drones

MDPI

Article

Frequency Point Game Environment for UAVs via Expert Knowledge and Large Language Model

Jingpu Yang [1], Hang Zhang [1], Fengxian Ji [2], Yufeng Wang [1,*], Mingjie Wang [3,*], Yizhe Luo [4] and Wenrui Ding [1]

[1] Institute of Unmanned Systems, Beihang University, Beijing 100191, China; jingpuyang@buaa.edu.cn (J.Y.); zhangh102@buaa.edu.cn (H.Z.)
[2] School of Computer and Communication Engineering, Northeastern University, Shenyang 110004, China; 202211234@stu.neu.edu.cn
[3] Academy for Network and Communications of Academy for Network and Communications, China Electronics Technology Group Corporation (CETC), Shijiazhuang 050081, China
[4] School of Electronics and Information Engineering, Zhengzhou University, Zhengzhou 450001, China
* Correspondence: wyfeng@buaa.edu.cn (Y.W.); dtwmingjie@163.com (M.W.)

Highlights

What are the main findings?

- We propose UAV-FPG, a novel reinforcement learning-based game environment that simulates dynamic signal interference and anti-interference confrontations between UAVs.
- Within UAV-FPG, the LLM-based opponent planner provides a practical, gradient-free mechanism to generate diverse, feedback-conditioned trajectories and often yields higher opponent rewards than fixed-path baselines, thereby strengthening simulator-side stress tests of ally anti-jamming policies (without implying real-world superiority).

What are the implications of the main findings?

- The UAV-FPG environment serves as a high-fidelity platform for systematically developing and validating anti-jamming decision-making strategies in complex electromagnetic scenarios.
- Our simulation results suggest that LLM-driven opponents can act as a stronger and more adaptive adversary *within UAV-FPG* , providing a practical, gradient-free way to generate diverse trajectories in high-dimensional decision spaces.

Abstract

Unmanned Aerial Vehicles (UAVs) have made significant advancements in communication stability and security through techniques such as frequency hopping, signal spreading, and adaptive interference suppression. However, challenges remain in modeling spectrum competition, integrating expert knowledge, and predicting opponent behavior. To address these issues, we propose UAV-FPG (Unmanned Aerial Vehicle–Frequency Point Game), a game-theoretic environment model that simulates the dynamic interaction between interference and anti-interference strategies of opponent and ally UAVs in communication frequency bands. The model incorporates a prior expert knowledge base to optimize frequency selection and employs large language models for episode-level opponent trajectory generation and planning within UAV-FPG, serving as an operationally more challenging simulator adversary for stress-testing anti-jamming policies under our evaluation protocol. Experimental results highlight the effectiveness of integrating the expert knowledge base and the large language model: relative to fixed-path baselines, iterative feedback-conditioned LLM planning tends to generate more adaptive trajectories and achieve higher opponent rewards in UAV-FPG. These findings are confined to the proposed simulation environment and are not intended as general claims about real-world jamming capability

Check for updates

Academic Editor: Ya Zhang

Received: 17 January 2026
Revised: 10 February 2026
Accepted: 12 February 2026
Published: 20 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

or onboard planning performance. UAV-FPG provides a robust platform for advancing anti-jamming strategies and intelligent decision-making in UAV communication systems.

Keywords: reinforcement learning; interference decision making; anti-interference strategy; expert knowledge base; LLM path planning

1. Introduction

In recent years, UAV communication technology has found widespread applications in various fields, such as military [1], agriculture [2], logistics [3], emergency rescue [4], environmental monitoring [5], urban management [6], and power inspection [7]. Using techniques such as frequency hopping [8,9], signal spreading [10,11], and adaptive interference mitigation [12,13], UAVs can maintain stable communication links even in complex electromagnetic environments, allowing them to effectively perform tasks such as reconnaissance, target tracking, and situational awareness. However, as UAVs are increasingly deployed in diverse and contested scenarios, ensuring secure and reliable communication links becomes more challenging. Consequently, research on interference and anti-interference mechanisms for UAV communication frequency bands now constitutes a pivotal area in the wireless communications domain [14]. In this context, the competition over frequency resources between opponent parties, coupled with the continuous evolution of interference techniques, has become a core issue that requires substantial attention.

Despite recent progress in UAV anti-jamming [15,16], many existing learning-based formulations focus on abstract channel/frequency selection with simplified jammer models and do not explicitly couple 3D UAV mobility, geometry-dependent propagation, and SINR/capacity-driven rewards in a step-by-step confrontation loop (e.g., CMAA [17], GPDS [18]). In addition, the FANETs framework [19] provides networking-oriented abstractions but lacks an executable spectrum confrontation game loop that models frequency-point decisions and geometry-coupled jamming effects in a unified environment. Conversely, high-fidelity UAV simulators (e.g., AirSim, RotorS, Flightmare, gym-pybullet-drones) provide realistic dynamics and RL interfaces, but do not instantiate a spectrum confrontation game with explicit frequency-point decision loops and jamming/anti-jamming interactions driven by link-quality metrics [20–23]. These gaps motivate UAV-FPG, which couples frequency-point jamming/anti-jamming with geometry-aware propagation and plug-in intelligence modules (expert knowledge base and an LLM-driven opponent planner) in one executable environment.

To address these challenges, we propose **UAV-FPG** , an executable UAV spectrum-confrontation environment that couples 3D mobility and geometry-dependent propagation with an explicit frequency-point decision loop driven by SINR/capacity-based evaluation (Figure 1). UAV-FPG also supports plug-in intelligence modules, including an expert knowledge base for guided frequency selection and an LLM-based planner for generating adaptive opponent trajectories for stress-testing anti-jamming policies. Throughout this paper, the term "LLM-driven strong adversary" is used in an *in-simulator* sense: it refers to an opponent planner that, under the same UAV-FPG rules and reward definitions, tends to produce more diverse, feedback-conditioned trajectories and often achieves higher opponent returns than our fixed-trajectory or non-LLM baselines. We do *not* interpret this as a general claim about real-world jamming capability or onboard UAV navigation performance.

Prior anti-jamming RL/MARL studies typically focus on channel/frequency selection under abstract jammer models and often treat mobility and propagation effects in a simplified manner, without an explicit 3D UAV confrontation loop. For example, collaborative

Markov-game formulations for anti-jamming mainly model multi-user channel selection and coordination in wireless networks rather than UAV kinematics and distance-dependent interference coupling [17]. GPDS formulates a multi-agent learning game for anti-jamming in MEC networks, but it is not designed as a UAV adversarial simulation platform with explicit geometry-aware jamming power and trajectory-level opponent behaviors [18]. On the other hand, widely used UAV simulators (e.g., AirSim, RotorS, Flightmare, gym-pybullet-drones) provide high-fidelity dynamics and RL interfaces, but they do not instantiate a spectrum confrontation game with explicit SINR/capacity-driven rewards and frequency-point decision loops [20–23]. In contrast, UAV-FPG bridges these two lines by coupling (i) a frequency-point jamming/anti-jamming game with explicit SINR/capacity computation, (ii) 3D geometry-dependent path loss and mobility, and (iii) two plug-in intelligence modules: an expert knowledge base that constrains/accelerates ally frequency selection, and an LLM-driven opponent planner that generates diverse, reward-aware adversarial trajectories inside the simulator for stress-testing anti-jamming policies. The novelty is *not* any single component alone, but the *joint instantiation* of these elements in one executable environment—i.e., a geometry-coupled spectrum confrontation loop where mobility, SINR/capacity evaluation, and adversarial behaviors co-evolve step-by-step—closing the gap between abstract anti-jamming games and high-fidelity UAV simulators. Table 1 provides a compact comparison between UAV-FPG and representative anti-jamming learning games and UAV simulation platforms.

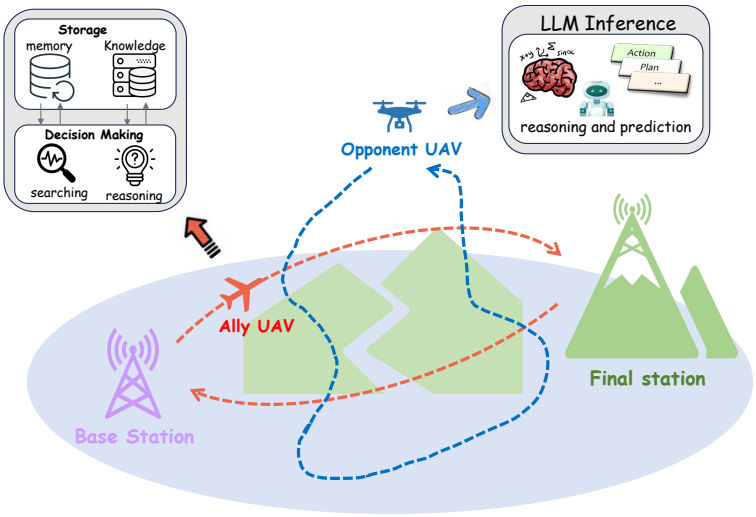


Figure 1. Illustration of the proposed UAV-FPG environment. UAV-FPG depicts the UAVs communication system interacting with a base station, ally and opponent UAVs, as well as interactions with an expert knowledge base and large language models to enhance decision-making capabilities.

The primary contributions of this paper are summarized as follows:

(1) We present **UAV-FPG** , an executable two-player Markov-game environment that *couples* (i) 3D UAV kinematics and geometry-dependent path loss with (ii) an explicit frequency-point jamming/anti-jamming loop, where link quality is evaluated via SINR/capacity and used to define step-wise rewards.

(2) We build an anti-jamming **expert knowledge base** and integrate it as a fixed guidance module that maps detected jamming types/frequency bands to safe frequency candidates, providing structured prior support for the ally UAV's frequency selection during hopping/spreading decisions.

(3) We develop an episode-level **LLM-based opponent planner** with feedback-conditioned prompting and feasibility constraints to generate adaptive adversarial

trajectories, and we benchmark it against fixed geometric trajectories and non-LLM baselines to quantify its effectiveness for stress-testing ally-side policies in UAV-FPG.

Table 1. Compact comparison with representative anti-jamming learning games and UAV simulation platforms (KB: expert knowledge base; LLM: large language model).

Work/Platform	Category	3D Mobility	Spectrum Confrontation	KB	LLM
Anti-jamming learning games					
Spectrum Waterfall DRL [24]	RL (single agent)	No	Partial	No	No
CMAA (Markov game) [17]	MARL (wireless channel selection)	No	Yes	No	No
GPDS [18]	MARL game (MEC security computing)	No	Yes	No	No
UAV simulation platforms					
AirSim [20]	High-fidelity UAV simulation	Yes	No	No	No
RotorS [21]	Gazebo-based MAV simulation	Yes	No	No	No
Flightmare [22]	Fast RL-oriented quadrotor simulator	Yes	No	No	No
UAV-FPG (ours)	UAV spectrum confrontation environment	Yes	Yes	Yes	Yes

Notes: "Spectrum confrontation" denotes an explicit frequency-point decision loop with SINR/capacity-based evaluation and jamming/anti-jamming interaction.

2. Related Work

2.1. Multi-Agent Game Theory

Multi-Agent Game Theory [25–28], as an essential branch of artificial intelligence and reinforcement learning, studies the interactions and strategy formulation of multiple agents in competitive and cooperative environments, and it achieves significant success in scenarios such as non-cooperative games, cooperative games, and zero-sum games. In recent years, the development of Deep Multi-Agent Reinforcement Learning provides powerful tools for this field [29], enabling agents to perform adaptive strategy optimization in dynamic and uncertain environments. These game-theoretic paradigms are directly relevant to spectrum signal confrontation: anti-jamming can be formulated as an adversarial Markov game, where the ally chooses anti-jamming actions (e.g., spreading/hopping) while the opponent chooses jamming and movement actions, and both sides adapt online according to link-quality feedback. In our setting, the feedback is instantiated by SINR/capacity-driven rewards and an explicit frequency-point decision loop, which makes standard opponent-learning/self-play and CTDE-style MARL techniques [30] applicable to the concrete jamming/anti-jamming process rather than abstract channel-selection only. In non-cooperative games, commonly used models include self-play and opponent learning, which allow agents to iteratively learn optimal strategies without cooperation. For example, AlphaGo and AlphaZero use self-play to iteratively optimize strategies, achieving success in complex game scenarios such as Go [31]. In cooperative games, frequently employed models include Centralized Training and Decentralized Execution (CTDE) [30] and game-theoretic strategy optimization methods, such as VDN [32] and QMIX [33]. These models enable agents to form effective cooperative strategies in multi-agent collaborative tasks, enhancing overall efficiency and success rates in scenarios like robotic cooperation [34] and multi-UAV control [35].

In zero-sum games, models often rely on opponent modeling and predictive strategies, such as Fictitious Play [36] and Counterfactual Regret Minimization (CFR) [37], allowing agents to dynamically adjust strategies in opponent environments. These models achieve successful applications across multiple fields. In autonomous driving [38], multi-agent game theory is used to address vehicle cooperation and collision avoidance, allowing the system to adapt well in dynamic traffic scenarios. In robotic collaboration [39], multi-agent

game theory enables multiple robots to allocate tasks effectively in warehouse management and rescue missions, improving efficiency and reducing conflicts. In financial markets [40], game-theoretic models simulate the behavioral interactions of different traders, helping researchers optimize trading strategies and gain a better understanding of market fluctuations. In military simulations [1], multi-agent games are employed to evaluate and optimize tactical strategies, enhancing the accuracy and effectiveness of decision-making.

In the current UAV domain, multi-agent game theory is widely applied to various aspects such as task allocation, path planning, and cooperative control. However, research related to signal games remains relatively limited [41–43]. To address this gap, we create a new environment for UAV signal games, aiming to explore signal optimization strategies under multi-agent collaboration and competition by simulating different signal game scenarios, which optimizes jamming and anti-jamming strategies between opponent parties through a non-cooperative game mechanism. This environment effectively enhances the adaptability and robustness of base stations when they face opponent signal interference, and it provides strong support for communication security in the UAV domain.

2.2. Incorporation of Expert Knowledge Bases

The integration of expert knowledge bases emerges as a crucial strategy for enhancing the performance and interpretability of intelligent systems, especially in complex tasks that require domain-specific knowledge support. Expert knowledge bases provide structured and domain-relevant information that effectively guides model behavior, improves learning efficiency, and supplies prior support in intricate decision-making environments. This integration achieves significant advancements across various fields, including natural language processing (NLP), healthcare [44–46], and robotics [47], enabling models to better align with established knowledge frameworks and expert rules.

In the realm of NLP, expert knowledge bases [48] such as ontologies and specialized terminological dictionaries play an essential role in enhancing tasks like named entity recognition, sentiment analysis, and question answering. By incorporating domain-specific prior knowledge, for instance, medical terminologies from the Unified Medical Language System (UMLS) [49,50], models attain higher accuracy and improved contextual understanding. Similarly, legal knowledge bases offer robust support for NLP models in processing legal documents, text classification, and information retrieval, which allows for more precise comprehension of specialized legal terminology and contexts [51]. In robotics, particularly within collaborative and autonomous systems, expert knowledge bases provide substantial support in areas such as path planning [52], task execution [53], and environmental interaction [54]. By integrating knowledge related to target identification, navigation constraints, and safety protocols, robotic systems operate more robustly and efficiently in dynamic environments, thereby significantly enhancing task completion efficiency and safety.

Building on these successful applications, we design a frequency selection model for UAV signal games that incorporates an opponent knowledge base of opponent drones under various interference types and frequencies. This knowledge base offers prior support for Ally UAVs' center frequency selection during frequency hopping, enabling rapid adjustments based on frequency management and interference avoidance strategies when faced with hostile signal interference. By embedding these expert strategies into the model, Ally UAVs swiftly adapt to opponent disruptions, effectively ensuring communication security and stability in complex opponent environments. This approach provides robust support for UAV communication systems under challenging conditions, fully demonstrating the significant value of prior knowledge in practical applications.

2.3. Path Planning with Large Language Models

Path planning, a core problem in unmanned aerial systems, increasingly benefits from the assistance of Large Language Models in recent years [55,56]. Leveraging their powerful capabilities for knowledge integration and reasoning, these models find widespread applications in various navigation and path planning tasks. Models such as GPT-4 [57], Vicuna [58], PaLM 2 [58,59], and LLaMA [59] are capable of interpreting environmental descriptions, states, and rewards to generate suitable path planning suggestions. In traditional path planning, the introduction of LLMs significantly enhances flexibility and adaptability, particularly in high-dimensional, dynamically changing complex environments where conventional methods struggle [60].

Recent studies, such as WayPoint, explore the potential of LLMs in generating goal-directed paths by combining natural language with visual information to create feasible navigation routes [61–63]. Similarly, multimodal models like LLaVA utilize both image and textual information to formulate more precise navigation strategies [64,65]. These studies indicate that LLMs possess considerable potential in complex decision-making tasks, progressively optimizing strategies through interaction with the environment and consideration of constraints. Moreover, certain projects combine LLMs with reinforcement learning to improve path planning performance. For instance, DeepMind's Gato model demonstrates outstanding performance in multitask navigation and control [66], while FLAN-T5 [67,68], through instruction fine-tuning, shows consistent performance in cross-task path planning.

We clarify that the LLM in this work is *not* used for low-level UAV control. Instead, it serves as an episode-level, high-level opponent planner that proposes motion directions inside an offline simulator, where feasibility is enforced by the environment constraints. Under our setting, the opponent objective is reward-driven and coupled with the spectrum confrontation loop, and we do not assume access to a fully specified, differentiable cost model. Conventional planners such as A*/RRT* or MPC typically require an explicit geometric goal and/or an accurate dynamics/cost formulation and may produce repetitive behaviors under similar initial conditions, which can limit trajectory diversity for stress-testing anti-jamming policies. In contrast, an LLM can condition on rich contextual summaries (past positions and rewards) to generate diverse, reward-aware adversarial motion suggestions without additional training. We emphasize that the goal here is to create a stronger and more diverse adversary *within UAV-FPG*, rather than to claim superiority for real-time onboard planning.

In our work, we employ LLMs for UAV path planning, aiming to enhance strategic behavior in multi-agent games. Specifically, we input the positions and rewards of opponent UAVs from the previous round into an LLM to infer and plan the movement directions of opponent UAVs in the next round, yielding a higher-reward adversary in UAV-FPG under our evaluation protocol. This approach effectively leverages LLMs for flight path planning, enhancing strategic flexibility in our simulated multi-agent setting. In particular, the generated trajectories provide evidence that an LLM can produce reward-aware, non-repetitive motion suggestions in UAV-FPG; however, we treat these findings as simulation-based observations and leave real-world generalization and onboard execution constraints to future work.

3. Environment Model

In the UAV-FPG communication game environment, we focus on signal attenuation between the base station and Ally UAV, as well as interference caused by opponent UAVs. We simulate interference power attenuation on UAVs and the intensities of various interference types. Ally UAV operates within a designated frequency band, while opponent

UAVs detect our communication center frequency and employ power suppression interference to degrade the signal-to-noise ratio (SNR), thereby reducing communication quality. Opponent UAVs adopt diverse interference strategies, including single-tone interference, narrowband targeting, wideband jamming, and comb-spectrum interference. Utilizing reinforcement learning agents, opponent UAVs select interference actions, adjust their center frequencies, and optimize their strategies based on reward signals.

We model UAV-FPG as a two-player Markov game $\mathcal{G} = \langle \mathcal{I}, \mathcal{S}, \{\mathcal{A}_i\}_{i\in\mathcal{I}}, \mathcal{T}, \{r_i\}_{i\in\mathcal{I}}, \gamma \rangle$, where $\mathcal{I} = \{\text{ally}, \text{opponent}\}$. At time step t, the environment state is $s_t \in \mathcal{S}$ and the agents simultaneously choose actions $a_t^{\text{ally}} \in \mathcal{A}_{\text{ally}}$ and $a_t^{\text{opponent}} \in \mathcal{A}_{\text{opponent}}$. The next state is sampled by $s_{t+1} \sim \mathcal{T}(\cdot \mid s_t, a_t^{\text{ally}}, a_t^{\text{opponent}})$, and rewards $(r_t^{\text{ally}}, r_t^{\text{opponent}})$ are returned (Equations (5) and (6)). Each agent maximizes its discounted return $J_i = \mathbb{E}\left[\sum_{t=0}^{T-1} \gamma^t r_t^i\right]$ with $\gamma \in (0, 1)$.

In our implementation, the state s_t is encoded as a 15-dimensional vector that jointly characterizes the UAV geometry and the spectrum environment, including relative spatial relationships among UAVs, the currently selected frequency band, observed jamming signals, and link-quality indicators such as the instantaneous SINR and its temporal variation. The action spaces model the ally's anti-jamming strategies, including spectrum spreading and frequency hopping, as well as the opponent's jamming behaviors and motion decisions. The transition function $\mathcal{T}$ is governed by UAV kinematics and the wireless link model in Equations (1)–(3), which jointly determine the evolution of SINR, achievable capacity, and the resulting rewards. Optionally, the expert knowledge base and the episode-level LLM planner can be interpreted as fixed policy components that restrict frequency choices and generate opponent motion primitives, while preserving the Markov-game formulation described above.

Our communication system consists of a base station with constant power and a fixed position, while Ally UAV follows a predetermined path. To mitigate interference, the UAV employs spread spectrum and frequency hopping techniques. By leveraging an expert knowledge base, the UAV selects interference-free frequencies during hopping and despreading processes, thereby enhancing its anti-interference capabilities. To further increase the complexity of the scenario, we integrate an LLM to infer and plan opponent UAV interference trajectories, thereby maximizing interference against Ally UAV. As illustrated in Figure 2, this setup ensures realistic and challenging opponent conditions, facilitating the evaluation and improvement of Ally UAV's resilience and communication performance under interference.

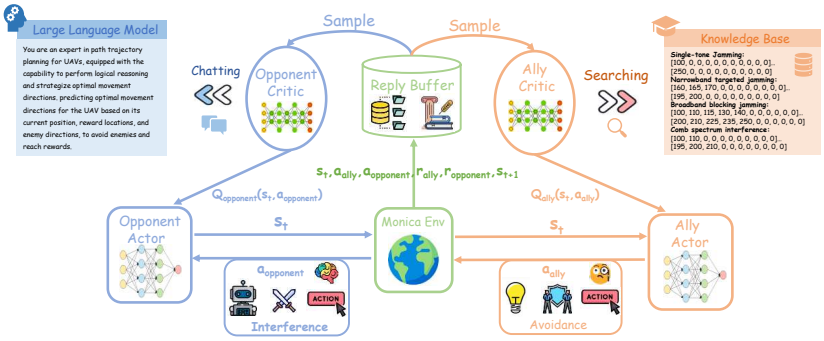


Figure 2. Framework of the multi-agent reinforcement learning-based UAV-FPG environment. UAV-FPG adopts an actor–critic architecture and integrates an expert knowledge base for anti-jamming frequency selection and an LLM module for opponent path planning. At each timestep, the state s_t is observed and the ally selects an avoidance action a_{ally}, while the opponent selects an interference action $a_{opponent}$. The transition $(s_t, a_{ally}, a_{opponent}, r_{ally}, r_{opponent}, s_{t+1})$ is stored in a replay buffer for policy optimization across episodes.

4. Methods

In wireless communication systems, suppressive power interference is a persistent challenge, with the power of interfering signals closely related to the distance between the opponent and ally UAV. To evaluate whether the opponent UAV's signal interferes with the ally's communication, we analyze whether the interference frequency range overlaps with the ally's center communication frequency. This chapter details the methods used to construct a signal interference and anti-interference game between the ally and opponent UAVs. In Section 4.1, we introduce a fundamental frequency point game model based on modern mobile communication principles. Section 4.2 explains how prior expert knowledge on the types of interference and interference frequencies is incorporated into the model for selecting the ally UAV's communication frequency. In Section 4.3, we explore and validate the potential of large language models for path inference and planning, as well as introduce prompt engineering techniques for effective path planning, thus creating the "strong opponent" effect.

4.1. Frequency Point Game in Wireless Communications

In the UAV-FPG environment, we construct a frequency game scenario involving ally and opponent UAVs to simulate the frequency interference and avoidance strategies between our side and the adversary in wireless communication. This environment consists of a base station and two UAVs (ally and opponent). The base station selects 15 frequency points within the range of 150 MHz to 250 MHz. We discretize the 150–250 MHz band into a finite set of candidate channels for a tractable Markov-game formulation. We use 15 frequency points as a default setting that balances spectral diversity and learning complexity. The number of frequency points N_f is configurable and can be adjusted for different experimental settings. The ally UAV, acting as the interference-avoidance agent, primarily employs spread spectrum and frequency hopping techniques to mitigate interference from the opponent UAV and maintain communication with the base station. Meanwhile, the opponent UAV is responsible for implementing suppressive signal interference, which includes techniques such as single-tone interference, narrowband targeted interference, broadband jamming, and comb spectrum interference, in an attempt to obstruct the communication between the ally UAV and the base station. Additionally, we employ the free-space path loss model to account for the energy attenuation of electromagnetic waves as they propagate through the air (see Equation (1)).

$$L_{\text{loss}} = 32.44 + 20\log d(\text{km}) + 20\log f(\text{MHz}), \tag{1}$$

where d denote the distance between the base station and ally UAV, and let f represent the center frequency of ally UAV's communication. Considering the attenuation from the base station and the opponent UAV, the instantaneous SINR at the ally UAV under opponent interference is given by:

$$\text{SINR}(\text{dB}) = 10\log_{10}\left(\frac{P_s}{P_n + P_i}\right), \tag{2}$$

where P_s, P_i, and P_n denote the received signal, interference, and noise powers (mW) over bandwidth B, respectively. They are converted from dBm/dB quantities as $P_s = 10^{\frac{P_{\text{base}} - L_{\text{base}} - 30}{10}}$, $P_i = 10^{\frac{P_{\text{opponent}} - L_{\text{opponent}} + 10\log_{10}(k) - 30}{10}}$, and $P_n = 10^{\frac{N_0 + 10\log_{10}(B) - 30}{10}}$. Here P_{base} and P_{opponent} are the transmit (jamming) powers in dBm, L_{base} and L_{opponent} are the corresponding path losses in dB, k is the jamming-intensity coefficient (Table 2), and N_0 is the noise power spectral density in dBm/Hz.

$$C = B\log_2(1 + \text{SINR}) = B\log_2\left(1 + 10^{\text{SINR}(\text{dB})/10}\right), \tag{3}$$

where C (bit/s) is the channel capacity and B (Hz) is the channel bandwidth. For brevity, we use the term "SNR" to refer to SINR(dB) throughout this section. In the case of spread spectrum, the increased bandwidth reduces the SNR of the ally UAV, thereby making it challenging for the opponent UAV to effectively observe the center frequency of the ally UAV. To mitigate this, we designed a spreading cost function, as presented in Equation (4), encouraging the ally UAV to de-spread early when necessary, while discouraging frequency hopping due to its high cost H.

$$D(t) = \begin{cases} 0.5t, & \text{if } t \leq 10, \\ 5 + 1.0(t-10), & \text{if } 10 < t \leq 20, \\ 15 + 1.2(t-20), & \text{if } 20 < t \leq 40, \\ 39 + 1.5(t-40), & \text{if } t > 40. \end{cases} \tag{4}$$

The opponent UAV attempts to detect the ally's center frequency every n seconds but fails if the ally's SNR falls below a threshold M. Detection resumes after de-spreading and SNR recovery. The ally UAV follows a fixed path, while the opponent UAV can either patrol along a preset route or plan its movements based on the previous round's position and reward using a large model. Considering factors such as SNR variations, the distance between UAVs, and action costs, we derived the reward functions for both ally and opponent UAVs (see Equations (5) and (6)).

$$r_{\text{ally}} = \text{SNR} - M - D(t) - H \cdot \begin{cases} 1, & \text{if } a_{\text{hopping}} > 0.5 \\ 0, & \text{if } a_{\text{hopping}} \leq 0.5 \end{cases}, \tag{5}$$

$$r_{\text{opponent}} = E \cdot \theta(30 - \|\mathbf{p}_{\text{opponent}} - \mathbf{p}_{\text{ally}}\|) + \alpha \cdot \Delta\text{SNR}, \tag{6}$$

where $a_{\text{hopping}} \in [0, 1]$ is the continuous hopping trigger output by the ally actor network (a hop is executed when $a_{\text{hopping}} > 0.5$), M is the minimum SNR threshold for reliable sensing/detection in our simulator, t is the spreading duration (in time steps) used in $D(t)$, H is the hopping-cost coefficient, and n is the opponent sensing interval (s). In Equation (6), $\mathbf{p}_{\text{opponent}}$ and $\mathbf{p}_{\text{ally}}$ are the 3D position vectors of the opponent and ally UAVs, respectively, and $\|\cdot\|$ denotes the Euclidean distance. The function $\theta(\cdot)$ is the unit step function; the constant 30 denotes the proximity threshold (meters), and E is the corresponding proximity-reward coefficient. The 30 m proximity threshold is a scenario-driven engineering setting that approximates the effective close-in engagement range under our default transmit-power and path-loss assumptions. It is used as reward shaping to avoid granting proximity reward at long distances and to encourage realistic pursuit-and-jam behavior. ΔSNR denotes the SNR change between consecutive time steps/rounds in our simulator, and α weights the SNR-degradation term. These terms jointly scalarize the objectives of both agents: the ally is encouraged to keep link quality above the threshold while avoiding excessive spreading/hopping overhead, whereas the opponent is rewarded for maintaining proximity and degrading the ally's SNR.

Thus, the entire game process is characterized by the opponent UAV continuously approaching the ally UAV to impose interference, while the ally UAV employs spread spectrum techniques and reselects communication frequencies to evade the interference. Once the opponent UAV detects the ally's central frequency again, it resumes its approach and interference, thereby establishing a persistent frequency-based game. The detailed pseudo-code for this environment is presented in Algorithm 1. The computational cost is dominated by the neural network forward/backward updates in MADDPG, which scales approximately as $O(N \cdot P)$, where N is the number of training steps and P is the

parameter size. In addition, the total runtime includes an environment-step cost $O(N \cdot C_{\text{env}})$ (computing path loss/SINR/rewards and state updates) and an episode-level external LLM-planning cost $O(E \cdot C_{\text{LLM}})$, where E is the number of episodes and C_{LLM} denotes the black-box API call cost (latency/token-dependent) per episode.

Table 2. The interference intensity coefficient k corresponding to different jamming techniques employed by the opponent [69].

Jamming Type	k
Single-tone Jamming	1.5
Narrowband Targeted Jamming	1.2
Broadband Blocking Jamming	0.4
Comb Spectrum Jamming	0.8

Algorithm 1 UAV-FPG

Require: Initialize actor networks $\pi_{\phi_{ally}}$, $\pi_{\phi_{opponent}}$ and critic networks $Q_{\theta_{ally}}$, $Q_{\theta_{opponent}}$ with random parameters θ_{ally}, $\theta_{opponent}$, ϕ_{ally}, $\phi_{opponent}$;
Require: prompt P, opponent locations $LLM_{opponent}\{\}$;
Require: Initialize environment state s, replay buffer $\mathcal{B} \leftarrow \emptyset$;
for $t = 1$ to N **do**
    Ally selects action a_{ally} using $\pi_{\phi_{ally}}$;
    $a_{ally} = \pi_{\phi_{ally}}(s_t) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$;
    Obtain the ally center frequency f_{ally} using an expert knowledge base.
    Opponent selects action a_{opponent} using $\pi_{\phi_{opponent}}$;
    $a_{opponent} = \pi_{\phi_{opponent}}(s_t) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$;
    Obtain new state s' and reward $r_{opponent}$, r_{ally};
Environment Step:
    Store transition $(s, a_{\text{ally}}, a_{\text{opponent}}, r_{\text{ally}}, r_{\text{opponent}}, s')$ in $\mathcal{B}$;
    Store the set of opponent positions in $LLM_{opponent}\{\}$;
    if episode ends **then**
        Reset the environment and state s;
        Update trajectory using LLM;
    end if
Sample mini-batch: $(s, a_{\text{ally}}, a_{\text{opponent}}, r_{\text{ally}}, r_{\text{opponent}}, s') \sim \mathcal{B}$;
    Critic Update:
        Compute Q targets and update Q_{ally}, Q_{opponent};
    Actor Update:
        Update π_{ally}, π_{opponent} by maximizing Q values;
    Visualize ally and opponent UAV trajectories;
end for

4.2. Optimized Frequency Selection with Expert Knowledge

UAVs employ frequency hopping and spread spectrum techniques to counter interference with limited adaptability. However, they struggle to optimize frequency selection under complex interference. This research integrates expert knowledge (e.g., interference strategies, central frequencies) and empirical criteria into datasets to improve decision-making, enhancing communication stability and anti-jamming capabilities. To address this issue, the study analyzes four major types of interference and their corresponding avoidance methods. These interference types include Single-tone Jamming, Narrowband Targeted Jamming, Broadband Blocking Jamming, and Comb Spectrum Jamming, each affecting different frequency ranges and requiring distinct frequency selection strategies to mitigate their impact. Through the examination of these interference types, effective anti-jamming solutions for UAVs in complex electromagnetic environments are developed.

Single-tone Jamming is a type of interference that focuses on specific frequency points by imposing a strong jamming signal at a particular frequency, thereby disrupting communication. The jamming model can be expressed as:

$$J_{\text{single-tone}}(f) = A \cdot \delta(f - f_j). \tag{7}$$

where A denotes the jamming amplitude, $\delta(\cdot)$ is the Dirac delta function, and f_j is the jamming center frequency. To address this issue, we adopt a fast frequency-hopping strategy to avoid interference at the center frequency. Real-time spectrum monitoring is employed to identify the interference frequency f_{opponent}, enabling dynamic selection of frequencies that are free from interference.

Narrowband Targeted Jamming is a form of interference that operates within a narrow frequency band, specifically designed to target the primary frequency range of a desired signal in order to maximize its disruptive effect. Compared to single-tone jamming, this approach covers a broader range of frequencies while remaining concentrated within a specific band. The mathematical representation of narrowband targeted jamming can be expressed as:

$$J_{Narrow}(f) = \begin{cases} A, & \text{if } f \in [f_c - B/2, f_c + B/2] \\ 0, & \text{otherwise} \end{cases}. \tag{8}$$

where A is the interference power level, f_c is the targeted center frequency, and B_{jam} denotes the jamming bandwidth. To address this, we opt to avoid the interference-affected range by employing a pseudo-random frequency hopping technique, wherein the transmission frequency rapidly switches across multiple channels. This approach effectively evades the interference band, making it difficult for narrowband jamming to continuously disrupt signal transmission.

Broadband Blocking Jamming is a type of jamming that spans a wide frequency band, aiming to degrade the SNR of the communication system by injecting noise or interference signals across a broad frequency range. This, in turn, disrupts the receiver's ability to decode the target signal. Unlike narrowband interference, broadband blocking interference is not confined to specific frequency bands but instead affects the entire or a significant portion of the communication spectrum, making it highly disruptive. The expression for this type of interference is given as:

$$J_{\text{Broadband Blocking}}(f) = \begin{cases} P_J, & \text{if } f \in [f_{\min}, f_{\max}] \\ 0, & \text{otherwise} \end{cases}. \tag{9}$$

where P_J denotes the jamming power level (or power spectral density in the blocked band), and $[f_{\min}, f_{\max}]$ is the jammed frequency range. To address this, we opt to implement frequency hopping to channels outside the interference range, thereby maximizing the distance from the interference band. Alternatively, the original signal's energy can be spread across a wider frequency band, effectively dispersing the impact of broadband interference by averaging its effects over a broader spectrum.

Comb Spectrum Jamming is a specialized form of jamming characterized by a spectrum with a comb-like distribution, where single-tone interference signals are injected at specific intervals of frequency points while maintaining no interference at other frequencies. This type of interference selectively occupies certain frequency points within the target communication band, causing significant disruption to frequency-selective signals such as those in OFDM systems. Its mathematical representation can be expressed as:

$$J_{\text{Comb}}(f) = \begin{cases} P_J, & \text{if } f = f_0 + n\Delta f,\ n \in \mathbb{Z} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where P_J denotes the power of each comb tone, f_0 is the starting comb frequency, Δf is the tone spacing, n is the tone index, and $\mathbb{Z}$ is the set of integers. To address this issue, we employ pseudo-random sequences to control frequency hopping, dynamically shifting the operating frequency either to the sensing band or away from the interference region. This approach ensures that the distribution of communication signal frequencies is effectively separated from the interference frequencies.

We have constructed an anti-jamming expert knowledge base incorporating identified interference strategies, center frequencies, and corresponding counter-strategies. By training with this knowledge base, Ally UAV learn the appropriate anti-jamming techniques, thereby enabling them to effectively counter adversarial interference. In practice, the expert knowledge base is a fixed, engineer-designed dataset that maps the detected jamming type and interference frequency (or band) to a recommended avoidance strategy and a set of interference-free candidate (safe) frequencies. The knowledge base does not evolve over time: before gameplay, we fit a linear layer on this dataset to obtain frequency-selection weights, and then use these learned parameters as an additional weighted guidance term during the game.

Concretely, we represent the expert knowledge base as a table of tuples $(\tau, f_{\text{opponent}}) \rightarrow (\mathcal{F}_{\text{safe}}, \text{strategy})$, where τ is the detected jamming type and f_{opponent} is the estimated interference center frequency (or band). We train a lightweight MLP policy g_ψ offline on this table to output a categorical distribution over the 15 candidate frequency points (i.e., frequency-selection weights). During gameplay, the query is dynamic: once interference is detected, if the ally policy triggers hopping ($a_{\text{hopping}} > 0.5$), we select $f_{\text{ally}} = \arg\max g_\psi(\tau, f_{\text{opponent}})$ (restricted to $\mathcal{F}_{\text{safe}}$); otherwise the ally keeps the current frequency. The KB module is fixed during RL training and only provides frequency candidates/guidance, while the RL policy learns when to spread/hop under jamming. Both the KB table and the trained predictor g_ψ are kept fixed throughout RL training and evaluation; the RL policy learns *when* to hop/spread, whereas g_ψ determines *which* safe frequency to select (restricted to $\mathcal{F}_{\text{safe}}$).

4.3. Toward a Stronger Simulator Adversary: Path Inference and Planning with LLMs

This section presents a path inference and planning approach based on a large language model API, aimed at enhancing the intelligent decision-making capabilities of opponent UAVs in game environments and constructing more challenging opponent scenarios. In this work, the LLM is used as a high-level planner inside the UAV-FPG simulator, and its contribution is evaluated only within this simulated environment. By integrating reinforcement learning with LLM-based inference mechanisms, we develop a path planning framework that leverages environmental variables as input, enabling opponent UAVs to optimize their action strategies even in complex communication interference environments, such as f_{opponent}.

Implementation details: the opponent UAV planner calls the iFLYTEK Spark Max-32K model via its API, and we do not perform any fine-tuning. The LLM is used as a high-level planner inside the UAV-FPG simulator to generate opponent motion directions conditioned on previous-round trajectories and rewards. LLM inference configuration: we query the LLM once per episode (round) at the episode boundary to generate the next-round motion directions. All decoding settings are fixed across runs and we did not tune generation hyperparameters. Since the planner is accessed as an external service in an offline simulator, we do not impose real-time latency constraints in this study; instead, the LLM call is treated as an episode-level planning step.

A systematically designed prompt is formulated to guide the LLM in generating task-specific future movement paths. The prompt incorporates historical positions of the opponent UAV along with their associated rewards, current positional information, and a standardized output format. This ensures that the generated path data is of high quality and accuracy, significantly improving the feasibility and reliability of path inference. Specifically, the core elements of the prompt design include the following:

(1) **Historical Positions and Rewards:** To help the model understand the behavioral performance of the opponent UAV in past environmental states, each historical position and its corresponding reward are explicitly linked. The expected format is as follows:

$$\text{``Positions and Rewards: } [x_i, y_i, z_i]: \quad r_{opponent}\text{''}$$

(2) **Current Position Information:** The real-time state of the UAV is conveyed to the model to facilitate the generation of rational paths based on its current status to intercept the ally UAV. The expected format is:

$$\text{``Current Position: [x, y, z]''}$$

(3) **Output Format:** A formatted example is provided to clearly define the structure and sequence length constraints for the generated path data, ensuring its usability and structural consistency. The expected format is:

$$\text{``Next directions: } [[x_1, y_1, z_1], [x_2, y_2, z_2], ..., [x_n, y_n, z_n]]\text{''}$$

Text-to-action mapping and fallback: the LLM outputs a sequence of direction vectors $\mathbf{d}_t = [d_x, d_y, d_z]$ in the format `Next directions: [[...], ...]`. We enforce the feasibility constraint $|d_x| + |d_y| + |d_z| \leq 1$. If the constraint is violated, we re-query the LLM up to R times; if it still fails (or the output is not parseable), we project the vector onto the feasible set (e.g., $\mathbf{d}_t \leftarrow \mathbf{d}_t / \max(1, |d_x| + |d_y| + |d_z|)$) as a fallback. The opponent position is then updated by $\mathbf{p}_{t+1} = \mathbf{p}_t + v\Delta t\, \mathbf{d}_t$, where v is the fixed UAV speed and Δt is the simulator time step. We report "stronger" only in terms of the opponent reward/pressure induced in UAV-FPG; we do not claim that the LLM constitutes a universally stronger adversary outside the proposed simulator.

5. Experiments

The present chapter provides a comprehensive description of the experimental design and analysis conducted in the UAV dynamic interaction simulation environment. Firstly, Section 5.1 details the reward-based experiments, wherein the dynamic behavior of allied and opponent UAVs in complex operational scenarios and under opponent conditions is established to evaluate the decision-making capabilities and operational efficiency of the UAVs. Subsequently, Section 5.2 introduces experiments involving path planning using a large language model, exploring the impact of fixed versus dynamic path strategies on enhancing the interference against opponent UAVs. Following this, Section 5.3 presents a dynamic game-theoretic analysis of frequency selection, revealing the evolution characteristics of frequency selection by both allied and opponent UAVs at various stages and demonstrating a significant improvement in the anti-jamming capability of the allied UAV in the latter stages of the game. Finally, Section 5.4 presents an ablation study, separately analyzing the individual contributions of interference methods, expert knowledge bases, and the large language model, thereby assessing their practical impact on decision-making and path planning during opponent engagements. The research presented in this chapter provides significant theoretical support and experimental evidence for the study of autonomous UAV decision-making and opponent strategies.

5.1. Environment Setting

This experiment is conducted in **UAV-FPG,** an executable 3D UAV spectrum-confrontation simulation environment that couples UAV kinematics, geometry-dependent propagation, and an explicit frequency-point decision loop with SINR/capacity-based rewards. The experiment leverages an RTX 4090 GPU for computation and utilizes the iFLYTEK Spark Max-32K model API as an *episode-level opponent trajectory planner* within UAV-FPG (rather than for real-time onboard control). The simulated operational space spans a three-dimensional region of $1500 \times 1500 \times 600$ meters, where the ally UAV follows a predefined Bezier curve trajectory for round-trip missions, while the opponent UAV operates under four distinct movement patterns [70]: `triangle`, `circle`, `rectangle`, and `AI-predicted adaptive trajectories`. The environment incorporates Gaussian noise and dynamic interference strategies to simulate realistic electromagnetic interference conditions; Table 3 summarizes the hyperparameter values of the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) model [30]. Unless otherwise stated, all learning curves and scalar results are averaged over $N_s = 5$ independent runs with different random seeds (different network initializations and environment stochasticity), and we report mean $\pm$ standard deviation across seeds. In reward plots, the shaded band corresponds to $\pm 2\sigma$ across seeds at each sampled step; frequency-overlap curves are reported with standard-deviation error bars across seeds. Table 4 summarizes the environmental parameters of UAV-FPG. Real-time environmental variables, such as channel capacity and SNR, are employed to dynamically adjust the ally UAV's strategies, including spreading, frequency hopping, and velocity control, while the opponent UAV adapts its jamming strategies based on the confrontation dynamics.

Table 3. Hyperparameters for the MADDPG Model.

Parameter	Value
Discount factor, γ	0.99
Total time, T	10^7
Batch size, $\lvert\mathcal{B}\rvert$	32
Learning rate, ϵ	0.001
Buffer capacity, C	10^6
State dimension, *state_dim*	15
Action dimension, *action_dim*	9
Max action value	5
Loss function	MSE
Actor/Critic network (hidden layers, activation)	2×256, ReLU
Exploration noise (Gaussian), σ	0.1

Table 4. Parameters of the UAV-FPG game environment. Note: the bandwidth values (5 MHz non-spread and 2400 MHz spread) are default simulation parameters chosen to create two clearly separated operating regimes in our simplified link/noise model. These bandwidth settings are configurable and can be replaced to match other experimental assumptions.

Parameter	Value
Ally bandwidth (non-spread spectrum)	5 MHz
Ally bandwidth (spread spectrum)	2400 MHz
Noise spectral density	-170 dBm
Opponent check interval	5 s
UAV speed range	10 m/s
Base station power, $\mathbf{P}_{\text{base}}$	45 dBm

Table 4. *Cont.*

Parameter	Value
Opponent power, $\mathbf{P}_{\text{opponent}}$	20 dBm
SNR threshold, M	8
Proximity threshold (Equation (6))	30 m
Number of random seeds, N_s	5
Curve smoothing window, w	200 training steps
LLM call frequency	Once per episode boundary

Note that UAV-FPG is designed as a decision-making/spectrum-confrontation environment where link quality is abstracted via SINR/capacity and used to define step-wise rewards; therefore, we primarily report task-aligned metrics (reward-based statistics, frequency overlap, and jamming success trends) rather than PHY/MAC packet-level metrics such as BER/outage/throughput.

5.2. Opponent Gameplay Performance in UAV-FPG Environment

We use an LLM to plan opponent UAV motion under two settings: fixed-path and dynamic-path gameplay. Fixed-path gameplay follows predefined geometric trajectories (triangle/circle/rectangle). Dynamic-path gameplay re-plans at each round by conditioning on the opponent's recent positions and rewards, yielding more adaptive trajectories and reducing repetition, which improves the opponent's interference effectiveness in UAV-FPG.

Figure 3 illustrates the performance of both static and dynamic path planning strategies executed by opponent UAVs using a LLM in game-theoretic scenarios. The results indicate that both strategies effectively engage in gameplay with allied UAVs, particularly highlighting the superior performance of dynamic path planning. Empirically in UAV-FPG, multi-round, feedback-conditioned prompting tends to improve the planner's trajectory consistency and opponent rewards compared with a single-round prompting baseline under the same decoding settings. For a broader comparison of opponent planners in UAV-FPG, Table 5 summarizes fixed-trajectory, RL-based navigation, and classical non-learning baselines. In addition to learning curves, Table 6 reports seed-averaged scalar metrics (Final-10% and AUC) to provide statistical comparison across methods.

Table 5. Baseline comparison of opponent planners in UAV-FPG using average episode rewards, including fixed trajectory, RL-based navigation, and classical non-learning baselines.

Model	Ally Average Reward	Opponent Average Reward
UAV-FPG (Triangular)	1965.01	1812.17
MASAC [71]	1413.57	1316.85
MAPPO [72]	1572.73	1559.42
UAV-FPG (LLM path planning)	2006.22	2257.48
Greedy Intercept (Pure Pursuit) [73]	3425.19	1899.98
MPC (Short-horizon Optimization) [74]	2252.49	1849.97

To address the reviewer's concern about the lack of benchmarking against prior anti-jamming techniques, we implement representative rule-based and learning-based ally baselines in UAV-FPG, including Random FHSS, blacklist-based Adaptive FH, Bandit-UCB, and a DQN-style channel-selection baseline. For a fair comparison, we fix the opponent to the LLM planner (a higher-reward adversary within UAV-FPG under Equation (6) and our baseline comparisons) and evaluate all methods using the same average episode reward. Table 7 shows that our full system achieves the highest ally reward, outper-

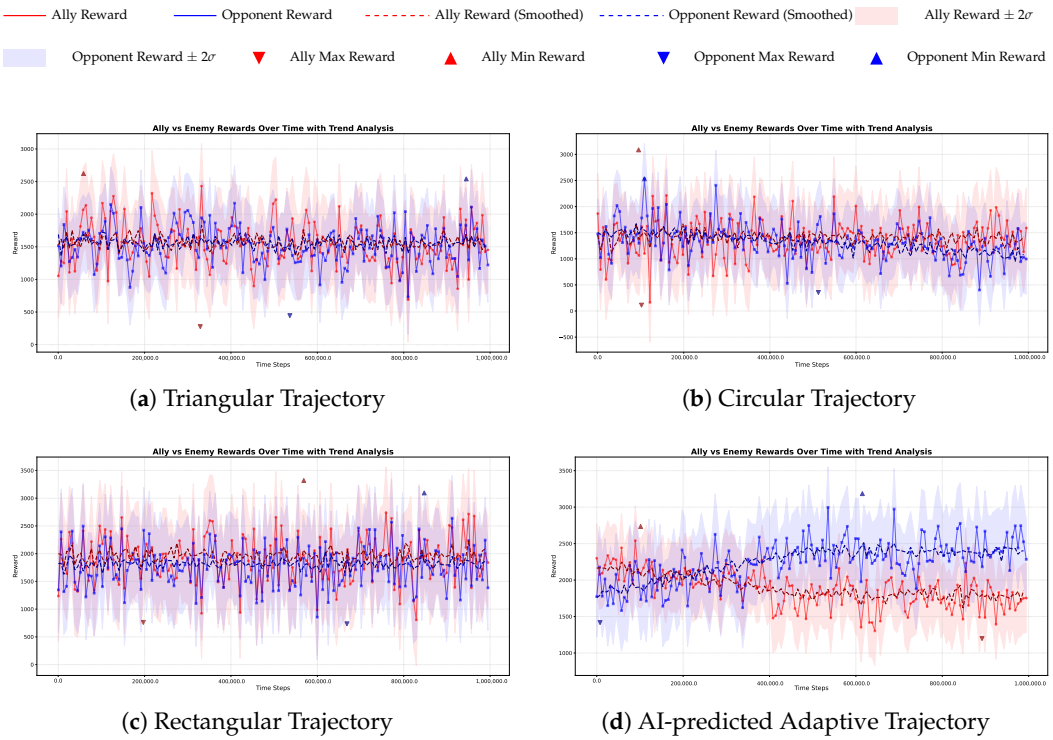
forming both heuristic hopping baselines and learning-based alternatives under the same environment setting.

Table 6. Statistical summary over $N_s = 5$ random seeds (mean $\pm$ std). Final-10% is the average over the last 10% training steps; AUC denotes the step-normalized area under the opponent-reward curve (same scale as reward); Late-stage overlap is computed over 9.5 M–10 M steps.

Method	Final-10% Opp. Reward	AUC (Opp. Reward)	Late-Stage Overlap (%)
UAV-FPG (Triangle)	1847.32 ± 112.45	1782.56 ± 98.73	53.42 ± 0.51
UAV-FPG (Circle)	1683.19 ± 94.28	1594.83 ± 87.62	53.87 ± 0.38
UAV-FPG (Rectangle)	1921.47 ± 105.36	1856.29 ± 93.15	54.15 ± 0.45
UAV-FPG (LLM planner)	2378.65 ± 143.82	2247.91 ± 128.54	54.63 ± 0.52

Table 7. Benchmarking ally-side anti-jamming strategies in UAV-FPG under a fixed LLM-planned opponent.

Model	Ally Avg Reward	Opponent Avg Reward
UAV-FPG (ours)	2006.22	2257.48
No defense (fixed freq, no spread)	952.37	3158.64
Random FHSS (pseudo-random hopping)	1486.53	2687.21
Adaptive FH (blacklist-based)	1753.82	2421.35
Bandit-UCB (15 arms)	1967.45	2198.73
DQN-based anti-jamming	2002.68	2039.56



(**a**) Triangular Trajectory

(**b**) Circular Trajectory

(**c**) Rectangular Trajectory

(**d**) AI-predicted Adaptive Trajectory

Figure 3. Reward trends under different opponent trajectories. Solid lines show raw rewards and dashed lines show smoothed rewards; shaded regions indicate $\pm 2\sigma$. Takeaway: compared with fixed geometric trajectories, the LLM-adaptive trajectory tends to yield higher opponent rewards (i.e., stronger adversarial pressure *as measured in UAV-FPG under Equation (6)*), improving stress-testing of ally policies within the proposed environment.

5.3. Analysis of Frequency Selection in Opponent Scenarios

In an opponent communication scenario, we conduct dynamic detection of the communication center frequencies of both the ally and opponent UAVs at different stages, aiming to analyze the dynamic game characteristics of frequency selection comprehensively. Specifically, we monitor the center frequencies of both sides during the initial, middle, and final stages of the game to evaluate the ally UAV's ability to evade opponent signal detection, as well as the opponent UAV's effectiveness in capturing and interfering with the ally's center frequency. This analysis provides theoretical support for understanding the frequency evolution characteristics in opponent communication environments.

In the different stages of the game illustrated in Figure 4, the distribution of the central frequencies between the opponent and ally parties exhibits significant variation. Notably, in the late stages of the game, the overlap between the ally communication central frequencies and the opponent interference frequencies (53.69%) decreases substantially compared to the early (54.46%) and middle (54.39%) stages. Both the overlap ratio and the maximum overlap value show a marked reduction. This indicates that, as the game progresses, particularly in its later stages, the probability of the adversary successfully detecting the allied communication central frequencies diminishes. Consequently, the allied party achieves more effective anti-interference capability in the selection of communication frequencies.

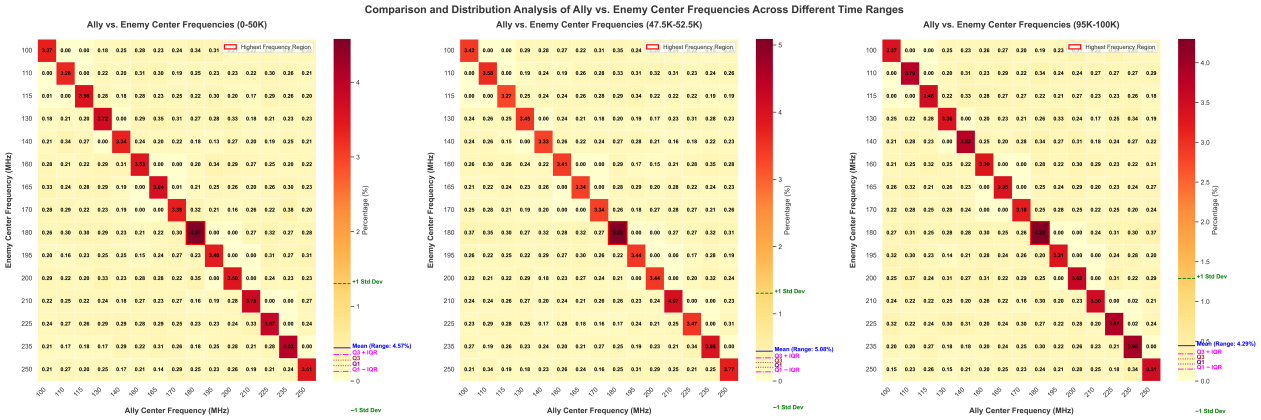


Figure 4. Center-frequency distributions of ally and opponent across training stages. The three panels show 0–50 K, 47.5 K–52.5 K, and 95 K–100 K steps; boxes indicate quartiles and variability. Takeaway: frequency overlap decreases in the late stage, suggesting improved anti-interference frequency selection in UAV-FPG.

Simultaneously, Figure 5 illustrates the dynamic evolution of the frequency overlap ratio and jamming success rate of opponent drones over time. The figure reveals that, although the frequency overlap between the adversary and our communication center gradually decreases, the jamming success rate exhibits a continuous upward trend. This observation suggests that opponent drones may employ strategies such as utilizing broader frequency bands or comb-like spectra to interfere with the communication of our drones. Consequently, they persist in countering our communication strategies throughout the game dynamics and maintain a relatively stable reward value. This further elucidates the evolutionary characteristics of the opponent drones' jamming strategies in contested communication environments. We note that different opponent planners may induce different absolute overlap levels; hence we mainly use overlap to characterize the within-setting stage-wise trend during training (Figure 4). To strengthen baseline comparisons beyond fixed geometric trajectories, Table 5 reports the average episode rewards under multiple opponent planners, including RL-based navigation, a heuristic intercept baseline, and a classical trajectory-optimization baseline. Overall, the LLM-based planner achieves a higher

opponent average reward than the fixed-trajectory and RL-based baselines, indicating stronger and more persistent adversarial pressure for stress-testing ally-side anti-jamming policies in UAV-FPG.

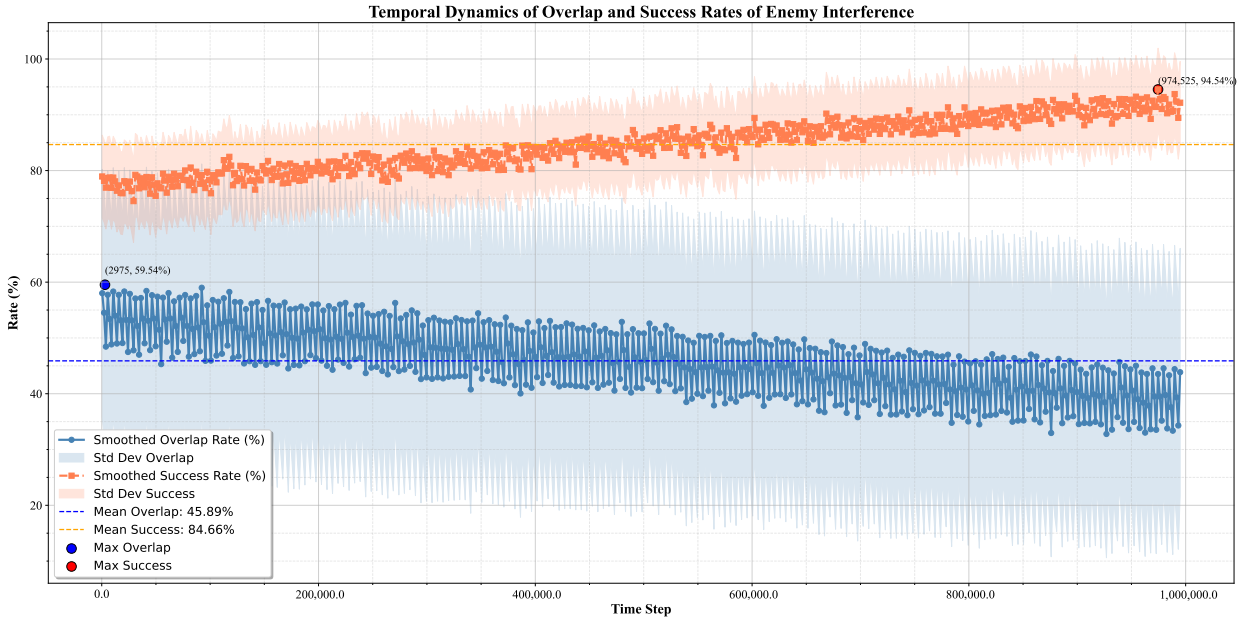


Figure 5. Evolution of frequency overlap and jamming success rate. Curves are smoothed by a moving average with sparse sampling; error bars indicate standard deviation. Takeaway: overlap gradually decreases while the jamming success rate increases, consistent with the opponent switching to broader-band or comb-spectrum strategies in later stages.

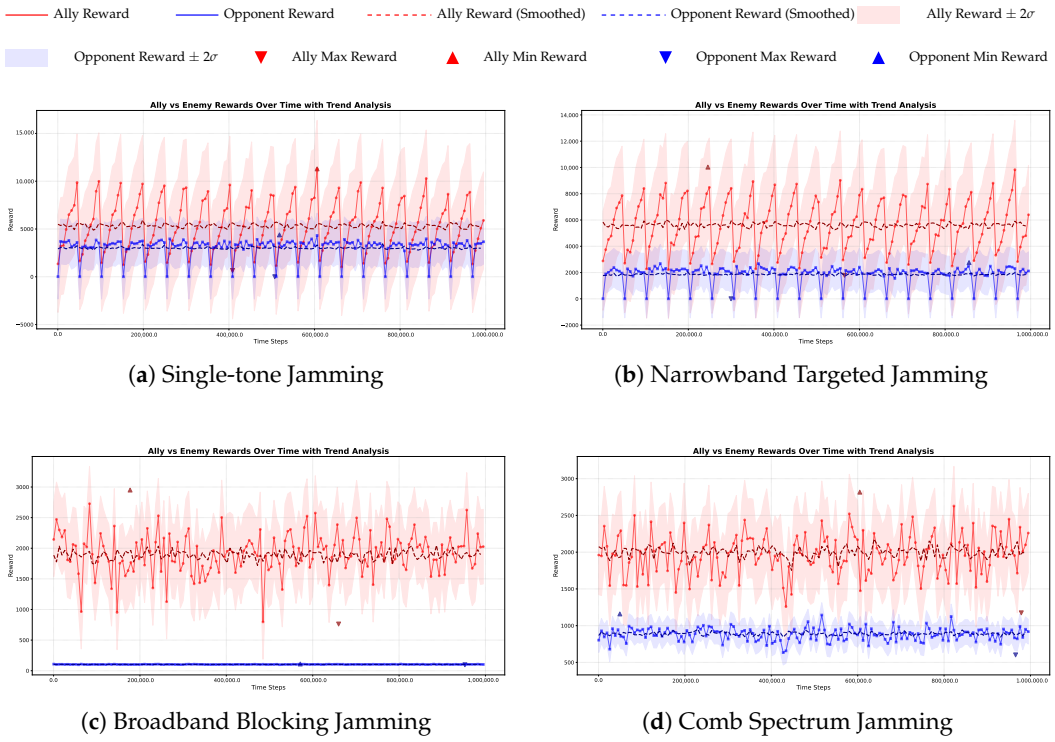
5.4. Ablation Study

To better disentangle the contributions of RL, the expert knowledge base, and the LLM support, we follow a one-factor-at-a-time protocol in all comparisons: only one module is changed while the remaining modules and environment settings are kept identical. In particular, the KB ablation disables the KB-based frequency selector while keeping the same RL training pipeline and opponent setting, whereas the LLM ablation replaces the episode-level LLM opponent planner with non-LLM planners under the same RL and KB configuration. In the ablation study, we divide our investigation into three main components. The first component focuses on the ablation of the opponent interference types, where a single interference method is employed to disrupt the communication between Ally UAV and the base station. This component aims to evaluate the independent effects of various interference strategies.

The second component involves the ablation of the expert knowledge base, where Ally UAV operates without reliance on expert knowledge, engaging directly in a game against the adversary UAV. This analysis intends to elucidate the contribution of expert knowledge to strategic decision-making. The third component pertains to the ablation of the large language model, where reinforcement learning algorithms are used for UAV path planning, enabling the adversary UAV to navigate based on the movement directions learned through reinforcement learning. This approach allows us to assess the contribution of the large language model to the overall task performance, particularly in enhancing the ability of the adversary UAV to approach and interfere with Ally UAV.

Figure 6 displays the results of our ablation experiments on the interference methods employ by opponent UAVs. The findings reveal that the singular interference strategies utilized by the adversary generally underperform compared to Ally UAVs. Notably, the single-tone interference has the least impact, whereas the broadband jamming interference

exerts the most significant effect. In the single-tone jamming scenario depicted in Figure 6a, it is observable that the allies' rewards fall below those of the adversaries during certain periods. This phenomenon can be attributed to the relatively high jamming coefficient, denoted as (k), associated with single-tone interference. Upon successful detection of the allies' communication center frequency and subsequent application of single-tone jamming by the enemy drones, significant disruption occurs in the normal communication among the allied drones, consequently leading to a reduction in their acquired rewards.



(**a**) Single-tone Jamming

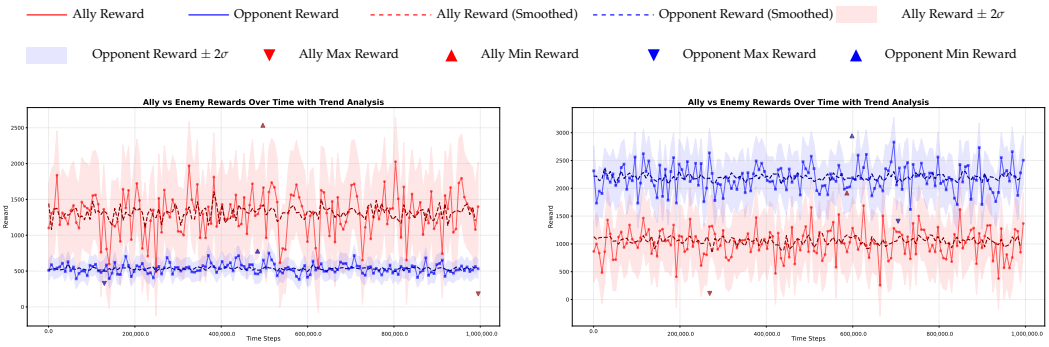
(**b**) Narrowband Targeted Jamming

(**c**) Broadband Blocking Jamming

(**d**) Comb Spectrum Jamming

Figure 6. Ablation study on individual jamming types. Rewards of ally (red) and opponent (blue) are shown with smoothed curves (dashed) and variability bands ($\pm 2\sigma$). Takeaway: broadband blocking generally imposes the strongest degradation in ally reward among single-type jamming baselines, while single-tone jamming shows comparatively weaker overall impact in our setting.

Figure 7a and Figure 7b, respectively present the ablation study results of the LLM for opponent UAV path planning and the expert knowledge base for ally UAV anti-interference strategies. The experimental results indicate that both the application of the LLM in path planning and the deployment of the expert knowledge base in anti-interference strategies play a critical role in the game process. Ablating the LLM significantly reduces the flexibility and adaptability of the opponent UAV's path, whereas ablating the expert knowledge base markedly weakens the ally UAV's anti-interference capability. These findings demonstrate that combining the LLM with the expert knowledge base effectively enhances the UAVs' performance in complex environments, improving the intelligence and adaptability of opponent encounters. Therefore, future research should focus on further optimizing the reasoning capabilities of LLMs and expanding the expert knowledge base's interference response strategies to achieve a more robust UAV opponent system.



(**a**) Without Using LLM to Plan Paths (**b**) Without Using Expert Knowledge Base

Figure 7. Ablation Study on the Contribution of LLM-Based Path Planning and Expert Knowledge Anti-Interference. The red and blue curves represent the rewards obtained by the ally and opponent sides in the game, respectively.

6. Conclusions

This paper introduces **UAV-FPG**, a drone communication environment based on frequency-point game theory that is designed to optimize frequency decision-making and path planning in complex electromagnetic settings. By establishing a signal game model, we investigate the dynamic interplay between interference and anti-interference strategies of ally and opponent drones within communication frequency bands. Leveraging a reinforcement learning-based game environment, we simulate various opponent signal interference strategies alongside our own frequency hopping and spread-spectrum interference avoidance techniques, which provides an effective platform to emulate threats encountered during signal transmission. Furthermore, we integrate an expert knowledge base to enhance the decision-making capabilities of Ally UAVs in frequency selection and management, enabling more effective countermeasures against opponent interference strategies. By introducing an LLM as a high-level planner for the opponent UAV inside UAV-FPG, our simulation results indicate that iterative, feedback-conditioned prompting can generate more adaptive trajectories than several fixed-path baselines, thereby strengthening the opponent pressure used to evaluate ally-side anti-jamming policies. We emphasize that these conclusions are drawn from the proposed simulation environment; validating transfer to higher-fidelity simulators and real-world flight/EM conditions remains an important direction for future work.

Experimental results demonstrate that the UAV-FPG environment significantly enhances the adaptability and robustness of UAVs under hostile signal interference. Specifically, in frequency-point selection experiments, as the game progresses, the overlap between Ally UAV's central frequency and the opponent's interference frequency markedly decreases, which evidences a progressive improvement in anti-interference capabilities during complex confrontations. Additionally, through multiple rounds of interaction with LLMs, we verify the substantial role of large language models in path planning; by integrating historical positions and reward information, they effectively strategize the next moves of opponent UAVs, substantially increasing the efficiency of opponent interference. Finally, through ablation studies, we further validate the independent contributions of different interference methods, expert knowledge bases, and LLMs within the UAV opponent environment. The findings indicate that expert knowledge and LLMs play pivotal roles in enhancing decision intelligence and adaptability in path planning.

7. Limitation and Future Work

Limitations of this work include the lack of real-world UAV/hardware deployment and field electromagnetic testing, the use of simplified propagation/interference assump-

tions in the simulator, and the reliance on an external LLM as an episode-level planner (with potential latency/cost/controllability considerations); addressing these issues via higher-fidelity simulation and real-world validation is an important direction for future work. In practice, the LLM module can be replaced by an edge/offboard planner or a lightweight distilled policy for deployment-oriented settings. A further extension is to incorporate more detailed PHY/MAC assumptions (e.g., modulation/coding and packetization) to enable packet-level metrics such as throughput, BER, and outage probability.

Author Contributions: Conceptualization, J.Y., Y.W. and M.W.; methodology, J.Y. and F.J.; software, J.Y., F.J. and H.Z.; validation, J.Y., H.Z. and Y.L.; formal analysis, J.Y. and Y.L.; investigation, J.Y. and F.J.; resources, Y.W., M.W. and W.D.; data curation, J.Y., H.Z. and M.W.; writing—original draft preparation, J.Y.; writing—review and editing, Y.W., M.W. and W.D.; visualization, J.Y.; supervision, Y.W., M.W. and W.D.; project administration, Y.W.; funding acquisition, Y.W. and W.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Defense Industrial Technology Development Program (JCKY2024601C023).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fan, B.; Li, Y.; Zhang, R.; Fu, Q. Review on the technological development and application of UAV systems. *Chin. J. Electron.* **2020**, *29*, 199–207. [CrossRef]
2. Radoglou-Grammatikis, P.; Sarigiannidis, P.; Lagkas, T.; Moscholios, I. A compilation of UAV applications for precision agriculture. *Comput. Netw.* **2020**, *172*, 107148. [CrossRef]
3. Rana, K.; Praharaj, S.; Nanda, T. Unmanned aerial vehicles (UAVs): An emerging technology for logistics. *Int. J. Bus. Manag. Invent.* **2016**, *5*, 86–92.
4. Wang, Q.; Li, W.; Yu, Z.; Abbasi, Q.; Imran, M.; Ansari, S.; Sambo, Y.; Wu, L.; Li, Q.; Zhu, T. An overview of emergency communication networks. *Remote Sens.* **2023**, *15*, 1595. [CrossRef]
5. Alsamhi, S.H.; Afghah, F.; Sahal, R.; Hawbani, A.; Al-qaness, M.A.; Lee, B.; Guizani, M. Green internet of things using UAVs in B5G networks: A review of applications and strategies. *Ad Hoc Netw.* **2021**, *117*, 102505. [CrossRef]
6. Gallacher, D. Drone applications for environmental management in urban spaces: A review. *Int. J. Sustain. Land Use Urban Plan.* **2016**, *3*, 1–14. [CrossRef]
7. Shi, L.; Marcano, N.J.H.; Jacobsen, R.H. A review on communication protocols for autonomous unmanned aerial vehicles for inspection application. *Microprocess. Microsyst.* **2021**, *86*, 104340. [CrossRef]
8. de Curtò, J.; de Zarzà, I.; Cano, J.C.; Calafate, C.T. Enhancing Communication Security in Drones Using QRNG in Frequency Hopping Spread Spectrum. *Future Internet* **2024**, *16*, 412. [CrossRef]
9. Wang, R.; Wang, S.; Zhang, W. Joint power and hopping rate adaption against follower jammer based on deep reinforcement learning. *Trans. Emerg. Telecommun. Technol.* **2023**, *34*, e4700. [CrossRef]
10. Rao, N.; Xu, H.; Qi, Z.; Wang, D. Fast adaptive jamming resource allocation against frequency-hopping spread spectrum in wireless sensor networks via meta deep reinforcement learning. *IEEE Trans. Aerosp. Electron. Syst.* **2024**, *60*, 7676–7693. [CrossRef]
11. Khan, M.T. A modified convolutional neural network with rectangular filters for frequency-hopping spread spectrum signals. *Appl. Soft Comput.* **2024**, *150*, 111036. [CrossRef]
12. Shakhatreh, H.; Sawalmeh, A.; Hayajneh, K.F.; Abdel-Razeq, S.; Al-Fuqaha, A. A Systematic Review of Interference Mitigation Techniques in Current and Future UAV-Assisted Wireless Networks. *IEEE Open J. Commun. Soc.* **2024**, *5*, 2815–2846. [CrossRef]
13. Pärlin, K.; Riihonen, T.; Turunen, M. Sweep jamming mitigation using adaptive filtering for detecting frequency agile systems. In *Proceedings of the Military Communications and Information Systems, Budva, Montenegro, 14–15 May 2019*; IEEE: Piscataway, NJ, USA; pp. 1–6.
14. Liu, T.; Huang, J.; Guo, J.; Shan, Y. Survey on anti-jamming technology of uav communication. In *Proceedings of the International Conference on 5G for Future Wireless Networks, Shanghai, China, 7–8 October 2022*; Springer: Berlin, Germany, 2022; pp. 111–121.
15. Wang, R.; Wang, S.; Zhang, W. *Cooperative Multi-UAV Dynamic Anti-Jamming Scheme with Deep Reinforcement Learning*; IEEE: Piscataway, NJ, USA, 2022; pp. 590–595.

16. Li, Z.; Lu, Y.; Li, X.; Wang, Z.; Qiao, W.; Liu, Y. UAV networks against multiple maneuvering smart jamming with knowledge-based reinforcement learning. *IEEE Internet Things J.* **2021**, *8*, 12289–12310. [CrossRef]
17. Yao, F.; Jia, L. A Collaborative Multi-Agent Reinforcement Learning Anti-Jamming Algorithm in Wireless Networks. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 1024–1027. [CrossRef]
18. Chen, Y.; Wang, Y.; Zhao, K.; Liang, H.; Liu, P.; Yang, Y. GPDS: A multi-agent deep reinforcement learning game for anti-jamming secure computing in MEC network. *Expert Syst. Appl.* **2022**, *210*, 118394. [CrossRef]
19. Liu, D.; Wang, J.; Xu, Y.; Ruan, L.; Zhang, Y. A coalition-based communication framework for intelligent flying ad-hoc networks. *arXiv* **2018**, arXiv:1812.00896.
20. Shah, S.; Dey, D.; Lovett, C.; Kapoor, A. AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles. In *Field and Service Robotics Springer Proceedings in Advanced Robotics*; Hutter, M., Siegwart, R., Eds.; Springer: Cham, Switzerland, 2018; Volume 5, pp. 621–635. [CrossRef]
21. Furrer, F.; Burri, M.; Achtelik, M.; Siegwart, R. RotorS—A Modular Gazebo MAV Simulator Framework. In *Robot Operating System (ROS): The Complete Reference (Volume 1)*; Studies in Computational Intelligence; Koubaa, A., Ed.; Springer: Cham, Switzerland, 2016; Volume 625, pp. 595–625. [CrossRef]
22. Song, Y.; Naji, S.; Kaufmann, E.; Loquercio, A.; Scaramuzza, D. Flightmare: A Flexible Quadrotor Simulator. In *Proceedings of the 2020 Conference on Robot Learning*; Kober, J., Ramos, F., Tomlin, C., Eds.; PMLR: Cambridge, MA, USA, 2021; Volume 155, pp. 1147–1157.
23. Panerati, J.; Zheng, H.; Zhou, S.; Xu, J.; Prorok, A.; Schoellig, A.P. Learning to Fly—A Gym Environment with PyBullet Physics for Reinforcement Learning of Multi-agent Quadcopter Control. *arXiv* **2021**, arXiv:2103.02142.
24. Liu, X.; Xu, Y.; Jia, L.; Wu, Q.; Anpalagan, A. Anti-jamming Communications Using Spectrum Waterfall: A Deep Reinforcement Learning Approach. *arXiv* **2017**, arXiv:1710.04830. [CrossRef]
25. Nowé, A.; Vrancx, P.; De Hauwere, Y.M. Game theory and multi-agent reinforcement learning. *Reinf. Learn.-State Art* **2012**, *15*, 441–470.
26. Yang, Y.; Wang, J. An overview of multi-agent reinforcement learning from game theoretical perspective. *arXiv* **2020**, arXiv:2011.00583.
27. Zhang, K.; Yang, Z.; Başar, T. Multi-agent reinforcement learning: A selective overview of theories and algorithms. In *Handbook of Reinforcement Learning and Control*; Springer: Berlin, Germany, 2021; pp. 321 –384.
28. Busoniu, L.; Babuska, R.; De Schutter, B. Multi-agent reinforcement learning: A survey. In *Proceedings of the Control, Automation, Robotics and Vision, Singapore, 5–8 December 2006;* IEEE: Piscataway, NJ, USA, 2006; pp. 1–6.
29. Yang, J.; Wang, H.; Zhao, Q.; Shi, Z.; Song, Z.; Fang, M. Efficient Reinforcement Learning via Decoupling Exploration and Utilization. In *Proceedings of the International Conference on Intelligent Computing, Zakopane, Poland, 7–11 June 2024*; Springer: Berlin, Germany, 2024; pp. 396–406.
30. Lowe, R.; Wu, Y.I.; Tamar, A.; Harb, J.; Pieter Abbeel, O.; Mordatch, I. Multi-agent actor-critic for mixed cooperative-competitive environments. *Adv. Neural Inf. Process. Syst.* **2017**, *30, 6379–6390*
31. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484–489. [CrossRef]
32. Sunehag, P.; Lever, G.; Gruslys, A.; Czarnecki, W.M.; Zambaldi, V.; Jaderberg, M.; Lanctot, M.; Sonnerat, N.; Leibo, J.Z.; Tuyls, K.; et al. Value-decomposition networks for cooperative multi-agent learning. *arXiv* **2017**, arXiv:1706.05296.
33. Rashid, T.; Samvelyan, M.; De Witt, C.S.; Farquhar, G.; Foerster, J.; Whiteson, S. Monotonic value function factorisation for deep multi-agent reinforcement learning. *J. Mach. Learn. Res.* **2020**, *21*, 1–51.
34. Zhao, N.; Ye, Z.; Pei, Y.; Liang, Y.C.; Niyato, D. Multi-agent deep reinforcement learning for task offloading in UAV-assisted mobile edge computing. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 6949–6960. [CrossRef]
35. Cui, J.; Liu, Y.; Nallanathan, A. Multi-agent reinforcement learning-based resource allocation for UAV networks. *IEEE Trans. Wirel. Commun.* **2019**, *19*, 729–743. [CrossRef]
36. Ganzfried, S. Fictitious play outperforms counterfactual regret minimization. *arXiv* **2020**, arXiv:2001.11165.
37. Brown, N.; Lerer, A.; Gross, S.; Sandholm, T. Deep counterfactual regret minimization. In *Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019;* PMLR: Cambridge, MA, USA, 2019; pp. 793–802.
38. Gipiškis, R.; Joaquin, A.S.; Chin, Z.S.; Regenfuß, A.; Gil, A.; Holtman, K. Risk Sources and Risk Management Measures in Support of Standards for General-Purpose AI Systems. *arXiv* **2024**, arXiv:2410.23472. [CrossRef]
39. Kulkarni, A.; Shivananda, A.; Manure, A. Decision Intelligence Overview. In *Introduction to Prescriptive AI: A Primer for Decision Intelligence Solutioning with Python*; Springer: Berlin, Germany, 2023; pp. 1–25.
40. Zhu, Q.; Başar, T. *Decision and Game Theory for Security*; Springer: Berlin, Germany, 2013.
41. Yu, A.; Kolotylo, I.; Hashim, H.A.; Eltoukhy, A.E. Electronic Warfare Cyberattacks, Countermeasures and Modern Defensive Strategies of UAV Avionics: A Survey. *IEEE Access* **2025**, *13*, 68660–68681. [CrossRef]
42. Wu, Q.; Wang, H.; Li, X.; Zhang, B.; Peng, J. Reinforcement learning-based anti-jamming in networked UAV radar systems. *Appl. Sci.* **2019**, *9*, 5173. [CrossRef]

43. Zhang, Z.; Zhou, Y.; Zhang, Y.; Qian, B. Strong electromagnetic interference and protection in uavs. *Electronics* **2024**, *13*, 393. [CrossRef]
44. Zhou, B.; Yang, G.; Shi, Z.; Ma, S. Natural language processing for smart healthcare. *IEEE Rev. Biomed. Eng.* **2022**, *17*, 4–18. [CrossRef]
45. Pee, L.G.; Pan, S.L.; Cui, L. Artificial intelligence in healthcare robots: A social informatics study of knowledge embodiment. *J. Assoc. Inf. Sci. Technol.* **2019**, *70*, 351–369. [CrossRef]
46. Wang, Y.H.; Lin, G.Y. Exploring AI-healthcare innovation: Natural language processing-based patents analysis for technology-driven roadmapping. *Kybernetes* **2023**, *52*, 1173–1189. [CrossRef]
47. Gyrard, A.; Tabeau, K.; Fiorini, L.; Kung, A.; Senges, E.; De Mul, M.; Giuliani, F.; Lefebvre, D.; Hoshino, I. Knowledge engineering framework for IoT robotics applied to smart healthcare and emotional well-being. *Int. J. Soc. Robot.* **2023**, *15*, 445–472. [CrossRef]
48. Piat, G.X. Incorporating Expert Knowledge in Deep Neural Networks for Domain Adaptation in Natural Language Processing. Ph.D. Thesis, Université Paris-Saclay, Paris, France, 2023.
49. Chanda, A.K.; Bai, T.; Yang, Z.; Vucetic, S. Improving medical term embeddings using UMLS Metathesaurus. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 114. [CrossRef]
50. McCray, A.T.; Aronson, A.R.; Browne, A.C.; Rindflesch, T.C.; Razi, A.; Srinivasan, S. UMLS knowledge for biomedical language processing. *Bull. Med. Libr. Assoc.* **1993**, *81*, 184.
51. El Ghosh, M. Automation of Legal Reasoning and Decision Based on Ontologies. Ph.D. Thesis, Normandie Université, Caen, France, 2018.
52. Karwowski, J.; Szynkiewicz, W.; Niewiadomska-Szynkiewicz, E. Bridging Requirements, Planning, and Evaluation: A Review of Social Robot Navigation. *Sensors* **2024**, *24*, 2794. [CrossRef]
53. Xiao, X.; Liu, B.; Warnell, G.; Stone, P. Motion planning and control for mobile robot navigation using machine learning: A survey. *Auton. Robot.* **2022**, *46*, 569–597. [CrossRef]
54. Sun, X.; Zhang, Y.; Chen, J. RTPO: A domain knowledge base for robot task planning. *Electronics* **2019**, *8*, 1105. [CrossRef]
55. Pan, H.; Huang, S.; Yang, J.; Mi, J.; Li, K.; You, X.; Tang, X.; Liang, P.; Yang, J.; Liu, Y.; et al. Recent Advances in Robot Navigation via Large Language Models: A Review. *Res. Gate* **2024**, *preprint*.
56. Yao, F.; Yue, Y.; Liu, Y.; Sun, X.; Fu, K. AeroVerse: UAV-Agent Benchmark Suite for Simulating, Pre-training, Finetuning, and Evaluating Aerospace Embodied World Models. *arXiv* **2024**, arXiv:2408.15511.
57. Andreoni, M.; Lunardi, W.T.; Lawton, G.; Thakkar, S. Enhancing autonomous system security and resilience with generative AI: A comprehensive survey. *IEEE Access* **2024**, *12*, 109470–109493. [CrossRef]
58. Liu, Y.; Yao, F.; Yue, Y.; Xu, G.; Sun, X.; Fu, K. NavAgent: Multi-scale Urban Street View Fusion For UAV Embodied Vision-and-Language Navigation. *arXiv* **2024**, arXiv:2411.08579.
59. Chen, Z.; Xu, L.; Zheng, H.; Chen, L.; Tolba, A.; Zhao, L.; Yu, K.; Feng, H. Evolution and Prospects of Foundation Models: From Large Language Models to Large Multimodal Models. *Comput. Mater. Contin.* **2024**, *80*, 1753–1808. [CrossRef]
60. Kang, J.; Liao, J.; Gao, R.; Wen, J.; Huang, H.; Zhang, M.; Yi, C.; Zhang, T.; Niyato, D.; Zheng, Z. Efficient and Trustworthy Block Propagation for Blockchain-enabled Mobile Embodied AI Networks: A Graph Resfusion Approach. *arXiv* **2025**, arXiv:2502.09624. [CrossRef]
61. Zheng, Z.; Bewley, T.R.; Kuester, F. Point cloud-based target-oriented 3D path planning for UAVs. In *Proceedings of the International Conference on Unmanned Aircraft Systems*; IEEE: Piscataway, NJ, USA, 2020; pp. 790–798.
62. Jin, Y.; Yue, M.; Li, W.; Shangguan, J. An improved target-oriented path planning algorithm for wheeled mobile robots. *J. Mech. Eng. Sci.* **2022**, *236*, 11081–11093. [CrossRef]
63. Zhao, X.; Cai, W.; Tang, L.; Wang, T. ImagineNav: Prompting Vision-Language Models as Embodied Navigator through Scene Imagination. *arXiv* **2024**, arXiv:2410.09874.
64. Zhao, R.; Yuan, Q.; Li, J.; Fan, Y.; Li, Y.; Gao, F. DriveLLaVA: Human-Level Behavior Decisions via Vision Language Model. *Sensors* **2024**, *24*, 4113. [CrossRef]
65. Zhang, Y.F.; Wen, Q.; Fu, C.; Wang, X.; Zhang, Z.; Wang, L.; Jin, R. Beyond LLaVA-HD: Diving into High-Resolution Large Multimodal Models. *arXiv* **2024**, arXiv:2406.08487. [CrossRef]
66. Reed, S.; Zolna, K.; Parisotto, E.; Colmenarejo, S.G.; Novikov, A.; Barth-Maron, G.; Gimenez, M.; Sulsky, Y.; Kay, J.; Springenberg, J.T.; et al. A generalist agent. *arXiv* **2022**, arXiv:2205.06175. [CrossRef]
67. Chen, X. One Step Towards Autonomous AI Agent: Reasoning, Alignment and Planning. Ph.D. Thesis, University of California, Los Angeles, CA, USA, 2024.
68. Jeong, H.; Lee, H.; Kim, C.; Shin, S. A Survey of Robot Intelligence with Large Language Models. *Appl. Sci.* **2024**, *14*, 8868. [CrossRef]
69. Poisel, R.A. *Introduction to Communication Electronic Warfare Systems*; Artech House, Inc.: Norwood, MA, USA, 2008.
70. Zhang, C.; Huang, G.; Liu, L.; Huang, S.; Yang, Y.; Wan, X.; Ge, S.; Tao, D. WebUAV-3M: A benchmark for unveiling the power of million-scale deep UAV tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 9186–9205. [CrossRef]

71. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In *Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018;* PMLR: Cambridge, MA, USA, 2018; Volume 80, pp. 1861–1870.
72. Yu, C.; Velu, A.; Vinitsky, E.; Gao, J.; Wang, Y.; Bayen, A.; Wu, Y. The Surprising Effectiveness of PPO in Cooperative, Multi-Agent Games. *arXiv* **2021**, arXiv:2103.01955.
73. Coulter, R.C. *Implementation of the Pure Pursuit Path Tracking Algorithm*; Technical Report CMU-RI-TR-92-01; Carnegie Mellon University, The Robotics Institute: Pittsburgh, PA, USA, 1992.
74. Mayne, D.Q.; Rawlings, J.B.; Rao, C.V.; Scokaert, P.O.M. Constrained model predictive control: Stability and optimality. *Automatica* **2000**, *36*, 789–814. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.