INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY TRENDS

Vol. 06 No. 01 (2026)
e. ISSN: 2409-5222
https://ijictt.ictt.info

Performance-Aware Reinforcement Learning Framework for LLM-Augmented Decision Agents in Complex Environments

Oliver Smith1, James Thompson2*, William Brown3
Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, United Kingdom
**Corresponding author: james.thompson@ox.ac.uk*

Abstract : *This study presents a performance-aware reinforcement learning framework for enhancing decision-making efficiency in LLM-augmented agent systems. The framework incorporates multi-objective reward functions that jointly consider task accuracy, computational cost and response latency. A large-scale experimental evaluation is conducted on 12,600 decision tasks, including financial analysis, operational planning and scenario simulation. The proposed method achieves a 20.9% increase in decision accuracy while reducing average computation time by 16.7%. Additionally, Pareto frontier analysis shows improved trade-offs between performance and efficiency compared to baseline approaches. The results indicate that performance-aware optimization can significantly enhance the practical applicability of intelligent agent systems.*
Keywords: *LLM agents; Reinforcement learning; Multi-objective optimization; Decision efficiency; Performance modelling*

1. INTRODUCTION

Large language model (LLM) agents are increasingly being developed for tasks involving multi-step reasoning, tool use, and adaptive decision making. Recent studies have shown their potential in software engineering, financial analysis, medical reasoning, and planning-oriented applications, indicating that agent-based LLM systems are becoming an important research direction with growing practical relevance [1,2]. At the same time, recent work on structured inference and representation learning has further suggested that performance in complex reasoning systems depends not only on model capacity, but also on how intermediate information is organized and used during decision processes [3]. This perspective is highly relevant to LLM agents, whose effectiveness often depends on the quality of multi-stage reasoning, action selection, and information integration across steps. Although these systems have demonstrated strong reasoning ability, their practical utility depends on more than final-task accuracy alone. In many real-world settings, an agent must operate under constraints related to computation time, tool invocation cost, and response delay. Under such conditions, the most useful decision is not always the most accurate one in isolation, but the one that achieves a better balance between quality and efficiency. This issue is especially important in time-sensitive environments, where a slightly less accurate decision may be preferable if it can be produced with substantially lower latency and cost [4,5]. As a result, recent studies have begun to examine the trade-off between effectiveness and efficiency, rather than evaluating agent performance only through success rate. Research on optimizing LLM agents has expanded rapidly in recent years. Existing surveys indicate that reinforcement learning has become one of the main approaches for improving agent behavior, especially in settings that require sequential reasoning, dynamic planning, and interactive tool use [6,7]. Compared with

static prompting strategies, trainable policies offer a more flexible way to support decision making over multiple steps, allowing agents to revise intermediate actions and adapt their execution process during task completion. This shift reflects a broader trend in agent research, where the emphasis is moving from prompt engineering toward policy optimization and decision control. Reinforcement learning has shown particular value in improving long-horizon decision processes in LLM-based agents. Prior studies suggest that RL can strengthen planning ability and improve the quality of tool use under complex task settings [8,9]. At the same time, these studies also indicate that reward design remains a central challenge. In many agent tasks, simple outcome-level rewards are insufficient because they fail to reflect the quality of intermediate decisions, tool efficiency, or execution overhead. More detailed reward mechanisms may improve learning performance, but they are often difficult to define consistently in real applications [10]. This makes reward construction a key issue in the development of robust and deployable LLM agent training frameworks. Another closely related challenge lies in evaluation. Recent benchmarks have revealed that current LLM agents still show limited capability in complex interactive environments [11,12]. However, benchmark settings vary substantially in task formulation, environment design, and evaluation criteria, which makes direct comparison across studies difficult. The problem becomes more pronounced when efficiency is taken into account, because many existing benchmarks focus primarily on task completion or answer correctness, while giving limited attention to latency, computational burden, or tool-use cost [13]. As a result, current empirical evidence remains insufficient for understanding whether an agent performs well because it reasons effectively, because it spends more resources, or because the evaluation framework overlooks important efficiency-related differences. LLM agents have already been explored in several high-value domains. In finance, they are used for analytical support and decision assistance. In medicine, they are applied to tasks such as diagnostic reasoning and risk assessment. These studies demonstrate that LLM agents can operate in complex and information-rich environments, but they also show that most existing systems are still designed primarily around accuracy improvement, while efficiency considerations are handled separately or treated as downstream engineering concerns [14]. This separation limits the practical value of many current methods, especially in applications where real-time responsiveness and resource control are essential. Despite recent progress, important gaps remain in the current literature. Many studies rely on relatively narrow datasets or domain-specific tasks, which restricts the assessment of general decision performance across broader scenarios [15]. A large portion of reinforcement learning approaches still optimize a single objective, usually task accuracy, without incorporating efficiency-related factors directly into the learning process [16]. Some methods also depend heavily on detailed reward annotations that are difficult to construct outside controlled benchmark environments. In addition, a number of existing experiments are conducted on limited task scales, making it difficult to determine whether the reported gains remain stable in larger and more diverse decision settings [17,18]. To address these limitations, this study proposes a performance-aware reinforcement learning framework for LLM-based decision agents. The framework introduces a multi-objective reward mechanism that jointly considers decision accuracy, computational cost, and response latency, so that the agent can learn to improve decision quality while remaining sensitive to execution efficiency. The proposed method is evaluated on 12,600 tasks covering financial analysis, operational planning, and scenario simulation, providing a broader testbed than that used in many prior studies. In addition to standard performance comparison, Pareto frontier analysis is employed to examine how different policies behave under competing objectives. The purpose of this study is to investigate whether performance-aware learning can produce LLM agents that are not only more accurate, but also more practical for deployment in complex decision environments. The significance of this work lies in moving LLM agent optimization beyond accuracy-centered training and toward a more balanced framework that better reflects the demands of real-world decision systems.

2.Materials and Methods

2.1. Samples and Study Scope

The experiments used 12,600 decision tasks from three domains: financial analysis, operational planning, and scenario simulation. Each task required the agent to process input information and produce a sequence of decisions. Financial tasks involved trend analysis, risk evaluation, and portfolio adjustment under uncertain conditions. Operational planning tasks included scheduling, resource allocation, and constraint handling. Scenario simulation tasks required multi-step reasoning with changing inputs and incomplete information. All tasks were generated under controlled conditions to cover different levels of difficulty, input size, and decision complexity. The dataset was divided into training, validation, and test sets, and no task appeared in more than one set.

2.2. Experimental Design and Baseline Settings

The proposed method used a reinforcement learning framework with a reward function that included multiple objectives. Its performance was compared with three baseline methods. The first baseline optimized only decision accuracy. The second baseline added fixed penalties for computation cost and response delay. The third baseline used supervised fine-tuning without reinforcement learning. All methods were built on the same base LLM and trained with similar computational budgets to ensure fair comparison. The experiments evaluated decision accuracy, computation time, and response latency. Each setting was repeated with different random seeds, and the final results were averaged.

2.3. Measurement and Quality Control

Three measures were used to evaluate model performance: accuracy, computation time, and response latency. Accuracy was defined as the proportion of correct decisions based on task-specific rules. Computation time was measured as the total processing time for each task. Response latency was defined as the delay between input and output. Each experiment was repeated five times under different initial conditions to improve reliability. Mean values and standard deviations were reported. Outliers were removed using an interquartile range rule. Hyperparameters such as learning rate and batch size were selected on the validation set and then fixed for all experiments.

2.4. Data Processing and Model Formulation

All input data were normalized before training to ensure a consistent scale. Text inputs were encoded by the LLM, and task features were converted into fixed-length vectors. The reward function combined multiple objectives:

$$R=w_1A-w_2C-w_3L$$

where A is decision accuracy, C is computation cost, L is response latency, and w_1,w_2,w_3 are weights. Policy learning followed the objective:

$$L=-E\left[\log \pi (a_t|s_t)\widehat{A}_t\right]$$

where $\pi(a_t|s_t)$ is the action probability and $\widehat{A}_t$ is the advantage estimate. The model parameters were updated using gradient descent to balance accuracy and efficiency.

2.5. Implementation and Evaluation Procedure

The model was implemented as a policy network built on a pre-trained LLM. Training used mini-batch updates over multiple episodes. At each step, the agent selected actions based on the current state and received feedback from the reward function. In the test stage, the model was evaluated on unseen tasks from all domains. Both single-step outputs and multi-step decision processes were examined. All methods were trained and tested under the same computational setting to ensure consistency. Final results were reported as mean values with standard deviations across repeated runs.

3.Results and Discussion

3.1. Overall decision performance

The proposed framework performed better than all baseline methods on the 12,600 tasks. Decision accuracy increased by 20.9%, which shows that the reward design helped the agent make better choices in complex tasks. The gain appeared in financial analysis, operational

planning, and scenario simulation, which suggests that the method was effective across different task types [19,20]. As shown in Fig.1, the proposed method achieved a better balance between decision quality and system cost than the comparison methods. This result is important because recent studies have shown that LLM agents may perform well on reasoning tasks, but their results are still unstable in interactive and multi-step environments.

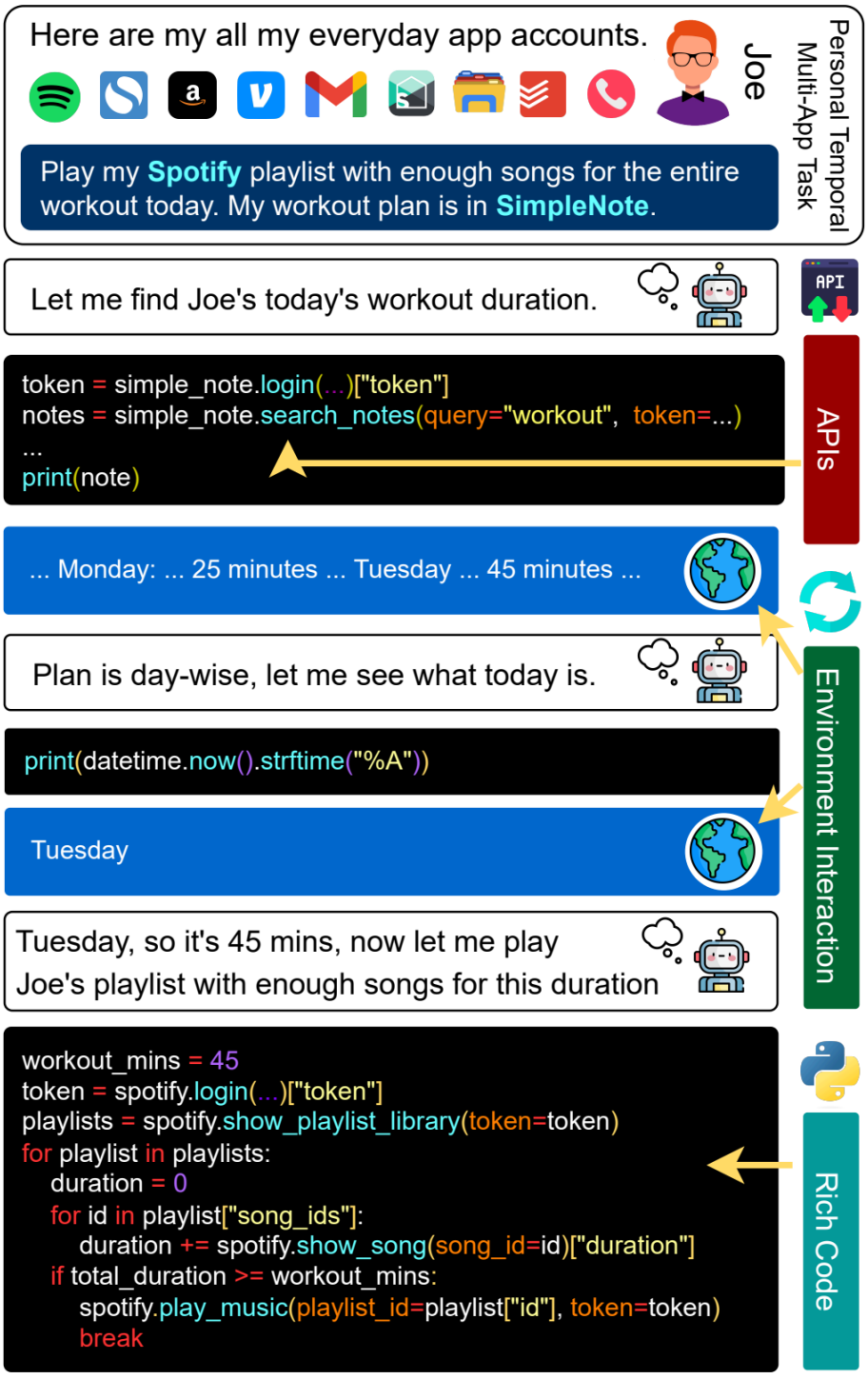


Figure 1Comparison of model performance, showing accuracy and efficiency across multiple task domains.

3.2. Efficiency improvement and latency reduction

The proposed method also reduced average computation time by 16.7%. This result suggests that the performance-aware reward guided the agent away from long and unnecessary reasoning paths. Lower latency is important in practical systems because many tasks require a timely answer, not only a correct one. As shown in Fig.2, the proposed framework moved the decision process to a better position in the performance-efficiency space. This finding agrees

with recent work showing that speed and accuracy should be studied together in LLM-based decision systems [21,22].

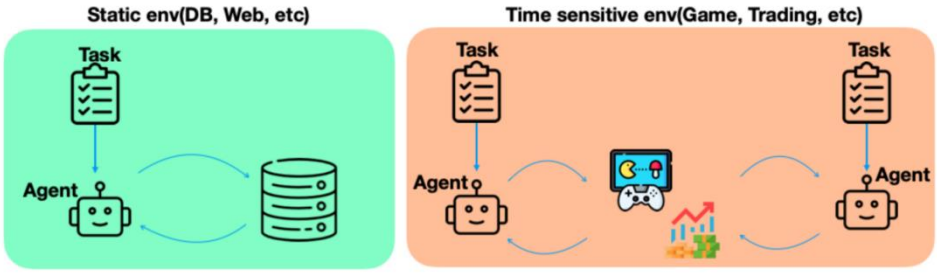


Figure 2Trade-off between computation time and decision accuracy for different methods.

3.3. Trade-off analysis and comparison with earlier studies

Pareto frontier analysis further showed that the proposed framework provided a better trade-off between decision accuracy and system cost than the baseline methods. This result is meaningful because many earlier studies focused mainly on task success, while computation time and latency were treated as secondary issues. In the present study, the Pareto frontier moved outward, which means that the model produced better decisions without a similar increase in computation cost. Compared with recent LLM-agent studies, the current framework places efficiency directly in the reward function instead of handling it only at the deployment stage [23,24]. This makes the training objective closer to real application needs.

3.4. Practical value, limitations, and future work

These results indicate that performance-aware reinforcement learning is a useful method for LLM-augmented decision agents in complex environments. A reward function that includes accuracy, computation cost, and latency can improve both decision quality and operating efficiency. This is important for practical systems in finance, planning, and simulation tasks, where time and resource limits often affect final value. However, several limitations remain. The experiments were conducted in controlled environments, and the reward weights were fixed during training. In real applications, the relative importance of speed and accuracy may change across users and task settings. In addition, current benchmarks still cover only part of the difficulty seen in real agent workflows. Future work should test adaptive reward weighting, broader task sets, and online settings with changing cost limits.

4.Conclusion

This study examined a performance-aware reinforcement learning framework for LLM-based decision agents in complex environments. The method used a reward function that combined decision accuracy, computation cost, and response latency. The results showed higher accuracy, lower computation time, and a better balance between quality and efficiency across financial analysis, operational planning, and scenario simulation tasks. These findings suggest that adding efficiency terms to the training objective can improve both decision quality and practical usability. The main contribution is a unified framework that treats accuracy and efficiency as joint goals, supported by evaluation on a large and diverse task set. The method can be applied in systems where both result quality and response speed are important, such as financial analysis tools, planning systems, and simulation-based decision support. However, the experiments were carried out in controlled environments, and the reward weights were fixed, which may limit flexibility in real applications. Future work should test adaptive reward settings and evaluate the method in real systems with changing conditions and resource limits.

References

Liu, S., Liu, X., & Feng, H. (2025, November). Research on AI-Driven Visual Design and Immersive Interactive Experiences Based on Multimodal Cognition and User. In Proceedings of the 2025 International Conference on Digital Society and Intelligent Computing (pp. 734-740).

Ferrag, M. A., Tihanyi, N., & Debbah, M. (2025). From llm reasoning to autonomous ai agents: A comprehensive review. arXiv preprint arXiv:2504.19678.

Xu, D., Liu, H., Qiu, D., & Ma, Q. (2026). Structured Modeling and Representation Methods for Post-Retrieval Inference Processes in Large Video Language Models.

Zanbouri, K., Noor-A-Rahim, M., John, J., Sreenan, C. J., Poor, H. V., & Pesch, D. (2024). A comprehensive survey of wireless time-sensitive networking (tsn): Architecture, technologies, applications, and open issues. IEEE Communications Surveys & Tutorials, 27(4), 2129-2155.

Zhang, Y., Gu, W., & Wang, J. (2026). Construction of Wind Farm Asset Health Index Based on Multi-Dimensional Indicators and Analytic Hierarchy Process and Its Correlation with Operational Performance. Authorea Preprints.

Milani, S., Topin, N., Veloso, M., & Fang, F. (2024). Explainable reinforcement learning: A survey and comparative review. ACM Computing Surveys, 56(7), 1-36.

Jiao, Y., Wang, A., Zhao, B., & Shi, T. (2026). The Impact of Visual Language Strategies in Public Art Creation on Community Spatial Perception and Public Behavior.

Singh, J., Magazine, R., Pandya, Y., & Nambi, A. (2025). Agentic reasoning and tool integration for llms via reinforcement learning. arXiv preprint arXiv:2505.01441.

Yue, L., Xu, D., Qiu, D., Shi, Y., Xu, S., & Shah, M. (2025, December). Sequential Cooperative Multi-Agent Online Learning and Adaptive Coordination Control in Dynamic and Uncertain Environments. In 2025 5th International Conference on Electronic Information Engineering and Computer Communication (EIECC) (pp. 692-697). IEEE.

Ibrahim, S., Mostafa, M., Jnadi, A., Salloum, H., & Osinenko, P. (2024). Comprehensive overview of reward engineering and shaping in advancing reinforcement learning applications. IEEE Access, 12, 175473-175500.

Gao, G., Ma, X., Lu, C., & Gao, R. (2026). Reliability Analysis and Application Research of SMS Communication Systems in Medical Notification Scenarios.

Mohammadi, M., Li, Y., Lo, J., & Yip, W. (2025, August). Evaluation and benchmarking of llm agents: A survey. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2 (pp. 6129-6139).

Qiu, D., Xu, D., & Yue, L. (2025, December). Reinforcement Learning-Augmented LLM Agents for Collaborative Decision Making and Performance Optimization. In 2025 7th International Conference on Frontier Technologies of Information and Computer (ICFTIC) (pp. 1337-1342). IEEE.

Axetorn, J., Edholm, F., Dobslaw, F., & Gren, L. (2026). Addressing trust requirements in the design of an open-source multi-agent LLM-based domain-specific chatbot. Requirements Engineering, 31(1), 3.

Xu, D., Gui, H., & Chen, H. (2026). Research on Layered Control and Fault Recovery Mechanisms for Fast Charging Safety Diagnosis of High Voltage Battery Systems Under Charging Network Interoperability Conditions.

Tholen, H. M., Mohamed, A. S., Wilson, V. I., Hall, D. M., & Najem, J. S. (2026). Design optimization of electric-fish-inspired power sources using statistical and machine learning approaches. Journal of Power Sources, 661, 238621.

Wang, Y., Chen, J., Wang, Y., & Yin, X. (2026). Application of Obtainable Biological Agent Characteristics in Efficacy Stratification of Oral Anti-Obesity Drugs.

Nabeel, M. Z. (2024). AI-enhanced project management systems for optimizing resource allocation and risk mitigation: Leveraging big data analysis to predict project outcomes and improve decision-making processes in complex projects. Asian Journal of Multidisciplinary Research & Review, 5(5), 53-65.

Jiao, Y., Wang, A., Zhao, B., & Shi, T. (2026). Quantitative Study on the Construction and Application Effectiveness of Graffiti Wall Painting Teaching Models in Public Space Contexts.

Cordova-Pozo, K., & Rouwette, E. A. (2023). Types of scenario planning and their effectiveness: A review of reviews. Futures, 149, 103153.

Wang, Y., Yin, X., Chen, J., & Wang, Y. (2026). Evidence-Based Study on Low-Burden Digital Phenotyping for Precision Screening of Oral Anti-Obesity Drug Efficacy.

Spatharioti, S. E., Rothschild, D., Goldstein, D. G., & Hofman, J. M. (2025, April). Effects of llm-based search on decision making: Speed, accuracy, and overreliance. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (pp. 1-15).

Zhang, Y., & Wang, J. (2026). Design and Implementation of a Computer-Aided Full Lifecycle Quality Management System for Wind Farms in Upgrades, Renovations, and Subcontractor Supervision.

Ferrag, M. A., Tihanyi, N., & Debbah, M. (2025). From llm reasoning to autonomous ai agents: A comprehensive review. arXiv preprint arXiv:2504.19678.