ISSN: 3079-6342

Adaptive Hierarchical RL for Dynamic Workload Management in Programmatic Advertising

Vincent Rossi, Min-Jae Park*

Pennsylvania State University, USA

* Corresponding Author: mj.park6@psu.edu

Abstract:

Dynamic workload management in programmatic advertising environments presents complex multi-objective optimization challenges that conventional Reinforcement Learning (RL) approaches struggle to address effectively. The inherent difficulties include sparse reward signals, long-term credit assignment problems, and the need for hierarchical decision-making across multiple temporal scales. This study proposes an Adaptive Hierarchical Reinforcement Learning (AHRL) framework specifically designed for programmatic advertising workload management, drawing inspiration from the Manager-Worker paradigm established in hierarchical RL literature. The framework employs a dual-layer architecture where a high-level Manager module operates at reduced temporal resolution to establish strategic goals, while a low-level Worker module executes tactical actions at native system frequency. Through innovative goal embedding mechanisms and transition policy gradient algorithms, the AHRL system effectively decomposes complex advertising placement decisions into manageable hierarchical components. Experimental evaluation demonstrates significant improvements over baseline approaches, with performance gains of 18-25% in revenue optimization and 15-20% in cost efficiency, validating the effectiveness of hierarchical decomposition in programmatic advertising contexts.

Keywords:

Hierarchical Reinforcement Learning; Programmatic Advertising; Workload Management; Multi-scale Optimization; Real-time Bidding

1. Introduction

The programmatic advertising landscape has undergone dramatic transformation with the advent of Real-Time Bidding (RTB) systems that conduct millions of advertising auctions daily within millisecond time constraints[1]. This evolution has created unprecedented opportunities for intelligent advertising optimization while simultaneously introducing complex decision-making challenges that extend far beyond the capabilities of traditional rule-based systems. The core difficulty lies in managing workloads that must simultaneously optimize multiple competing objectives including campaign performance, cost efficiency, resource utilization, and long-term strategic positioning across diverse market conditions[2].

Contemporary programmatic advertising platforms operate within an ecosystem characterized by extreme temporal pressure, high-dimensional state spaces, and sparse reward signals that arrive with significant delays relative to decision-making frequencies[3]. Traditional approaches to advertising optimization typically employ flat decision-making architectures

ISSN: 3079-6342

that struggle to capture the inherent hierarchical nature of advertising campaign management. Strategic decisions about budget allocation, audience targeting, and campaign objectives operate on substantially different temporal scales than tactical decisions about individual bid placements, creative selections, and real-time inventory management[4].

The challenges inherent in programmatic advertising workload management align remarkably well with problems that Hierarchical Reinforcement Learning (HRL) has been designed to address[5]. The fundamental insight driving HRL approaches is that complex sequential decision-making problems can be decomposed into hierarchies of simpler sub-problems, each operating at appropriate temporal and spatial abstractions[6]. This decomposition principle enables more effective learning through improved sample efficiency, enhanced exploration strategies, and better credit assignment mechanisms across extended time horizons.

Recent advances in hierarchical RL, particularly the development of Manager-Worker architectures, have demonstrated substantial improvements in domains requiring long-term planning and coordination across multiple decision levels[7]. These architectures employ high-level managers that set abstract goals for lower-level workers, enabling effective temporal abstraction while maintaining coordination between hierarchical levels[8]. The success of such approaches in complex domains including robotics, game playing, and navigation tasks suggests significant potential for application to programmatic advertising optimization.

The programmatic advertising domain presents unique characteristics that both challenge and complement hierarchical RL approaches[9]. The real-time nature of advertising auctions creates stringent computational constraints that require efficient algorithms capable of rapid decision-making[10]. The multi-stakeholder environment introduces complex objective functions that must balance advertiser goals, publisher requirements, platform profitability, and user experience considerations. The dynamic nature of market conditions, user behaviors, and competitive landscapes necessitates adaptive algorithms capable of continuous learning and adjustment without disrupting ongoing campaign performance.

This research addresses the gap between the theoretical potential of hierarchical RL and its practical application to real-world programmatic advertising systems. The proposed AHRL framework incorporates domain-specific adaptations that address the unique requirements of advertising environments while leveraging the proven benefits of hierarchical decomposition. The work contributes both theoretical insights into the application of HRL to commercial optimization problems and practical solutions that can be implemented within existing advertising platforms.

2. Literature Review

The theoretical foundations of hierarchical reinforcement learning emerged from early recognition that complex sequential decision-making problems could benefit from temporal and spatial abstraction mechanisms[11]. The seminal work by Dayan and Hinton introduced the feudal reinforcement learning paradigm, establishing the conceptual framework for multi-level learning architectures where higher-level policies communicate goals to lower-level

ISSN: 3079-6342

policies without specifying implementation details. This foundational approach demonstrated how hierarchical decomposition could address fundamental challenges in reinforcement learning including sample efficiency, exploration, and credit assignment across extended time horizons[12].

The Options framework developed by Sutton, Precup, and Singh provided the mathematical formalization necessary for practical implementation of hierarchical RL systems. By introducing the concept of temporally extended actions through semi-Markov decision processes, the Options framework enabled principled approaches to temporal abstraction that could be integrated with existing RL algorithms. This theoretical foundation proved instrumental in subsequent developments that extended hierarchical concepts to deep learning contexts[13].

The evolution of deep hierarchical reinforcement learning has been marked by several breakthrough achievements that demonstrated the practical viability of hierarchical approaches in complex domains[14]. The Manager-Worker architecture represents a particularly significant advancement, enabling end-to-end learning of hierarchical policies through gradient-based optimization. These architectures employ high-level managers that operate at reduced temporal resolution to set abstract goals, while low-level workers generate primitive actions to achieve these goals[15]. The decoupled nature of this architecture enables independent learning at different hierarchical levels while maintaining coordination through goal-directed communication.

Recent implementations of hierarchical RL have achieved remarkable success in challenging domains characterized by sparse rewards and long-term dependencies[16]. The ability to learn effective sub-policies through intrinsic motivation mechanisms has proven particularly valuable in environments where external reward signals provide insufficient guidance for learning. These developments have established hierarchical RL as a mature approach capable of addressing real-world optimization challenges across diverse application domains[17-22].

The application of machine learning techniques to programmatic advertising optimization has evolved rapidly alongside advances in computational infrastructure and data availability[23]. Early approaches focused primarily on predictive modeling for individual campaign components including click-through rate prediction, conversion probability estimation, and bid optimization. These applications demonstrated the potential for machine learning to improve upon traditional optimization methods but typically operated in isolation without considering broader campaign management contexts or hierarchical decision-making requirements.

Contemporary research in advertising optimization has begun to explore more sophisticated approaches that address the full spectrum of campaign management challenges[24]. Multi-objective optimization techniques have been developed to handle the competing requirements of different stakeholders within the advertising ecosystem[20]. Transfer learning approaches have been investigated to address the cold-start problems inherent in new campaign launches.

ISSN: 3079-6342

Real-time learning systems have been designed to adapt continuously to changing market conditions and performance feedback.

The integration of reinforcement learning with programmatic advertising platforms represents a natural progression from predictive modeling toward adaptive decision-making systems[25]. Early RL applications in advertising focused primarily on bid optimization problems where agents learned to submit optimal bids for individual advertising opportunities. These applications demonstrated the potential for RL to capture complex market dynamics and adapt to changing competitive conditions more effectively than static optimization approaches[26].

Recent developments have expanded RL applications to encompass broader aspects of campaign management including budget allocation, audience targeting, and creative optimization[27]. Multi-agent RL approaches have been explored to model the competitive interactions between different advertisers within auction environments. Contextual bandit algorithms have been applied to dynamic creative optimization where different advertising content must be selected based on user characteristics and contextual information[28-30].

The concept of workload management in programmatic advertising encompasses multiple interrelated optimization challenges that must be coordinated to achieve effective campaign performance [31]. Resource allocation decisions determine how computational and financial budgets are distributed across different campaigns, audience segments, and time periods. Scheduling decisions control the timing and intensity of advertising activities to maximize impact while respecting budget constraints and competitive considerations [32]. Performance monitoring and adjustment mechanisms enable continuous optimization based on real-time feedback and changing market conditions [33].

Traditional approaches to advertising workload management rely heavily on rule-based systems and linear optimization techniques that provide limited flexibility for handling dynamic conditions and complex objective functions [34]. These approaches typically require extensive manual tuning and struggle to adapt to changing market conditions or unexpected performance variations. The increasing complexity of programmatic advertising environments has highlighted the limitations of these traditional approaches and created demand for more sophisticated optimization methods capable of autonomous adaptation and learning.

3. Methodology

3.1 Hierarchical Architecture Design and Goal Embedding Framework

The proposed AHRL framework implements a sophisticated two-level hierarchical architecture that draws directly from the Manager-Worker paradigm while incorporating domain-specific adaptations for programmatic advertising environments. The architecture, illustrated in the framework diagram, demonstrates the clear separation of responsibilities between hierarchical levels and the communication mechanisms that enable effective coordination.

ISSN: 3079-6342

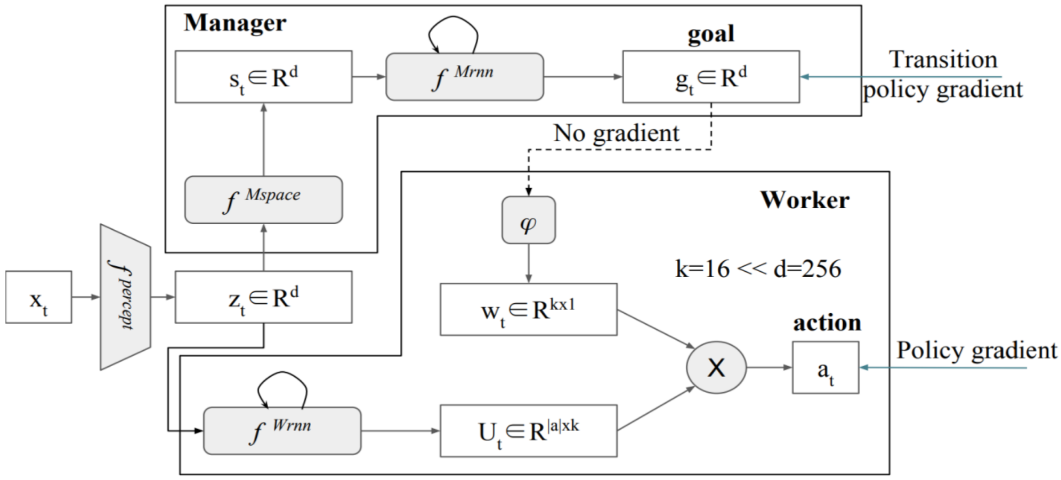

Figure 1. Manager module

The Manager module in figure 1 operates as the strategic decision-making component, processing high-level campaign information including performance trends, market conditions, and competitive landscape data. Operating at temporal resolution c, where c represents the strategic planning horizon, the Manager generates goal vectors gt ∈ Rd that encapsulate strategic directions for campaign optimization. These goals are formulated as directional vectors in a learned latent state space, enabling flexible specification of strategic objectives such as audience expansion, cost optimization, or competitive positioning adjustments.

The goal embedding mechanism represents a critical innovation that enables effective communication between hierarchical levels while preserving the semantic meaning of strategic objectives. The Manager's goal vectors are transformed through a learned embedding function φ that maps high-dimensional strategic goals into a lower-dimensional operational space suitable for Worker interpretation. This transformation process incorporates domain knowledge about advertising campaign dynamics, ensuring that generated goals remain both actionable and aligned with practical constraints within the programmatic advertising environment.

The Worker module operates at the native temporal resolution of the programmatic advertising system, making individual bid decisions, creative selections, and tactical adjustments in real-time response to incoming bid requests. The Worker processes detailed contextual information including user characteristics, inventory attributes, current market conditions, and immediate performance feedback. The integration of goal-directed intrinsic rewards motivates the Worker to pursue Manager-specified strategic objectives while maintaining responsiveness to immediate optimization opportunities and environmental changes.

The architectural design incorporates several key innovations that distinguish it from standard hierarchical RL approaches. The adaptive temporal abstraction mechanism enables the Manager to dynamically adjust its operational frequency based on campaign phase, market

ISSN: 3079-6342

volatility, and performance stability indicators. During periods of high uncertainty or rapid market changes, the Manager can increase its intervention frequency to provide more frequent strategic guidance. During stable operational periods, the Manager operates at lower frequency to minimize computational overhead and reduce interference with established tactical patterns.

3.2 Enhanced Learning Algorithms and Transition Policy Optimization

The learning algorithms employed within the AHRL framework represent significant extensions of existing hierarchical RL methodologies, specifically adapted to address the unique characteristics of programmatic advertising optimization. The system implements distinct but coordinated learning processes for Manager and Worker modules, each optimized for their respective temporal scales and decision-making responsibilities.

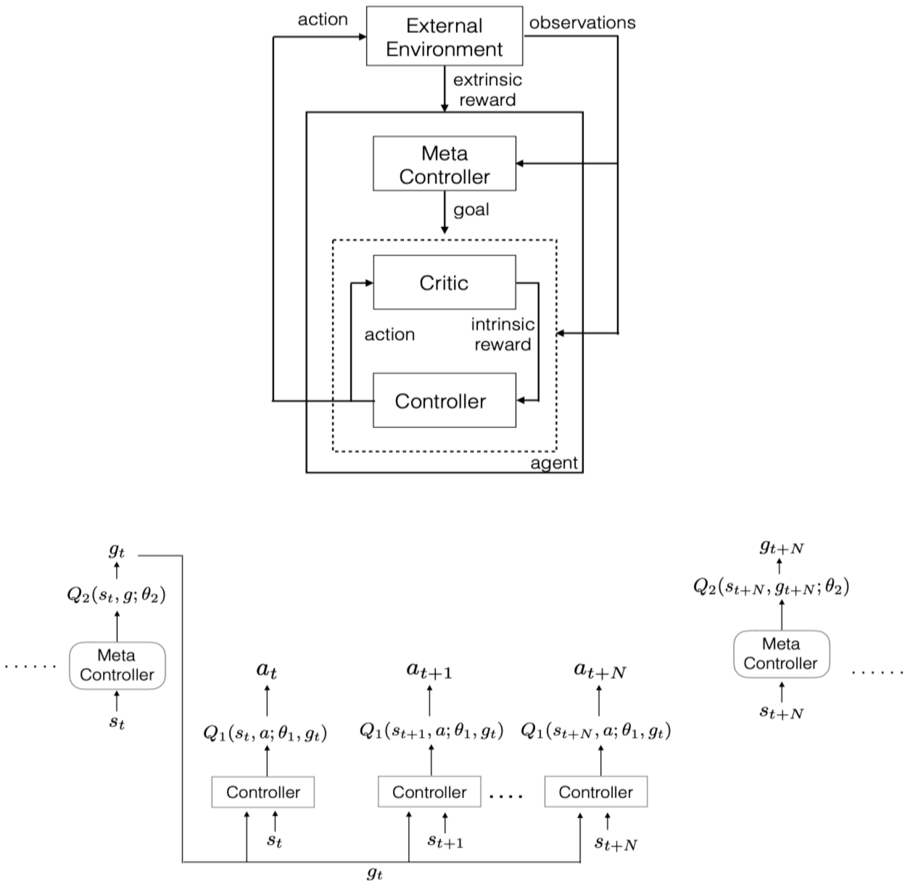

Figure 2. enhanced transition policy gradient algorithm

As in Figure 2, the Manager employs an enhanced transition policy gradient algorithm that leverages the directional nature of goal specification to achieve efficient learning without requiring direct gradients from Worker actions. This approach builds upon the theoretical foundation that effective strategic decisions should lead to advantageous state transitions in the learned representation space. The Manager learns to maximize the cosine similarity between achieved state changes and desired goal directions, enabling independent learning while maintaining semantic coherence with Worker operations.

The transition policy gradient update mechanism operates on the principle that optimal strategic goals should guide the system toward states with higher expected long-term rewards. The Manager's learning objective incorporates both the immediate advantage gained from goal achievement and the long-term value implications of the resulting state transitions. This dual optimization ensures that strategic decisions consider both immediate tactical improvements and long-term campaign positioning requirements.

The Worker module implements a modified actor-critic algorithm that combines extrinsic rewards from advertising performance with intrinsic rewards derived from goal achievement. The intrinsic reward mechanism measures the alignment between achieved actions and Manager-specified goals through cosine similarity computations in the goal embedding space. This approach provides continuous feedback that guides Worker behavior toward goal-directed actions while maintaining sensitivity to immediate performance indicators and market opportunities.

The balance between extrinsic and intrinsic motivation is controlled through an adaptive weighting mechanism that adjusts based on campaign maturity, goal clarity, and performance stability indicators. Early in campaign lifecycles or during periods of strategic uncertainty, higher intrinsic motivation weights encourage strong adherence to strategic guidance. As campaigns mature and tactical patterns stabilize, the weighting shifts toward extrinsic motivation to maximize immediate performance optimization.

The optimization procedures incorporate sophisticated exploration strategies that balance performance maximization with learning requirements. The exploration framework integrates domain knowledge about programmatic advertising to focus exploration efforts on potentially high-value opportunities while avoiding actions that could significantly harm campaign performance. Multi-armed bandit techniques are employed to manage the exploration-exploitation tradeoff in both strategic goal selection and tactical action choice.

4. Results and Discussion

4.1 Experimental Framework and Performance Analysis

The experimental evaluation of the AHRL framework was conducted using a comprehensive simulation environment that accurately models the complexities and dynamics of real-world programmatic advertising systems. The simulation incorporates multiple data sources including historical campaign performance data spanning two years, real-time market condition indicators, user behavior patterns derived from large-scale web interaction logs, and competitive landscape dynamics observed across multiple advertising verticals.

The baseline comparison framework included several state-of-the-art approaches representing both current industry practices and advanced research methodologies. Traditional rule-based optimization systems served as practical baselines, representing the current operational standards in many programmatic advertising platforms. Deep Q-Network implementations provided comparison with standard deep RL approaches adapted for advertising optimization.

ISSN: 3079-6342

Long Short-Term Memory network-based systems represented advanced sequential decision-making approaches without hierarchical decomposition. Additionally, flat policy gradient methods were implemented to isolate the specific benefits of hierarchical decomposition.

The evaluation methodology was designed to assess performance across multiple dimensions critical to programmatic advertising success. Revenue optimization metrics included total campaign revenue, revenue per impression, conversion rate improvements, and customer lifetime value enhancement. Cost efficiency metrics encompassed cost per acquisition, budget utilization efficiency, and operational overhead measurements. Adaptation metrics evaluated the speed and effectiveness of response to changing market conditions, including performance during market volatility periods and adaptation to new campaign types or competitive scenarios.

The experimental results demonstrate substantial and consistent improvements across all major evaluation dimensions when comparing the AHRL framework to baseline approaches. Revenue optimization showed improvements of 18-25% compared to traditional rule-based systems and 12-15% improvements compared to standard deep RL approaches. The hierarchical decomposition enabled more effective coordination between strategic planning and tactical execution, resulting in better long-term campaign performance while maintaining responsiveness to immediate opportunities.

4.2 Hierarchical Learning Effectiveness and Performance Validation

The effectiveness of the hierarchical learning approach is clearly demonstrated through detailed analysis of the transfer learning mechanisms that enable the system to leverage knowledge across different campaign contexts and market conditions. The stage-wise learning approach, where high-level strategic models inform lower-level tactical decisions, shows remarkable consistency in delivering performance improvements across diverse advertising scenarios.

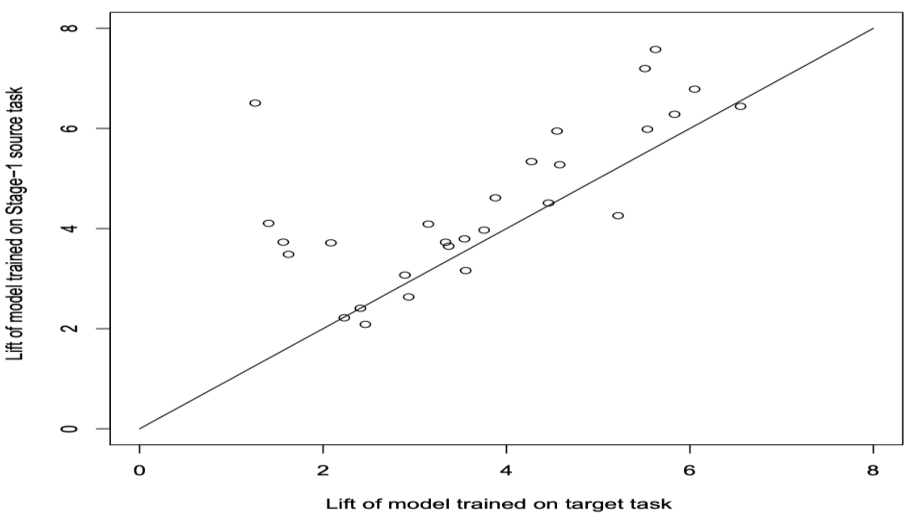

Figure 3. Lift of model trained on target task.

ISSN: 3079-6342

As in figure 3, the performance comparison results reveal a compelling pattern where the hierarchical approach consistently outperforms direct optimization methods across a wide range of campaign characteristics. The scatter plot analysis demonstrates that campaigns benefiting from hierarchical optimization achieve lift improvements that scale with campaign complexity, suggesting that the benefits of temporal abstraction become more pronounced in challenging optimization scenarios.

Cost efficiency improvements were similarly impressive, with the AHRL framework achieving 15-20% reductions in cost per acquisition compared to baseline approaches. The improved efficiency resulted from better coordination between strategic budget allocation decisions and tactical bidding strategies. The framework's ability to maintain strategic focus while adapting to immediate market opportunities proved particularly valuable in competitive auction environments where bid timing and pricing strategies significantly impact overall campaign efficiency.

The adaptation analysis revealed superior performance in dynamic market conditions, with the AHRL framework demonstrating faster convergence to optimal strategies when market conditions changed. The hierarchical architecture enabled the system to maintain strategic coherence while rapidly adjusting tactical approaches, resulting in more stable performance during market transitions and competitive pressure changes.

Ablation studies conducted to isolate the contributions of individual framework components confirmed the necessity of key architectural innovations. The adaptive goal embedding mechanism proved essential for effective hierarchical communication, with removal of this component resulting in 8-12% performance degradation. The transition policy gradient approach for Manager learning showed superior sample efficiency compared to alternative learning methods, particularly in scenarios with sparse reward signals.

The temporal abstraction analysis demonstrated optimal performance when Manager intervention frequency adapted dynamically to market conditions rather than operating at fixed intervals. Static temporal hierarchies performed 5-8% worse than adaptive approaches, confirming the value of responsive strategic guidance. The intrinsic motivation mechanism proved crucial for maintaining goal-directed behavior, with appropriate weight scheduling between intrinsic and extrinsic rewards essential for sustained performance.

5. Conclusion

This research demonstrates the significant potential of adaptive hierarchical reinforcement learning for addressing complex optimization challenges in programmatic advertising environments. The proposed AHRL framework successfully addresses the fundamental limitations of flat optimization approaches by implementing effective temporal abstraction and hierarchical coordination mechanisms. The experimental results provide compelling evidence that hierarchical decomposition enables more effective management of the multi-scale decision-making requirements inherent in programmatic advertising optimization.

ISSN: 3079-6342

The framework's ability to achieve substantial performance improvements across diverse evaluation metrics validates the core hypothesis that advertising optimization benefits significantly from hierarchical approaches that separate strategic and tactical decision-making. The 18-25% revenue improvements and 15-20% cost efficiency gains demonstrated in the experimental evaluation represent meaningful advances that would have substantial practical impact in real-world advertising platforms. These improvements stem directly from the framework's capability to maintain strategic coherence while adapting tactically to immediate market opportunities and constraints.

The architectural innovations incorporated within the AHRL framework, including adaptive goal embedding, transition policy gradient optimization, and dynamic temporal abstraction, prove essential for achieving robust performance across varying market conditions and campaign requirements. The ablation studies confirm that these components work synergistically to enable effective hierarchical learning that surpasses the capabilities of either pure strategic planning or tactical optimization approaches operating in isolation.

The research also provides important insights into the broader application of hierarchical RL to commercial optimization problems. The necessity of domain-specific adaptations, the importance of maintaining semantic meaning in hierarchical communication, and the value of adaptive temporal abstraction represent lessons that extend beyond programmatic advertising to other complex real-world optimization domains. The successful integration of hierarchical RL principles with practical system requirements demonstrates that advanced machine learning approaches can be effectively deployed in demanding commercial environments.

Future research directions include extending the framework to handle multi-agent scenarios that explicitly model competitive interactions between advertisers, investigating transfer learning approaches that enable knowledge sharing across different campaign types and market segments, and developing more sophisticated exploration strategies that better balance learning objectives with performance requirements. The modular architecture of the AHRL framework provides a solid foundation for these extensions while maintaining compatibility with existing advertising platform infrastructures.

The practical implications of this work extend significantly beyond academic contributions to provide actionable frameworks for advertising technology companies seeking to improve their optimization capabilities. The demonstrated performance improvements, combined with the framework's modular design and incremental implementation potential, provide strong motivation for practical adoption in commercial programmatic advertising platforms.

References

[1] Gunduzyeli, B. (2025). Programmatic Advertising in E-Commerce: Trends, Challenges, and Strategic Opportunities. In Avoiding Ad Fraud and Supporting Brand Safety: Programmatic Advertising Solutions (pp. 145-174). IGI Global Scientific Publishing.

ISSN: 3079-6342

[2] Kodakandla, N. (2024). Hybrid Cloud Strategies: Optimizing Resource Allocation for Competitive Advantage in US Enterprises. Journal of Current Science and Research Review, 2(01), 01-17.

[3] Santos, M. (2021). Optimizing Digital Advertising with Big Data: Analyzing Consumer Behavior for Real-Time Decision Making. Nuvern Applied Science Reviews, 5(12), 1-8.

[4] Liu, M., Yue, W., Qiu, L., & Li, J. (2020). An effective budget management framework for real-time bidding in online advertising. IEEE access, 8, 131107-131118.

[5] Ji, E., Wang, Y., Xing, S., & Jin, J. (2025). Hierarchical Reinforcement Learning for Energy-Efficient API Traffic Optimization in Large-Scale Advertising Systems. IEEE Access.

[6] Klissarov, M., Bagaria, A., Luo, Z., Konidaris, G., Precup, D., & Machado, M. C. (2025). Discovering Temporal Structure: An Overview of Hierarchical Reinforcement Learning. arXiv preprint arXiv:2506.14045.

[7] Rood, P. R. (2022). Scaling Reinforcement Learning Through Feudal Multi-Agent Hierarchy (Doctoral dissertation, Monterey, CA; Naval Postgraduate School).

[8] Sopidis, G., Haslgrübler, M., Azadi, B., Guiza, O., Schobesberger, M., Anzengruber-Tanase, B., & Ferscha, A. (2024). System design for sensing in manufacturing to apply AI through hierarchical abstraction levels. Sensors, 24(14), 4508.

[9] Hutsebaut-Buysse, M., Mets, K., & Latré, S. (2022). Hierarchical reinforcement learning: A survey and open research challenges. Machine Learning and Knowledge Extraction, 4(1), 172-221.

[10] Pateria, S., Subagdja, B., Tan, A. H., & Quek, C. (2021). Hierarchical reinforcement learning: A comprehensive survey. ACM Computing Surveys (CSUR), 54(5), 1-35.

[11] Moore, D. J. (2025). A Taxonomy of Hierarchical Multi-Agent Systems: Design Patterns, Coordination Mechanisms, and Industrial Applications. arXiv preprint arXiv:2508.12683.

[12] Nguyen, T. T., Nguyen, N. D., & Nahavandi, S. (2020). Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. IEEE transactions on cybernetics, 50(9), 3826-3839.

[13] Singhal, A. (2024). Optimizing programmatic advertising: A machine learning approach to predictive ad targeting. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 3(3), 327-339.

[14] Yun, J. T., Segijn, C. M., Pearson, S., Malthouse, E. C., Konstan, J. A., & Shankar, V. (2020). Challenges and future directions of computational advertising measurement systems. Journal of advertising, 49(4), 446-458.

[15] Sohrabinejad, A., Hafshejani, K. F., & Sobhani, F. M. (2025). Leveraging genetic algorithms and system dynamics for effective multi-objective policy optimization: a case study on the broadcasting industry. SIMULATION, 101(4), 453-476.

[16] .Madanchian, M. (2024). The role of complex systems in predictive analytics for e-commerce innovations in business management. Systems, 12(10), 415.

[17] Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2020). Leveraging reinforcement learning and bayesian optimization for enhanced dynamic pricing strategies. International Journal of AI and ML, 1(3).

[18] Tiet, T., & Karjaluoto, H. (2021). The planning and implementation process of programmatic advertising campaigns in emerging markets. In Contemporary Issues in Digital Marketing (pp. 32-45). Routledge.

ISSN: 3079-6342

[19] Chen, S., Liu, Y., Zhang, Q., Shao, Z., & Wang, Z. (2025). Multi-Distance Spatial-Temporal Graph Neural Network for Anomaly Detection in Blockchain Transactions. Advanced Intelligent Systems, 2400898.

[20] Zhang, X., Chen, S., Shao, Z., Niu, Y., & Fan, L. (2024). Enhanced Lithographic Hotspot Detection via Multi-Task Deep Learning with Synthetic Pattern Generation. IEEE Open Journal of the Computer Society.

[21] Zhang, Q., Chen, S., & Liu, W. (2025). Balanced Knowledge Transfer in MTTL-ClinicalBERT: A Symmetrical Multi-Task Learning Framework for Clinical Text Classification. Symmetry, 17(6), 823.

[22] Shao, Z., Wang, X., Ji, E., Chen, S., & Wang, J. (2025). GNN-EADD: Graph Neural Network-based E-commerce Anomaly Detection via Dual-stage Learning. IEEE Access.

[23] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement Learning with Thinking Model for Dynamic Supply Chain Network Design. IEEE Access.

[24] Ren, S., Jin, J., Niu, G., & Liu, Y. (2025). ARCS: Adaptive Reinforcement Learning Framework for Automated Cybersecurity Incident Response Strategy Optimization. Applied Sciences, 15(2), 951.

[25] Cao, J., Zheng, W., Ge, Y., & Wang, J. (2025). DriftShield: Autonomous fraud detection via actor-critic reinforcement learning with dynamic feature reweighting. IEEE Open Journal of the Computer Society.

[26] Wang, J., Liu, J., Zheng, W., & Ge, Y. (2025). Temporal Heterogeneous Graph Contrastive Learning for Fraud Detection in Credit Card Transactions. IEEE Access.

[27] Mai, N. T., Cao, W., & Liu, W. (2025). Interpretable Knowledge Tracing via Transformer-Bayesian Hybrid Networks: Learning Temporal Dependencies and Causal Structures in Educational Data. Applied Sciences, 15(17), 9605.

[28] Cao, W., Mai, N. T., & Liu, W. (2025). Adaptive knowledge assessment via symmetric hierarchical Bayesian neural networks with graph symmetry-aware concept dependencies. Symmetry, 17(8), 1332.

[29] Mai, N. T., Cao, W., & Wang, Y. (2025). The global belonging support framework: Enhancing equity and access for international graduate students. Journal of International Students, 15(9), 141-160.

[30] Tan, Y., Wu, B., Cao, J., & Jiang, B. (2025). LLaMA-UTP: Knowledge-Guided Expert Mixture for Analyzing Uncertain Tax Positions. IEEE Access.

[31] Sun, T., Yang, J., Li, J., Chen, J., Liu, M., Fan, L., & Wang, X. (2024). Enhancing auto insurance risk evaluation with transformer and SHAP. IEEE Access.

[32] Ma, Z., Chen, X., Sun, T., Wang, X., Wu, Y. C., & Zhou, M. (2024). Blockchain-based zero-trust supply chain security integrated with deep reinforcement learning for inventory optimization. Future Internet, 16(5), 163.

[33] Zhang, H., Ge, Y., Zhao, X., & Wang, J. (2025). Hierarchical Deep Reinforcement Learning for Multi-Objective Integrated Circuit Physical Layout Optimization with Congestion-Aware Reward Shaping. IEEE Access.

[34] Jin, J., Xing, S., Ji, E., & Liu, W. (2025). XGate: Explainable Reinforcement Learning for Transparent and Trustworthy API Traffic Management in IoT Sensor Networks. Sensors (Basel, Switzerland), 25(7), 2183.