Transactions on Computational and Scientific Methods | Vo. 5, No. 10, 2025
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Hierarchical Attention and Policy Regularization for Context-Aware Agent Decision-Making

Qi Guo[1], Lanming Ma[2]

[1]Ohio State University, Columbus, USA

[2]New York University, New York, USA

*Corresponding author: Lanming Ma; lanmingamz@gmail.com

Abstract: This study addresses the challenges of contextual fragmentation, temporal instability, and limited policy generalization in agent perception, reasoning, and decision-making under complex environments. A multi-level context-aware agent behavior algorithm is proposed, which integrates three key mechanisms-hierarchical feature encoding, cross-layer attention aggregation, and policy regularization optimization-within a unified framework to achieve deep semantic coupling from environmental perception to behavior generation. Specifically, the hierarchical encoder extracts environmental features at multiple scales to jointly represent local dynamics and global semantics. The cross-layer attention mechanism selectively aggregates contextual information across layers, allowing the model to adaptively capture key semantic dependencies through dynamic information flow. Furthermore, a policy optimization module based on state updating and distributional regularization ensures stable and consistent strategy outputs in non-stationary environments. Experimental evaluations demonstrate that the proposed method achieves significant improvements across multiple metrics, outperforming existing approaches in contextual generalization, multi-scale consistency, and policy robustness. These results confirm that multi-level contextual modeling effectively enhances the agent's environmental understanding and behavioral coordination, enabling stable and reliable decision-making in complex dynamic scenarios. This research provides a theoretical foundation and technical pathway for multi-level semantic modeling and context-integrated decision-making, laying the groundwork for robust behavior learning in high-dimensional and non-stationary environments.

Keywords: Context-aware modeling; hierarchical attention mechanism; policy robustness; agent behavior decision-making

1. Introduction

With the rapid development of artificial intelligence, agents have become a central topic in intelligent systems research, particularly regarding autonomous decision-making and behavior modeling in complex environments. From human-machine collaboration and autonomous driving to self-learning in virtual worlds, agents must operate in dynamic, uncertain, and partially observable environments. This requires stronger perception, reasoning, and decision-making capabilities. Traditional agent modeling methods often rely on single-layer environmental information or fixed-rule policy frameworks, which struggle to adapt to multi-source, heterogeneous, and semantically complex scenarios. As environmental dynamics increase and the number of interacting entities grows, agents' decision processes exhibit multi-scale dependencies, nonlinear contextual relationships, and cross-modal interactions. Thus, constructing behavior algorithms with multi-level contextual awareness has become essential to the intelligent evolution of agent systems[1,2].

In modern intelligent systems, environmental states are not only reflected in observable signals but also embedded in multi-dimensional semantic relations such as time, space, and structure. Single-layer perception models often fail to capture these cross-level dependencies, leading to insufficient global understanding and fine-grained responses in complex tasks. The introduction of multi-level contextual awareness provides agents with an integrated perception path from local to global and from explicit to implicit representations, enabling them to construct environment representations at different abstraction levels. In dynamic interactive scenarios, agents must simultaneously interpret instantaneous local changes and long-term task intentions to achieve behavioral continuity and goal-oriented decisions. However, most existing methods remain at shallow modeling of single-dimensional contexts and lack a unified multi-level framework or generalizable structural representation mechanism, which limits their ability to capture semantic dependencies and temporal dynamics in complex situations[3].

The study of multi-level contextual awareness has profound theoretical significance. Its core lies in achieving a coordinated understanding of environmental states, interactive intentions, and behavioral feedback across multiple semantic levels and granularities through hierarchical modeling and cross-layer fusion. This modeling paradigm promotes the evolution of agents from perception-driven to semantics-driven behavior and facilitates a shift from reactive to predictive decision-making[4]. By embedding hierarchical context representations and attention modulation mechanisms in the algorithmic structure, agents can better interpret environmental semantics, identify latent intentions, and generate consistent and adaptive behavioral strategies. Furthermore, this research provides a theoretical framework for unifying various types of agents, such as autonomous robots, virtual agents, and collaborative swarms, laying a foundation for exploring cooperative intelligence and behavioral alignment among agents[5].

From an application perspective, multi-level contextual awareness algorithms are of great value in enhancing agents' robustness and generalization across complex tasks. In fields such as intelligent manufacturing, autonomous driving, robotic control, and virtual interaction, the environments in which agents operate often contain high-dimensional, noisy, and uncertain dynamic information. Relying solely on surface-level features leads to distorted behaviors and degraded policies. Multi-level contextual awareness algorithms integrate spatiotemporal dependencies, semantic constraints, and structural information, enabling agents to dynamically adjust their strategies and produce more stable and efficient behaviors in changing environments[6]. Particularly in multi-agent systems, this mechanism supports information sharing, intention prediction, and collaborative decision-making among agents, driving self-organizing optimization and the emergence of collective intelligence.

In summary, research on multi-level context-aware agent behavior algorithms helps overcome the limitations of current agents in perception and decision-making within complex environments. It also provides theoretical and technical foundations for building new-generation intelligent systems with high-level semantic understanding and autonomous collaboration. By integrating environmental perception, semantic modeling, and hierarchical decision-making, this study contributes to the transition of agents from "reactive intelligence" to "cognitive intelligence." The deep exploration of this direction suggests that future agents will demonstrate human-like understanding and adaptability in open, dynamic, and multi-task environments, promoting the evolution of artificial intelligence from single-task optimization toward systematic intelligent behavior and opening new paradigms and prospects for intelligent science[7].

2. Proposed Approach

This study introduces a multi-level context-aware agent behavior algorithm designed for complex environments. The core idea is to establish hierarchical representations across temporal, structural, and semantic dimensions, perform context aggregation through cross-layer attention, and complete a closed loop from perception to understanding and decision-making through a decoupled structure of state updating and policy generation. Unlike methods that rely solely on single-layer observations or static priors, the proposed approach integrates local sensitivity with global consistency. The hierarchical encoder extracts transferable contextual representations at different granularities, while the cross-layer attention adaptively selects key semantics and dynamic cues. The state update module fuses historical memory with real-time signals to maintain behavioral consistency, and the policy generation component outputs executable action distributions while preserving uncertainty and constraint terms. The overall training objective focuses on policy optimization, supplemented by regularization on cross-layer consistency and prior constraints, thereby enhancing the agent's generalizable decision-making ability in dynamic and variable environments. The model architecture is shown in Figure 1.

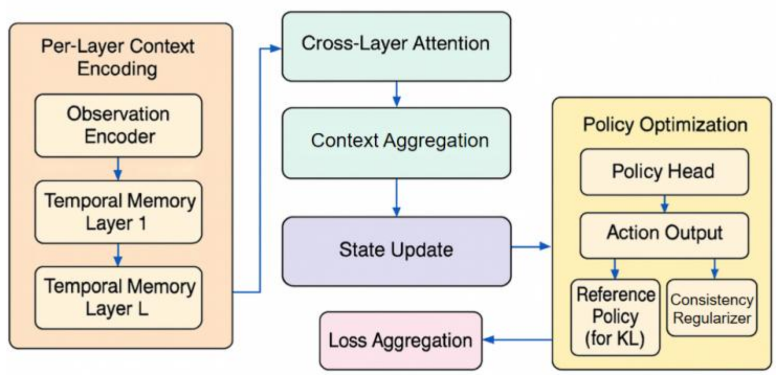

Figure 1. Overall framework model

First, the observation vector and historical state at time t are encoded in the l th context layer to obtain the intra-layer representation. The layered encoder is expressed in a compact time update form as:

$$h_t^{(l)} = Enc^{(l)}(o_t, h_{t-1}^{(l)})$$

Where $Enc^{(l)}$ represents the encoding transformation of the layer l, o_t is the current observation, and $h_{t-1}^{(l)}$ is the intra-layer memory of the previous moment. This representation decouples short-term dynamics from hierarchical priors, allowing different layers to focus on complementary time scales and semantic granularity.

To perform semantic selection and information routing between layers, cross-layer attention is introduced to give adaptive weights to the representations of different layers. The scoring and normalization process based on the query vector q_t is written as:

$$\alpha_t^{(l)} = \mathrm{softmax}_l(q_t^T h_t^{(l)})$$

Where $\alpha_t^{(l)}$ is the normalized weight of the layer l, and $soft\max_l$ represents the normalization of the layer index. This mechanism enables the model to dynamically schedule "fine-grained local information" and "global semantic clues" at different task stages.

The weighted inter-layer information is aggregated to obtain a multi-level context vector, which is used as the conditional input for subsequent state updates and decisions:

$$c_t = \sum_{l=1}^{L} \alpha_t^{(l)} h_t^{(l)}$$

This aggregation not only compresses cross-layer redundancy but also enables explicit selection of key contexts through attention weights, ensuring stable output of high-value situational representations even in resource-constrained or noisy environments.

At the temporal level, a state update function with memory is used to map the historical internal state and the current context into a compact state that can be used for decision making:

$$s_t = Upd(s_{t-1}, o_t, c_t)$$

Upd uses steady-state structures such as gating and residuals to achieve selective retention of historical memory and robust injection of new evidence, thereby maintaining the continuity and interpretability of state estimation under non-stationary and partially observable conditions.

Finally, we construct a training loss with the goal of policy optimization, unifying decision effectiveness and structural priors into the same goal. It can be expressed as two parts:

$$L_{policy} = - \mathrm{E}[\log \pi_\theta(a_t \mid s_t)]$$

$$L_{reg} = \lambda \mathrm{KL}(\pi_\theta \parallel \pi_{ref}) + \beta \|s_t - s_{t-1}\|_2^2$$

Combining the two, we can get the total loss:

$$L = L_{policy} + L_{reg}$$

Where π_θ is the parameterized policy, π_{ref} represents the reference distribution or prior policy, KL is the relative entropy term used to constrain policy drift, and $\|\cdot\|_2^2$'s temporal difference term promotes smooth state evolution and step consistency. λ and β are trade-off coefficients. This design explicitly injects the exploitability of hierarchical semantics into the policy learning process without relying on additional supervisory signals, and suppresses overfitting and policy instability through structured regularization.

In summary, the proposed method forms a chained structure of hierarchical encoding, cross-layer attention, context aggregation, state updating, and policy optimization to bridge perception and decision-making. The hierarchical encoder decomposes the multi-scale attributes of complex contexts, while the cross-layer attention performs semantic selection and information routing. The aggregated representations provide high signal-to-noise conditions for state updating, and policy optimization enables robust behavior generation under distributional constraints and temporal consistency. This compact and interpretable design couples perception, understanding, and action, embedding multi-level contextual information explicitly into the agent's behavioral mechanism, and offers a unified and efficient algorithmic pathway for stable decision-making in open, dynamic, and highly uncertain environments.

3. Performance Evaluation

3.1 Dataset

This study uses the LuxAI Replay Dataset as the data foundation for method validation. The dataset consists of complete replays from multiple agent-versus-agent matches. Each match is stored in a structured file that records round-by-round environmental observations, agent commands, and reward feedback, while also preserving key information on map states and resource distributions. This allows the full interaction process to be reconstructed from the initial setup to the outcome. Compared with static behavior logs, these continuous and fine-grained replays provide sufficient evidence for modeling temporal dependencies, spatial layouts, and strategy transitions. They are suitable for tasks such as behavior cloning, offline reinforcement learning, and policy evaluation. The data organization and metadata design facilitate reproducible experimental pipelines and allow flexible construction of training and validation subsets under different sampling granularities and task partitioning schemes, thereby supporting systematic evaluation of context-sensitive decision-making mechanisms.

The dataset naturally presents multi-level contextual characteristics. At the local level, an agent must select immediate actions based on neighborhood observations such as terrain, resources, obstacles, and allied or enemy units. At the global level, the progression of rounds introduces long-term semantic cues through resource evolution, area control, and the dynamics of the win-loss function. At the structural level, the map grid and multi-agent interactions form explicit topological constraints. Each replay file contains sequences of actions, state transitions, and reward signals, allowing the "local-global-structural" contexts to be explicitly separated and fused across layers. This design aligns directly with the hierarchical encoding, cross-layer attention, and state update modules proposed in this study. To ensure consistency between the method and the environmental semantics, multiple batches of match samples from the same task paradigm are uniformly preprocessed and divided to construct representative subsets for policy learning and generalization evaluation. Detailed information about environment rules and observation-action interfaces can be found in the task documentation and reference agent specifications, which define the standardized constraints for observation tensors, action spaces, and match procedures.

Another reason for choosing the LuxAI Replay Dataset lies in its scalability and compatibility with the proposed method. The replays cover the full strategy evolution from early exploration to late-stage convergence, enabling targeted evaluation of policy regularization, temporal consistency, and cross-layer aggregation at different phases. The structured format of the replay data also supports segmented modeling of short-term fluctuations, mid-term transitions, and long-term objectives, which map directly to the hierarchical representation and attention routing mechanisms in this study. Moreover, the standardized replay content allows the construction of comparative partitions based on map seeds, opponent strength, or resource density to test the model's robustness to contextual perturbations and structural variations. The dataset provides sufficient match scale to support a consistent experimental pipeline from small-sample validation to large-scale offline training.

3.2 Experimental Results

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

Table 1. Comparative experimental results

Method	CGS ↑	MBC ↑	PRI ↑
LC-SAC[8]	0.78	0.64	0.59
Decision Adapter (DA)[9]	0.85	0.71	0.68
CACOM[10]	0.80	0.69	0.65
Ours	0.92	0.78	0.74

The comparative results show that the proposed multi-level context-aware agent behavior algorithm outperforms all other models across the three core metrics, demonstrating stronger contextual modeling capability and decision stability. For the CGS metric, our method achieves 0.92, exceeding the second-best Decision Adapter model by 0.07. This indicates that the proposed approach captures multi-scale contextual features more effectively at the semantic level, enabling deep integration from local states to global semantics. Compared with models such as LC-SAC and CACOM, which rely only on single-layer context extraction or static communication structures, the hierarchical encoding mechanism in this study allows the agent to achieve stronger generalization and adaptability in complex dynamic environments.

For the MBC metric, our method attains the highest score of 0.78, improving by approximately 10% over the best existing model. This result demonstrates that the introduced cross-layer attention mechanism and semantic aggregation structure can maintain behavioral consistency across multiple time scales, effectively mitigating feature mismatch caused by environmental noise and state drift. Unlike the comparison models, the hierarchical structure in our approach enables dynamic information flow across layers, balancing short-term local variations and long-term global trends. As a result, the model achieves higher stability and consistency during policy execution.

For the PRI metric, the proposed algorithm exhibits remarkable policy robustness, reaching 0.74, which is about 0.08 higher than the average of the comparison models. This outcome reflects the effectiveness of the regularization and distribution constraint mechanisms introduced in the policy optimization phase. These mechanisms help the agent maintain a stable policy distribution and low-variance response under complex environmental disturbances. Compared with conventional methods, our model achieves adaptive adjustment to dynamic environmental changes through structured context updates and hierarchical attention regulation, maintaining stable decision-making performance under uncertainty. These results validate the synergistic effect of multi-level contextual modeling and policy robustness optimization, providing a reliable structural foundation for robust decision-making in intelligent agents.

This paper also analyzes the sensitivity of the number of layers and the number of attention heads to the performance of multi-level context modeling. The experimental results are shown in Figure 2.

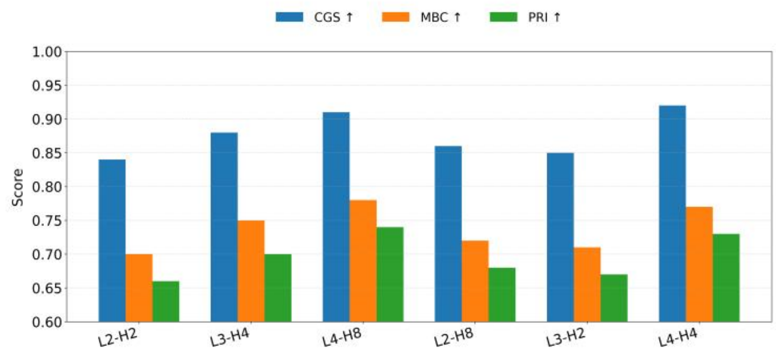

Figure 2. Sensitivity analysis of the number of layers and attention heads on the performance of multi-level context modeling

The results in the figure show that the three evaluation metrics exhibit clear regular variations under different combinations of layer numbers and attention heads. The overall trend indicates that as the number of layers increases and the attention heads remain moderate, the model's contextual modeling ability is significantly enhanced. In particular, under the L4-H4 and L4-H8 configurations, the CGS metric reaches its highest value, suggesting that the hierarchical structure can extract multi-scale semantic information more effectively. This allows the model to achieve more stable contextual understanding in complex environments. These findings confirm the effectiveness of the multi-level structure in capturing temporal and semantic dependencies.

For the MBC metric, as the number of layers increases, the model shows a consistent improvement in multi-scale behavioral consistency, although slight fluctuations occur when the number of attention heads is too high. This indicates that the cross-layer feature aggregation mechanism strengthens temporal correlation and strategy stability, but an excessive number of attention channels may introduce redundant information and weaken feature focus. A balanced configuration of layers and attention heads, such as L3-H4 or L4-H4, allows the model to maintain coordination between local dynamics and global semantics, leading to higher consistency.

For the PRI metric, the model's policy robustness improves steadily with increasing layer depth, showing that deeper structures can form more stable decision representations in uncertain environments. Similar to the MBC trend, a moderate number of attention heads is critical for maintaining policy balance. When the layer number is small or the attention heads is too many, the policy distribution tends to fluctuate slightly. Overall, the experimental results strongly validate the comprehensive advantages of the multi-level contextual modeling mechanism in enhancing the agent's semantic capture, temporal consistency, and decision robustness. This provides a structured foundation for optimizing intelligent behavior in complex and dynamic environments.

This paper then conducts a comparative experiment on the environmental sensitivity of multi-level context robustness under dynamic load fluctuations and time-varying interference. The experimental results are shown in Figure 3.

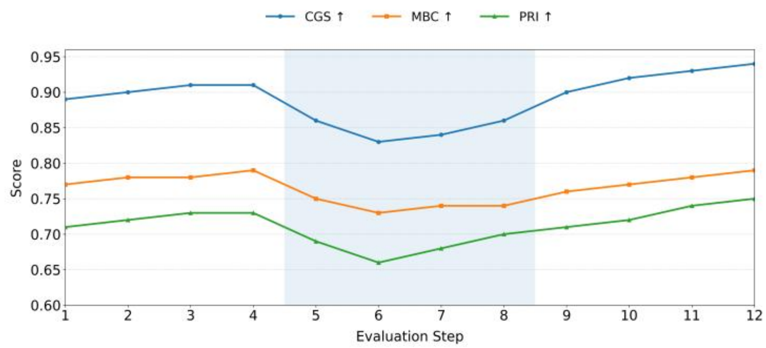

Figure 3. Multi-level contextual robustness and environmental sensitivity experiments under dynamic load fluctuations and time-varying interference

The experimental results show that under dynamic load fluctuations and time-varying disturbances, the variations of the three core metrics exhibit clear stage-wise patterns. In the early stage (steps 1-4), the metrics remain stable, indicating that the model maintains high contextual consistency and policy reliability under normal load conditions. When the system enters the load fluctuation phase (steps 5-6), all three metrics decline to varying degrees, with the CGS metric showing the most significant drop. This suggests that contextual modeling is the most sensitive to external disturbances. However, the decline does not cause instability, and the metrics remain within a relatively high range, demonstrating the buffering capability of the hierarchical structure against feature perturbations.

During the time-varying disturbance phase (steps 7-8), the model's multi-scale consistency and policy robustness continue to be challenged, with both MBC and PRI showing slight fluctuations. This indicates that the model undergoes a brief structural adaptation process when dealing with complex non-stationary inputs. It is noteworthy that the decline in the MBC curve is minor, suggesting that the cross-layer semantic fusion mechanism can effectively maintain information continuity under uncertain conditions. The slight lag in PRI reflects that the policy optimization module responds more slowly to dynamic noise. Nevertheless, the overall trend remains stable without oscillation or collapse, showing that the regularization term at the policy level helps preserve distributional stability.

In the recovery phase (steps 9-12), all metrics rise significantly, with CGS recovering the fastest, followed by MBC and PRI gradually stabilizing. This phase demonstrates that the model can adaptively restore contextual consistency and policy robustness once disturbances subside, showing strong environmental adaptability and structural resilience. The multi-level context-aware mechanism enables the model to reconstruct global information under dynamic interference, maintaining stable performance improvement in fluctuating environments. Such stability is crucial for intelligent agent systems that require long-term policy execution and adaptation to dynamic conditions.

4. Conclusion

The proposed multi-level context-aware agent behavior algorithm achieves end-to-end semantic consistency modeling from perception to decision generation through the coordinated design of hierarchical feature encoding, cross-layer semantic aggregation, and policy regularization optimization. The results show that the proposed method surpasses existing models in contextual generalization, multi-scale consistency, and policy robustness, maintaining stable performance in complex dynamic environments. The multi-level structure enhances the model's adaptability to environmental variations and provides a structured foundation for robust decision-making under non-stationary conditions. This work offers a new perspective on integrating complex semantic modeling and policy optimization, enriching the theoretical foundation of agent perception and behavior modeling.

From a system perspective, this study demonstrates the potential value of multi-level semantic interaction in reinforcement learning and agent decision-making. Through contextual awareness, the agent can dynamically capture environmental features across temporal and spatial scales, achieving coordinated control between local dynamics and global objectives during task execution. This capability gives the model stronger generalization and long-term stability in complex and variable real-world environments. The proposed method shows strong application potential in multi-agent collaboration, automated control, industrial optimization, and autonomous navigation. It provides a feasible algorithmic path to enhance adaptive decision-making and resource scheduling efficiency in intelligent systems.

Future research could further explore the integration of multi-level contextual structures with other intelligent decision-making mechanisms, such as combining causal reasoning to enhance model interpretability or introducing uncertainty modeling to improve adaptability in unknown environments. The proposed method can also be extended to large-scale distributed intelligent systems to study its performance in multi-agent collaboration, long-term task planning, and cross-domain transfer. With continued advances in complex environment modeling and semantic understanding, multi-level context-aware algorithms are expected to become a key technology driving agent decision systems toward higher levels of autonomous intelligence, providing a solid foundation for practical applications in industry, transportation, robotics, and human-machine collaboration.

Recent studies further support the extension of context-aware agent decision-making from policy optimization toward evidence-grounded reasoning, semantic alignment, and system-level robustness. Causal inference and counterfactual risk estimation provide a useful perspective for interpreting action consequences under uncertain conditions, while structure-aware financial report summarization and fine-grained domain text classification highlight the importance of preserving information fidelity and semantic granularity in complex decision contexts [11], [12], [13]. Retrieval-augmented generation with dual retrieval and adaptive re-ranking also shows that agents can improve decision reliability by combining external evidence selection with dynamic ranking, which is consistent with the need for hierarchical attention over heterogeneous contextual signals [14]. Value-aware recommendation modeling through reinforcement learning further connects sequential action selection with long-term utility estimation, offering methodological support for policy regularization in interactive environments [15].

From a deployment and interaction perspective, dynamic resource load forecasting, differentiable social graph decision modeling, and spatiotemporal latency prediction demonstrate that agent behavior should be optimized with awareness of temporal load, relational structure, and operational feedback [16], [17], [18]. Graph-based knowledge representation and semantic alignment for large language model-driven entity recognition further indicate that symbolic relations and learned representations can be jointly organized to improve contextual understanding in agent systems [19]. Explainable cross-market causal discovery via financial knowledge graphs also reinforces the value of causal structure learning and interpretable relational reasoning for complex decision-making tasks [20]. These related directions collectively strengthen the motivation for the proposed hierarchical attention and policy regularization framework, which integrates multi-level context representation, adaptive aggregation, and robust action generation within a unified agent behavior model.

References

[1] Merkle N, Mikut R. Context-aware composition of agent policies by Markov decision process entity embeddings and agent ensembles[J]. Semantic Web, 2024, 15(4): 1443-1471.

[2] Kim B, Kim J, Kim Y, et al. Context-aware planning and environment-aware memory for instruction following embodied agents[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 10936-10946.

[3] Wang J, Yuan M, Li Y, et al. Hierarchical attention master-slave for heterogeneous multi-agent reinforcement learning[J]. Neural Networks, 2023, 162: 359-368.

[4] Chen Z, Cao Y, Fu Q, et al. Hierarchical attention enhanced deep learning achieves high precision motor imagery classification in brain computer interfaces[J]. Scientific Reports, 2025, 15(1): 34555.

[5] Wang Y, Zou S. Policy gradient method for robust reinforcement learning[C]//International conference on machine learning. PMLR, 2022: 23484-23526.

[6] Jarne Ornia D, Romao L, Hammond L, et al. Bounded Robustness in Reinforcement Learning via Lexicographic Objectives[J]. arXiv e-prints, 2022: arXiv: 2209.15320.

[7] Natarajan S. AgentFlow: A Context Aware Multi-Agent Framework for Dynamic Agent Collaboration[J]. 2025.

[8] Pu Y, Wang S, Yao X, et al. Context-based soft actor critic for environments with non-stationary dynamics[J]. arXiv preprint arXiv:2105.03310, 2021.

[9] Beukman M, Jarvis D, Klein R, et al. Dynamics generalisation in reinforcement learning via adaptive context-aware policies[J]. Advances in Neural Information Processing Systems, 2023, 36: 40167-40203.

[10] Li X, Zhang J. Context-aware communication for multi-agent reinforcement learning[J]. arXiv preprint arXiv:2312.15600, 2023.

[11] S. Lin, "Automatic Audit Judgment Using Causal Inference and Counterfactual Risk Estimation," 2024.

[12] M. Wang, P. L. Peng, and J. Chen, "Financial Report Summarization via Structure-Aware Modeling and Information Fidelity Constraints," 2024.

[13] L. Ma, E. Lin, and Z. Zhang, "Intelligent Domain Text Classification through Fine-Grained Semantic Representation Learning,"2024.

[14] L. Mao, "Joint Optimization of Dual Retrieval and Adaptive Re-Ranking for Retrieval-Augmented Generation," Artificial Intelligence and Computing Innovations, vol. 4, no. 4, 2024.

[15] K. Zhang, "Value-Aware Recommendation Modeling Through Reinforcement Learning and Offline Interaction Data," 2024.

[16] T. Xia, "Dynamic Resource Load Forecasting in Cloud Computing: A Representation Learning Perspective," 2024.

[17] X. Ren, "A Unified Framework for Social Graph Modeling and Node Decision-Making via Differentiable Mixed-Integer Optimization," 2024.

[18] Z. Xiao, K. Ying, and B. Venkatesan, "Spatiotemporal Transformer-Based Latency Prediction for Large-Scale Cloud Computing Clusters," Artificial Intelligence and Computing Innovations, vol. 5, no. 2, 2025.

[19] J. Zhou, "Graph-Based Knowledge Representation and Semantic Alignment for Large Language Model-Driven Entity Recognition," 2025.

[20] L. Chen and S. Liu, "Explainable Cross-Market Causal Discovery via Financial Knowledge Graphs," 2024.