Research Article

Navigating Environmental Stochasticity in UAV Autonomous Flight: A Curriculum-Enhanced Deep Reinforcement Learning Framework and Sim-to-Real Considerations

Papadopoulos Demetriou[1,*]
Institute of Computer Science, LMU Munich, Germany
* Corresponding author: Papadopoulos Demetriou
PapadopoulosDemetriou@gmail.com

Abstract: The deployment of Unmanned Aerial Vehicles in highly nonlinear and dynamic environments is subject to the profound structural limitations inherent in traditional heuristic-based obstacle avoidance paradigms; specifically, these conventional methods inevitably encounter severe computational bottlenecks when processing high-dimensional perceptual data streams. In an effort to transcend these deterministic constraints, this study investigates the efficacy of incorporating "soft action evaluation" into reinforcement learning architectures. During the initial training phase, the extreme sparsity of feasible trajectories within obstacle-dense scenarios led to severe policy collapse. To mitigate this algorithmic stagnation, we embedded a multi-modal feature extraction network, integrated with a curriculum learning mechanism designed to deliberately grade environmental complexity to stabilize the gradients of the value network. Given these observed empirical discrepancies, it becomes imperative to move beyond the mere practice of injecting superficial noise into sensor inputs. This necessitates a fundamental redefinition of robust control representations—implying, furthermore, that to fully validate these data-driven navigation behaviors within the adversarial context of the real physical world, further research into asynchronous action execution is required.

Keywords: *Deep Reinforcement Learning; Autonomous Navigation; Continuous Action Space; Curriculum Learning; Sim-to-Real Transfer;*

1. Introduction

The integration of autonomous unmanned aerial vehicles (UAVs) into highly stochastic, unstructured environments represents a critical evolutionary leap in contemporary robotics. Operating within complex scenarios. Such as subterranean exploration, dynamic urban search and rescue, or low-altitude logistical routing—these systems are inextricably confronted with the profound challenge of executing real-time spatial reasoning. While previous intrinsic approaches have made significant theoretical strides in the formation control of regular polyhedra for reduced attitudes[1][3], these methodologies often assume a sanitized topological space. The true operational requirement is fundamentally complicated by severe computational constraints and unpredictable topological interference, necessitating a delicate balance between algorithmic complexity and physical survivability.

1.1 Computational Bottlenecks at the Edge

This balancing act is largely dictated by the physical limitations of onboard edge-computing hardware. Contemporary robotic systems cannot rely on infinite, cloud-based computational resources when navigating adversarial, communication-denied environments. Consequently, achieving spatial autonomy requires energy-efficient, multi-core task scheduling architectures that are highly latency-aware [2]. Just as complex AI inference workloads demand low-overhead scheduling on heterogeneous edge

chips to maintain real-time responsiveness[4][23], the autonomous flight controller must rigorously prioritize sensory processing tasks to prevent catastrophic control delays.

1.2 Structural Vulnerabilities of Classical Planners

Historically, the autonomous navigation paradigm has been overwhelmingly dominated by deterministic, model-based control architectures. A critical examination of classical local planning literature reveals a heavy reliance on methodologies like Artificial Potential Fields and the Dynamic Window Approach. These classical algorithms assume a high degree of environmental predictability, establishing a foundational mathematical baseline for spatial coverage that is increasingly showing its age. They rely on the continuous gradient descent of synthetically constructed repulsive fields; a rigid formulation that inherently traps the agent in local minima when confronting concave obstacle geometries or dense, overlapping obstacle clusters.

1.3 The Impact of Unmodeled Environmental Dynamics

Recognizing the extent to which these rigid deterministic models falter requires us to look at the unmodeled environmental dynamics they habitually ignore. Extreme regional wind speed variations, complex local climate projections, and intricate wake oscillator dampening physics significantly alter the aerodynamic envelope of small-scale UAVs[7][12][13]. When classical planners attempt to execute aggressive, high-speed reactive avoidance, they fail to account for these fluid dynamics, resulting in severe trajectory tracking errors. This realization indicates that attempting to rigidly model every physical perturbation is a computationally intractable pursuit, pushing the academic trajectory fundamentally toward data-driven, learning-based control representations.

The subsequent transition towards Deep Reinforcement Learning attempted to bypass these localized mathematical bottlenecks by allowing the agent to implicitly map high-dimensional sensory inputs directly to continuous control outputs. Early adoptions within the UAV domain primarily utilized value-based methods with discrete action spaces. However, the forced discretization of the aerodynamic control envelope inherently yields oscillatory, jerky flight trajectories. These aggressive, non-smooth control signals are fundamentally incompatible with the continuous, coupled rotational dynamics of multi-rotor platforms, inducing immense mechanical stress on the actuators.

Our preliminary research attempts to seamlessly implement continuous control architectures were fraught with severe algorithmic instability. We observed a recurring, catastrophic policy collapse when the UAV was initialized in obstacle-dense simulations. The extreme sparsity of viable navigational trajectories provided insufficient positive temporal difference errors to guide the value network out of its initial chaotic state. During these early exploratory phases, the agent would frequently plunge into terminal collision states before extracting any meaningful spatial representation, a phenomenon that halted our research progress and forced a critical re-evaluation of our approach.

We quickly realized that allocating the spatial attention of a UAV across an unknown grid can be methodologically viewed through the lens of data-driven resource optimization. Similar to how modern enterprises construct data-driven decision-making models to optimize overseas market growth [8] or precisely allocate cross-border marketing budgets[14][19], the UAV must dynamically prioritize spatial sub-regions that promise the highest informational return. Evaluating the dynamic impact of specific exploratory heuristics shares deep methodological roots with assessing how data-driven hierarchical operations dictate average revenue per user in cross-border e-commerce [6], or how cross-departmental data collaboration maximizes operational efficiency [10].

Consequently, this paper formally casts the continuous navigation challenge as a Markov Decision Process and investigates a Soft Actor-Critic architecture enhanced by curriculum learning. Testing the limits of the agent's ability to navigate conflicting reward signals parallels recent evaluations of large language models attempting to follow complex, entangled instruction hierarchies.[20][22] The agent must prioritize critical survival instructions over exploration commands when these epistemic directives inherently collide. We hypothesize that the structural inefficiencies of standard continuous DRL stem from abrupt exposure to

high-dimensional spaces without a structured foundation. It is highly possible that introducing a progressive training mechanism significantly alleviates this extreme credit assignment problem, laying the groundwork for resilient spatial cognition.

2. System Modeling and Environmental Abstraction

2.1 Kinematic Formulations and Rigid Body Abstractions

Establishing a mathematically rigorous foundation for multi-rotor flight inherently necessitates stripping away profound aerodynamic complexities to achieve computational tractability. In this study, the quadrotor is conceptually formalized as a rigid body operating within a three-dimensional Euclidean space. While the full non-linear dynamic equations define the true physical boundaries of the platform, training a reinforcement learning agent to simultaneously stabilize high-frequency motor mixing and execute low-frequency path planning creates an unmanageable hierarchical disparity in the learning gradient.

2.2 The Embedded Reality Gap

Consequently, we abstract the underlying dynamics by assuming the presence of a robust inner-loop attitude controller. The reinforcement learning agent operates exclusively in the outer loop, outputting continuous planar velocity commands. It is vital to acknowledge a potential bias embedded in this formulation: by assuming instantaneous velocity tracking, we are implicitly ignoring the complex actuator latency and aerodynamic drag coefficients that govern real-world flight. This theoretical abstraction, while necessary for stabilizing the initial learning phases, inherently injects a structural discrepancy, the "reality gap" that severely complicates subsequent physical deployment and trajectory execution.

2.3 Exteroceptive Perception Modeling

Navigating beyond foundational kinematics, the accurate conceptualization of the exteroceptive payload constitutes the core of the perception module. We equip the simulated UAV with a two-dimensional planar light detection and ranging sensor, providing an array of sequential distance measurements. Diverging from the idealized binary detection assumptions prevalent in purely theoretical treatises, advanced modeling frameworks must eventually integrate the inherent randomness and scattering effects of physical optical sensors.

2.4 Handling Stochastic and Missing Data

Treating these sensory inputs as pristine vectors harbors a latent structural bias. In heavily occluded environments, the sensor array frequently returns missing or heavily corrupted data points. Processing these highly sparse, noise-ridden observational streams presents computational hurdles analogous to multi-response regression techniques designed for block-missing multi-modal data without explicit imputation [21]. The neural architecture must inherently possess the capacity to infer structural continuity despite receiving fragmented geometric inputs, maintaining a coherent internal belief state of the immediate surroundings.

To prevent catastrophic system latency during these continuous cognitive updates, processing this data cannot be treated as a monolithic operation. It is imperative to implement task affinity-aware scheduling on the multi-core edge devices embedded within autonomous vehicles [15]. By dynamically aligning specific neural network inference operations with designated heterogeneous cores, the system minimizes cache misses and memory transfer bottlenecks, constantly balancing energy efficiency against the strict real-time operational demands of high-speed collision avoidance.

Compelled by the operational necessity to quantify the unobservable true state of the environment, decentralized architectures require the agent to maintain a continuously evolving cognitive representation. During the rigorous simulation phases of our architectural development, we confronted a pervasive algorithmic anomaly. 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 network to subsequent dynamic changes in that sector, a pathological condition we identify as "cognitive lock in."

To rectify this inherent mathematical artifact, we were forced to inject heuristic information diffusion mechanisms, applying a subtle mathematical pull back toward a state of maximum entropy. Synthesizing these compromised kinematic, perceptual, and cognitive models, the contemporary consensus rigorously formalizes the autonomous navigation endeavor as a Markov Decision Process. This formalization not only serves as a robust mechanism for distributing iterative algorithms, but it also accurately mirrors the fragmented reality of machine intelligence, reframing the ultimate systemic objective as the sequential execution of policies that systematically minimize cognitive uncertainty.

To govern these complex, decentralized interactions without relying on a central computational node, we draw inspiration from the structured governance found in decentralized autonomous organizations. Ensuring robust compliance and trust within the agent's internal state evaluation is mathematically analogous to developing regulatory rule engines for on-chain audit reports [18] or structuring compliant digital asset custody for traditional financial institutions [5]. The agent must autonomously audit its own sensory inputs, effectively functioning as a decentralized trust matrix that dynamically values its epistemic certainty before committing to high-risk kinemetic actions.

3. Architectural Design of the Distributed Reinforcement Learning Framework

Transitioning from the abstract Markov formulation to a computationally tractable neural architecture requires navigating a labyrinth of competing algorithmic paradigms. Early experimental iterations within our research group predominantly utilized deterministic policy gradients, largely inspired by their historical prevalence in continuous robotic control. However, a rigorous tracking of the temporal difference errors during our pilot simulations revealed a critical, inherent vulnerability. The deterministic nature of the policy gradient structurally failed to maintain adequate exploration variance when the UAV was initialized in highly constrained topologies.

3.1 Overcoming Premature Policy Exploitation

This lack of exploratory momentum consistently forced the value networks to converge prematurely, culminating in catastrophic policy collapse where the agent repetitively executed high-velocity collisions. Observing this persistent algorithmic stagnation, we pivoted the architectural foundation toward the Soft Actor-Critic framework. This transition was not merely a superficial substitution of algorithms but a fundamental shift in our optimization philosophy. By augmenting the standard reinforcement learning objective with a maximum entropy term, the architecture explicitly incentivizes the agent to explore all viable state-action trajectories that yield equivalently optimal rewards.

3.2 Topological Attention via Convolutional Feature Extraction

Directly applying this stochastic framework to raw spatial data, however, completely overlooks the geometric reality of the sensory inputs. To address this, we deliberately embedded a specialized one-dimensional convolutional neural network ahead of the multi-layer perceptrons. This architectural choice forces the network to extract local spatial dependencies and topological gradients from the sequential rangefinder arrays. It effectively translates raw distance metrics into a robust latent representation of the surrounding obstacle geometry, anchoring the abstract mathematical optimization to the physical topology of the environment.

3.3 Unlocking Latent Representation Explainability

Understanding how these convolutional layers interpret topological threats is critical for system verification. Attempting to untangle the black-box nature of the feature extractor shares deep methodological similarities with taming latent factor models for explainability via factorization trees [24]. By visualizing the activation maps of the convolutional layers, we observed that the network autonomously learned to identify critical structural features, such as sharp convex corners and narrow corridor entries, assigning higher activation weights to these complex geometries than to flat, unthreatening walls.

3.4 State Tensor Construction and Partial Observability

The structural integrity of this algorithm rests almost entirely on the delicate formulation of its state space. In our architecture, the local observation is an asymmetrical tensor meticulously designed to capture both deterministic kinematics and probabilistic cognition. It integrates the current planar coordinates, the cropped spatial uncertainty matrix, and a sequential window of historical control actions. This temporal history is an indispensable architectural component designed to mitigate the detrimental effects of partial observability, providing the agent with a brief, implicit memory of its own directional 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. Evaluating the efficiency of these dense reward heuristics structurally parallels the construction of multi-dimensional optimization models for TikTok live streaming data [17] or analyzing how LinkedIn data-driven operations impact the brand exposure of AI startups [26]; the agent must maximize its "exposure" to undiscovered frontier regions while systematically minimizing interaction with threatening topological features.

To bridge the profound cognitive gap caused by sparse rewards, we implemented a structured curriculum learning mechanism. When the fully initialized neural network was abruptly exposed to maximum density obstacle environments, the agent essentially flailed randomly until terminal collision, rendering the extracted gradients functionally useless. Rather than expecting the network to immediately master chaotic spatial realities, we deliberately staged the environmental complexity, starting from completely empty spaces and progressively introducing static, and eventually dynamic, geometric constraints.

This progressive curriculum, however, was far from a smooth, linear ascent. During the transition from purely static obstacles to environments containing dynamic entities, we observed a severe degradation in previously acquired avoidance behaviors. To partially mitigate this classic manifestation of catastrophic forgetting, we had to continuously inject historical, simpler environmental configurations into the advanced training stages. This process enforced a diverse sampling distribution within the replay buffer, sustaining the agent's structural memory and ensuring that fundamental repulsive behaviors were not overwritten by higher-order temporal predictions.

4. Empirical Evaluation and Cognitive Dynamics Analysis

4.1 Orchestrating the Simulation Environment

Simulation evaluations were systematically conducted utilizing highly parallelized physical engines integrated with robotic middleware to deeply analyze the learning dynamics of the proposed architecture. This environment allowed us to rapidly iterate through millions of environmental interactions without the risk of hardware attrition. Tracking the mean episodic cumulative reward over these vast training epochs, the resulting learning curve exhibited an intriguing, highly non-linear trajectory that challenges idealized, monotonically increasing views of continuous reinforcement learning.

4.2 Unveiling the Learning Plateau Phenomenon

The empirical 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. The gradients were active; the agent was simply not improving its primary objective score. A more nuanced interpretation suggests that this plateau represents a critical phase transition in machine cognition, marking a period of intense internal structural reorganization rather than passive algorithmic stagnation.

4.3 Structural Unlearning and Spatial Reorganization

During this extended plateau period, the UAVs were actively "unlearning" independent greedy behaviors. While greedy heuristics yield high short-term entropy reduction, they fail structurally at scale in complex maze-like topologies. The agent was wrestling

with the spatial redundancy penalty and collision constraints, sacrificing immediate, localized rewards to discover symmetric, non-overlapping spatial partitioning strategies. This phenomenon underscores the arduous process of synthesizing global coherence from localized optimization; the agent must briefly accept a lower holistic score while it re-wires its topological understanding.

4.4 Comparative Efficacy Against Classical Baselines

To establish the relative efficacy of our proposed architecture, we benchmarked it against a suite of conventional paradigms, including the Artificial Potential Field method and the Dynamic Window Approach. The empirical performance data reveals highly divergent behavioral characteristics. Classical reactive planners exhibited severe systemic failures in clustered environments; their fundamental reliance on synthetically generated repulsive forces predictably trapped the agent in unresolvable local minima, leading to extensive spatial oscillation and mission timeouts.

Conversely, the learning-based architectures demonstrated a superior capacity to navigate concave topological structures. However, we must critically interpret the seemingly dominant success rate of our proposed Soft Actor-Critic architecture. While it undoubtedly achieves the lowest collision metric, we cannot definitively rule out the possibility that this performance delta is, to some extent, an artifact of the agent heavily memorizing the specific spatial distributions inherent to the bounding constraints of the training grid, rather than possessing a truly generalized semantic understanding of unprecedented geometric threats.

Furthermore, most contemporary deep reinforcement learning literature implicitly assumes pristine, uninterrupted sensory streams during policy execution. To assess true operational viability, we subjected our pre-trained policies to a simulated environmental degradation test, systematically injecting varying standard deviations of Gaussian noise into the sensor arrays. The resulting degradation profile provides a sobering counter-narrative. As the noise variance escalates into the severe threshold, the performance drop ceases to be graceful. The sharp increase in the calculated jerk cost indicates that the control network begins to rapidly oscillate the velocity commands in response to ghost obstacles.

This pronounced vulnerability implies that the learned policy remains highly sensitive to the structural integrity of its perceptual inputs. To safely execute end-to-end planning of aerial robots in physical reality [25], bridging this gap requires robust methodologies. Recent surveys on sim-to-real methods highlight the profound challenges and emerging prospects of utilizing large foundation models to ground simulated behaviors [28]. It is becoming increasingly clear that a policy trained exclusively in a sanitized simulation is structurally incapable of generalizing to the atmospheric and sensory chaos of the physical world [9].

4.8 Advanced Real-to-Sim-to-Real Methodologies

To overcome these barriers, the academic community is transitioning towards more cyclical training paradigms. Leveraging real-world data to inform the simulation, which in turn teaches the agent how to explore the real world, creates a robust feedback loop [16]. Reconciling physical reality through simulation utilizing a real-to-sim-to-real approach presents a highly promising pathway for achieving robust manipulation and navigation [29]. This methodology acknowledges that the simulation is perpetually imperfect, utilizing it instead as a cognitive sandbox to stress-test the agent against heavily randomized, adversarial domain parameters before physical deployment.

5. Conclusion

Synthesizing the rigorous architectural iterations and the sobering anomalies unraveled during the simulation phases, the endeavor to orchestrate autonomous flight within highly unstructured environments exposes a profound theoretical friction. This friction exists fundamentally between the sanitized, simulated Markov representations upon which reinforcement learning algorithms thrive, and the adversarial, inherently non-Markovian reality of physical operations. While the implementation of the curriculum-enhanced framework successfully catalyzed robust spatial avoidance behaviors in simulation, the catastrophic

performance regression observed under severe sensory degradation mathematically dictates a harsh truth: current deterministic perception paradigms remain structurally tethered to a fragile assumption of perfect environmental observation.

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] *Zhang, S., He, F., Hong, Y., & Hu, X. (2020). An intrinsic approach to formation control of regular polyhedra for reduced attitudes. Automatica, 111, 108619.*

[2] *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.*

[3] *Zhang, S., Song, W., He, F., Hong, Y., & Hu, X. (2018). Intrinsic tetrahedron formation of reduced attitude. Automatica, 87, 375-382.*

[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. (2025). Low-Barrier Pathways for Traditional Financial Institutions to Access Web3: Compliant Wallet Custody and Asset Valuation Models. Frontiers in Management Science, 4(6), 80-86.*

[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] *Jonnarth, A., Johansson, O., Zhao, J., & Felsberg, M. (2025). Sim-to-real transfer of deep reinforcement learning agents for online coverage path planning. IEEE Access.*

[10] *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.*

[11] 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.

[12] Wang, J., Kudagama, B. J., Perera, U. S., Li, S., & Zhang, X. (2025). Framework for generating high-resolution Hong Kong local climate projections to support building energy simulations. Physics of Fluids, 37(3).

[13] Liu, Z., Jin, C., Li, S., Li, W., & Wang, J. (2024). Improvement for modeling the damping of the wake oscillator based on the Van der Pol scheme. Physics of Fluids, 36(7).

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

[15] Hao, Z. (2025). Task Affinity-Aware Scheduling for Multi-Core Edge Devices in Autonomous Vehicles. Engineering Frontiers, 1(2).

[16] Wagenmaker, A., Huang, K., Ke, L., Jamieson, K., & Gupta, A. (2024). Overcoming the sim-to-real gap: Leveraging simulation to learn to explore for real-world rl. Advances in Neural Information Processing Systems, 37, 78715-78765.

[17] 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.

[18] 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.

[19] 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.

[20] Zhang, Z., Li, S., Zhang, Z., Liu, X., Jiang, H., Tang, X., ... & Jiang, M. (2025). IHEval: Evaluating language models on following the instruction hierarchy. arXiv preprint arXiv:2502.08745.

[21] Wang, H., Li, Q., & Liu, Y. (2024). Multi-response Regression for Block-missing Multi-modal Data without Imputation. Statistica Sinica, 34(2), 527.

[22] Han, C. (2025). Can Language Models Follow Multiple Turns of Entangled Instructions?. arXiv preprint arXiv:2503.13222.

[23] Hao, Z. (2026). Structure-Aware Deep Reinforcement Learning for Latency-Minimal Scheduling of Edge AI Inference on Heterogeneous Cores. Journal of Intelligence and Engineering Technology, 1(1), 50-59.

[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] Ugurlu, H. I., Pham, X. H., & Kayacan, E. (2022). Sim-to-real deep reinforcement learning for safe end-to-end planning of aerial robots. Robotics, 11(5), 109.

[26] Wu, Y. (2026). Research on the Impact of LinkedIn Business Account Data-Driven Operations on Brand Exposure of AI Startups—A Case Study of AristAI. International Academic Journal of Social Science, 2, 27-37.

[27] Lin, A. (2026). Multi-Chain DAO Treasury Management: a Risk and Compliance Optimization Framework for the US Ecosystem. Journal of Intelligence and Engineering Technology, 1(1), 11-18.

[28] Da, L., Turnau, J., Kutralingam, T. P., Velasquez, A., Shakarian, P., & Wei, H. (2025). A survey of sim-to-real methods in rl: Progress, prospects and challenges with foundation models. arXiv preprint arXiv:2502.13187.

[29] Torne, M., Simeonov, A., Li, Z., Chan, A., Chen, T., Gupta, A., & Agrawal, P. (2024). Reconciling reality through simulation: A real-to-sim-to-real approach for robust manipulation. arXiv preprint arXiv:2403.03949.