Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Agentic Ai-Based Multi-Agent Reinforcement Learning for Autonomous Task Allocation and Collaborative Decision-Making

Dhavalkumar Thakar[1] , Mahesh kumar Gaddam[2]

[1]Senior Systems and Software Engineer, Capgemini America Inc, Sandy, Utah, USA, Email: d10thakar@gmail.com, Orcid ID – 0009-0002-6470-865X
[2]Enterprise Architect, Federal Express Corporation, Little Elm Dallas, USA, Email: mahi.ramuk@gmail.com, Orcid ID – 0009-0005-3938-5379

Abstract
Autonomous task allocation in multi-agent systems requires effective coordination among multiple agents with different capabilities, workloads, and operating conditions. Conventional task-allocation methods often depend on predefined rules or centralized decision mechanisms, which can limit adaptability in dynamic environments, while standard multi-agent reinforcement learning approaches may experience inefficient coordination, unstable learning, and limited high-level reasoning. This study proposes an Agentic AI-based multi-agent reinforcement learning framework for autonomous task allocation and collaborative decision-making. The framework combines task reasoning, agent capability assessment, task-agent matching, collaborative coordination, and adaptive policy learning to support efficient and context-aware decisions. A simulation-based environment is developed with heterogeneous agents, dynamically arriving tasks, varying task priorities, and changing workload conditions. The proposed framework is evaluated in terms of task allocation quality, coordination performance, learning performance, and decision efficiency using task completion rate, allocation success rate, coordination success rate, cumulative reward, convergence behavior, workload balance, communication overhead, and decision latency. The results show that the proposed method achieves a task completion rate of 94.6%, coordination success rate of 92.8%, and allocation success rate of 93.7%, while reducing decision latency by 21.4% compared with conventional MARL. These findings demonstrate the potential of Agentic AI-MARL for efficient, adaptive, and collaborative multi-agent decision-making.

Keywords: Agentic AI; Multi-Agent Reinforcement Learning; Autonomous Task Allocation; Collaborative Decision-Making; Multi-Agent Systems; Reinforcement Learning.

1. Introduction

Autonomous multi-agent systems are increasingly applied in dynamic environments where several intelligent agents must cooperate to complete tasks under changing priorities, workloads, and resource constraints. Effective task allocation is challenging because agents may differ in capability, availability, and workload, while tasks may vary in complexity and urgency. Conventional rule-based or centralized allocation methods often lack the flexibility required for such changing conditions.
Multi-Agent Reinforcement Learning (MARL) enables agents to learn task-allocation and coordination policies through repeated interaction with the environment. It supports adaptive and decentralized decision-making, but standard MARL methods may face slow convergence, unstable learning, communication overhead, and limited reasoning about task priority and agent suitability. Collaborative decision-making also requires effective information sharing, conflict resolution, and workload distribution among agents.

Agentic AI provides autonomous reasoning, planning, goal interpretation, and adaptive action selection, which can strengthen MARL-based task allocation. However, many existing approaches do not fully integrate high-level reasoning, agent selection, collaborative coordination, and reinforcement learning within a unified framework. This creates a research gap for an approach capable of jointly improving task allocation quality, coordination, learning performance, and decision efficiency in dynamic multi-agent environments.
The key contributions of this study include the development of an Agentic AI-based MARL framework for autonomous task allocation, an agent-selection mechanism that considers task requirements, agent capability, workload, and availability, and a collaborative coordination strategy for conflict resolution and joint decision-making. The framework also incorporates adaptive policy learning to improve task

completion, allocation success, coordination performance, convergence behavior, workload balance, communication efficiency, and decision latency.

2. Related Work

Multi-agent task allocation has been studied extensively in autonomous systems where agents must coordinate tasks according to capability, workload, resource availability, and task priority. Recent work has explored hierarchical planning and scheduling for robot-operated manufacturing systems [4], workforce-constrained task scheduling [5], and adaptive control across multiple robotic embodiments [15]. Cooperative interaction is also important when several agents share resources or pursue common objectives, and prior studies have shown that effective coordination can improve collective behavior and reduce decision conflicts in multi-agent environments [3].

Reinforcement learning has become an important approach for adaptive task allocation because it allows agents to improve their decisions through interaction with the environment. Cooperative MARL focuses on learning policies that support shared goals and coordinated behavior, with representation learning playing an important role in improving cooperation [1]. Policy-gradient methods such as PPO have also demonstrated strong performance in cooperative multi-agent games [14]. However, conventional MARL may still face non-stationarity, communication overhead, unstable convergence, and limited reasoning about dynamic task priorities and heterogeneous agent capabilities.

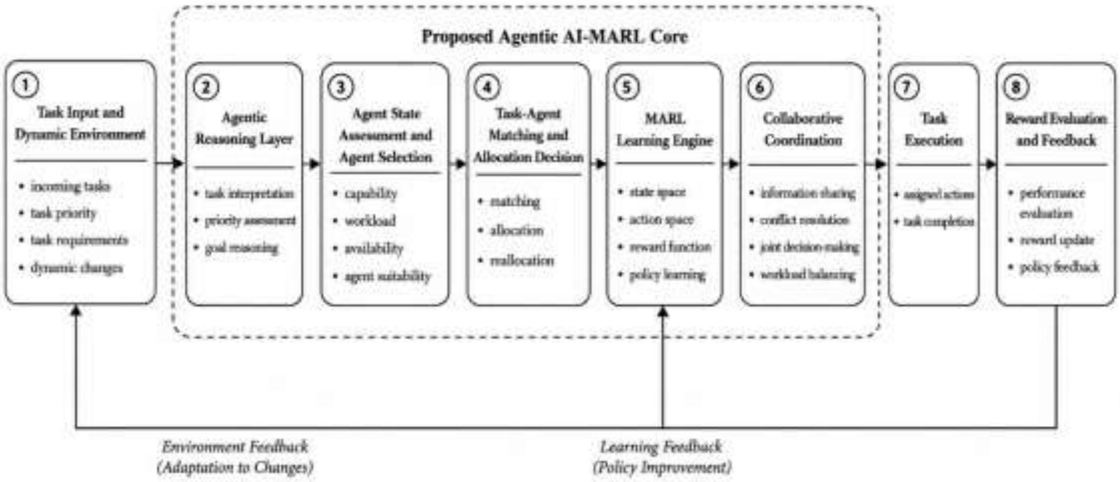
Agentic AI introduces autonomous reasoning, planning, reflection, communication, and tool-supported action, which can complement MARL-based decision-making. Generative agents with fast and slow reasoning have demonstrated improved performance in complex interactive tasks [6], while communicative multi-agent systems have shown the value of structured collaboration between specialized agents [9]. Tool-using language models [11], reflection-based reinforcement mechanisms [12], and deliberate reasoning strategies [13] further demonstrate how autonomous agents can plan, evaluate, and refine their actions. Similar capabilities are increasingly being explored in intelligent manufacturing, self-organizing multi-agent production, embodied decision-making, and human-robot collaboration [2], [7], [8], [10].

Despite these advances, an important research gap remains between cooperative MARL and Agentic AI-based autonomous reasoning. Existing MARL studies primarily emphasize policy learning and cooperation but often provide limited explicit reasoning about task priority, agent capability, workload, availability, and dynamic reallocation [1], [3], [14], [15]. In contrast, recent Agentic AI approaches provide stronger planning, communication, reflection, and reasoning capabilities but are not always integrated with reinforcement-based task allocation and quantitative coordination optimization [2], [6]-[13]. Scheduling-oriented studies also address task assignment and resource constraints but generally lack a unified integration of agentic reasoning and MARL [4], [5]. Therefore, the proposed framework combines Agentic AI reasoning with MARL to support capability-aware task allocation, collaborative coordination, adaptive policy learning, and efficient decision-making in dynamic multi-agent environments.

3. Proposed Agentic AI-Based MARL Framework

The proposed framework integrates Agentic AI with Multi-Agent Reinforcement Learning to support autonomous task allocation and collaborative decision-making in dynamic multi-agent environments. It combines task interpretation, agent-state assessment, task-agent matching, cooperative policy learning, and feedback-driven adaptation within a unified decision process. As illustrated in Fig. 1, incoming tasks are first analyzed by the agentic reasoning layer, after which suitable agents are identified according to their capabilities, workloads, availability, and current task requirements. The selected allocation is then processed through the MARL-based coordination mechanism, followed by task execution, reward evaluation, and continuous policy feedback.

Figure 1. Proposed Agentic AI-MARL Framework



3.1 Multi-Agent Task Environment

The multi-agent environment consists of a set of autonomous agents and dynamically arriving tasks. Each task is characterized by its priority, complexity, resource requirement, and execution demand, while each agent is represented by its capability, workload, current availability, and suitability for particular task categories. The system continuously observes changes in task requirements and agent states so that allocation decisions reflect the current operating condition rather than relying on fixed assignments. This representation enables the framework to assign urgent or resource-intensive tasks to agents with appropriate capability while avoiding excessive workload concentration.

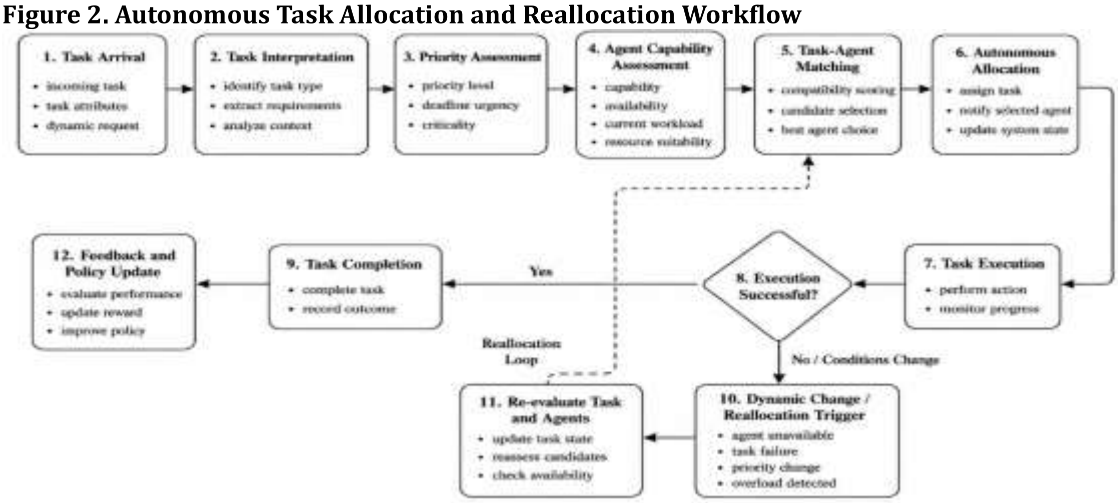
3.2 Agentic Reasoning and Task Allocation

The agentic reasoning mechanism interprets each incoming task and evaluates its priority, requirements, and expected resource demand before initiating allocation. Available agents are assessed according to capability, current workload, availability, and compatibility with the task. To quantify the suitability of agent i for task j, a task-agent suitability score is defined as

$$S_{ij} = w_1 C_{ij} + w_2 A_i + w_3 (1 - W_i) + w_4 P_j + w_5 R_{ij} \_\_\_\_\_\_\_ (1)$$

where S_{ij} represents the suitability score of agent i for task j, C_{ij} denotes capability compatibility, A_i represents agent availability, W_i is the current workload, P_j denotes task priority, R_{ij} represents task-requirement compatibility, and w_1–w_5 are weighting coefficients. A higher S_{ij} indicates a more suitable agent-task assignment. Based on this score, the framework identifies candidate agents and selects an allocation that balances task priority with operational efficiency. The complete reasoning and allocation sequence is presented in Fig. 2, including task interpretation, priority assessment, agent evaluation, task-agent matching, autonomous allocation, and dynamic reallocation. When an assigned agent becomes unavailable or task conditions change, the framework re-evaluates the current state and reallocates the task to a more suitable agent.

Figure 2. Autonomous Task Allocation and Reallocation Workflow



3.3 Multi-Agent Reinforcement Learning

The MARL component enables agents to improve their allocation and coordination policies through repeated interaction with the environment. The state space incorporates task characteristics, agent capability, workload, availability, and current allocation conditions, while the action space represents possible task-selection and assignment decisions. At each decision step, agent ii selects an action according to its current policy as

$$a_i^t \sim \pi_i(a_i^t \mid s_i^t) \qquad (2)$$

where a_i^t denotes the action selected by agent i at time t, s_i^t represents the corresponding state, and π_i denotes the learned policy of the agent. After each action, the environment provides a reward reflecting task completion, allocation quality, coordination effectiveness, execution efficiency, and communication cost. Individual agents update their policies using these rewards while considering the decisions of cooperating agents, allowing the system to progressively improve collective task-allocation behavior and adapt to changing operating conditions.

3.4 Collaborative Decision-Making

Collaborative decision-making enables agents to exchange relevant task and status information before finalizing assignments. The coordination mechanism reduces conflicting allocations by identifying cases in which multiple agents compete for the same task or when several tasks require shared resources. Conflicts are resolved by considering task priority, agent suitability, workload, and expected contribution to overall system performance. Joint decision-making also supports workload balancing by redistributing tasks when individual agents become overloaded, thereby improving coordination success and collaborative task completion while limiting unnecessary communication among agents.

3.5 Reward and Policy Optimization

The reward mechanism guides MARL policy learning toward effective and efficient collaborative behavior. Positive rewards are assigned for successful task completion, accurate task-agent matching, balanced workload distribution, efficient coordination, and reduced execution time, whereas penalties are applied for failed allocations, excessive communication, task conflicts, delays, and inefficient resource utilization. The overall reward at time t is formulated as

$$R_t = \alpha R_{TA} + \beta R_C + \gamma R_D - \lambda C_{comm} \qquad (3)$$

where R_{TA} represents task-allocation quality, R_C denotes coordination performance, R_D represents decision efficiency, and C_{comm} denotes communication cost. The coefficients α, β, γ, and λ control the contribution of each component to the overall reward. The policy is iteratively optimized using the accumulated reward so that agents increasingly prefer actions that improve task allocation quality and cooperative performance. By jointly considering allocation effectiveness, coordination, decision efficiency, and communication cost, the optimization process supports stable learning and efficient autonomous decision-making.

4. Experimental Setup

4.1 Simulation Configuration

The proposed Agentic AI-MARL framework was evaluated in a controlled simulation environment containing 20 heterogeneous agents and 500 dynamically arriving tasks distributed across five task categories and three priority levels. The agents differed in capability, workload, and availability, while tasks varied in urgency and execution requirements. Training was conducted for 1000 episodes using a learning rate of 0.001 and a discount factor of 0.95, with dynamic task arrival enabled throughout the simulation. Twenty independent simulation runs were performed to reduce the influence of random variation, and the same experimental conditions were maintained for all comparative approaches. The main simulation parameters and training settings are summarized in Table 1.

Table 1. Simulation Parameters and Training Configuration

Parameter	Value
Number of agents	20
Number of tasks	500
Task categories	5
Task-priority levels	3
Training episodes	1000

Learning rate	0.001
Discount factor	0.95
Dynamic task arrival	Enabled
Simulation runs	20

4.2 Baseline Methods and Comparative Evaluation

The proposed Agentic AI-MARL framework was compared with Random Task Allocation, Priority-Based Allocation, Independent Reinforcement Learning, and Conventional MARL. Random Task Allocation served as a basic reference by assigning tasks without considering capability or workload, while Priority-Based Allocation used predefined urgency-based rules. Independent Reinforcement Learning allowed agents to learn separately without explicit cooperation, whereas Conventional MARL enabled coordinated policy learning among agents. The proposed method additionally incorporated agentic reasoning, capability-aware agent selection, collaborative coordination, and dynamic reallocation. All methods were tested under identical simulation conditions and compared using task allocation quality, coordination performance, learning performance, and decision efficiency.

5. Results and Discussion

5.1 Task Allocation Quality

The proposed Agentic AI-MARL framework achieved the best task-allocation performance, reaching a 94.6% task completion rate and 93.7% allocation success rate, compared with 88.5% and 86.9% for Conventional MARL. The workload balance index also improved to 0.92, indicating more even task distribution among agents. The comparative results are summarized in Table 2 and visually supported in Fig. 4.

5.2 Coordination Performance

Table 2 shows that coordination performance improved progressively across the compared methods. Coordination success increased from 61.7% with Random Allocation to 69.8% with Priority-Based Allocation, 78.5% with Independent RL, 86.4% with Conventional MARL, and 92.8% with the proposed Agentic AI-MARL. Collaborative task completion similarly reached 93.2%, while the conflict rate decreased to 4.8%. Although communication overhead increased slightly from 2.8 to 3.1 messages per task compared with Conventional MARL, the higher coordination success and lower conflict rate indicate that the additional communication was beneficial.

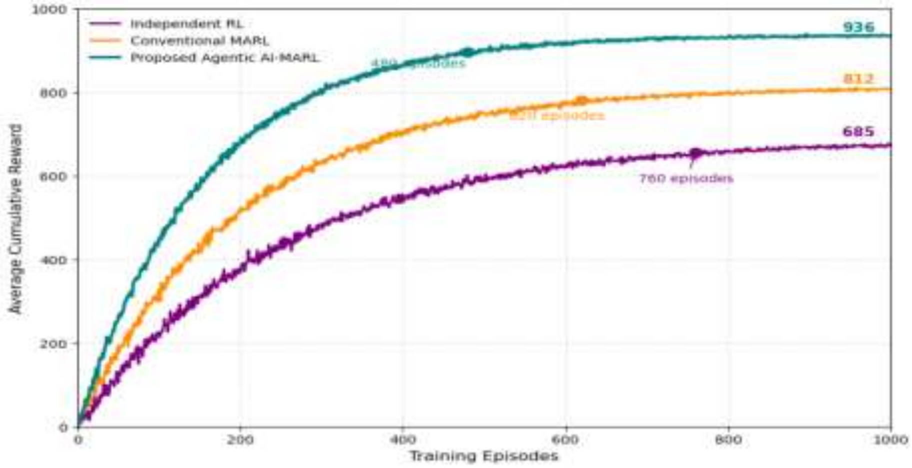
Table 2. Overall Performance Comparison of Methods

Metric	Random	Priority-Based	Independent RL	Conventional MARL	Agentic AI-MARL
Task Completion Rate (%)	71.8	78.9	83.7	88.5	94.6
Allocation Success Rate (%)	69.5	77.6	82.8	86.9	93.7
Coordination Success Rate (%)	61.7	69.8	78.5	86.4	92.8
Average Cumulative Reward	N/A	N/A	685	812	936
Decision Latency (ms)	68.5	61.2	57.6	53.8	42.3

5.3 Learning Performance

Fig. 3 demonstrates clear differences in learning speed and final reward among the reinforcement-learning methods. Independent RL converged at approximately 760 episodes with an average cumulative reward of 685, while Conventional MARL improved convergence to about 620 episodes and increased the reward to 812. The proposed Agentic AI-MARL converged much earlier, at approximately 480 episodes, and achieved the highest cumulative reward of 936. This corresponds to a reduction of about 140 convergence episodes compared with Conventional MARL and 280 episodes compared with Independent RL, indicating faster policy stabilization. The higher final reward and reduced convergence time suggest that agentic reasoning, collaborative state sharing, and coordinated task selection enabled the proposed framework to learn more effective allocation policies with less unnecessary exploration.

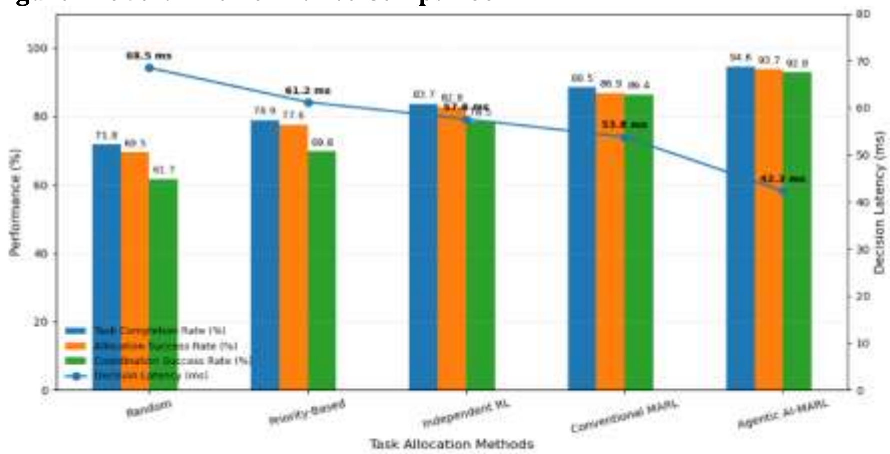
Figure 3. Average Cumulative Reward and Convergence Performance



5.4 Decision Efficiency

Fig. 4 shows that the proposed Agentic AI-MARL framework achieved the strongest overall performance across the main evaluation metrics. Task completion increased from 71.8% for Random Allocation, 78.9% for Priority-Based Allocation, 83.7% for Independent RL, and 88.5% for Conventional MARL to 94.6% for Agentic AI-MARL. Allocation success similarly improved from 69.5% to 93.7%, while coordination success increased from 61.7% to 92.8% across the compared methods. In contrast, decision latency decreased from 68.5 ms for Random Allocation to 61.2 ms, 57.6 ms, 53.8 ms, and finally 42.3 ms for the proposed framework. These results indicate that Agentic AI-MARL not only improves task allocation and coordination effectiveness but also reduces decision time, demonstrating a better balance between performance quality and execution efficiency.

Figure 4. Overall Performance Comparison



5.5 Overall Discussion

Overall, the results demonstrate that integrating Agentic AI with MARL improves autonomous task allocation, collaborative coordination, adaptive learning, and decision efficiency. Agentic reasoning improves task-agent matching and dynamic reallocation, while MARL supports continuous policy adaptation and collaborative decision-making. The main trade-off is slightly higher communication overhead, but this is compensated by improved completion, coordination, convergence, and latency performance. However, the findings remain simulation-based and require real-world validation before broader deployment claims can be made.

6. Conclusion

This study proposed an Agentic AI-based Multi-Agent Reinforcement Learning framework for autonomous task allocation and collaborative decision-making in dynamic multi-agent environments. The framework integrated task reasoning, capability-aware agent selection, cooperative coordination, adaptive policy learning, and dynamic reallocation to improve overall system performance. Simulation results demonstrated strong task allocation quality, achieving a 94.6% task completion rate and 93.7% allocation success rate, while coordination success reached 92.8%. The proposed method also achieved the highest average cumulative reward of 936 and converged at approximately 480 episodes, indicating faster and more stable learning than conventional MARL approaches. Decision latency was reduced to 42.3 ms, showing improved decision efficiency alongside stronger coordination and allocation performance. Overall, the results demonstrate that integrating Agentic AI reasoning with MARL can enhance adaptive, collaborative, and efficient task management. Future work will focus on validation in real-world multi-agent systems, heterogeneous autonomous agents, communication-constrained environments, and large-scale dynamic task scenarios.

References

1. Castellini, J. (2022). Improved Representations for Cooperative Multi-Agent Reinforcement Learning. The University of Liverpool (United Kingdom).
2. Chen, C., Li, L., Wang, T., Wang, Z., Leng, J., Shao, H., & Cheng, L. (2026). Leveraging large language models for smart manufacturing: Reviews, enablers, challenges, and opportunities. Journal of Manufacturing Systems, 85, 593-612.
3. Hughes, E., Anthony, T. W., Eccles, T., Leibo, J. Z., Balduzzi, D., & Bachrach, Y. (2020). Learning to resolve alliance dilemmas in many-player zero-sum games. arXiv preprint arXiv:2003.00799.
4. Li, C., Bhatta, K., Waseem, M., & Chang, Q. (2025). Demand-driven hierarchical integrated planning-scheduling control for a mobile robot-operated flexible smart manufacturing system. Robotics and Computer-Integrated Manufacturing, 95, 103015.
5. Li, J., Li, X., Gao, L., Wang, C., & Chen, H. (2025). A multitasking workforce-constrained flexible job shop scheduling problem: An application from a real-world workshop. Journal of Manufacturing Systems, 83, 196-215.
6. Lin, B. Y., Fu, Y., Yang, K., Brahman, F., Huang, S., Bhagavatula, C., ... & Ren, X. (2023). Swiftsage: A generative agent with fast and slow thinking for complex interactive tasks. Advances in Neural Information Processing Systems, 36, 23813-23825.
7. Liu, C., Tang, D., Zhu, H., Wang, L., Cai, Q., & Nie, Q. (2026). LLM-enhanced embodied multi-agent manufacturing system: A novel self-organizing production paradigm for embodied perception, embodied analysis and embodied decision. Journal of Manufacturing Systems, 84, 357-382.
8. Ma, Y., Tang, D., Zhu, H., Cai, Q., Zhang, Z., Wang, L., & Liu, C. (2026). Probing AR-assisted seamless HRC assembly for industry 5.0: Multi-modal mutual cognition and LLM-driven knowledge reasoning. Robotics and Computer-Integrated Manufacturing, 97, 103112.
9. Qian, C., Liu, W., Liu, H., Chen, N., Dang, Y., Li, J., ... & Sun, M. (2024, August). Chatdev: Communicative agents for software development. In Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers) (pp. 15174-15186).
10. Ren, Y., Liu, Y., Ji, T., & Xu, X. (2025). AI Agents and Agentic AI–navigating a plethora of concepts for future manufacturing. Journal of Manufacturing Systems, 83, 126-133.
11. Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Hambro, E., ... & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. Advances in neural information processing systems, 36, 68539-68551.
12. Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. Advances in neural information processing systems, 36, 8634-8652.
13. Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. Advances in neural information processing systems, 36, 11809-11822.
14. Yu, C., Velu, A., Vinitsky, E., Gao, J., Wang, Y., Bayen, A., & Wu, Y. (2022). The surprising effectiveness of ppo in cooperative multi-agent games. Advances in neural information processing systems, 35, 24611-24624.
15. Yu, C., Zhang, W., Lai, H., Tian, Z., Kneip, L., & Wang, J. (2022). Multi-embodiment legged robot control as a sequence modeling problem. arXiv preprint arXiv:2212.09078.