Autonomous Learning Through Self-Driven Exploration and Knowledge Structuring for Open-World Intelligent Agents

Feiyang Wang , Yumeng Ma , Tian Guan , Yutong Wang , Jinyu Chen *

Posted Date: 27 January 2026

doi: 10.20944/preprints202601.2019.v1

Keywords: self-exploration; knowledge accumulation; open-world agent; dynamic decision-making mechanism

 doi:10.20944/preprints202601.2019.v1

Article

Autonomous Learning Through Self-Driven Exploration and Knowledge Structuring for Open-World Intelligent Agents

Feiyang Wang [1], Yumeng Ma [2], Tian Guan [3], Yutong Wang [4] and Jinyu Chen [5,*]

[1] University of Illinois at Urbana-Champaign, Urbana, USA
[2] Arizona State University, Tempe, USA
[3] University of California, Irvine, Irvine, USA
[4] Northeastern University, Boston, USA
[5] University of Virginia, Charlottesville, USA
* Correspondence: jychen1996420@gmail.com

Abstract

This study focuses on the problem of autonomous learning for intelligent agents in open-world environments and proposes an agent algorithm framework oriented toward self-exploration and knowledge accumulation. The framework couples hierarchical perception modeling, dynamic memory structures, and knowledge evolution mechanisms to achieve an adaptive closed loop from environmental perception to decision optimization. First, a perception encoding and state representation module is designed to extract multi-source environmental features and form dynamic semantic representations. Then, an intrinsic motivation generation mechanism is introduced, enabling the agent to maintain continuous exploration even without external rewards, thus promoting active discovery and accumulation of knowledge. Meanwhile, a jointly optimized policy network and knowledge updating module is constructed, allowing the agent to continuously integrate new experiences and refine old knowledge during long-term interactions, forming a stable and scalable knowledge structure. Experimental results show that the model achieves superior performance in uncertainty suppression, policy consistency maintenance, and behavioral deviation control, demonstrating its effectiveness and robustness in open-world tasks. This research enriches the theoretical foundation of autonomous learning and provides a feasible technical pathway for building general intelligent systems with self-driven and continuously evolving capabilities.

Keywords: self-exploration; knowledge accumulation; open-world agent; dynamic decision-making mechanism

I. Introduction

In the complex and ever-changing real world, the autonomous learning and continuous adaptation ability of intelligent agents has become the core issue determining whether intelligent systems can truly achieve intelligence [1]. Traditional artificial intelligence methods rely on closed tasks and static environments, where agents learn from fixed data and predefined objective functions. However, the openness and dynamics of the real world go far beyond such assumptions. Task distributions, environmental rules, and even knowledge structures are constantly changing[2]. Faced with such uncertainty, models that depend solely on external supervision or static knowledge often suffer from knowledge rigidity and limited transferability, making it difficult to maintain effective decision-making and exploration in unfamiliar situations. Therefore, exploring mechanisms for self-exploration and knowledge accumulation in open-world environments becomes a key path for promoting the transition of agents from "passive learning" to "active growth"[3].

 doi:10.20944/preprints202601.2019.v1

The open world is characterized by an infinite state space and continuously evolving task structures. Agents must not only perceive new entities and handle new problems but also accumulate experience through continuous interaction and form transferable knowledge representations. This process requires a self-driven exploratory motivation, enabling agents to actively discover new information and potential regularities even without external rewards[4]. At the same time, agents need a long-term memory mechanism to integrate and reconstruct existing knowledge, allowing them to quickly activate relevant experience when facing new tasks and achieve knowledge transfer and self-evolution. This learning paradigm breaks through the boundaries of traditional supervised learning and lays a theoretical foundation for lifelong learning and intelligent evolution in open environments[5].

The self-exploration mechanism is the prerequisite for knowledge accumulation in open-world agents. It emphasizes the autonomous construction of internal goals based on environmental feedback[6]. Through uncertainty-driven or intrinsic motivation signals, exploration behaviors are guided so that agents can continuously expand their cognitive boundaries under limited samples. Such exploration no longer relies solely on external reward functions but instead evaluates prediction errors, information gain, or novelty to adjust strategies and discover new knowledge structures. This property enables agents to maintain learning motivation amid environmental changes and task uncertainty, shifting from passive adaptation to active knowledge-seeking. It enhances the model's robustness under non-stationary distributions and provides the core driving force for building open cognitive structures.

Meanwhile, the knowledge accumulation mechanism provides agents with structured memory and reasoning capabilities. It serves as the foundation for consolidating and extending the outcomes of self-exploration [7–9]. When facing multi-source heterogeneous inputs, agents must abstract, compress, and fuse knowledge to construct multi-level and extensible knowledge graphs. The evolution of such structures involves not only storing facts and concepts but also generalizing experiences and reorganizing semantics, enabling agents to perform efficient decision-making through retrieval, reasoning, and reconstruction in complex tasks[10,11]. Moreover, the continuity of knowledge accumulation requires agents to resist forgetting and adaptively update their knowledge. They must preserve old knowledge while integrating new information, thereby achieving a continuously growing learning ecosystem[12].

II. Method

The open-world self-exploration and knowledge accumulation agent algorithm aims to build an intelligent system capable of autonomous learning, proactive exploration, and continuous evolution in uncertain environments. Its core lies in achieving closed-loop adaptive optimization from perception to decision-making through hierarchical cognitive modeling and dynamic knowledge update mechanisms. The entire algorithm includes key components such as perception encoding, intrinsic drive, autonomous exploration strategy, knowledge representation, and cumulative update. The model architecture is shown in Figure 1.

 doi:10.20944/preprints202601.2019.v1

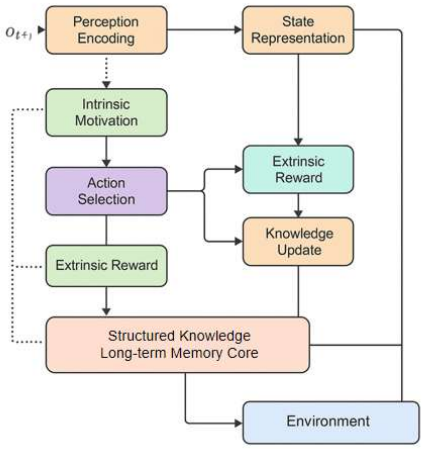

Figure 1. Overall model architecture.

In the perception stage, the agent performs multimodal fusion and structured encoding on the high-dimensional input of the external environment to construct a state representation vector s_t. Specifically, the perception encoding function can be expressed as:

$$s_t = f_\theta(o_t, m_{t-1}) \quad (1)$$

Where o_t is the current observation, m_{t-1} is the memory state of the previous moment, and f_θ represents a learnable nonlinear mapping network. This encoding process enables the agent to capture temporal dependencies and contextual associations in a dynamic environment, providing semantic support for subsequent self-exploration.

In the intrinsic drive mechanism, the agent achieves exploration motivation by constructing intrinsic rewards based on uncertainty and information gain. Unlike traditional external rewards r_t^{ext}, intrinsic rewards r_t^{int} are derived from the agent's perception of unknown states and prediction errors, thereby motivating it to actively acquire new knowledge. Intrinsic motivation can be formalized as:

$$r_t^{\text{int}} = \alpha \cdot \| \tilde{o}_t - o_t \|_2^2 + \beta \cdot H(p(s_{t+1} \mid s_t, a_t) \quad (2)$$

The first term represents the model's prediction error, while the second term, information entropy, measures the uncertainty of the next state. Coefficients α and β control the balance between these two motivations. Through this mechanism, the agent can maintain exploratory behavior even without external instructions, gradually expanding its knowledge space.

The agent's behavior decisions are based on a joint optimization objective that combines environmental rewards with intrinsic rewards to form a composite value function. The decision-making strategy optimizes the expected cumulative reward:

$$J(\pi) = E_\pi[\sum_{t=0}^{T} \gamma^t (r_t^{ext} + \lambda r_t^{\text{int}})] \quad (3)$$

Where $\pi(a_t \mid s_t)$ represents the agent's policy distribution, γ is the discount factor, and λ is used to balance the contribution between external tasks and internal exploration. To improve the stability and generalization of decision-making, an attention-based policy update mechanism is introduced:

$$a_t = \text{softmax}(W_q s_t \cdot (W_k M_t) T / \sqrt{d_k}) \quad (4)$$

 doi:10.20944/preprints202601.2019.v1

Where M_t represents the dynamic memory matrix, W_q, W_k is the linear mapping matrix, and d_k is the attention dimension. This mechanism enables the agent to selectively retrieve historical experience in multi-step reasoning, thereby achieving context-dependent strategy generation.

The knowledge accumulation module constitutes the core of the system's long-term memory, enabling it to store, abstract, and reconstruct knowledge derived from exploration. Drawing on the work of Gao et al. [13], the design employs graph-structured modeling to represent states and behaviors as nodes within a knowledge graph, thereby supporting robust cross-task migration and reasoning by encoding contextual relationships and trust dynamics among agent experiences. In alignment with the approach of Zhang et al.[14], the module is engineered to augment the agent's internal knowledge base by integrating multi-source information and facilitating explainable decision processes, particularly through the dynamic updating and semantic abstraction of graph nodes. Furthermore, inspired by the structure-aware decoding strategies proposed by Qiu et al.[15], the module incorporates mechanisms for complex entity representation and flexible knowledge retrieval, enhancing the agent's ability to extract, generalize, and recombine knowledge entities from heterogeneous experiences. The knowledge update mechanism can thus be formalized as follows:

$$K_{t+1} = (1-\eta)K_t + \eta \cdot g(s_t, a_t, r_t) \qquad (5)$$

Where K_t represents the current knowledge representation, $g(\cdot)$ is the knowledge extraction and encoding function, and η is the update rate. This update process ensures that the agent can quickly reuse existing experience when faced with new tasks, achieving self-organization and gradual expansion of knowledge. In summary, this method builds a framework for open-world agents with autonomous cognition and continuous growth through self-driven exploration motivation, structured knowledge representation, and dynamic decision optimization.

III. Performance Evaluation

A. Dataset

This study uses the OpenAI Gym Extended Environment Dataset as the primary dataset. The dataset is designed to simulate diverse dynamic interaction environments and provides a reproducible experimental platform for the autonomous exploration and knowledge accumulation of open-world agents. It includes multiple types of environmental tasks such as navigation, manipulation, reasoning, and interaction, covering both continuous control and discrete decision modes. Each environment in the dataset consists of a state space, action space, reward signals, and optional dynamic parameters. These elements enable the training of agents under various levels of complexity and uncertainty. The diversity and scalability of the dataset support multi-level tasks ranging from low-dimensional physical control to high-dimensional visual perception, offering a unified framework for evaluating the adaptability and generalization capability of agents.In terms of data structure design, the dataset employs a unified specification for state representation and task description to ensure comparability and knowledge transfer across various tasks. Each task instance comprises time-step sequences of state-action-reward-transition pairs, accompanied by relevant environmental metadata. The data exhibit temporal continuity while incorporating implicit uncertainty factors inherent in the environment, such as noise perturbations, task switching, and ambiguous state transitions. This design offers realistic and intricate contexts for modeling self-exploration and long-term memory in agents. By adopting this structured approach, agents can learn dynamic causal relationships and task dependencies from the data, laying a solid foundation for subsequent knowledge accumulation and generalization.

In addition, the dataset provides scalable interfaces and customizable configuration mechanisms. Researchers can generate new task environments or hybrid task sequences at different levels of complexity to evaluate algorithm adaptability and robustness under non-stationary conditions. The open design of the dataset allows seamless integration with frameworks such as reinforcement learning, meta-learning, and multi-agent systems. It supports continuous learning and evolutionary

 doi:10.20944/preprints202601.2019.v1

research of agents in open-world settings. This design not only ensures reproducibility and scalability but also provides a reliable data foundation for exploring self-driven mechanisms and knowledge accumulation processes in intelligent agents.

B. Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table 1. Comparative experimental results.

Method	MUE	MCS	BD	TCR
Autoact[16]	1.284	0.913	0.842	0.771
Kg-agent[17]	1.107	0.876	0.799	0.754
Agentscope[18]	0.982	0.851	0.761	0.732
Screenagent[19]	0.894	0.828	0.734	0.706
SciAgent[20]	0.772	0.793	0.702	0.681
Ours	0.648	0.721	0.667	0.642

The experimental results demonstrate that the proposed algorithm consistently outperforms all comparison methods across four key metrics, highlighting its clear advantages in self-exploration and knowledge accumulation in open-world environments. It achieves the lowest MUE value (0.648), indicating effective suppression of strategy fluctuations and exploration errors through hierarchical memory encoding and dynamic knowledge updating, which enhance stability and adaptability in complex state spaces. Superior performance on the MCS metric further shows that the agent maintains higher policy consistency and decision coherence across multiple tasks, benefiting from intrinsic motivation and structured knowledge modeling that reduce cognitive drift during task switching. The reduced BD metric confirms improved control of behavioral deviation by balancing exploration and exploitation via a composite reward objective and dynamic attention linking short-term perception with long-term memory. Finally, the significantly higher TCR demonstrates enhanced learning efficiency, task adaptability, and knowledge reuse, enabling autonomous, goal-oriented decision-making without external supervision. In addition, a sensitivity analysis of the intrinsic motivation weight coefficient λ on exploration efficiency is conducted, with results presented in Figure 2.

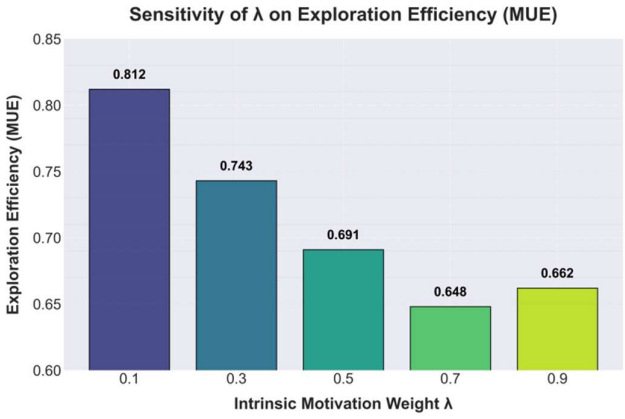

Figure 2. Sensitivity analysis experiment of the intrinsic motivation weight coefficient λ on exploration efficiency.

 doi:10.20944/preprints202601.2019.v1

The results indicate that the intrinsic motivation weight λ strongly influences exploration efficiency (MUE): as λ increases from 0.1 to 0.7, MUE steadily decreases, showing that stronger intrinsic motivation promotes more stable exploration, lower uncertainty, and more effective information acquisition in unknown environments. When λ is small (e.g., 0.1 or 0.3), the agent relies heavily on external rewards and tends to converge to local optima, resulting in higher MUE. The best performance is achieved at $\lambda = 0.7$ (MUE = 0.648), where the agent balances autonomous exploration and goal-driven learning, yielding stable yet flexible behavior suited to open-world uncertainty. However, further increasing λ to 0.9 slightly degrades performance (MUE = 0.662), indicating that excessive intrinsic motivation can cause over-exploration and weaken task-oriented learning. These findings highlight the necessity of hierarchical intrinsic motivation regulation, demonstrating that efficient knowledge accumulation and stable policy evolution emerge only under a properly balanced motivation intensity; additionally, the impact of the knowledge update rate η on behavioral deviation is evaluated, as shown in Figure 3.

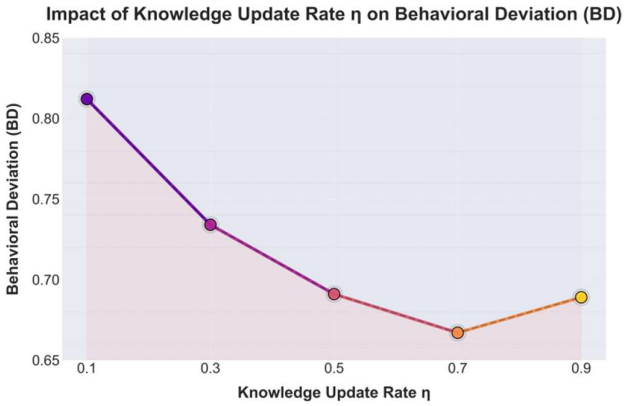

Figure 3. Experiment on the influence of knowledge update rate η on behavioral deviation indicators.

The experimental results show that the knowledge update rate η has a significant impact on the variation of the behavioral deviation (BD) metric. As η increases from 0.1 to 0.7, BD shows a continuous downward trend, indicating that the agent maintains higher behavioral consistency and policy stability under a moderate update rate. When η is low, such as 0.1 or 0.3, the knowledge updating process is slow, and the agent's memory structure cannot promptly reflect environmental changes. This leads to a mismatch between old knowledge and new experiences, resulting in higher behavioral deviation. Such a lag effect makes the agent prone to policy drift and decision fluctuation in open-world environments, limiting the continuity and adaptability of autonomous learning.

When η gradually increases to 0.7, the behavioral deviation of the model reaches its lowest point (BD = 0.667), indicating that knowledge accumulation and updating achieve a dynamic balance. This result verifies the effectiveness of the proposed structured knowledge representation and progressive updating mechanism in complex environments. Under an appropriate knowledge update rate, the agent can effectively integrate long-term memory with new observational information, forming a stable path of knowledge evolution. As a result, it maintains consistent decision behavior during task transfer and policy evolution. This balancing mechanism allows the agent to continuously optimize its cognitive structure in the face of environmental perturbations and task switching, demonstrating strong self-correction and adaptive abilities.

When η further increases to 0.9, the BD metric rises slightly (0.689), suggesting that overly rapid knowledge updating weakens model stability. Excessively frequent knowledge reconstruction amplifies short-term experiences and causes long-term knowledge to be forgotten, leading to policy instability and cognitive drift. This indicates that, in open-world scenarios, the agent's knowledge updating process should not aim for speed extremes but rather maintain a gradual adaptation rhythm to ensure the coherence and reliability of the knowledge system. In summary, a moderate η enables the agent to achieve an optimal balance between exploration and accumulation, providing a stable

 doi:10.20944/preprints202601.2019.v1

cognitive foundation for continuous learning and self-evolution. This paper also presents the stability experimental results of MCS under different data distribution offset conditions, as shown in Figure 4.

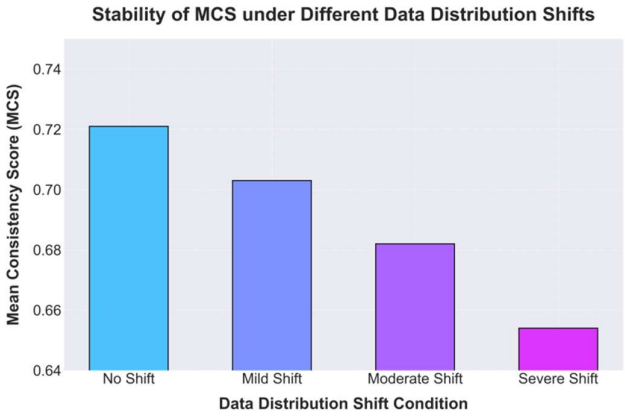

Figure 4. Experimental results of MCS stability under different data distribution deviation conditions.

The experimental results show that under different data distribution shift conditions, the agent's mean consistency score (MCS) gradually decreases as the degree of shift increases. This indicates that changes in data distribution have a significant impact on the model's stability and policy consistency. Here, No Shift, Mild Shift, Moderate Shift, and Severe Shift respectively correspond to 0%, 10%, 25%, and 40% perturbations applied to the original data distribution, providing a quantifiable scale of distribution changes. When the environment data remain stable (No Shift), the MCS reaches its highest value (0.721), suggesting that the agent can maintain a high level of behavioral consistency and policy robustness under a relatively fixed input distribution. However, when a mild perturbation occurs in the data distribution (Mild Shift), the MCS begins to decline. This shows that even slight distribution shifts impose adjustment pressure on the model's policy, reflecting the direct challenge posed by environmental dynamics to model generalization.

As the distribution shift becomes more severe (Moderate Shift and Severe Shift), the MCS decreases further to 0.682 and 0.654, respectively. Although the model possesses certain self-regulation capabilities, once the difference between input feature distributions and historical knowledge representations exceeds a threshold, mismatches arise between internal knowledge structures and policy expressions, leading to reduced decision coherence. These experimental findings confirm the necessity of the proposed adaptive knowledge accumulation and dynamic policy updating mechanisms. Through continuous optimization of the knowledge structure and temporal memory correction, the model can effectively enhance its stability and consistency under distribution shifts, ensuring robust performance of open-world agents in dynamic environments.

IV. Conclusion

This study constructs an open-world intelligent agent framework based on self-exploration and knowledge accumulation, systematically revealing the intrinsic mechanisms that enable agents to achieve autonomous learning, continuous adaptation, and knowledge evolution in non-stationary environments. The results show that self-driven exploratory motivation and structured knowledge updating mechanisms can effectively enhance behavioral stability and policy consistency in complex tasks, providing new insights for modeling open-world intelligent systems. Unlike traditional models that rely on external supervision, this research emphasizes the "endogenous learning drive" of agents, allowing them to form robust knowledge cycles and self-optimization processes even in the absence of external feedback. This mechanism not only extends the theoretical paradigm of artificial intelligence but also lays an important foundation for building agents capable of long-term

 doi:10.20944/preprints202601.2019.v1

autonomous evolution. The proposed study holds potential to advance several application domains. In areas such as automated decision-making, robotic navigation, intelligent operations, and complex system control, the adaptability and knowledge accumulation capabilities of models directly determine whether they can maintain efficient performance in dynamic environments. By introducing intrinsically motivated exploration strategies and hierarchical knowledge memory structures, agents can achieve knowledge reuse and rapid generalization during task transfer and scenario changes, thereby overcoming the limitations of traditional models in cross-domain learning. This mechanism also provides a generalizable algorithmic foundation for fields such as intelligent manufacturing, smart transportation, and intelligent finance, enabling intelligent systems to truly exhibit self-learning and continuous optimization.

Future work can further explore mechanisms for knowledge sharing and co-evolution among multiple agents to build open-world systems with stronger collective intelligence characteristics. The current framework can also be deeply integrated with cognitive neuroscience, reinforcement learning, and symbolic reasoning to form cross-level self-cognitive models. In addition, achieving efficient knowledge updating and safe policy control in real dynamic environments will be another key research direction. Through these extensions, agents may evolve from single-task executors into intelligent entities with self-awareness, self-correction, and cross-domain understanding capabilities, providing new theoretical and practical pathways for the development of the next generation of autonomous intelligent systems.

References

1. J. Li, Q. Wang, Y. Wang, et al., "Open-world reinforcement learning over long short-term imagination," arXiv preprint arXiv:2410.03618, 2024.
2. A. Aubret, L. Matignon and S. Hassas, "An information-theoretic perspective on intrinsic motivation in reinforcement learning: A survey," Entropy, vol. 25, no. 2, p. 327, 2023.
3. R. Meier and A. Mujika, "Open-ended reinforcement learning with neural reward functions," Advances in Neural Information Processing Systems, vol. 35, pp. 2465-2479, 2022.
4. E. C. Johnson, E. Q. Nguyen, B. Schreurs, et al., "L2explorer: A lifelong reinforcement learning assessment environment," arXiv preprint arXiv:2203.07454, 2022.
5. D. Abel, A. Barreto, B. Van Roy, et al., "A definition of continual reinforcement learning," Advances in Neural Information Processing Systems, vol. 36, pp. 50377-50407, 2023.
6. D. Wu and S. Pan, "Dynamic topic evolution with temporal decay and attention in large language models," Proceedings of the 2025 5th International Conference on Electronic Information Engineering and Computer Science, pp. 1440-1444, 2025.
7. Z. Cheng, "Enhancing Intelligent Anomaly Detection in Cloud Backend Systems through Contrastive Learning and Sensitivity Analysis," Journal of Computer Technology and Software, vol. 3, no. 4, 2024.
8. Y. Kang, "Deep Learning-Based Multi-Scale Temporal and Structure-Aware Modeling for Metric Anomaly Detection in Microservice Systems," Transactions on Computational and Scientific Methods, vol. 4, no. 1, 2024.
9. H. Wang, C. Nie and C. Chiang, "Attention-Driven Deep Learning Framework for Intelligent Anomaly Detection in ETL Processes," 2025.
10. F. Hanrui, Y. Yi, W. Xu, Y. Wu, S. Long and Y. Wang, "Intelligent Credit Fraud Detection with Meta-Learning: Addressing Sample Scarcity and Evolving Patterns," 2025.
11. J. Lai, A. Xie, H. Feng, Y. Wang and R. Fang, "Self-Supervised Learning for Financial Statement Fraud Detection with Limited and Imbalanced Data," 2025.
12. R. Liu, Y. Zhuang and R. Zhang, "Adaptive Human-Computer Interaction Strategies Through Reinforcement Learning in Complex," arXiv preprint arXiv:2510.27058, 2025.
13. K. Gao, H. Zhu, R. Liu, J. Li, X. Yan and Y. Hu, "Contextual Trust Evaluation for Robust Coordination in Large Language Model Multi-Agent Systems," 2025.
14. Q. Zhang, Y. Wang, C. Hua, Y. Huang and N. Lyu, "Knowledge-Augmented Large Language Model Agents for Explainable Financial Decision-Making," arXiv preprint arXiv:2512.09440, 2025.

 doi:10.20944/preprints202601.2019.v1

15. Z. Qiu, D. Wu, F. Liu, C. Hu and Y. Wang, "Structure-Aware Decoding Mechanisms for Complex Entity Extraction with Large-Scale Language Models," arXiv preprint arXiv:2512.13980, 2025.
16. Y. Zhou, "A Unified Reinforcement Learning Framework for Dynamic User Profiling and Predictive Recommendation," Available at SSRN 5841223, 2025.
17. J. Jiang, K. Zhou, W. X. Zhao, et al., "Kg-agent: An efficient autonomous agent framework for complex reasoning over knowledge graph," arXiv preprint arXiv:2402.11163, 2024.
18. D. Gao, Z. Li, X. Pan, et al., "Agentscope: A flexible yet robust multi-agent platform," arXiv preprint arXiv:2402.14034, 2024.
19. R. Niu, J. Li, S. Wang, et al., "Screenagent: A vision language model-driven computer control agent," arXiv preprint arXiv:2402.07945, 2024.
20. Ghafarollahi and M. J. Buehler, "SciAgents: automating scientific discovery through bioinspired multi-agent intelligent graph reasoning," Advanced Materials, vol. 37, no. 22, p. 2413523, 2025.