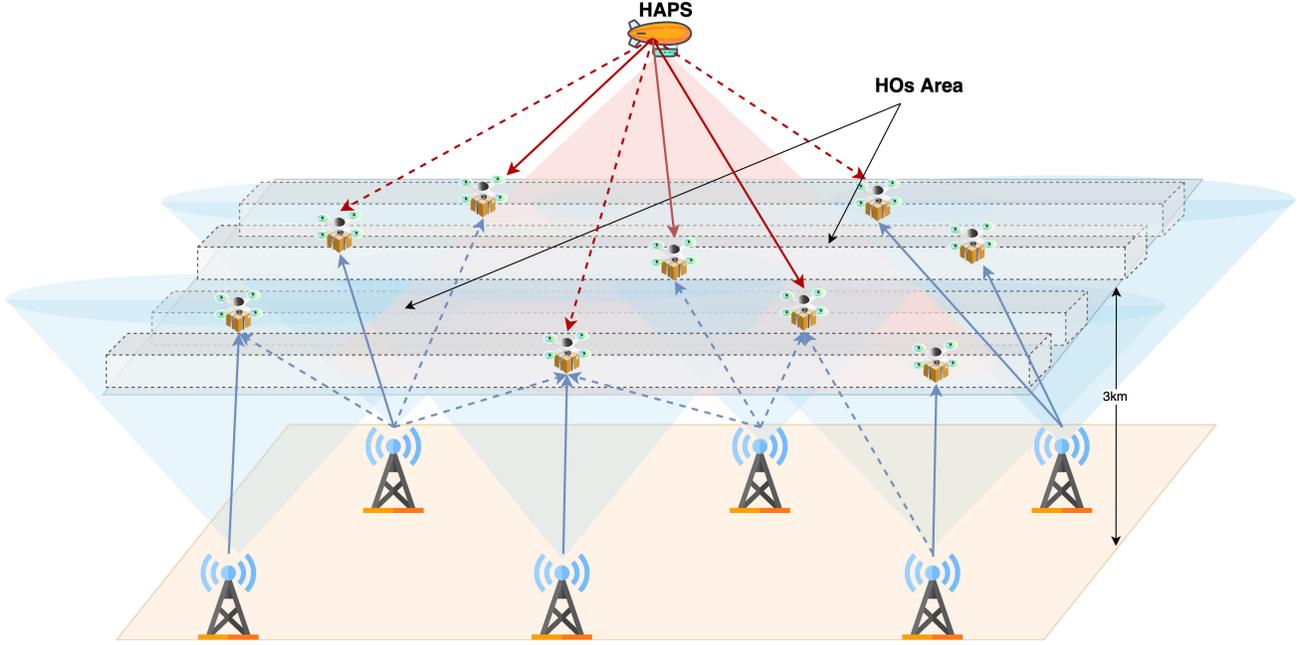


Figure 1. Illustration of the proposed aerial network model.

Assuming that one BS equipped with N antennas, and the array factor of the ULA of BS i and UAV j is

$$F(\zeta_t^{ij}) = \frac{\sin(\frac{N\pi}{2}(\sin \zeta_t^{ij} - \sin \zeta_i^d))}{\sqrt{N} \sin(\frac{\pi}{2}(\sin \zeta_t^{ij} - \sin \zeta_i^d))}, \quad (4)$$

where ζ_i^d is the down-tilt of BS i 's ULA. Finally, the array radiation pattern from BS i to UAV j is

$$B_t^{ij} = B(\zeta_t^{ij}, \phi_t^{ij}) + F(\zeta_t^{ij}). \quad (5)$$

Then, the probability of UAV j has a LoS connection with BS i depends on the altitude of the UAV and the surrounding environment, and the LoS probability is given by

$$P_{\text{LoS}}(q_t^{ij}) = \begin{cases} 1, & d_t^{ij} \leq d_1 \\ d_1/d_t^{ij} + e^{-\frac{d_t^{ij}}{p_1}} (1 - d_1/d_t^{ij}), & \text{otherwise,} \end{cases} \quad (6)$$

where $d_1 = \max\{460 \log_{10}(h_j) - 700, 18\}$ and $p_1 = 4300 \log_{10}(h_j) - 3800$. If $h_j \in [100, 300]$ m, we assume the LoS probability is 1. For the path loss between BS i and UAV j , the probabilistic mean path loss value is

$$L_t^{ij} = L_i^{\text{LoS}} P_{\text{LoS}}(r_t^{ij}) + L_i^{\text{NLoS}} P_{\text{NLoS}}(r_t^{ij}), \quad (7)$$

where L_i^{LoS} and L_i^{NLoS} indicate the LoS and NLoS path loss, respectively. Finally, the average received power from BS i is

$$P_t^{ij} = P_T + B_t^{ij} - L_t^{ij}, \quad (8)$$

where P_T is BS transmit power. The BS-UAV link is determined by the strength of the received signal P_t^{ij} , and the related signal-to-interference-plus-noise-ratio (SINR) is

$$\text{SINR}_t^{ij} = \frac{P_t^{ij}}{\sum_{i'=1, i' \neq i}^{N_R} P_t^{i'j} + \sigma_0^2}, \quad (9)$$

where σ_0^2 is the additive white Gaussian noise power.

2.2. UAV-HAPS Channel Model

We assume each UAV carries a single antenna (Ren et al., 2023), whereas the HAPS platform employs an array of multiple antennas. Then the HAPS can generate a large number of spot-beams. We fix the carrier frequency at 2 GHz for every UAV-HAPS connection and assign each link to an orthogonal channel. Given the total available bandwidth B_{max} , we seek an optimal partitioning among these channels: For UAV j , the fraction of bandwidth allocated to its link with the HAPS is denoted b^{Hj} , subject to $\sum_{j \in \mathcal{M}} b^{Hj} \leq 1$. Similarly, power distribution must respect the UAVs' transmission constraints. Let P_{max} be the maximum transmit power for any UAV; we denote the fraction of this power allocated to UAV j 's HAPS link by p^{Hj} and $0 \leq p^{Hj} \leq 1, \forall j \in \mathcal{M}$.

The UAV-HAPS links are modeled as LoS, and their large-scale attenuation follows the free-space path-loss model. Accordingly, the instantaneous channel gain can be written

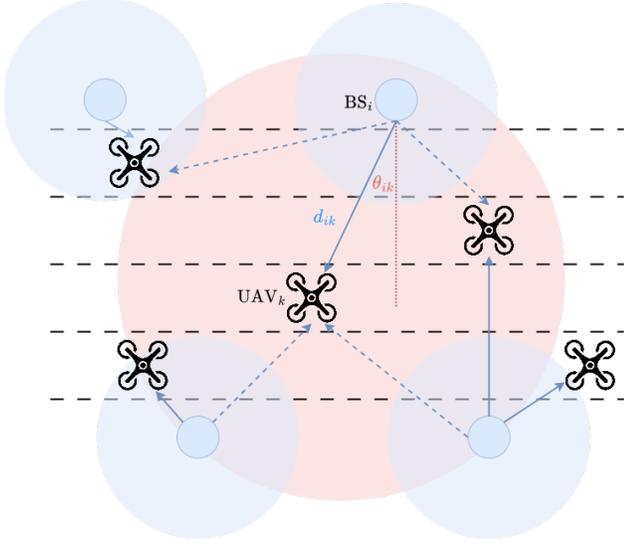


Figure 2. Illustration of the proposed aerial network model (top view). Blue circles represent BSs; Solid/dash lines represent desired/interference link.

as (Alsharoa & Alouini, 2020):

$$B_t^{Hj} = B \left(\frac{c}{4\pi d_t^{Hj} f_c} \right)^2 |h^{Hj}|^2, \quad (10)$$

where c denotes the speed of light, d_t^{Hj} is the UAV-HASP distance, and f_c is the carrier frequency. B represents the directional antenna gain, and h^{Hj} is the small-scale Rician-fading coefficient accounting for the LoS component.

Therefore, the resulting transmission rate over the UAV-HAPS link is

$$R_t^{Hj} = b^{Hj} B_{\max} \log_2 \left(1 + \frac{p^{Hj} P_{\max} B_t^{Hj}}{b^{Hj} B_{\max} N_0} \right). \quad (11)$$

2.3. Weighted Data Rate with Handovers Constraints

As UAVs navigate the aerial corridor, we distinguish two categories of handoffs: 1) Horizontal handoff: the UAV's link is transferred between two terrestrial BSs; 2) Vertical handoff: the UAV's connection shifts between terrestrial BSs and the HAPS. Excessive handoff events can degrade the UAV's achievable data rate, owing to the delays and potential failures associated with each transition. We assume each BS can access the possible UAV associations at each time instance, denoted by n_i . Then, based on the quota of each BS i , $Q_i \in [Q_R, Q_H]$, each UAV computes a *weighted data rate metric* that encourages traffic load balancing and discourages unnecessary handoffs:

$$WR_t^{ij} = \frac{R_t^{ij}}{\min(Q_i, n_i)} (1 - \mu), \quad (12)$$

where μ denotes the coefficient to discourage unnecessary handoffs. For instance, $\mu = 0$ if keeping the previous BS, and $\mu = 0.5$ if switching from terrestrial BS to HAPS and its traversal.

3. Collaborative LLM Framework for Multi-UAV Control

As depicted in Fig. 1, we employ a hierarchical two-level control architecture in which two collaborative LLM-based agents coordinate the UAV networks. The first LLM resides on the HAPS as a meta-controller overseeing global network operations. The second LLM is deployed at the network edge on each UAV, handling that individual UAV's real-time motion decisions. The HAPS-level LLM continuously monitors the overall system state, including the total available HAPS bandwidth and each UAV's instantaneous data rate, and issues high-level directives to the UAVs. In turn, the on-board UAV LLM controllers execute fine-grained flight control actions (e.g., speed and lane adjustments) in accordance with the meta-controller's directives, while also responding to local environment conditions. This two-tier LLM framework enables cooperative decision-making across meta-control and edge-control levels, addressing both communication network performance and aerial traffic safety in a unified manner.

3.1. HAPS Meta Decision Making Phase

At the meta-control layer, a LLM deployed on the HAPS functions as a global network orchestrator. Whenever the aggregate demand on the HAPS backhaul approaches its capacity, the LLM selects a subset of UAVs with the poorest HAPS link quality and instructs them to hand over to terrestrial BSs, thereby relieving congestion. Once bandwidth becomes available, the meta-controller authorizes the offloaded UAVs to re-associate with the HAPS. This policy maximizes the number of UAVs benefiting from the high-throughput HAPS link while avoiding capacity violations and unnecessary handovers.

The decision process is modelled as an MDP $(\mathcal{S}_{\text{meta}}, \mathcal{A}_{\text{meta}}, \mathcal{P}_{\text{meta}}, r_{\text{meta}})$. The state $\mathbf{s}_{\text{meta}} \in \mathcal{S}_{\text{meta}}$ captures the current HAPS load, per-UAV data rates, and the availability of ground BS coverage. An action $\mathbf{a}_{\text{meta}} \in \mathcal{A}_{\text{meta}}$ specifies the set of UAVs to offload or recall. The transition kernel $\mathcal{P}_{\text{meta}}$ reflects the evolution of HAPS load and individual link qualities following these handovers. The reward of the HAPS is defined as

$$r_{\text{meta}} = \underbrace{\eta_1 \sum_{j \in \mathcal{M}} WR_t^{ij}}_{\text{throughput and saturation}} - \eta_2 \text{Sat}_{\text{HAPS}} - \underbrace{\eta_3 \mu}_{\text{handover penalty}}, \quad (13)$$

where the first term $\sum_{j \in \mathcal{M}} WR_t^{ij}$ weights the total through-

put, the second term $\text{Sat}_{\text{HAPS}} \in \{0, 1\}$ penalizes HAPS saturation events, and the third term $\eta_3 \mu$ imposes a cost proportional to the number of enforced handovers. The LLM therefore learns a policy that maintains load balance, sustains high throughput, and minimizes disruptive switching. The overall procedure is summarized in Algorithm 1.

Algorithm 1 Meta-controller LLM for UAV Network Reconfiguration

Input : Initial meta-state \mathbf{s}_{meta} , UAV set \mathcal{U} , HAPS bandwidth limit C , ground BS availability, threshold parameters

Output : Optimal UAV association actions \mathbf{a}_{meta}
Initialize LLM parameters and policy network π_{meta} **for** each timestep t **do**

Observe current meta-state \mathbf{s}_t (UAV data rates, HAPS load, BS availability)

Compute bandwidth usage:

$$B_t \leftarrow \sum_{j \in \mathcal{J}} \text{WR}_t^{ij}$$

if $B_t > C$ **then**

Identify UAV j with lowest data rates or least critical tasks

Action $\mathbf{a}_t \leftarrow$ Offload UAV j to terrestrial BSs

else

Identify offloaded UAV eligible for reattachment to HAPS

Action $\mathbf{a}_t \leftarrow$ Reattach UAV j to HAPS

Execute action \mathbf{a}_t

UAVs update associations accordingly

Observe next meta-state \mathbf{s}_{t+1} and compute reward r_t by Equation (13)

Update policy π_{meta} using $(\mathbf{s}_t, \mathbf{a}_t, r_t, \mathbf{s}_{t+1})$

return learned optimal association policy π_{meta}^*

3.2. UAV Motion and Telecommunication Action Phase

At the **edge-control** layer, each UAV embeds an onboard LLM responsible for real-time 3D trajectory and radio access decisions. In addition to longitudinal speed adjustments and lateral lane shifts, the UAV may alter its altitude whenever authorized by the HAPS meta-controller, yielding a unified horizontal-vertical maneuver set. Consequently, each UAV addresses its own Markov decision process (MDP), whose action space is influenced by the HAPS controller, while state transitions and rewards are observed locally.

3.2.1. STATE SPACE

Let M_1 denote the number of target UAVs within the decision horizon. Each UAV j at time t is characterized by

$$\{x_t^j, y_t^j, z_t^j, v_t^j, \psi_t^j, n_t^{R,j}, n_t^{H,j}\},$$

where (x_t^j, y_t^j, z_t^j) is the 3D position, v_t^j the forward velocity, ψ_t^j the heading angle, and $n_t^{R,j}$ ($n_t^{H,j}$) the count of terrestrial BSs (HAPS channels) within service range that meet the target rate. Stacking these row vectors for all $j = 1, \dots, M_1$ forms the global state

$$\mathcal{S}_t = \begin{bmatrix} x_t^1 & y_t^1 & z_t^1 & v_t^1 & \psi_t^1 & n_t^{R,1} & n_t^{H,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_t^{M_1} & y_t^{M_1} & z_t^{M_1} & v_t^{M_1} & \psi_t^{M_1} & n_t^{R,M_1} & n_t^{H,M_1} \end{bmatrix}.$$

3.2.2. TWO-DIMENSIONAL ACTION SPACE

At each decision instant, a **joint** action $(a_{\text{tele}}, a_{\text{tran}}) \in \mathcal{A}_{\text{tele}} \times \mathcal{A}_{\text{tran}}$ is selected. The set

$$\mathcal{A}_{\text{tran}} = \{a_{\text{tran}}^1, \dots, a_{\text{tran}}^5\}$$

corresponds to **left-lane change**, **keep lane**, **right-lane change**, **accelerate**, and **decelerate**. Acceleration values follow the kinematic model in (Yan et al., 2023), so identical symbolic actions can produce different numeric rates.

The telecommunication choices

$$\mathcal{A}_{\text{tele}} = \{a_{\text{tele}}^1, a_{\text{tele}}^2, a_{\text{tele}}^3\}$$

denote BS selection strategies: a_{tele}^1 selects the BS maximizing WR_{ij} in (12); a_{tele}^2 uses the same metric with load threshold $\mu = 0$, iteratively moving to the next candidate if saturated; and a_{tele}^3 opts for the BS offering the highest instantaneous rate $R_{ij,t}$.

3.2.3. REWARD DESIGN

The transportation reward for UAV j is

$$r_t^{j,\text{tran}} = w_1 \left(\frac{v_t^j - v_{\min}}{v_{\max} - v_{\min}} \right) - w_2 \delta_c - w_3 \chi_t^j, \quad (14)$$

where $\delta_c \in \{0, 1\}$ indicates a collision and χ_t^j is the ratio of total lane changes to elapsed time. Negative returns are avoided to discourage premature termination.

For communications, the reward is

$$r_t^{j,\text{tele}} = w_4 \text{WR}_{i^*,j,t} (1 - \min(1, \xi_t^j)), \quad (15)$$

where ξ_t^j is the empirical handover probability up to time t . The weights w_1, \dots, w_4 prioritize safety (w_2) and connectivity (w_4) above secondary objectives.

Although the HAPS meta-controller operates on a slower timescale, its altitude offloading commands restrict each UAV's vertical choices and, via $n_t^{H,j}$, influence the local state and rewards. This hierarchical LLM framework jointly optimizes network utilization and collision-free traffic flow, yielding a Pareto-efficient trade-off among throughput, handover overhead, and flight safety.

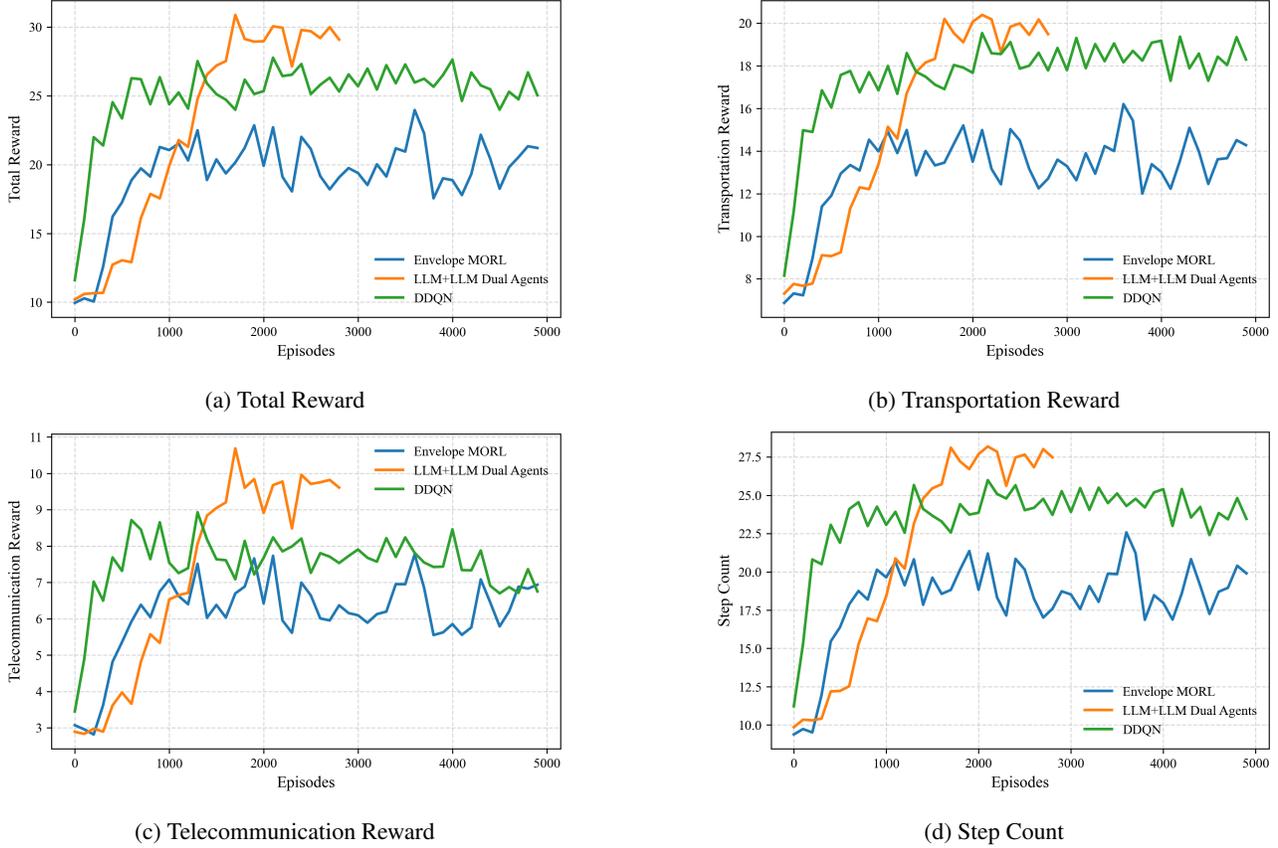


Figure 3. Performance Comparison

4. Numerical Results and Discussions

4.1. Experiment Setup

We consider a scenario with a total of five UAVs ($N_{\text{UAV}} = 5$) navigating a five-lane aerial highway, with flying speeds ranging from 5 m/s to 20 m/s. GBSs operate at a carrier frequency of 2.1 GHz and transmit at a power of $P_T = 40$ dBm. Each BS supports a maximum of three UAV users and is deployed in varying quantities (5, 10, 15, or 20) to evaluate network scalability.

Channel propagation is modeled using both LoS and NLoS path loss coefficients, set to $\eta_{\text{LoS}} = 1$ and $\eta_{\text{NLoS}} = 20$, respectively. The received signal power is constrained within the range $P_{\text{min}} = -100$ dBm to $P_{\text{max}} = -80$ dBm.

All LLM-driven simulations are conducted using the Ollama framework on a Linux host equipped with dual Intel E5-2650 v4 Broadwell CPUs (2.2 GHz) and two NVIDIA P100 GPUs, each with 16 GB of HBM2 memory. The maximum duration of each episode is capped at 30 time steps.

We benchmark our proposed framework against the following state-of-the-art deep reinforcement learning (DRL) and

LLM-enhanced hybrid baselines:

- 1) **DDQN**: A buffer-aware Double Deep Q-Network that incorporates Lyapunov optimization to select bitrate levels under buffer occupancy constraints (Yan & Tabassum, 2022).
- 2) **Envelope-MORL**: A generalized multi-objective reinforcement learning approach that leverages envelope updates to manage trade-offs among latency, throughput, and handover frequency in vehicular networks (Yan & Tabassum, 2024).
- 3) **LLM+DDQN**: A hybrid system where a LLM offers semantically-informed guidance for autonomous driving decisions, while a DDQN agent concurrently optimizes V2I communication strategies (Yan et al., 2025).

4.2. Training-Phase Convergence

Fig. 3 confirms that our *LLM-LLM Dual Agents* outperform all benchmark methods across multiple metrics.

- (a) **Total Reward**: All methods exhibit steady improvements, but the Dual-Agent variant converges approximately an order of magnitude faster (within $\sim 1.5 \times 10^3$ episodes)

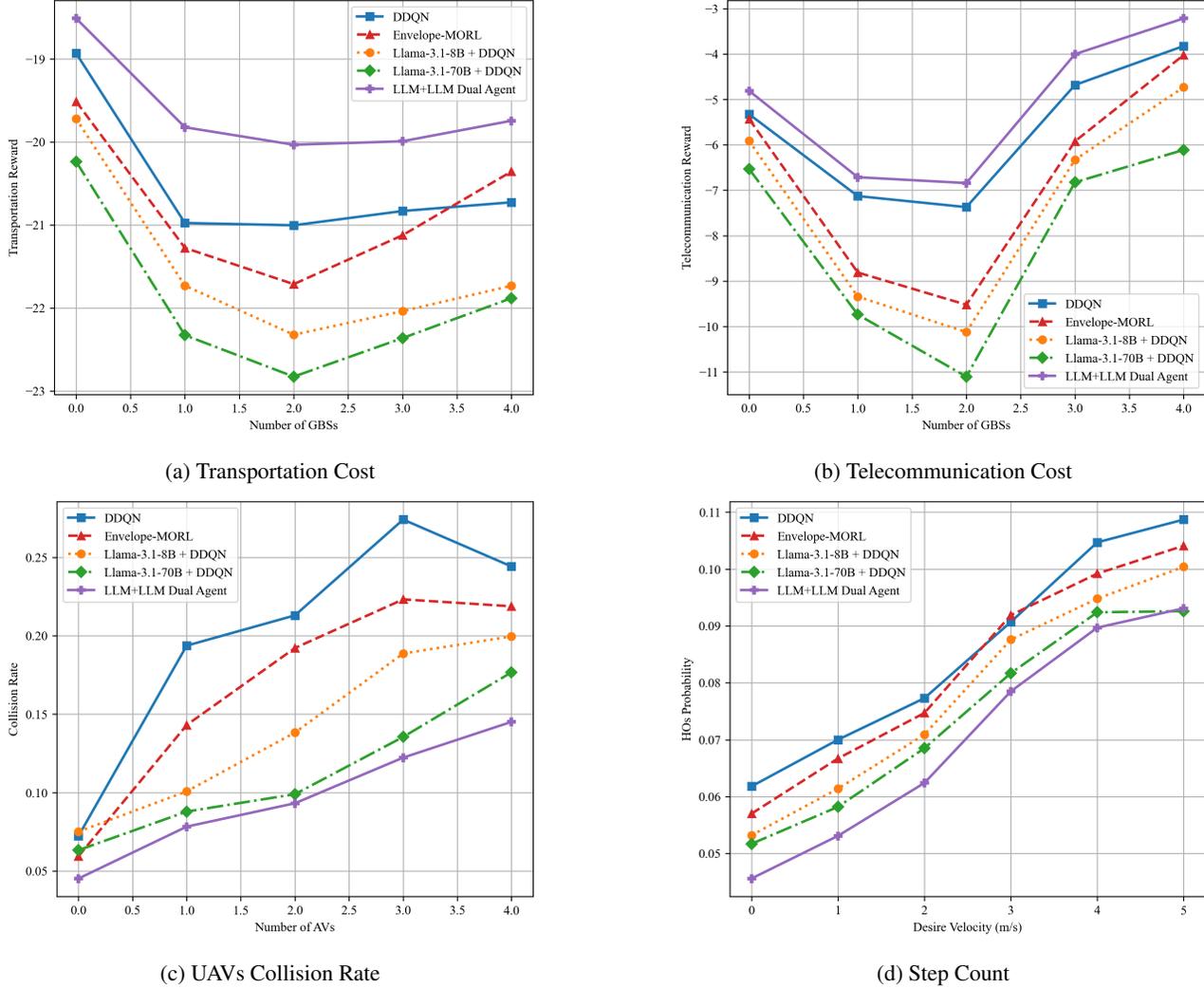


Figure 4. Evaluation Comparison

and peaks at a reward of 30. In contrast, DDQN saturates near 23, while Envelope-MHRL plateaus below 20.

(b) Transportation Reward: A similar performance ranking is observed. After convergence, the Dual-Agent model consistently achieves $\sim 25\%$ higher transportation rewards than DDQN, indicating more assertive—yet safe—speed and lane-selection strategies.

(c) Telecommunication Reward: The language-driven policies lead to enhanced V2I throughput. The Dual-Agent curve maintains a 2–3 unit advantage over DDQN throughout most of the training horizon, while Envelope-MHRL fluctuates between the two baselines.

(d) Step Count: All algorithms show an increasing trend in average episode length, suggesting smoother and collision-free trajectories. The Dual-Agent policy again reaches the

maximum episode length the fastest, highlighting its efficient exploration and stable decision-making early in training.

4.3. Evaluation under Varying Traffic Density

Fig. 4 presents cost-oriented metrics averaged over the final 200 episodes as the number of UAVs increases from 5 to 40. Five schemes are evaluated: *DDQN*, *Envelope MHRL*, *Llama 3.1 8B + DDQN*, *Llama 3.1 70B + DDQN*, and the proposed *LLM-LLM Dual Agent*.

(a) Total reward: The reward initially decreases as the density increases from 5 to 20 UAVs, due to platooning effects, and then increases again as congestion begins to dominate. Throughout the range, the LLM-LLM framework consistently yields the highest total rewards, with an average

improvement of approximately 16.3% over the strongest baseline.

(b) Telecommunication reward: A similar U-shaped pattern is observed. The LLM-LLM approach maintains the best performance, whereas DDQN suffers significantly when the number of UAVs exceeds 25, primarily due to uncoordinated and frequent handovers.

(c) Collision rate: While all methods keep the collision rate below 0.12 at low densities, it increases sharply when $M > 30$. The proposed dual LLM agent effectively limits the rate to below 0.08, surpassing DDQN by up to 21%.

(d) Average steps: The average number of steps grows nearly linearly with traffic density, reflecting increased queuing and decreased effective velocity. The dual agent system requires fewer additional steps compared to DDQN, demonstrating the robustness of language-guided exploration under congested conditions.

Overall, across both the training and evaluation phases, the proposed dual LLM agent framework achieves a clear Pareto improvement. It delivers higher cumulative rewards and reduced operational costs while maintaining lower collision rates in dense traffic. The purely language-driven control architecture consistently obtains the best episodic performance in both transportation and telecommunication domains.

5. Conclusion and Future Work

LLMs have demonstrated remarkable potential across various domains, and this work explores their capabilities in enhancing UAV mobility and communication control. We propose a novel hierarchical and collaborative decision-making framework that leverages LLMs to guide UAVs in dynamic multi-agent environments. The experimental results validate the superiority of our approach over several state-of-the-art baselines in terms of both transportation efficiency and telecommunication quality.

In future work, we aim to extend this framework by incorporating more accurate UAV localization and positioning mechanisms, especially in GPS-denied or urban canyon environments. We are also interested in integrating multi-modal sensory inputs and exploring online adaptation strategies to further improve the robustness and generalizability of LLM-driven decision-making in intelligent aerial transportation systems.

References

3GPP. Study on enhanced LTE support for aerial vehicles (release 15), TR 36.777. Jun. 2018.

Alsharoa, A. and Alouini, M.-S. Improvement of the global

connectivity using integrated satellite-airborne-terrestrial networks with resource optimization. *IEEE Transactions on Wireless Communications*, 19(8):5088–5100, 2020.

Chen, Y., Lin, X., Khan, T., and Mozaffari, M. Efficient drone mobility support using reinforcement learning. In *Proc. IEEE Wireless Commun. Network. Conf. (WCNC)*, pp. 1–6, 2020.

Cherif, N. *Cellular-connected UAV in Next-Generation Wireless Networks*. PhD thesis, Université d’Ottawa/University of Ottawa, 2022.

Cherif, N., Jaafar, W., Yanikomeroglu, H., and Yongacoglu, A. 3d aerial highway: The key enabler of the retail industry transformation. *IEEE Communications Magazine*, 59(9):65–71, 2021.

Cherif, N., Jaafar, W., Yanikomeroglu, H., and Yongacoglu, A. RL-based cargo-uav trajectory planning and cell association for minimum handoffs, disconnectivity, and energy consumption. *IEEE Transactions on Vehicular Technology*, 2023.

Kotarski, D., Piljek, P., and Kasać, J. Design considerations for autonomous cargo transportation multirotor uavs. *Self-Driving Vehicles and Enabling Technologies*, 2020.

Kurt, G. K., Khoshkholgh, M. G., Alfattani, S., Ibrahim, A., Darwish, T. S., Alam, M. S., Yanikomeroglu, H., and Yongacoglu, A. A vision and framework for the high altitude platform station (haps) networks of the future. *IEEE Communications Surveys & Tutorials*, 23(2):729–779, 2021.

Li, X., Tan, J., Liu, A., Vijayakumar, P., Kumar, N., and Alazab, M. A novel uav-enabled data collection scheme for intelligent transportation system through uav speed control. *IEEE Transactions on Intelligent Transportation Systems*, 22(4):2100–2110, 2020.

Qiu, K., Bakirtzis, S., Wassell, I., Song, H., Zhang, J., and Wang, K. Large language model-based wireless network design. *IEEE Wireless Communications Letters*, 2024.

Ren, Q., Abbasi, O., Kurt, G. K., Yanikomeroglu, H., and Chen, J. Handoff-aware distributed computing in high altitude platform station (haps)-assisted vehicular networks. *IEEE Transactions on Wireless Communications*, 2023.

Treiber, M. and Kesting, A. *Traffic Flow Dynamics: Data Models and Simulation*. Springer-Verlag Berlin Heidelberg, Sensors, 2013.

Xu, M., Niyato, D., Kang, J., Xiong, Z., Mao, S., Han, Z., Kim, D. I., and Letaief, K. B. When large language model agents meet 6g networks: Perception, grounding, and alignment. *IEEE Wireless Communications*, 2024.

- Yan, Z. and Tabassum, H. Reinforcement learning for joint v2i network selection and autonomous driving policies. In *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, pp. 1241–1246, 2022. doi: 10.1109/GLOBECOM48099.2022.10001396.
- Yan, Z. and Tabassum, H. Generalized multi-objective reinforcement learning with envelope updates in urllc-enabled vehicular networks. *arXiv preprint arXiv:2405.11331*, 2024.
- Yan, Z., Jaafar, W., Selim, B., and Tabassum, H. Multi-uav speed control with collision avoidance and handover-aware cell association: Drl with action branching. In *GLOBECOM 2023-2023 IEEE Global Communications Conference*, pp. 5067–5072. IEEE, 2023.
- Yan, Z., Zhou, H., Tabassum, H., and Liu, X. Hybrid llm-ddqn-based joint optimization of v2i communication and autonomous driving. *IEEE Wireless Communications Letters*, 14(4):1214–1218, 2025. doi: 10.1109/LWC.2025.3539638.
- Yu, J., Liu, X., Gao, Y., Zhang, C., and Zhang, W. Deep learning for channel tracking in irs-assisted uav communication systems. *IEEE Transactions on Wireless Communications*, 21(9):7711–7722, 2022.
- Zhang, R., Zong, Q., Zhang, X., Dou, L., and Tian, B. Game of drones: Multi-uav pursuit-evasion game with online motion planning by deep reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- Zhou, H., Hu, C., Yuan, D., Yuan, Y., Wu, D., Liu, X., and Zhang, C. Large language model (llm)-enabled in-context learning for wireless network optimization: A case study of power control. *arXiv preprint arXiv:2408.00214*, 2024.
- Zhou, H., Hu, C., Yuan, D., Yuan, Y., Wu, D., Chen, X., Tabassum, H., and Liu, X. Large language models for wireless networks: An overview from the prompt engineering perspective. *IEEE Wireless Communications*, 2025.

The following presents some supplementary experiments and discussions for “*Hierarchical and Collaborative LLM-Based Control for Multi-UAV Motion and Communication in Integrated Terrestrial and Non-Terrestrial Networks*”. In particular, we first present detailed example prompts that we used in the experiments to guide LLMs for UAV motion planning and communication decision-making. This document serves as supplementary material for readers to better understand how LLMs can be applied for UAV communication and motion planning dual optimization.

A. HAPS Meta Action Selection

Here is the example input on ChatGPT-4o for HAPS meta action Selection

Task Description: You are the **HAPS meta-controller** overseeing UAV associations in an integrated terrestrial+HAPS network. Your job is to keep aggregate HAPS load under its capacity while maximising overall system throughput and minimising unnecessary handovers.

Task Goal:

- Maintain total HAPS bandwidth usage $B_t \leq C$ (capacity).
- Offload the least-efficient UAVs when $B_t > C$; recall them when spare capacity exists.
- Minimise the number of enforced handovers and avoid service saturation.

Environment Features:

- **UAV ID:** Index of the UAV under control.
- **Link:** Current serving station (HAPS or TBS).
- **Data Rate [Mbps]:** Instantaneous throughput R_{ij}^t .
- **Task Priority:** Integer 1–5 (1 = mission-critical, 5 = delay-tolerant).

HAPS capacity: $C = 100$ Mbps.

Current Meta-state:

UAV ID	Link	R_{ij}^t (Mbps)	Priority
1	HAPS	28	2
2	HAPS	26	3
3	HAPS	24	1
4	HAPS	22	4
5	TBS	18	3

Total HAPS load $B_t = 28 + 26 + 24 + 22 = 100$ Mbps.

Experience Replay: The previous step was flagged as a **good** meta-decision.

Here are some examples of **good** previous experiences (high reward):

- **State:** $B_t = 98$, UAV-4 offloaded → **Action:** `Offload{4}` → **Reward:** +1.20
- **State:** $B_t = 85$, UAV-5 recalled → **Action:** `Recall{5}` → **Reward:** +0.95

Here are some examples of **poor** previous experiences (negative reward):

- **State:** $B_t = 112$, no action taken → **Action:** `Idle` → **Reward:** -1.50

Rules: Choose **one** of the following meta-actions:

- `Offload{ID1, ID2, ...}` – detach the listed UAVs from HAPS to TBSs.
- `Recall{ID1, ID2, ...}` – move the listed UAVs back to HAPS.
- `Idle` – keep all current associations.

Decision: Provide the chosen action wrapped in `<meta_action></meta_action>` and briefly justify your choice.

LLM agent respond as

HAPS load is exactly at the limit ($B_t = 100 \text{ Mbps} = C$). A single network fluctuation could push us beyond capacity, so pre-emptive relief is prudent. UAV-4 has the **lowest data rate (22 Mbps)** and **lowest priority (4)** among HAPS-attached nodes, making it the best candidate for offloading. No handover penalty will be incurred for UAV-5 because it is already on a terrestrial BS.

`<meta_action>Offload{4}</meta_action>`

B. UAV Motion and Communication Dual Actions Selection

Task Description: You are an agent that **simultaneously** decides **(i)** the ego-UAV’s driving maneuver on a multi-lane highway and **(ii)** which base-station policy ($t1/t2/t3$) to follow for V2I communication.

Task Goal:

- **Transport:** Maximise velocity while avoiding collisions and reducing unnecessary lane changes.
- **Telecom:** Maximise weighted data-rate, minimise handovers (HOs), and keep load balanced across BSs.

Environment Features (per UAV row):

- `'x'` – Horizontal offset (m).
- `'y'` – Vertical offset (lane index).
- `'vx'` – Longitudinal speed (m/s).
- `'vy'` – Lateral speed (m/s); non-zero means a lane change.

Additional telecom counters for the ego UAV: `gbs_cnt`, `haps_cnt`. Normalised bounds: $[100, 100, 20, 20]$ for x, y, vx, vy .

Observations: Current discretised state for **three** UAVs (row 1 = ego):

$$\underbrace{\begin{bmatrix} 2 & 4 & 3 & 1 \\ 3 & 1 & 0 & 3 \\ 4 & 5 & 2 & 2 \end{bmatrix}}_{\text{transport matrix}} \quad [\text{gbs_cnt} = 2, \text{haps_cnt} = 1]$$

Experience Replay: Last step was good.

Good examples (top-3):

- **State:** $[2, 4, 3, 1, 3, 1, 0, 3, 4, 5, 2, 2, 2, 1]$, **Action:** $\{\text{FASTER}, t1\}$, **Reward:** $+1.05$
- **State:** $[3, 4, 3, 1, \dots]$, **Action:** $\{\text{IDLE}, t1\}$, **Reward:** $+0.92$

Bad examples (top-1):

- **State:** [5, 6, 6, 2, ...], **Action:** {LANE_LEFT, t3}, **Reward:** -1.20

Rules:

- Transport action \in {FASTER, SLOWER, LANE_RIGHT, LANE_LEFT, IDLE}.
- Telecom action \in {t1, t2, t3}.
- Output must contain `<tran_action>...</tran_action>` and `<tele_action>...</tele_action>`.

Decision: Provide both actions and justify briefly.

The prompt is organised into six intuitive blocks from the viewpoint of a **UAV decision agent**:

1. **Task Description** declares the overarching mission: maintain high-speed, collision-free flight along the aerial highway.
2. **Task Goal** – details the three optimisation targets for each UAV: (i) maximise forward velocity, (ii) avoid collisions, and (iii) minimise unnecessary lateral manoeuvres.
3. **State Definition** – enumerates all environmental features inspected by the agent, namely the 3-D position, velocity, heading, and the counts of viable **BS/HAPS** links.
4. **Observations** – present a discretised post-processing matrix of those features for the ego UAV and surrounding aircraft.
5. **Experience Memory** – supplies two reference tables (“good” vs. “bad” cases) that offer precedents to guide the LLM when it encounters unseen states.
6. **Reply Rules** – constrain the model’s output to exactly one transport action and one telecom policy selected from the predefined action set, ensuring decisions remain focused and comparable across time-steps.

This structured natural-language template encapsulates the entire optimisation problem while remaining fully compatible with the meta-controller logic in Algorithm 1.

LLM agent respond is

The ego UAV is already centred in its lane and moving at a moderate speed ($v_x=3$). Vehicle 2 ahead is slower, but there is ample gap on its left; a controlled acceleration keeps throughput high without risking collision.

For telecom, the load is light ($gbs_cnt=2$, $haps_cnt=1$) and recent good experiences indicate t1 yields higher long-term reward when HO penalties matter.

```
<tran_action>FASTER</tran_action>
```

```
<tele_action>t1</tele_action>
```

- **Edge deployment via OLLAMA.** The LLM is hosted on nearby edge servers using the OLLAMA framework. Locating inference close to the autonomous vehicles eliminates wide-area round-trip delays, keeping response times within the strict real-time bounds required for autonomous driving.
- **Targeted experience prompts.** To curb computational load, the agent supplies the LLM with only the five **most similar good** and **bad** experiences, determined by Euclidean distance in state space, rather than the entire replay buffer. This focused context preserves decision quality while markedly accelerating inference.