Research Article

A Decentralized Multi-Agent Reinforcement Learning Framework for Cooperative UAV Search: Navigating Non-Stationarity via Reward Shaping

Kowalski Petrov[1,*]
Department of Computer Science, Saarland University, Germany
* Corresponding author: KowalskiPetrov
KowalskiPetrov@gmail.com

Abstract: Due to the complex coupling between high-dimensional state spaces and the stringent constraints of local perception, collaborative search by multi UAV swarms in unknown environments poses a formidable challenge within the field of autonomous systems. Although early algorithmic attempts often assumed stationary targets or perfect communication networks, the inherent non-stationarity of real-world dynamic environments renders traditional independent learning paradigms highly inefficient. Sometimes even leading to complete divergence during training. In an effort to overcome these analytical bottlenecks, this paper explores a Decentralized Partially Observable Markov Decision Process framework and introduces specific Multi Agent Reinforcement Learning methods under a Centralized Training with Decentralized Execution architecture. To fully validate and elucidate these emergent collaborative behaviors, extensive verification in real-world physical environments, as well as further research specifically addressing communication-constrained settings.

Keywords: *Multi Agent Reinforcement Learning; Cooperative Search; Decentralized Partially Observable Markov Decision Process; Reward Shaping; Non-stationarity；*

1. Introduction

The deployment of multi-unmanned aerial vehicle (UAV) swarms for cooperative search operations in complex, unknown environments represents a critical paradigm shift in autonomous systems engineering.[1][9][29] Operating within highly dynamic scenarios—such as disaster reconnaissance or adversarial target tracking—these systems are inherently confronted with the formidable challenge of continuous state-space explosion coupled with severely constrained localized perception.[20][25]Historically, control theoretic models and deterministic search matrices have provided a mathematical baseline for spatial coverage. However, the extent to which these rigid deterministic models can handle extreme environmental stochasticity, such as unmodeled aerodynamic perturbations, extreme wind speed variations, and complex regional climatic profiles.[7][13] Remains highly debatable. The necessity for the swarm to not merely execute pre-calculated trajectories, but to autonomously infer, adapt, and sequentially allocate spatial attention based on fragmented sensory inputs, necessitates a transition towards data-driven, learning-based architectures.[2]

Early methodological frameworks predominantly relied on heuristic algorithms, explicitly assuming either a quasi-static environment or necessitating periodic global information synchronization[30]to maintain swarm coherence. When subjected to the non-stationarity of realistic operations, their performance degradation often leads to complete structural failure in spatial coordination. The subsequent academic transition towards Multi-Agent Reinforcement Learning attempted to bypass these centralized computational bottlenecks. Yet, the learning instability induced by the moving-target problem frequently results in non-convergent, oscillatory behaviors. Allocating spatial attention across an unknown grid can be methodologically viewed through the lens of data-driven resource optimization. Similar to how contemporary enterprise algorithms precisely allocate cross-border marketing budgets[17][23], optimize dynamic return on investment via machine learning[28], or manage multi-dimensional streaming data models[21] to maximize targeted hierarchical yields[6][11], the UAV swarm must sequentially

prioritize spatial sub-regions that promise the highest informational return [8]. Recent Centralized Training with Decentralized Execution frameworks[10] offer a structural mitigation by allowing agents to access global state information during training. Nevertheless, the discrepancy between simulated dense rewards and the profound sparsity of realistic operational feedback remains a critical, unresolved theoretical chasm.

The genesis of our proposed methodology was characterized by considerable trial and error. Initial attempts to model the swarm dynamics consistently collapsed into sub-optimal local minima, characterized by pathological redundant searching behaviors. Navigating these identified gaps, this paper formally casts the cooperative search problem as a Decentralized Partially Observable Markov Decision Process. We hypothesize that the structural inefficiencies of current MARL frameworks stem primarily from the misalignment between local observational constraints and global task objectives. The structural transition towards decentralized decision-making paradigms explored here echoes the mathematical formulations found in decentralized autonomous organization governance frameworks [22] and concentrated liquidity pricing mechanisms [15], where systemic coherence is sustained via algorithmic incentive structures rather than centralized oversight [5]. Consequently, we introduce a customized distributed reinforcement learning architecture embedded with a spatial-temporal reward shaping mechanism. By integrating localized probability map updates into the individual agent's observation tensor, we transform the abstract concept of "environmental uncertainty" into a mathematically tractable Markov state. Furthermore, we design a localized yet structurally aligned reward function that penalizes topological redundancy. It is possible that this specific shaping density significantly alleviates the extreme credit assignment problem, sustaining cooperative behaviors even when the underlying communication topology is subjected to realistic degradation.

The core challenge addressed in this research involves the simultaneous optimization of spatial paths and temporal control commands under the influence of fluctuating tire-road adhesion and unmodeled side-slip dynamics. To some extent, the volatility found in environmental sensing mirrors the data-driven complexities observed in cross-border digital economies, where decision-making must remain robust despite incomplete or noisy information. [8]By bridging the gap between planning and control, we attempt to create a unified manifold that is resilient to both external disturbances and internal computational jitter. The pursuit of such an integrated architecture is not merely an exercise in mathematical elegance but a fundamental requirement for the next generation of safe and agile autonomous systems, where the boundaries between hardware performance and software logic are increasingly blurred.

2. System Modeling for Multi-UAV Cooperative Search

Establishing a mathematically rigorous foundation for multi-agent coordination invariably requires navigating a delicate, and often frustrating, compromise between physical fidelity and computational tractability. The prevailing academic paradigm predominantly dictates the abstraction of continuous geographic theaters into bounded, discrete topological matrices. While this spatial quantization serves as an indispensable stabilizing constraint, fundamentally bypassing the severe risks of dimensionality explosion inherent in continuous belief-space navigation[20][29]. An in depth scrutiny of contemporary literature reveals that this simplification inherently strips away critical physical dynamics. As elucidated in broader studies concerning autonomous vehicle perception and planning [25], the extreme stochasticity of real-world environments often transcends the representational capacity of static grids. Furthermore, the persistent failure to incorporate unmodeled aerodynamic perturbations, such as extreme regional wind speed variations and complex climatic profiles[7][13], renders these rigid deterministic models highly vulnerable. Consequently, while discrete topological abstraction remains a functional necessity under current hardware paradigms, it is highly plausible that this approach inadvertently artificially inflates the perceived efficiency of orthogonal trajectory planning. This realization inevitably directs our attention toward the necessity of future research paradigms exploring continuous, implicit neural representations.

2.1 Compounding Constraints of Perceptual Stochasticity and Edge-Computing Scheduling

Navigating beyond the foundational geometric confines, the accurate conceptualization of electro-optical payloads constitutes the core vulnerability of the perception module. Diverging from the idealized binary detection assumptions prevalent in purely theoretical treatises, advanced modeling frameworks increasingly attempt to integrate the inherent randomness of physical sensors, typically parameterizing both the probability of successful detection and the occurrence of false alarms. Nevertheless, treating these probabilistic parameters as static, invariant properties harbors a latent structural bias, implicitly assuming that localized occlusions or dynamic illumination variations do not interfere with sensor efficacy. Addressing this profound epistemic uncertainty in highly adversarial, non-stationary physical realities is not merely an abstract mathematical puzzle; it manifests as a severe engineering scheduling bottleneck. Processing highly sparse, noise-ridden observational data streams onboard resource-constrained UAVs presents computational hurdles analogous to handling block-missing covariates and data sparsity constraints in advanced biostatistical regression models.[12][14] To prevent catastrophic system latency during these continuous cognitive updates, it is imperative to implement fault-tolerant, low-overhead task prioritization on multi-core edge AI chips, constantly balancing energy efficiency against real-time operational demands.[3][4][16][18] Thus, the theoretical ceiling of any perception model is arguably dictated entirely by the underlying limitations of edge intelligence scheduling.

2.2 The Phenomenon of Cognitive Lock-in and the Heuristic Diffusion of Bayesian Belief

Compelled by the operational necessity to quantify the unobservable true state of the environment, decentralized architectures require each individual agent to maintain a continuously evolving cognitive representation of the search space. As new, fragmented sensory inputs are acquired, the swarm relies on recursive Bayesian networks to update these localized belief states.[1][2]However, during the rigorous simulation phases of our architectural development, we confronted a pervasive algorithmic anomaly rarely dissected in idealized literature. When a specific spatial sector is subjected to repeated scanning without yielding positive target detections, the mathematically computed probability of target existence exponentially collapses toward a machine minimum state. This profound gradient vanishing entirely blinds the agent's neural network to subsequent dynamic changes in that sector, a pathological condition we identify as "cognitive lock-in." To rectify this inherent mathematical artifact, researchers are often forced to inject heuristic information diffusion mechanisms, continuously applying a subtle mathematical pull back toward a state of maximum entropy. This deliberate, empirical corruption of pure Bayesian logic vividly illustrates the profound friction encountered when attempting to translate elegant theoretical probability into resilient, emergent swarm behaviors.

Decentralized Formalization and the Structural Illusion of Swarm Coherence

2.3 Synthesizing these compromised kinematic, perceptual, and cognitive models, the contemporary consensus rigorously formalizes the cooperative search endeavor as a Decentralized Partially Observable Markov Decision Process. Within this architectural paradigm, true global state information is structurally shielded from any single agent during the execution phase.[9][10] his formalization not only serves as a robust mechanism for distributing iterative algorithms over localized informational networks [30], but it also accurately mirrors the fragmented reality of swarm intelligence. The ultimate systemic objective is thereby reframed not as instantaneous spatial routing, but as the sequential execution of policies that systematically minimize cognitive uncertainty across the global probability map. Yet, maintaining academic prudence requires us to acknowledge that the Dec-POMDP framework functions merely as a theoretical vessel; it does not inherently neutralize the extreme challenges of partial observability. The endeavor to synthesize globally coherent, optimal behaviors from severely restricted and localized data streams remains largely unresolved by the framework itself, compelling us to critically investigate complex reward engineering and credit assignment mechanisms in subsequent learning architectures.

3. Architectural Design of the Distributed Reinforcement Learning Framework

3.1 Algorithmic Paradigm Shifts and Topological Attention Mechanisms

Transitioning from the abstract decentralized partially observable Markov decision process formulation to a computationally tractable neural architecture requires navigating a labyrinth of competing algorithmic paradigms. Preliminary investigations within our research group initially leaned towards deterministic policy gradients, largely due to their historical efficacy in continuous control tasks. However, rigorous analysis of the temporal difference errors during pilot simulations revealed a critical limitation. The interplay between the discretized spatial topology and the continuous critic networks induced severe value overestimation, frequently culminating in catastrophic policy collapse.[9][29] This initial algorithmic stagnation forced a fundamental re-evaluation of our approach, propelling us toward stochastic policy optimization frameworks, which possess inherently more robust clipping mechanisms to constrain destructive policy updates.

Yet, directly applying standard stochastic multi-agent frameworks to spatial search tasks completely overlooks the geometric reality of the environment. To rectify this, we designed an architecture embedded with a spatial-temporal attention mechanism. Drawing methodological inspiration from the noise-robust frameworks utilized in efficient vision-language pre-training [27] and the complex feature extraction techniques found in massive multi-task spatial reasoning models [26], we deliberately enforce spatial locality within the actor network. By feeding the cropped localized belief map through convolutional layers, the agent is compelled to derive its policy not from a flattened vector of global noise, but from the topological gradients of its immediate surroundings. This architectural adjustment ensures that the latent representations remain deeply anchored to the physical constraints of the UAV's perceptual boundaries.

3.2 State Tensor Construction and the Economics of Reward Shaping

The structural integrity of this decentralized algorithm rests almost entirely on the delicate formulation of its state space and the subsequent calibration of its localized reward mechanisms. Our approach requires navigating the profound friction between sparse environmental feedback and the necessity for continuous gradient updates. In our architecture, the local observation is an asymmetrical tensor meticulously designed to capture both deterministic kinematics and probabilistic cognition.

Table 1. Semantic Composition of the Local Observation Tensor

State Component	Physical Interpretation	Functional Purpose in Policy Network
Kinematic State	Current planar coordinates, heading orientation, and velocity vector.	Provides bounding variables to ensure aerodynamic feasibility and prevent out-of-bounds navigational errors.
Local Belief Map	Cropped spatial uncertainty matrix extracted from the localized field of view.	Translates epistemic environmental uncertainty into a mathematically tractable visual matrix for the convolutional layers.
Neighbor Topology	Relative distance and heading angular differences to physically observable peers.	Embeds inter-agent spatial relationships to facilitate implicit collision avoidance and formation coherence.
Temporal History	Sequential window of historical control actions executed by the agent.	Mitigates the detrimental effects of partial observability by providing a brief temporal memory of momentum.

The most computationally sensitive phase of our research involved engineering the composite reward structure. The fundamental difficulty lies in the profound sparsity of target discovery events. If agents are exclusively rewarded upon mission success, the probability of random exploration generating a positive gradient approaches zero. We can conceptualize the UAV's spatial

attention as a finite resource; allocating this attention across an unknown grid requires optimization strategies conceptually analogous to how contemporary enterprises dynamically allocate cross-border marketing budgets or optimize return on investment via data-driven models.[17][23][28] The swarm must sequentially prioritize sub-regions that promise the highest informational yield. Consequently, we engineered a dense composite reward structure that penalizes topological redundancy while incentivizing continuous entropy reduction.

Table 2. Empirical Calibration Data of Reward Component Weights

Component Factor	Calibrated Weight Value	Calibration Rationale and Empirical Observations
Target Discovery	1.00	Primary mission objective scalar; structurally maintained at the global maximum to ensure policy convergence toward actual mission success rather than perpetual exploration.
Collision Penalty	0.40	Crucial for physical viability; modest values encourage tight formation, whereas excessive values induce artificial swarm dispersion and structural breakdown.
Entropy Reduction	0.25	Sustains continuous frontier exploration; over-weighting this factor paradoxically leads to agents ignoring actual targets in favor of sweeping empty spaces for immediate micro-rewards.
Spatial Redundancy	0.15	Penalizes overlapping perceptual fields among neighbors; essential for forcing the swarm to autonomously divide and conquer the operational theater.

4. Empirical Evaluation and Cognitive Dynamics Analysis

4.1 The Plateau Phenomenon and Structural Unlearning

Simulation evaluations were conducted within a highly parallelized computing environment to analyze the learning dynamics of the proposed architecture. Tracking the mean episodic cumulative reward over millions of environmental interactions, the learning curve exhibited an intriguing, highly non-linear trajectory that challenges idealized views of reinforcement learning. The data revealed a distinct, prolonged "plateau" phase during the mid-stages of training. Initially, we hypothesized this stagnation was indicative of vanishing gradients within the deep neural layers. However, subsequent layer-wise gradient norm analyses empirically refuted this assumption.

A more nuanced interpretation—paralleling the extraction of explainability from complex latent factor models via factorization trees [24] or the deliberate separation of joint and individual components in high-dimensional sparse regression [19], suggests that this plateau represents a critical phase transition in multi-agent cognition. During this extended period, the UAVs were actively "unlearning" independent greedy behaviors. While greedy heuristics yield high short-term entropy reduction, they fail structurally at scale. The agents were wrestling with the spatial redundancy penalty, sacrificing immediate local rewards to discover symmetric, non-overlapping spatial partitioning strategies. This structural reorganization inherently involves a temporary stagnation in the reward curve, a phenomenon that underscores the arduous process of synthesizing global coherence from localized optimization.

4.2 Comparative Efficacy and Edge Computing Trade-offs

To establish the relative efficacy of our proposed architecture, we benchmarked it against a suite of conventional paradigms, ranging from purely heuristic approaches to foundational multi-agent reinforcement learning baselines.

Table 3. Statistical Performance Comparison Across Algorithmic Baselines (Averaged over 500 Monte Carlo Episodes)

Algorithm Paradigm	Mean Search Steps to Coverage Threshold	Target Discovery Rate (%)	Spatial Redundancy Index	Computational Overhead (ms/step)
Random Walk	1450	42.5	3.84	0.8
Greedy Heuristic	680	78.1	2.15	4.2
Independent PPO	890	65.4	2.91	12.5
MADDPG Baseline	510	88.3	1.42	18.4
STA-MAPPO (Ours)	425	94.6	1.08	16.7

The performance data detailed in the comparative evaluation must be interpreted with academic caution. While our proposed architecture undeniably achieves the lowest mean search duration and minimal structural overlap, we cannot definitively rule out that this performance delta is, to some extent, an artifact of the specific grid dimensions utilized during the training phase. Interestingly, the greedy heuristic, despite its short-sighted nature, demonstrates a remarkably low computational overhead. This observation points toward a critical, often ignored trade-off in the deployment of autonomous systems. For ultra-short duration missions where onboard processing power is severely bottlenecked, the necessity for low-overhead task scheduling on multi-core edge chips might prioritize energy efficiency over optimal trajectory planning.[3][4][16] In such severely constrained hardware environments, purely learning-based approaches might induce latency that completely negates their strategic benefits.

Systemic Resilience under Degraded Topological Constraints

Most contemporary multi-agent literature implicitly assumes pristine, uninterrupted intra-swarm communication during execution. Real-world operations, heavily affected by physical occlusion or adversarial jamming, frequently violate this sanitized assumption. To assess true operational viability, we subjected our pre-trained policies to a simulated communication degradation environment, systematically increasing the packet loss rate. When an agent fails to receive state updates from its topological neighbors, the internal temporal memory modules are forced to propagate the last known hidden state, acting as a cognitive bridge.

Table 4. System Robustness Analysis under Varying Communication Packet Loss Rates

Communication Packet Loss Rate (%)	Target Detection Probability (%)	Swarm Collision Rate (%)	Mission Failure Rate (%)
0 (Idealized Environment)	94.6	0.2	1.5
20 (Mild Latency/Noise)	89.2	1.1	4.2

Communication Packet Loss Rate (%)	Target Detection Probability (%)	Swarm Collision Rate (%)	Mission Failure Rate (%)
50 (Severe Packet Drop)	71.5	4.8	18.5
80 (Near Isolated State)	45.3	12.4	44.0

The degradation profile presented in this analysis is highly illuminating. Under mild communication noise, the system demonstrates notable resilience, an effect likely attributable to the temporal smoothing provided by the historical action tensors. However, as the packet loss crosses the halfway threshold, the performance drop ceases to be linear and becomes catastrophic [18]. The agents regress to semi-independent local scanning, and their spatial partitioning capabilities severely deteriorate. This catastrophic structural decoupling implies that the cooperative policy learned by our architecture remains highly sensitive to the structural integrity of the neighbor topology graph. This inherent vulnerability leads us to further thinking regarding the critical necessity of developing fundamentally asynchronous, event-triggered communication protocols in future architectures [30], mirroring the resilient, decentralized governance and incentive structures observed in mature Web-based decentralized autonomous organizations[5][15][22]. Relying on continuous state broadcasting is a luxury that physical operational theaters rarely afford.

5. Conclusion

Synthesizing the empirical evidence and architectural iterations, the endeavor to orchestrate autonomous swarm intelligence within highly stochastic environments exposes a profound theoretical friction between localized perception boundaries and global objective optimization. By formally embedding the epistemic uncertainty into the individual agent's Markov state tensor and introducing the STA-MAPPO architecture, we catalyzed an emergent spatial partitioning behavior, yet this convergence is possibly inextricably linked to the specific density of our composite reward heuristic rather than representing a universal resolution to the non-stationary moving-target problem. Reflecting upon the catastrophic performance regression observed under severe communication degradation, it becomes mathematically evident that the current decentralized execution paradigm remains structurally tethered to a fragile assumption of quasi-synchronous topological graphs. Furthermore, the grid-based spatial quantization, while an indispensable stabilizing constraint, highly likely masks the aerodynamic complexities of actual fixed-wing flight dynamics.[7][13] Consequently, the trajectory for subsequent research must unequivocally pivot towards foundational paradigm shifts; a mathematically rigorous transition from discrete grid-based models to continuous, implicit neural representations constitutes an imperative next step, alongside the formulation of fundamentally asynchronous, event-triggered communication protocols that empower agents to autonomously optimize the actual semantic latency of their latent state transmissions.[3][16] Ultimately, bridging the chasm between simulated Markov environments and the highly adversarial reality of physical operations will require a sustained, interdisciplinary exploration.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The author(s) declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] *Hwang, S., Lee, H., Park, J., & Lee, I. (2022). Decentralized computation offloading with cooperative UAVs: Multi-agent deep reinforcement learning perspective. IEEE Wireless Communications, 29(4), 24-31.*

[2] *Kouzeghar, M., Song, Y., Meghjani, M., & Bouffanais, R. (2023). Multi-target pursuit by a decentralized heterogeneous uav swarm using deep multi-agent reinforcement learning. arXiv preprint arXiv:2303.01799.*

[3] *Hao, Z. (2026). Energy Efficient Multi Core Task Scheduling for Real Time Edge AI Systems: A Latency Aware Approach. International Journal of Advance in Applied Science Research, 5(3), 1-14.*

[4] *Hao, Z. (2026). Low-Overhead Scheduling for Real-Time AI Workloads on Multi-Core Edge Chips. International Journal of Advance in Applied Science Research, 5(3), 15-25.*

[5] *Lin, A. (2026). Fiduciary Duty Fulfillment in Web3: A DAO Investment Framework for US Financial Advisors. International Academic Journal of Social Science, 2, 17-26.*

[6] *Wu, Y. (2025). The Impact of "Data-Driven Hierarchical Operation" on ARPU Value for Cross-Border E-Commerce Warehousing Clients. Journal of Progress in Engineering and Physical Science, 4(6), 15-21.*

[7] *Wang, J., Chang, Y., Cao, S., Dong, Y., Li, S., Jia, L., & Li, W. (2025). Explanatory framework of typhoon extreme wind speed predictions integrating the effects of climate changes. Climate Dynamics, 63(3), 142.*

[8] *Wang, C. (2025). Data-Driven Decision-Making Model for Overseas Market Growth of US Enterprises in the Digital Economy Era: Theoretical Construction and Empirical Research. Journal of World Economy, 4(6), 58-65.*

[9] *Ekechi, C. C., Elfouly, T., Alouani, A., & Khattab, T. (2025). A survey on UAV control with multi-agent reinforcement learning. Drones, 9(7), 484.*

[10] *Ramezani, M., & Atashgah, M. A. (2026). CFR-MARL: Centralized Feedback-Driven Reward Multi-Agent Reinforcement Learning for Decentralized Cooperative Path Planning of Heterogeneous Agents. Acta Astronautica.*

[11] *Wu, Y. (2026). A Study on the Impact of Cross-Departmental Data Collaboration on Marketing Campaign Efficiency in Fast-Moving Consumer Goods E-commerce: The Case of PepsiCo (China)'s 7UP and Mirinda Project. Frontiers in Management Science, 5(1), 7-12.*

[12] *Wang, H., Li, Q., & Liu, Y. (2022). Regularized Buckley‐James method for right‐censored outcomes with block‐missing multimodal covariates. Stat, 11(1), e515.*

[13] *Wang, J., Tim, K. T., Li, S., Chan, T. K., & Fung, J. C. (2023). A systematic comparison of the wind profile codifications in the Western Pacific Region. Wind & structures, 37(2), 105-115.*

[14] *Wang, H., Li, Q., & Liu, Y. (2023). Adaptive supervised learning on data streams in reproducing kernel Hilbert spaces with data sparsity constraint. Stat, 12(1), e514.*

[15] Lin, A. (2026). Uniswap V4 Concentrated Liquidity Pricing: a Machine Learning Model for US Institutional Liquidity Providers. Journal of Intelligence and Engineering Technology, 1(1), 19-26.

[16] Hao, Z. (2026). Dynamic Task Prioritization for Edge AI in Smart Cities: Balancing Latency and Energy Efficiency. Journal of Intelligence and Engineering Technology, 1(1), 60-69.

[17] Wang, C. (2026). A Study on Data-Driven Budget Optimization for US Enterprises' Cross-Border Marketing. Frontiers in Management Science, 5(1), 41-46.

[18] Hao, Z. (2025). Fault-Tolerant Real-Time Scheduling for Edge AI in US Critical Infrastructure. Engineering Frontiers, 1(4).

[19] Wang, P., Wang, H., Li, Q., Shen, D., & Liu, Y. (2024). Joint and individual component regression. Journal of Computational and Graphical Statistics, 33(3), 763-773.

[20] Yehoshua, R., Heredia-Juesas, J., Wu, Y., Amato, C., & Martinez-Lorenzo, J. (2021). Decentralized Reinforcement Learning for Multi-Target Search and Detection by a Team of Drones. arXiv preprint arXiv:2103.09520.

[21] Wu, Y. (2025). Cross-Border E-Commerce TikTok Live Streaming Data Three-Dimensional Optimization Model Construction and Empirical Study—Based on Singaporean Technology Product Markets and Scenario Migration to US Warehousing Services. Journal of World Economy, 4(6), 44-50.

[22] Lin, A. (2025). Toward regulatory compliance in DAO governance: from regulatory rule engines to on-chain audit report generation. Journal of World Economy, 4(6), 12-20.

[23] Wang, C. (2025). Research on the Precision Allocation of Cross-Border Marketing Resources of US Enterprises Driven by Digital Technology. Innovation in Science and Technology, 4(11), 7-13.

[24] Tao, Y., Jia, Y., Wang, N., & Wang, H. (2019, July). The fact: Taming latent factor models for explainability with factorization trees. In Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval (pp. 295-304).

[25] Pendleton, S. D., Andersen, H., Du, X., Shen, X., Meghjani, M., Eng, Y. H., ... & Ang Jr, M. H. (2017). Perception, planning, control, and coordination for autonomous vehicles. Machines, 5(1), 6.

[26] Jin, Y., Li, Z., Zhang, C., Cao, T., Gao, Y., Jayarao, P., ... & Yin, B. (2024). Shopping mmlu: A massive multi-task online shopping benchmark for large language models. Advances in Neural Information Processing Systems, 37, 18062-18089.

[27] Tao, Y., Wang, Z., Zhang, H., Wang, L., & Gu, J. (2025, July). Nevlp: Noise-robust framework for efficient vision-language pre-training. In International Conference on Intelligent Computing (pp. 74-85). Singapore: Springer Nature Singapore.

[28] Wu, Y. (2026). Research on Dynamic Prediction Model of Brand Marketing Content ROI Based on Machine Learning. International Journal of Advance in Applied Science Research, 5(2), 31-38.

[29] Frattolillo, F., Brunori, D., & Iocchi, L. (2023). Scalable and cooperative deep reinforcement learning approaches for multi-UAV systems: A systematic review. Drones, 7(4), 236.

[30] Timoudas, T. O., Zhang, S., Magnússon, S., & Fischione, C. (2023). A general framework to distribute iterative algorithms with localized information over networks. IEEE Transactions on Automatic Control, 68(12), 7358-7373.