

LLM-Driven Multi-Agent Decision Systems with Reinforcement Learning-Based Adaptive Communication

James Smith¹, Emily Brown², Daniel Williams^{3*}

School of Computer Science and Engineering, University of New South Wales, Sydney, NSW
2052, Australia

Corresponding author: d.williams@unsw.edu.au

Abstract.

Effective communication among LLM-driven agents is critical for achieving high-quality collaborative decision-making. This study proposes an adaptive communication framework that leverages reinforcement learning to dynamically adjust message frequency, content granularity and routing strategies among agents. The system is trained on a multi-domain dataset comprising 9,800 tasks across negotiation, planning and data reasoning scenarios. A communication-aware reward function is designed to balance performance gains and communication cost. Experimental results demonstrate a 25.6% improvement in task success rate and a 30.2% reduction in communication overhead compared to fixed communication protocols. The proposed approach highlights the importance of adaptive communication policies in large-scale agent systems..

Keywords:

LLM agents; Adaptive communication;
Reinforcement learning;
Multi-agent systems;
Decision optimization

INTRODUCTION

Large language model (LLM)-based multi-agent systems have recently emerged as an important paradigm for collaborative problem solving in tasks such as planning, negotiation, and data reasoning. In these systems, multiple agents interact through natural language messages to exchange intermediate results, coordinate decisions, and refine collective outputs. Compared with conventional multi-agent systems, LLM-driven collaboration offers stronger reasoning flexibility and broader task applicability, but it also introduces new challenges related to communication efficiency. Because language-based interaction is costly in both computation and latency, uncontrolled message exchange can easily lead to repetitive responses, redundant information flow, and excessive inference overhead. These issues become more pronounced as the number of agents increases and as task processes become longer and more interdependent. Recent studies have shown that multi-agent LLM systems can improve task performance in complex collaborative settings, yet communication inefficiency remains a persistent bottleneck that limits scalability and practical deployment [1]. Recent studies on structured modeling of intermediate inference

processes show that explicitly organizing post-retrieval reasoning can significantly improve downstream decision quality, highlighting the importance of structured information flow in complex agent systems [2]. This line of evidence indicates that effective collaboration depends not only on the reasoning capability of individual agents, but also on how communication is structured, filtered, and adapted during the interaction process [3]. Research in multi-agent reinforcement learning has long recognized communication as a core component of collaborative intelligence. Existing studies have investigated communication from several directions, including structured message passing, communication topology design, and coordination under bandwidth or resource constraints [4,5]. These studies consistently show that communication can enhance coordination by allowing agents to share local observations and align their decisions. At the same time, they also demonstrate that excessive or poorly organized exchange may reduce efficiency, introduce instability, and even weaken policy performance. To address this issue, several methods have explored adaptive communication strategies that regulate interaction frequency, modify connectivity structure, or selectively activate communication links according to the current state of the environment [6,7]. Although these approaches provide useful insights, most of them were developed for conventional agents that communicate through compact symbols, latent vectors, or low-cost signals. Their assumptions do not fully match LLM-based settings, where messages are longer, semantically richer, and substantially more expensive to generate and process. Recent work has therefore begun to examine communication specifically in LLM-based multi-agent systems [8]. Current studies have focused on interaction protocols, message design, role allocation, and feedback-based collaboration mechanisms [9,10]. Benchmark results further suggest that communication quality is one of the main determinants of collaborative performance in large language agent systems [11]. Some approaches allow agents to refine subsequent decisions by incorporating communication-derived feedback, thereby improving task completion under complex reasoning requirements [12,13]. Even so, many existing frameworks still rely on fixed communication rules, such as predefined turn-taking schedules, static agent routing, or manually specified interaction templates. These designs can be effective in relatively simple tasks, but they often become inefficient when task requirements vary across domains or when the amount of relevant information changes over time. In heterogeneous environments, a fixed communication pattern may either underutilize useful coordination opportunities or encourage unnecessary exchanges that increase cost without improving decision quality. This limitation is especially important because communication in LLM systems is not merely a coordination channel; it is also a major source of computational expense. Each additional message may require prompt construction, context maintenance, decoding, and response integration, all of which contribute directly to runtime and resource consumption. In this sense, communication cost should not be treated only as an external implementation concern. It is part of the decision process itself and should be considered when designing collaborative policies. A communication strategy that improves raw task accuracy while generating large volumes of unnecessary interaction may not be suitable for real-world deployment, particularly in large-scale agent systems where message overhead accumulates quickly. As a result, efficiency and collaboration quality must be optimized jointly rather than independently. Current research still leaves several important gaps. Many studies evaluate communication strategies on a narrow range of tasks, which makes it difficult to judge whether the learned behavior generalizes across domains with different reasoning structures and coordination demands [14,15]. Communication-related factors such as message frequency, content granularity, and routing path are also often studied in isolation, even though these factors interact strongly in practice. A system that reduces message frequency, for example, may still

remain inefficient if it transmits overly detailed content or routes information through unnecessary agents. In addition, communication cost is commonly treated as a constraint or post hoc metric rather than being incorporated directly into the learning objective. This is a significant omission in LLM-based systems, where the cost of message generation is intrinsic to the collaboration mechanism. Another limitation is that many experimental settings remain relatively small or homogeneous, which restricts conclusions about scalability and real-world applicability [16,17]. These observations motivate the need for a more integrated approach to communication learning in LLM-driven multi-agent collaboration. Rather than relying on static communication templates, an effective system should be able to adjust whether to communicate, what level of detail to transmit, and where information should be routed according to the evolving task state. Such a mechanism would allow the communication policy to respond to differences in task complexity, agent dependency, and intermediate uncertainty, thereby improving both coordination quality and efficiency. From this perspective, adaptive communication is not simply an auxiliary control mechanism. It is a central component of scalable multi-agent reasoning, especially in environments where task demands and interaction patterns vary substantially. Based on this motivation, the present study investigates adaptive communication in LLM-driven multi-agent systems using reinforcement learning. The proposed framework dynamically adjusts message frequency, content granularity, and routing behavior according to the current task state, while introducing a communication-aware reward to jointly balance task performance and interaction cost. This design is intended to move beyond fixed communication rules and enable more flexible coordination policies that can better match the needs of heterogeneous collaborative tasks. The framework is evaluated on 9,800 tasks covering negotiation, planning, and data reasoning scenarios, allowing assessment across multiple types of interaction structure and reasoning demand. The purpose of this study is to examine whether reinforcement learning can be used to learn communication policies that are both effective and efficient in LLM-based multi-agent environments. Beyond improving task success rates, this work aims to clarify how adaptive control over communication frequency, message detail, and routing can reduce redundant exchange while preserving coordination quality. Methodologically, the study contributes an integrated perspective in which communication behavior is treated as an optimizable component of collaborative reasoning rather than a fixed protocol. Practically, it provides evidence for designing large-scale agent systems that can maintain strong collaborative performance without incurring unnecessary language interaction cost. In this way, the study contributes to the development of more scalable, communication-aware LLM multi-agent systems for complex real-world applications.

2. Materials and Methods

2.1. Task Samples and Environment Description

The experiments were conducted on a multi-domain task set containing 9,800 samples. The tasks covered negotiation, planning, and data reasoning scenarios. Each task required cooperation among 3–7 LLM-based agents. Agents exchanged textual messages to share local information and coordinate decisions. The environment was partially observable. Each agent could access only part of the task state at each step. Task settings differed in reasoning depth, communication dependence, and decision length. To reduce bias from repeated patterns, task conditions were changed across episodes. The full dataset was split into training, validation, and test sets at 70%, 15%, and 15%, respectively.

2.2. Experimental Design and Baseline Setting

Two communication settings were examined. The first used a fixed communication protocol. Under this setting, agents sent messages at preset intervals, and message structure and routing

remained unchanged during the task. The second used an adaptive communication policy. In this setting, communication decisions were selected by a reinforcement learning model. The model determined whether a message should be sent, how detailed it should be, and which agent should receive it. Both settings used the same LLM backbone, the same training budget, and the same optimization setup. This design kept the comparison fair and allowed the effect of adaptive communication to be examined directly. Each experiment was repeated five times with different random seeds, and the reported results were averaged across runs.

2.3. Measurement Methods and Quality Control

Three indicators were used for evaluation: task success rate, communication load, and communication efficiency. Task success rate was defined as the proportion of tasks completed within a fixed number of interaction rounds. Communication load was measured by the total number of exchanged tokens during one task. Communication efficiency was used to describe the balance between performance and cost. To reduce random effects, all results were averaged over repeated runs, and standard deviations were also reported. The same hardware and software environment was used in all experiments. Hyperparameter values were kept unchanged across different settings. Early stopping based on validation results was applied to reduce the risk of overfitting.

2.4. Data Processing and Model Formulation

Training data were organized as message-state-action-reward sequences collected from agent interaction. Text messages were converted into dense vector representations before being used by the policy model. Message length was normalized to reduce large variation in token count across tasks. The adaptive communication policy was trained to maximize task performance while limiting unnecessary exchange. The communication efficiency score was defined as

$$E_c = \frac{S}{1 + \gamma M}$$

where S denotes the task success indicator, M is the total communication cost, and γ is a scaling factor that controls the penalty on message usage.

Policy updates were carried out with PPO. The optimization target was written as

$$L(\theta) = E_t[\rho_t(\theta) \hat{A}_t - \alpha D_t]$$

where $\rho_t(\theta)$ is the policy ratio at t , $\hat{A}_t$ is the estimated advantage, D_t denotes the communication decision cost, and α controls the weight of the communication penalty. This form encourages useful communication while suppressing excessive message passing.

2.5. Training Procedure and Implementation Details

Training was performed for 1.4 million interaction steps. Communication trajectories were collected from parallel environments and stored for policy update. Mini-batches were sampled during training. The learning rate was set to 3×10^{-4} , and the discount factor was 0.99. The policy network was updated after each batch, while the underlying LLM reasoning module remained fixed. Parallel rollout was used to improve sampling speed and training stability. The full framework was implemented in PyTorch and trained on a GPU platform. After training, the learned policy was tested on unseen tasks with different communication demands to examine generalization ability.

3. Results and Discussion

3.1. Overall task performance

The adaptive communication policy performed better than the fixed protocol on all tasks. The gain was clear in negotiation, planning, and data reasoning settings. It was more evident in tasks that required multiple interaction steps. The task success rate increased by 25.6%. This result shows

that fixed communication rules are not suitable for dynamic task conditions. As shown in Fig. 1, the proposed method maintained a stable advantage across different domains, while the fixed protocol showed a drop in performance as task complexity increased. Similar patterns have been reported in recent work, where communication quality strongly affects coordination outcomes in multi-agent systems [18,19].

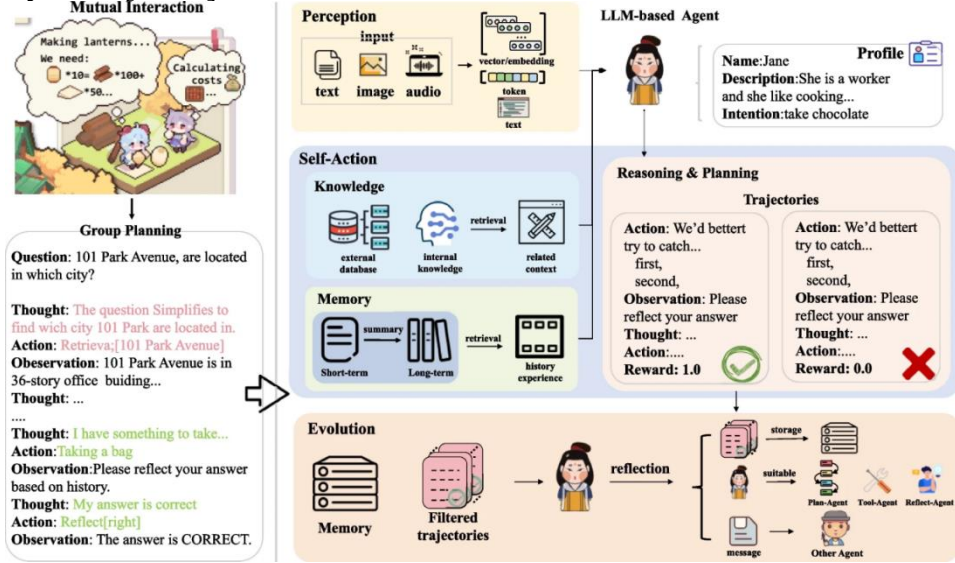


Figure 1. Task success rates for adaptive and fixed communication across tasks.

3.2. Communication overhead and efficiency

The proposed method reduced communication overhead while improving performance. Compared with the fixed protocol, communication cost decreased by 30.2%. This suggests that many messages in static settings do not contribute to task progress. The communication-aware reward helped limit unnecessary exchanges by adjusting message frequency and detail level. This is important in LLM-based systems, where message generation is costly. Earlier studies in communication-aware reinforcement learning also note that excessive communication can reduce efficiency [20,21]. The present results extend this observation to LLM-driven systems.

3.3. Effect of routing and content control

Performance gains were also linked to improved message routing and content control. The adaptive policy reduced unnecessary broadcasts and focused communication on relevant agents. It also adjusted message length based on task needs. As shown in Fig. 2, communication became more targeted, which reduced repeated exchanges and improved coordination consistency. This behavior is important in multi-step tasks, where poor routing can lead to information overload. Recent studies suggest that learned communication structure is more effective than fixed protocols. The current results support this view and show that routing and content decisions should be learned together [22].

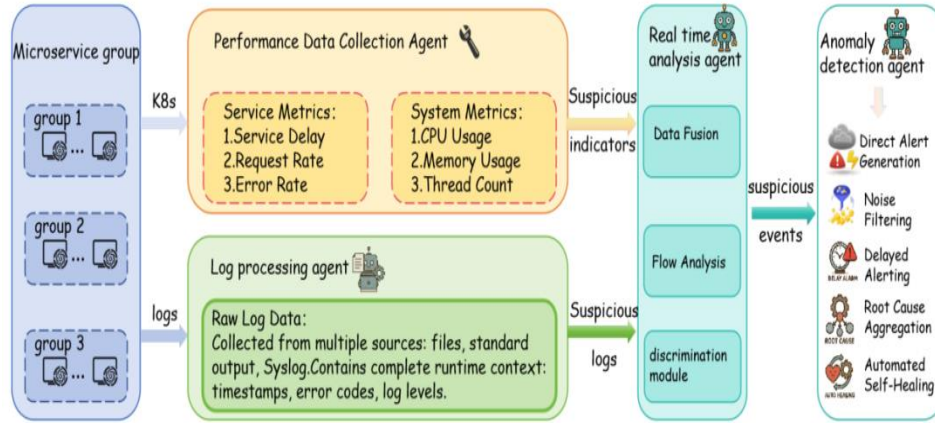


Figure 2. Message flow under the adaptive communication policy.

3.4. Comparison with existing studies and implications

Compared with existing studies, this framework treats communication as a joint decision problem. Many earlier methods focus on one aspect, such as communication topology or frequency, while other factors remain fixed. In this work, frequency, content, and routing are adjusted within one policy. The evaluation covers 9,800 tasks across different domains, which provides a broader test setting. However, some limits remain. The experiments were conducted in a simulated environment, and the LLM backbone was not updated during training. In addition, communication quality was measured mainly by task outcome and cost. Future work may include larger systems, dynamic model updates, and more detailed measures of message usefulness. Despite these limits, the results support adaptive communication as a practical way to improve coordination and reduce cost in LLM-based multi-agent systems.

4. Conclusion

This study examined adaptive communication in LLM-based multi-agent systems using a reinforcement learning framework. The policy adjusted when agents communicate, how much information they send, and where messages are directed. This setting improved coordination across tasks with different interaction patterns. Compared with fixed protocols, the method achieved a higher task success rate and lower communication cost. These results show that communication should be treated as part of the decision process rather than a fixed rule. The communication-aware reward helped balance task performance and message cost, which is important in systems where text generation is expensive. The approach can be used in tasks such as negotiation, planning, and data reasoning, where agents exchange information over multiple steps. However, some limitations remain. The experiments were conducted in a simulated environment, and the language model was not updated during training. Communication quality was evaluated mainly by task outcome and cost. Future work may consider larger systems, joint training of reasoning and communication, and more direct measures of message usefulness to improve efficiency and robustness.

References

- Shaikh, S. H. (2025, October). LLM-Based Multi-agent Systems: Frameworks, Evaluation, Open Challenges, and Research Frontiers. In International Joint Conference on Computational Intelligence (pp. 149-170). Cham: Springer Nature Switzerland.
- Xu, D., Liu, H., Qiu, D., & Ma, Q. (2026). Structured Modeling and Representation Methods for Post-Retrieval Inference Processes in Large Video Language Models.

- Schröder, F., Heinrich, F., & Kopp, S. (2025). Towards fluid human-agent collaboration: From dynamic collaboration patterns to models of theory of mind reasoning. *Frontiers in Robotics and AI*, 12, 1532693.
- Liu, S., & Yim, J. (2025). Research on Generative AI Creation Systems Based on Visual Language Modeling: Human-Machine Collaboration and Cognitive Feedback Mechanisms. Available at SSRN 6139770.
- Aboagye, S., Saeidi, M. A., Tabassum, H., Tayyar, Y., Hossain, E., Yang, H. C., & Alouini, M. S. (2024). Multi-band wireless communication networks: Fundamentals, challenges, and resource allocation. *IEEE Transactions on Communications*, 72(7), 4333-4383.
- 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.
- Sanjalawe, Y., Fraihat, S., Abualhaj, M., Makhadmeh, S., & Alzubi, E. (2025). A review of 6G and AI convergence: Enhancing communication networks with artificial intelligence. *IEEE Open Journal of the Communications Society*.
- Jiao, Y., Zhao, B., Wang, A., & Shi, T. (2026). Construction and Empirical Study of a Modularized Teaching System for Art Courses Based on a Unified Training Pathway.
- Yildiz Durak, H. (2024). Feedforward-or feedback-based group regulation guidance in collaborative groups. *Journal of Computer Assisted Learning*, 40(2), 410-436.
- 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.
- Marconi, L., Longo, L., & Cabitza, F. (2026). Assessing Interaction Quality in Human–AI Dialogue: An Integrative Review and Multi-Layer Framework for Conversational Agents. *Machine Learning and Knowledge Extraction*, 8(2), 28.
- Kausar, F., Muallem, A., Sattar, N. S., & Kurdi, M. Z. (2026). Integrating Multi-Agent Simulation, Behavioral Forensics, and Trust-Aware Machine Learning for Adaptive Insider Threat Detection. *arXiv preprint arXiv:2601.04243*.
- 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.
- Reith-Hall, E., & Montgomery, P. (2023). Communication skills training for improving the communicative abilities of student social workers: A systematic review. *Campbell systematic reviews*, 19(1), cl2-1309.
- Gao, G., Gao, R., Gao, R., & Zhou, H. (2026). Engineering Analysis and Quantitative Research on the Platform-Based Evolution of Enterprise Communication Systems.
- Iroegbu, A. O. C., Teffo, M. L., Sadiku, E. R., Meijboom, R., & Hlangothi, S. P. (2025). Advancing wastewater treatment with green and scalable metal–organic frameworks: from synthetic strategies to real-world deployment. *npj Clean Water*, 8(1), 85.
- Xu, D., Chen, H., & Gui, H. (2026). Unified Online Estimation Method for SOC, SOH, and Power Capacity Considering Safety Boundary Consistency in Battery Management Systems.
- Ponniah, J., & Dantsker, O. D. (2022). Strategies for scaleable communication and coordination in multi-agent (uav) systems. *Aerospace*, 9(9), 488.
- 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.

- Choi, Y., Di Marco, P., & Park, P. (2025). Communication-Aware Graph Neural Network for Multi-Agent Reinforcement Learning. IEEE Access.
- Zhang, Y., Gu, W., & Wang, J. (2025). Research on First Article Inspection (FAI)-Driven Quality Assurance Methods for Wind Turbine Installation and Operation & Maintenance and Their Effect on Reliability Improvement. Available at SSRN 6094206.
- Priyadarshi, R. (2024). Exploring machine learning solutions for overcoming challenges in IoT-based wireless sensor network routing: a comprehensive review. *Wireless Networks*, 30(4), 2647-2673.