Transactions on Computational and Scientific Methods | Vo. 5, No. 9, 2025
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Unified Modeling of Autonomous Exploration and Goal Discovery for Open-Environment Agents

Yicheng Chen[1], Shunqi Liu[2], Chen Qian[3]
[1]Ohio State University, Columbus, USA
[2]University of Southern California, Los Angeles, USA
[3]Brown University, Providence, USA
[*]Corresponding Author: Chen Qian; qianc1992@gmail.com

Abstract: This paper addresses the absence of explicit goal specification and the tendency toward fragmented exploration in agents operating within open environments. It proposes a unified modeling approach that integrates autonomous exploration and goal discovery. Goals are treated as internal and evolving variables and are incorporated together with environmental states into the decision process. Exploration, therefore, no longer depends on external task constraints or static reward design. Instead, it is organized around continuous generation and revision of goal structures. At the modeling level, a goal update mechanism captures how interaction experience shapes goal representations. Environmental feedback and intrinsic drivers are coordinated within a unified optimization framework. This enables joint evolution of exploration direction and decision preference. Systematic evaluation based on comparative results shows that the unified approach improves task completion and goal coverage in open environments. It also leads to more focused exploration with stronger structural consistency. Overall, the findings confirm that explicitly introducing goal discovery into exploration modeling enhances long-term decision stability and behavioral organization. The approach provides a feasible modeling strategy for building agents with intrinsic goal-driven mechanisms in open environments.

Keywords: Autonomous exploration, goal discovery, open environment, unified modeling

1. Introduction

As artificial intelligence systems move from closed task settings toward open real-world environments, agents no longer operate under predefined objectives and fixed rules. Instead, they must continuously perceive, decide, and act in highly uncertain and dynamically evolving contexts. In such environments, goals are often not explicitly specified at the outset. They may gradually emerge through interaction, or be reshaped as conditions change [1]. This shift fundamentally challenges traditional agent modeling paradigms that rely on prior goals, static rewards, or clear task boundaries. It also redirects research focus from completing known objectives to discovering meaningful goals and conducting autonomous exploration in unknown environments.

Autonomous exploration and goal discovery constitute the core capabilities required for long-term autonomy in open environments. Exploration enables agents to move beyond limited local experience and actively encounter potentially valuable states and behaviors [2]. Goal discovery determines how exploratory outcomes are structured, abstracted, and transformed into persistent drivers of decision making. Without effective goal discovery, exploration can degenerate into aimless random behavior and fail to support stable learning. If goals are overly dependent on external specification or short-term feedback, the agent's understanding of deeper environmental structure is constrained. Exploration and goal discovery are therefore tightly coupled processes that jointly shape the evolution of agent cognition and behavior.

Existing studies often treat exploration and goal modeling as separate components [3]. Exploration is typically enhanced through extrinsic incentives, intrinsic rewards, or uncertainty-driven signals, while goals are assumed to be fixed or stage-wise optimization targets. This fragmented modeling strategy simplifies analysis but struggles to capture how goals are generated, revised, and discarded in open environments. It also provides a limited explanation of how agents gradually form behavior directions with semantic coherence and temporal continuity from raw interactions. In complex settings, environmental feedback is often sparse, delayed, and ambiguous. Local signals alone are insufficient to support stable long-horizon decision-making. This makes unified modeling of exploration behavior and goal structure increasingly necessary.

At a higher level, unified modeling of autonomous exploration and goal discovery concerns how agents construct internal representations of the environment under incomplete information. It also concerns how evolving decision preferences are formed on this basis. These issues directly affect learning efficiency, behavioral stability, and long-term adaptability [4]. A unified framework allows exploration to be viewed as a dynamic process of expanding and reorganizing the goal space, rather than simple state coverage. It also treats goals as experience-driven and revisable mediating variables, rather than externally imposed endpoints. This perspective offers a more natural theoretical basis for continual learning, policy transfer, and cross-context generalization.

From an application perspective, agents with autonomous exploration and goal discovery capabilities are critical for complex systems and intelligent decision making in open environments [5, 6]. In domains such as autonomous systems, intelligent scheduling, long-term interactive services, and complex virtual worlds, it is impractical to enumerate all possible tasks and objectives in advance. Heavy reliance on manual design is costly and inflexible. A unified modeling approach enables agents to identify valuable behavioral directions during operation and to continuously refine their decision structures. This can substantially improve autonomy, scalability, and long-term stability [7]. Systematic study of unified modeling for autonomous exploration and goal discovery, therefore, holds significant theoretical importance and provides a key foundation for advancing intelligent systems toward higher-level autonomous cognition and decision making.

Recent research on intelligent decision systems further indicates that autonomous goal discovery should be modeled together with collaborative reasoning and adaptive operational control. Multi-agent stock trading driven jointly by collaborative decision making and large language models shows how agents can coordinate heterogeneous observations, high-level reasoning, and action selection under complex market dynamics [8]. Learning-based intelligent agents for backend resource scheduling similarly emphasize that operational goals should be adjusted according to changing resource states and decision constraints [9]. Retrieval-augmented long-text reasoning with semantic conflict detection and cross-paragraph evidence integration provides another methodological reference for maintaining coherent goal representations when agents must synthesize long-horizon observations and conflicting environmental cues [10].

Robust exploration also depends on the agent's ability to recognize abnormal events, maintain stable representations, and adapt policy behavior across distributed environments. Log semantic graph construction and event anomaly detection demonstrate that structured relations among events can improve the interpretation of changing system states [11]. Mamba-LSTM collaborative modeling for multi-table ETL anomaly detection extends this view by combining sequence modeling with cross-source process information, which is useful for representing goal-relevant signals in non-stationary environments [12]. Reinforcement learning-based decision modeling for large-scale distributed systems highlights the value of learning policies that adapt to system-level feedback rather than relying on static rules [13]. Collaborative scheduling for multi-tenant distributed inference services further supports the need to coordinate agent decisions under shared service constraints [14]. Stable fine-tuning guided by semantic energy also suggests that controllable representation adjustment can reduce distributional drift and improve the reliability of goal-conditioned generation [15]. These studies collectively support the proposed view that autonomous exploration and goal discovery should be unified through evolving representations, intrinsic signals, and adaptive policy optimization.

2. Proposed Framework

In an open environment, to achieve a unified modeling of autonomous exploration and goal discovery by an agent, this paper formalizes the interaction process between the agent and the environment as a continuous-time decision-making process. The environmental state is represented as s_t. At time t, the agent takes action a_t and transitions to the next state s_{t+1}. Unlike traditional settings, the goal is not externally given but is modeled as a latent variable g_t within the agent, characterizing the behavioral direction that the agent deems "worth pursuing" at the current stage. The overall state representation is extended to a joint state z_t, defined as:

$$z_t = (s_t, g_t)$$

This representation allows the objective to be incorporated into the decision-making context, influencing strategy selection in conjunction with the environmental state, thereby providing structural guidance for exploratory behavior. This paper also presents the overall model architecture, as shown in Figure 1.

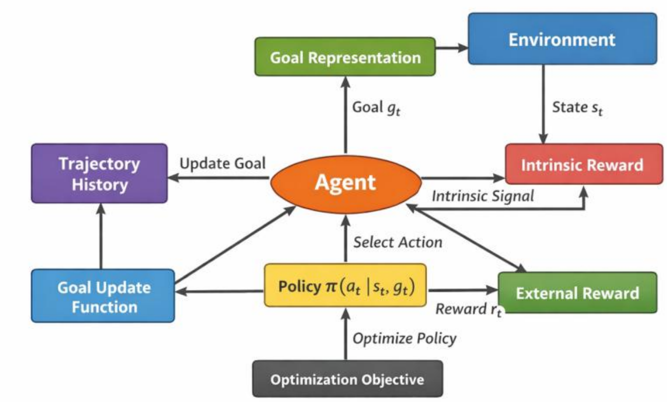

Figure 1. Overall model architecture

Based on this, the agent's policy is defined as a conditional policy, which is modeled on the joint state space:

$$\pi(a_t/z_t) = \pi(a_t/s_t, g_t)$$

The target variable g_t is not static but is continuously updated based on interaction experience. To characterize the dynamic evolution of the target, a target update function $G(\bullet)$ is introduced, which generates a new target representation based on historical trajectory information:

$$g_{t+1} = G(g_t, s_t, a_t, s_{t+1})$$

This update mechanism makes the target flexible, allowing it to be gradually adjusted as the environment changes and experience accumulates, thereby avoiding the decision failure problem caused by fixed targets in non-stationary environments.

To coordinate exploration and target discovery within a unified framework, this paper introduces an intrinsic driving signal to measure the information gain of state transitions in the target space. An intrinsic evaluation function r_t^{int} is defined as the norm of the target representation change:

$$r_t^{int} = //g_{t+1} - g_t//$$

This signal reflects whether the current interaction has facilitated an effective update of the target structure, thus guiding the agent to prioritize exploring behavioral paths that can enrich or modify the target's cognition. Therefore, exploration is no longer a blind overlay of the state space, but a directed process unfolding around the evolution of the target structure.

At the decision optimization level, the agent balances environmental feedback and intrinsic motivation by maximizing long-term cumulative utility. The comprehensive utility function is defined as follows:

$$J(\pi) = E[\sum_{t=0}^{\bowtie} \gamma^t (r_t^{ext} + \lambda r_t^{int})]$$

Where r_t^{ext} represents environmental feedback, γ is the discount factor, and λ is the intrinsic driving weight. This objective function unifies autonomous exploration and goal discovery into a single optimization objective, enabling the policy update process to consider both behavioral outcomes and goal structure evolution. Through the above modeling, the agent can form an experience-driven, continuously adjusting goal system in an open environment, and on this basis, achieve stable and efficient autonomous decision-making.

3. Experimental Analysis

3.1 Dataset

This paper uses Crafter as a unified evaluation data source for autonomous exploration and target discovery tasks in open environments. Crafter is an open-source, open-world interactive environment where agents continuously interact in a dynamic world comprised of terrain, resources, and survival constraints. Unlike traditional single-objective tasks, this environment simultaneously contains multiple achievable achievements and behavioral cues, and effective feedback is often sparse and has long-term dependency characteristics. Therefore, it is more suitable for research settings in open environments where targets are not entirely explicit and need to be gradually discovered and organized through interaction. The environment's code and runtime interface are publicly available, facilitating reproduction and supporting large-scale generation of interaction trajectories.

In terms of data organization, this paper records the interaction process between the agent and the Crafter environment as trajectory sequence data to uniformly characterize the coupling relationship between the exploration process and the target discovery process. Each trajectory consists of a time series (s_t, a_t, s_{t+1}) and environmental event records, where state s_t can be represented by observed images or structured features, action a_t Comes from a discrete action set, and environmental events reflect key behavioral consequences such as resource gathering, tool crafting, and changes in survival status. Because Crafter's open-world structure features diverse reachable sub-goals and a multi-path policy space, trajectory data naturally presents the entire process of goal generation, goal correction, and exploration strategy evolution, thus providing a consistent data carrier for unified modeling of "autonomous exploration and goal discovery."

From a thematic fit perspective, Crafter's core value lies in its ability to simultaneously provide the non-stationarity and long-range dependency characteristics of open environments, and to offer a sufficiently rich potential signal space for goal discovery through a multi-achievement structure. Agents must make choices under resource constraints and survival pressures, avoiding policy collapse due to short-sighted behavior while continuously expanding the reachability boundary through exploration, and forming an internal goal representation of "which behaviors are worth pursuing" in the process. This data and task structure directly supports the unified modeling setting of this paper, namely, viewing exploration behavior as a process of continuous evolution of the goal structure within the same framework, and using goal discovery as the intrinsic driving force for long-term autonomous decision-making.

3.2 Experimental Results

This article first presents the results of the comparative experiments, as shown in Table 1.

Table 1: Comparative experimental results

Method	TSR	GDC	EC	LTR
SciAgents [16]	0.843	0.742	18.6	142.7
M3builder [17]	0.861	0.758	17.9	136.4
OmniNova [18]	0.875	0.771	17.2	131.8
G-designer [19]	0.889	0.789	16.6	126.9
Autodefense [20]	0.902	0.803	16.1	123.4
Ours	0.936	0.842	15.0	115.2

From an overall perspective, performance differences among methods in open environments are not reflected by improvements in a single metric. They instead arise from systematic differences in how goal discovery and exploration are organized. Our method shows concurrent advantages in TSR and GDC. This indicates that behavior sequences are more effectively guided toward verifiable task completion. It also shows stronger capability in identifying and stabilizing multiple potential goals during interaction. This observation aligns with the unified modeling principle emphasized in the paper. Autonomous exploration and goal discovery are embedded within a single decision loop. As a result, the agent can form clearer internal objectives under non-predefined goal settings. This contributes to improved long-term autonomy and scalability.

At the same time, the overall trends of EC and LTR reflect a structural trade-off that is more consistent with open environment characteristics. Compared with baseline methods, our approach achieves lower EC. This suggests that exploration is not driven by indiscriminate expansion of the visited space. Instead, it is organized around discovered goal structures with more directed extension and efficient coverage. Such focused exploration can be understood as the effect of goal representations constraining and guiding exploration strategies. Exploration shifts from random probing to goal-centered active information acquisition. This helps avoid high coverage with low effective return in open worlds.

The reduction in LTR does not necessarily indicate weaker long-term capability. It may instead reveal differences in optimization preferences over reward signals. In open environments, external rewards are often sparse and strongly context dependent. Excessive emphasis on short-term cumulative return can induce reliance on locally high reward trajectories. This can limit the expansion of the goal space. The stronger performance in TSR and GDC suggests that our method prioritizes the extraction of transferable goal structures and behavioral sub-skills from interaction. These are converted into more stable achievement capabilities. In this sense, the relative decrease in LTR reflects a strategic allocation favoring goal discovery and task completion, rather than a reduction in adaptability to open environments.

Taken together, these comparative results support the core claim of the paper. In open environments, the key challenge is not simply to expand exploration or maximize reward. It is to construct a unified mechanism that continuously generates and updates goals. Exploration then becomes controllable, interpretable, and accumulative under this mechanism. Baseline methods may maintain strong performance on certain dimensions. However, they rely more on fixed preferences or local heuristics. Our method demonstrates stronger systemic capability. It transforms goal discovery into a sustained driver of decision-making. This leads to advantages in overall autonomy and task achievement that better match the demands of open environments.

The target update frequency determines how quickly an agent re-evaluates and modifies its internal target representation during interactions, thus affecting the organization of exploration paths and the long-term consistency of decisions. By conducting a single-index sensitivity analysis on the target update frequency, we can more clearly characterize the boundary and trade-offs of this mechanism's effect on different capability dimensions in an open environment. The experimental structure is shown in Figure 2.

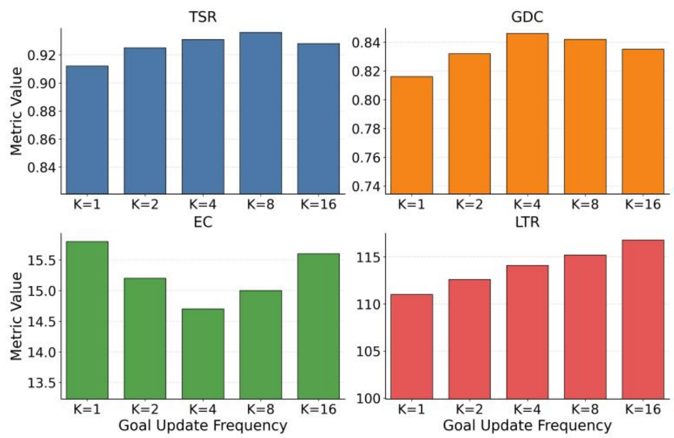

Figure 2. The impact of target update frequency on experimental results

Overall trends indicate that the goal update frequency exerts a systematic influence on how agent behavior is organized in open environments. As the update pace changes, policy performance does not evolve monotonically. Instead, different capability dimensions exhibit distinct responses to the chosen temporal scale. This observation suggests that goals are not static decision conditions. They must be updated at an appropriate granularity to effectively participate in exploration and decision-making. Such a setting supports the formation of relatively stable behavior patterns during long-term interaction. This is highly consistent with the unified modeling view that treats goals as internal regulatory variables.

For metrics related to goal achievement and goal discovery, moderate update frequencies appear more favorable for fully expressing agent capabilities. Excessively frequent goal revisions can interrupt established exploration directions. This prevents the policy from accumulating long-horizon experience. In contrast, overly slow updates may cause goals to lag behind environmental changes. This limits the agent's ability to perceive and absorb new potential goals. These observations indicate that the goal update mechanism balances exploration stability and adaptability. Its proper design is therefore a key prerequisite for efficient autonomous exploration.

Changes in exploration coverage and reward-related dimensions further show that goal update frequency affects both exploration breadth and decision depth. When updates are slower, policies tend to persist along existing goals. This supports coherent long-term behavior but may reduce proactive probing of new regions. Faster updates encourage more frequent restructuring of exploration paths. This increases behavioral diversity but can weaken long-term consistency. This trade-off highlights that exploration behavior is explicitly constrained by the speed of goal evolution. It is not driven solely by randomness or immediate feedback.

Taken together, these results show that goal update frequency is not an isolated engineering parameter. It directly shapes the cognitive structure of the agent. Its effects include filtering exploration directions, maintaining goal stability, and regulating long-term decision consistency. By explicitly incorporating and tuning this factor within a unified modeling framework, agents can avoid blind exploration and goal drift in open environments. This enables more coherent and sustainable autonomous behavior. It also provides further support for the necessity and rationality of unifying exploration and goal discovery within a single modeling framework.

Offline data quality noise can alter the reliability of an agent in extracting target cues and behavioral patterns from historical trajectories, thus affecting the coupling effect between target discovery and autonomous exploration within a unified modeling framework. By conducting single-index sensitivity analysis on offline data quality noise, we can more directly characterize the robustness boundary of the model to data perturbations and the differences in response across different capability dimensions. The experimental results are shown in Figure 3.

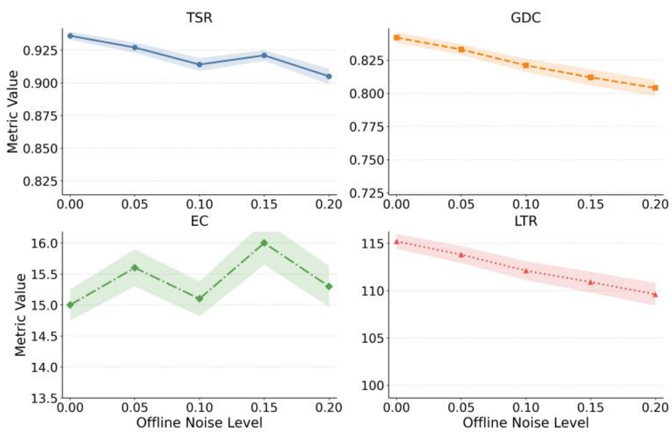

Figure 3. The impact of offline data quality noise on experimental results

Variations in different metrics under changing noise levels in offline data quality indicate that data reliability directly affects internal goal modeling and exploration organization. As noise increases, capabilities related to goal achievement and goal discovery show an overall weakening trend. This suggests that when historical interaction data contain substantial disturbance, agents struggle to stably extract generalizable goal cues from offline trajectories. This phenomenon indicates that, within a unified modeling framework, goal representations depend not only on online interaction feedback but also strongly on the structural consistency of offline experience. Noise perturbs the goal update process and indirectly degrades the quality of subsequent exploration decisions.

At the same time, changes in exploration coverage and long-term behavioral return reveal a more nuanced trade-off. Exploration-related metrics do not decline monotonically with increasing noise. Instead, they exhibit noticeable fluctuations. This reflects that agents actively adjust exploration strategies when facing unreliable data in order to compensate for insufficient goal information. Such adjustments, however, often come at the cost of decision consistency. Long-term behavior becomes more dependent on immediate feedback rather than stable goal structures. Overall, these results further support the central argument of this work. In open environments, autonomous exploration and goal discovery must be jointly modeled within a unified framework. Only with high-quality experience can goal mechanisms effectively guide exploration and maintain long-term decision stability.

4. Conclusion

This work focuses on the core problem of autonomous exploration and goal discovery for agents in open environments. It proposes a systematic approach that unifies these two processes within a single modeling framework. This approach moves beyond decision paradigms that rely on externally specified tasks and static goal assumptions. By treating goals as internal and evolving cognitive variables and explicitly embedding them into decision making and learning, this study offers a more general perspective for describing how agent behavior is organized during long-term interaction. Comparative results demonstrate that the unified framework effectively guides exploration from unstructured trial behavior toward goal-driven and organized evolution. As a result, agents exhibit more stable decision performance and stronger task completion capability in complex open environments.

From a methodological perspective, this study further clarifies the critical role of goal modeling in intelligent decision making under open environment conditions. Compared with strategies that depend solely on reward signals or exploration heuristics, explicit goal discovery provides agents with a medium to long-term behavioral reference. This allows decision consistency to be maintained under uncertainty and non-stationarity. By jointly considering exploration benefits and goal evolution within a unified optimization objective, agents gradually develop an internal understanding of environmental structure. Fragmented experiences are integrated into reusable behavior patterns. This mechanism improves policy stability and enhances adaptability to environmental change.

At the application level, the findings offer important implications for a range of long-horizon autonomous decision scenarios. In autonomous system operation, complex resource scheduling, continuous interactive services, and open world virtual environments, it is often impractical to enumerate all possible goals in advance. Excessive reliance on manual design is neither realistic nor flexible. The proposed unified modeling framework provides a feasible path for building self-guided intelligent systems. These systems can continuously identify, adjust, and reinforce their behavioral objectives during operation. This improves overall autonomy and operational efficiency. Such capability directly contributes to reducing human intervention costs and enhancing long-term system reliability.

Looking ahead, research on unified modeling of autonomous exploration and goal discovery retains substantial potential for further development. On one hand, hierarchical and abstract modeling of goal structures may deepen agent understanding of complex environments. This would support planning and decision-making over longer time scales. On the other hand, combining this framework with multi-agent cooperation, cross-environment transfer, or real-world constraints may promote a transition from individual autonomy to higher-level collective intelligence and social decision making. Overall, this study establishes an important foundation for building agents with continual learning capability and intrinsic goal-driven mechanisms in open environments. It also provides forward-looking theoretical support for intelligent upgrading in related application domains.

References

[1] Kim S, Lee K, Choi J. Variational curriculum reinforcement learning for unsupervised discovery of skills[J]. arXiv preprint arXiv:2310.19424, 2023.

[2] Bae J, Park K, Lee Y. Tldr: Unsupervised goal-conditioned rl via temporal distance-aware representations[J]. arXiv preprint arXiv:2407.08464, 2024.

[3] Matthews M, Beukman M, Ellis B, et al. Craftax: A lightning-fast benchmark for open-ended reinforcement learning[J]. arXiv preprint arXiv:2402.16801, 2024.

[4] Du Y, Kosoy E, Dayan A, et al. What can ai learn from human exploration? intrinsically-motivated humans and agents in open-world exploration[C]//Neurips 2023 workshop: Information-theoretic principles in cognitive systems. 2023.

[5] Agarwal S, Chuck C, Sikchi H, et al. A unified framework for unsupervised Reinforcement Learning algorithms[C]//Workshop on Reinforcement Learning Beyond Rewards@ Reinforcement Learning Conference 2025.

[6] Wang W W, Han D, Luo X, et al. Toward open-ended embodied tasks solving[J]. arXiv preprint arXiv:2312.05822, 2023.

[7] Janner M, Du Y, Tenenbaum J B, et al. Planning with diffusion for flexible behavior synthesis[J]. arXiv preprint arXiv:2205.09991, 2022.

[8] S. Sun, "Adaptive Stock Trading Algorithms Jointly Driven by Multi-Agent Collaborative Decision Making and Large Language Models in Complex Financial Market Environments," 2025.

[9] L. Ren, Y. Liu, and Y. Zhao, "Learning-Based Intelligent Agents for Backend Resource Scheduling and Operational Decision Making," 2025.

[10] S. Y. Huang, "Retrieval-Augmented Long-Text Reasoning with Semantic Conflict Detection and Cross-Paragraph Evidence Integration," 2024.

[11] Y. Yang and A. Xu, "Log Semantic Graph Construction and System Event Anomaly Detection via Convolutional Neural Networks," 2024.

[12] Y. Zhang and E. Cheng, "Anomaly Detection in E-Commerce Multi-Table ETL Processes through Mamba-LSTM Collaborative Modeling," 2025.

[13] R. Wei and L. Tan, "Reinforcement Learning-Based Intelligent Decision Modeling for Large-Scale Distributed Systems," 2024.

[14] Z. Huang, J. Luo, and C. Chen, "Learning-Driven Collaborative Scheduling for Multi-Tenant Distributed Inference Services," 2025.

[15] H. Zhuang and A. Shen, "Semantic Energy-Guided Stable Fine-Tuning for Distribution-Controlled Generation in Large Language Models," 2025.

[16] Ghafarollahi A, Buehler M J. SciAgents: automating scientific discovery through bioinspired multi-agent intelligent graph reasoning[J]. Advanced Materials, 2025, 37(22): 2413523.

[17] Feng J, Zheng Q, Wu C, et al. M3builder: A multi-agent system for automated machine learning in medical imaging[C]//International Workshop on Agentic AI for Medicine. Cham: Springer Nature Switzerland, 2025: 115-124.

[18] Du P. OmniNova: A General Multimodal Agent Framework[J]. arXiv preprint arXiv:2503.20028, 2025.

[19] Zhang G, Yue Y, Sun X, et al. G-designer: Architecting multi-agent communication topologies via graph neural networks[J]. arXiv preprint arXiv:2410.11782, 2024.

[20] Zeng Y, Wu Y, Zhang X, et al. Autodefense: Multi-agent llm defense against jailbreak attacks[J]. arXiv preprint arXiv:2403.04783, 2024.