Enhancing LLM Reasoning with Multi-Path Collaborative Reactive and Reflection agents

Chengbo He[a], Bochao Zou[a,*], Xin Li[a], Jiansheng Chen[a], Junliang Xing[b,*], Huimin Ma[a,*]

[a] *University of Science and Technology Beijing, 30 Xueyuan Road, Haidian District, Beijing, 100083, Beijing, China*
[b] *Tsinghua University, 30 Shuangqing Road, Haidian District, Beijing, 100084, Beijing, China*

Abstract

Agents have demonstrated their potential in scientific reasoning tasks through large language models. However, they often face challenges such as insufficient accuracy and degeneration of thought when handling complex reasoning tasks, which impede their performance. To overcome these issues, we propose the Reactive and Reflection agents with Multi-Path Reasoning (RR-MP) Framework, aimed at enhancing the reasoning capabilities of LLMs. Our approach improves scientific reasoning accuracy by employing a multi-path reasoning mechanism where each path consists of a reactive agent and a reflection agent that collaborate to prevent degeneration of thought inherent in single-agent reliance. Additionally, the RR-MP framework does not require additional training; it utilizes multiple dialogue instances for each reasoning path and a separate summarizer to consolidate insights from all paths. This design integrates diverse perspectives and strengthens reasoning across each path. We conducted zero-shot and few-shot evaluations on tasks involving moral scenarios, college-level physics, and mathematics. Experimental results demonstrate that our method outperforms baseline approaches, highlighting the effectiveness and advantages of the RR-MP framework in managing complex scientific reasoning tasks.

Keywords:
Multi-agent systems, Human–Machine systems, Large language model

*Corresponding authors: Bochao Zou - Email: zoubochao@ustb.edu.cn; Junliang Xing - Email: jlxing@tsinghua.edu.cn; Huimin Ma - Email: mhmpub@ustb.edu.cn

1. Introduction

Large Language Models-based agents have demonstrated significant potential in scientific reasoning tasks. However, when faced with complex scientific reasoning challenges, these models often exhibit limited performance due to insufficient accuracy [1, 2, 3]. For instance, in tasks involving moral judgment or multi-level knowledge integration (such as university-level scientific problems), agents are capable of generating preliminary and comprehensible outputs but frequently struggle to provide comprehensive and accurate solutions [4, 5, 6]. Although step-by-step reasoning has somewhat enhanced the capabilities of agents [7], fundamental issues such as hallucination persist when addressing these complex tasks, leading agents to generate content that appears reasonable but is inherently illogical [8].

Relevant studies have proposed solutions, among which self-correction, the simplest form of post-hoc adjustment, has garnered significant attention in recent years [9, 10]. This approach leverages Large Language Models (LLMs) to generate feedback and optimize their own outputs, enabling LLMs to automatically rectify their generated content under zero-shot or few-shot prompts [11]. Although error detection is a prerequisite for self-correction, effectively implementing it remains a challenge. Previous research indicates that LLMs, similar to humans, do not always produce optimal outputs on the first attempt. Consequently, researchers have introduced the SELF-REFINE method, which assists agents in continuously improving their performance on specific tasks through iterative feedback and optimization [6] .However, despite the potential of self-reflection to enhance answer quality, it relies on the LLM's self-assessment capabilities, which have yet to be fully validated [12]. Moreover, the reflection process of a single agent may lead to the Degeneration-of-Thought (DoT). Specifically, once a Large Language Model-based agent establishes confidence in its responses, it becomes incapable of generating novel insights through subsequent self-reflection, even if its initial stance is erroneous [3].

We propose the Reactive and Reflection agents with Multi-Path Reasoning (RR-MP) framework to address the issues of insufficient accuracy and DoT faced by LLMs-based agents in complex scientific reasoning tasks. As illustrated in Figure 1, The RR-MP framework employs a multi-path reasoning mechanism, analogous to human reasoning—complex reasoning tasks

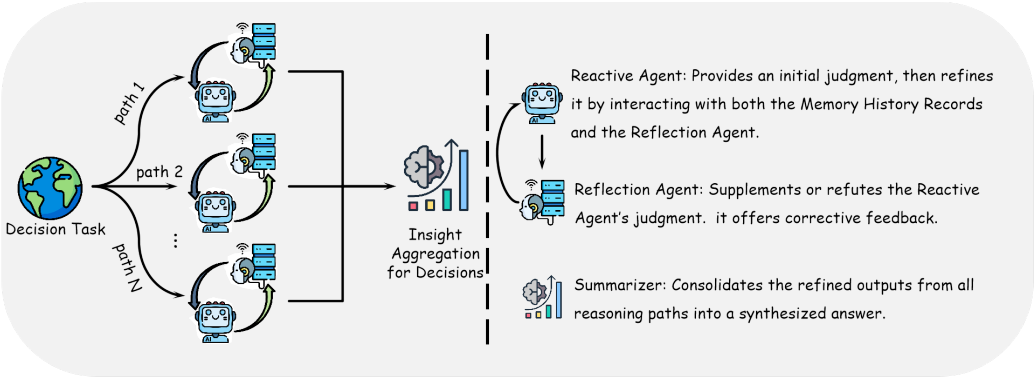

Figure 1: Reactive and Reflection agents with Multi-Path Reasoning

typically require multiple reasoning paths to arrive at correct answers [13]. In each path, to enhance reasoning accuracy, we optimize each pathway through iterative reflection, thereby preventing the occurrence of DoT during the iterative process. The framework integrates reactive agents and reflection agents working collaboratively; the reflection agents stimulate reactive agents to perform self-correction. The dual-system model, comprising reactive agents and reflection agents, is reminiscent of the two systems in human cognition—System 1 (fast and intuitive) and System 2 (slow and deliberative)—thereby effectively enhancing decision-making performance [14, 15, 16]. We validated our approach in zero-shot and few-shot scenarios across three complex scientific reasoning tasks—moral scenarios, college physics, and college mathematics. The results indicate that self-correction through external stimulation and optimization of reasoning paths achieves higher accuracy. Notably, our method significantly outperforms strong baseline methods in zero-shot settings [7, 17, 4, 18, 6].

In summary, our contributions are summarized as follows:

- We propose the Reactive and Reflection agents with Multi-Path Reasoning (RR-MP) framework as an effective post-hoc error-correction approach, aimed at significantly enhancing agents' reasoning capabilities in complex scientific tasks.

- We propose a multi-path reasoning mechanism that enables multiple reactive agents to generate parallel reasoning paths, thereby improving accuracy and robustness in complex scientific reasoning.

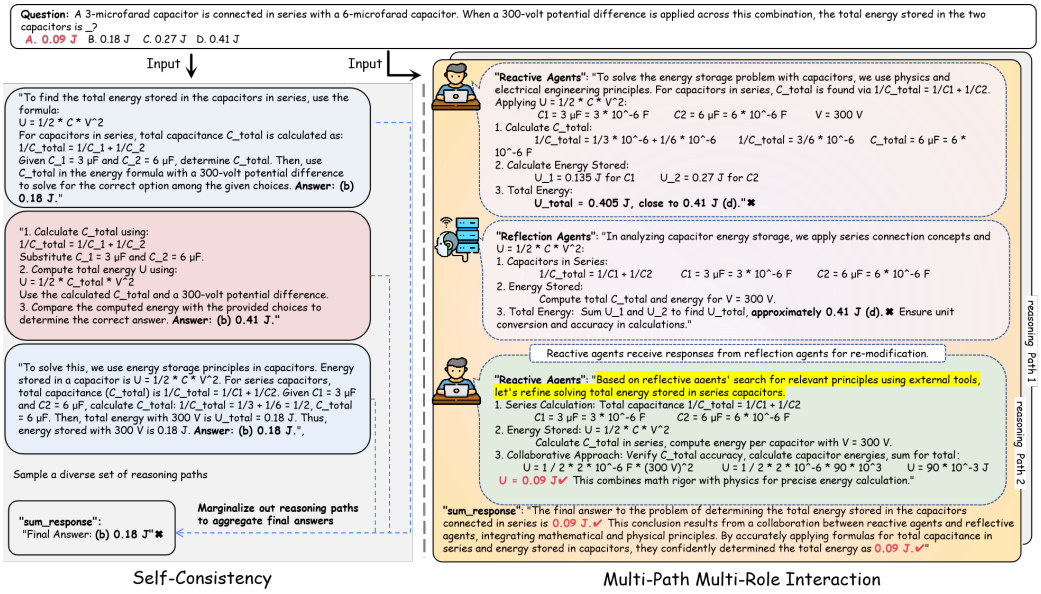

Figure 2: A comparison of the self-consistency method and our approach. Using the College Physics datasets as an example, our multi-path, multi-role interactive framework effectively mitigates errors caused by the majority of incorrect judgments in majority voting and leverages accurate reflection stimulated by the reactive agents' input on the reflection agents' reasoning. Even if the first path yields an incorrect result, the final answer is achieved through reflection analysis of the second path. Refer to B.1 in the appendix for details.

- We conduct a comparative analysis of the performance of reactive and reflection agents under various prompt types, and further investigate how different communication modes (e.g., collaboration and debate) influence scientific reasoning outcomes.

2. Related Work

Self-Correction of a Single agent.

Current LLMs still exhibit limitations in scientific reasoning, with accuracy often compromised due to hallucinations. Developing a simple yet effective approach to enhance the self-correction capabilities of intelligent agents remains a critical challenge [19, 20, 21, 22, 23]. Wei et al. [17] proposed the chain-of-thought method, which improves the model's complex reasoning ability through intermediate reasoning steps. Additionally, researchers have suggested decomposing complex problems into simpler subproblems to enable

LLMs to plan in a manner similar to the human brain [24, 25, 26]. These works lay the foundation for subsequent self-correction mechanisms. Wang et al. [18] introduced self-consistency decoding, which addresses repetition and local optimum issues in chain-of-thought prompts, reducing randomness during the generation process. Madaan et al. [6] proposed the SELF-REFINE method, where the agent first generates an output based on a given input and passes it back to the same model for feedback. The feedback is then returned to the model for optimization. This process iterates until a stopping condition is met. However, a single agent often lacks sufficient decision-making and planning abilities when dealing with complex tasks [4, 5]. One aspect of our work is to optimize iterative output and feedback among multiple agents, effectively avoiding the DoT that occurs during self-reflection in a single agent.

Collaborative Error Correction in Multi-agent Systems. The outputs of multi-agent systems can effectively correct errors, thereby enhancing the efficiency and accuracy of solving complex problems [27, 28]. A multi-agent system consists of multiple autonomous agents that interact with each other. By leveraging a shared environment or tasks, it facilitates the distributed resolution of decision-making problems. This collaborative approach can significantly improve the efficiency and accuracy of multi-agent systems in solving complex problems [29, 30, 31, 32]. For example, agents proficient in physical models can perform physical logical reasoning more effectively but are prone to calculation errors when dealing with formulas. In contrast, agents skilled in mathematical computations can reflect and correct the calculation structure of physical model agents, thereby solving complex university-level physics problems [33]. Additionally, critical interactions among agents are another effective pathway to enhance the ability of multi-agent systems to solve complex problems. Related studies have shown that utilizing multiple agents for critical debates can enhance problem-solving capabilities and mitigate the DoT through debate [34, 3]. Multi-agent systems based on LLMs have already demonstrated encouraging collective intelligence. However, current multi-agent systems still face limitations in demand responsiveness, as tasks are often handled by fixed agents, and feedback mechanisms for intermediate tasks remain insufficient. These shortcomings restrict the adaptability and decision-making efficiency of multi-agent systems in complex scenarios.

Our research is closely related to the field of multi-agent systems, with a focus on exploring the effectiveness of the RR-MP framework. We guide

LLMls to generate diverse reasoning paths, simulating the human experience of observing the world from different reasoning perspectives to derive accurate answers [14]. This enables multiple agents to dynamically collaborate and achieve diversified demand responsiveness, thereby improving system performance. To address the issue of insufficient feedback mechanisms, we design an interaction framework between reactive and reflection agents, enhancing the timeliness and effectiveness of reasoning feedback through collaborative correction and information sharing. This method leverages the collaborative capabilities of agents to achieve efficient self-correction and optimization.

3. Methods

We introduces our proposed RR-MP framework, which is divided into two parts. Section 3.1, *Multi-Path Reasoning for Enhanced Cognitive Flexibility*, demonstrates the effectiveness of multi-path reasoning through theoretical analysis. Section 3.2, *Multi-agent Interactions for Collaborative Cognitive Task Solving*, describes the communication mechanisms between reactive and reflection agents and provides a detailed analysis in the experimental section.

3.1. Multi-Path Reasoning for Enhanced Cognitive Flexibility

We adopt a multi-path reasoning approach to emulate the collaborative behavior of human teams. Specifically, when different members of the team produce consistent answers, it increases our confidence in the correctness of the solution. Unlike self-consistent methods that rely on aggregating multiple reasoning paths to achieve consensus [18], our approach not only integrates decision outcomes from multiple paths but also conducts in-depth analyses to derive the final decision. This enables the timely and effective evaluation and correction of the reasoning process, even when most initial paths are incorrect, potentially allowing the corrected answer to be output as the final result, as illustrated in Figure 2.

The core of our RR-MP framework is to achieve optimal solutions through diverse reasoning pathways. We assigned specific roles to the agents in each reasoning path to encourage collaboration among agents with diverse roles to solve the target task. This approach represents a simple yet effective prompting technique [35], and our design principles follow those of Chen et al. [33], with detailed implementation provided in Appendix B. By leveraging multiple diverse and reasonable reasoning paths generated by different

roles, we ultimately achieved the optimal solution. To validate this, we conducted a theoretical proof. Following Sel et al. [5], we view the reasoning process in complex problem-solving as an implicit optimization (a.k.a. mesa-optimization [36]) of the overall welfare function contributed by multi-path reasoning roles. We now perform a theoretical analysis of multi-path reasoning, assuming we have a problem datasets Q, an action space A, and a prompt system p. For a single query $q \in Q$, there is a specific action decision $a \in A$ that yields the optimal $F^S(q)$. We can consider the decision process for $F^S(q)$ as an implicit optimization process, where the function $F^S(q)$ represents the decision function F of the decision-maker S, who is responsible for making the final decision. We formalize this process as:

$$F^S(q) = \arg\max_{a \in A} \prod_{i=1}^{n} \mathbb{E}_{x \sim h^{\mathrm{m}_i}(q, F^{\mathrm{p}_i}(a))} h_u^{\mathrm{p}_i}(x) \tag{1}$$

where $h^m : Q \times A \to \mathcal{P}(\mathcal{X})$ serves as a logic generator within a multi-agent interaction in a multi-path Framework, inferring the logic of possible decision processes based on a given query $q \in Q$ and the prompt from the prompting system p. $\mathcal{P}(\mathcal{X})$ is the set of all probability distributions over the decision space $\mathcal{X}$. The term $\arg\max_{a \in A}$ represents maximizing the expected value of all paths under a specific action a. The symbol $\prod$ indicates the product over all possible paths, denoted by Π^n, where i represents the i-th path in the set and n is the total number of paths. The expectation operator $\mathbb{E}$ represents the expected value of the random variable x, and x is the random variable representing outcomes generated by different reasoning logics. The notation $\sim$ signifies that the distribution of x follows a probability distribution. The method $h^{\mathrm{m}}(q, F^{\mathrm{p_i}}(a))$ generates the random variable x for the question q using the decision $F^{\mathrm{p}}(a)$. The term $F^{\mathrm{p_i}}(a)$ denotes the optimal decision for the question q along the i-th path. The utility function $h_u^{\mathrm{p}}(x)$ represents the utility of the outcome x along this path. Overall, h represents the utility function, reflecting the effectiveness of the method, which is manifested in the correctness of the final answer. The symbol p denotes the specific path, and u represents the overall utility or effectiveness value of the method.

We assume the utility function $h_u^p(x)$ is consistent. Let $X_1^{q,a}, \ldots, X_n^{q,a}$ be i.i.d. samples from the distribution $h_s^p(q, a)$. The true utility $G_p(x)$ we want to optimize through the prompt system p is consistent, i.e., $\mathbb{E}\left[G_p(x)\right] = \mathbb{E}\left[\prod_{i=1}^{n} h_u^p(x_i)\right]$. Define the total variation distance between two distribu-

tions as $D_{TV}(Z_1 \| Z_2) = \sup_{A \subseteq Z} |Z_1(A) - Z_2(A)|$. We obtain the following inequality:

$$P\left(\left|\mathbb{E}_{x \sim h^{m_i}(q, F^{p_i}(q))} \prod_{i=1}^{n} h_u^{p_i}(x) - \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{n} h_u^{p_i}(X_i^{q,a})\right]\right| \geq t\right) \leq \frac{\sigma^2}{nt^2} \tag{2}$$

where, for any query $q \in Q$, any decision $a \in A$, and error bound $t \in \mathbb{R}^+$, can be defined as:

$$t = \|G\|_\infty D_{TV}\left[X^{q,a} \middle\| h^{\mathrm{m}_i}(q, a)\right] + \epsilon \tag{3}$$

where $\|G\|_\infty$ provides the maximum oscillation range of G under all inputs. $D_{TV}\left[X^{q,a} \middle\| h^{\mathrm{m}_i}(q, a)\right]$ gives the maximum discrepancy between the empirical distribution of the samples and the theoretical distribution. ϵ is a small adjustment parameter used to add extra tolerance for error.

Furthermore, we have the equation:

$$P\left(\left|\mathbb{E}_{x \sim h^{\mathrm{m}_i}(q, F^{\mathrm{p}_i}(q))} G_P(x) - \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{n} h_u^{\mathrm{p}_i}(x_i^{q,a})\right]\right| \geq \|G\|_\infty D_{TV}\left[X^{q,a} \middle\| h^{\mathrm{m}_i}(q, a)\right] + \epsilon\right) \leq \frac{\sigma^2}{nt^2} \tag{4}$$

Based on Chebyshev's inequality, as n increases, the probability that the deviation exceeds a fixed value t decreases, which means the probability of an error occurring decreases. The formula is as follows:

$$\left|\mathbb{E}_{x \sim h^{m_i}(q, F^{p_i}(q))} G_P(x) - \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \prod_{i=1}^{n} h_u^{p_i}(x_i^{q,a})\right]\right| \to 0 \quad \text{as} \quad n \to \infty \tag{5}$$

Therefore, we conclude that under the given assumptions, the optimization result of the formula $G_p(q)$ is proven through the aforementioned inequality. This demonstrates that by combining the expected utilities of different agents' paths and methods, we can identify the optimal decision that maximizes the utility function. Furthermore, we theoretically prove that as the number of agents increases, the generated multi-path reasoning significantly enhances decision quality. This conclusion is consistent with the experimental results of Wang et al. [18].

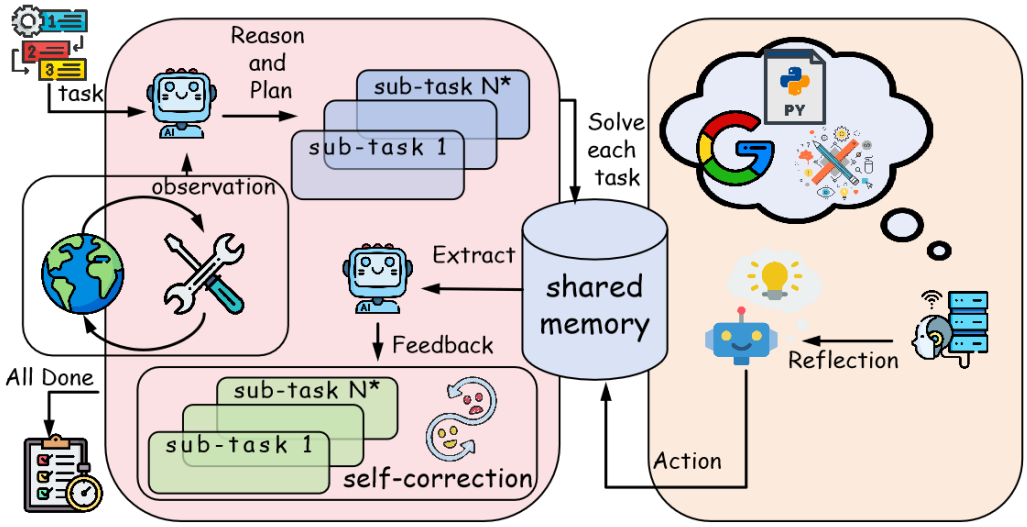

Figure 3: The reasoning process of the reactive agent and reflection agent. The reactive agent receives information from the external environment, decomposes it into sub-tasks, and stores them in the database. The reflection agent performs each sub-task through a process of supplementation or critique and returns the results to the reactive agent. Based on the feedback, the reactive agent refines its reasoning and completes the scientific reasoning process.

3.2. Multi-agent Interactions for Collaborative Cognitive Task Solving

In this chapter, we introduce the interaction process between reactive agents and reflection agents within a specific path of a multi-path reasoning framework. As illustrated in Figure 3, the primary interaction between the reactive agent and the reflection agent occurs through the shared memory module Shared memory (retrieved and stored in list format). The preliminary responses generated by the reactive agent are stored in shared memory , from which the reflection agent retrieves these responses for further analysis and processing. Before the final answer is obtained, the reactive agent awaits the completion of the reflection agent's analysis until the final answer is generated. The following sections provide a detailed description of this process.

The reactive agent maintains a partially observable understanding of the environment. Upon receiving a question datasets Q, it generates a specific action decision a'. Through its actions, the reactive agent produces a preliminary answer s' to address the problem, which is stored in memory as a dictionary entry, awaiting extraction by the reflection agent. Once the re-

flection agent retrieves the preliminary answer s' from the reactive agent, it undergoes multi-step reasoning and, with the assistance of relevant tools and external knowledge, formulates the reasoning strategy π.

The language-based agent selects and extends the initial action a' based on the strategy π implemented by a LLM with parameters Θ, adhering to a set of instructions p provided via prompts. The reflection agent's inputs include the instructions p, the preliminary response s', and the original question Q. We formalize this process as follows: during the update phase, the language-based agent selects an action $a \in A$ based on the strategy π implemented by the LLM with parameters Θ:

$$a \sim \pi(a'|p, s'; \Theta) \tag{6}$$

Consequently, after the interaction between the reactive agent and the reflection agent, the original action decision a' is expanded to action a, a process referred to as an "augmented action". By partially observing the task information b', and utilizing the LLM with parameters Θ to invoke tools or obtain information from external knowledge bases, and under the constraints of instructions p to formulate the final strategy π, the newly augmented action a is executed. This process effectively enhances decision-making performance.

4. Experiments

4.1. datasets and Baseline Methods

We select three datasets—College Physics, College Mathematics, and Moral Scenarios—from the Massive Multitask Language Understanding (MMLU) benchmark [37] to evaluate the performance of large language models in scientific reasoning tasks. The College Physics datasets evaluates mastery of domain-specific physical knowledge, the College Mathematics datasets focuses on logical reasoning and complex computation, and the Moral Scenarios datasets examines ethical decision-making and abstract reasoning. Together, these datasets capture the core requirements of scientific reasoning tasks and present significant challenges to large language models [4, 33]. With its broad coverage of key areas in scientific reasoning, the MMLU benchmark serves as a powerful tool for identifying model blind spots in domain knowledge, causal reasoning, and value-based judgment, offering a comprehensive evaluation of reasoning capabilities.

To evaluate the effectiveness of the *Reactive and Reflection agents with Multi-Path Reasoning* method in scientific reasoning tasks, we compared five baseline methods in zero-shot and few-shot settings. Each method represents a different paradigm of reasoning and decision-making for agents, as detailed below:

1. **Standard** [7]: This method simulates the traditional approach where the agent directly generates an output from the input without engaging in any reasoning or self-reflection. It is suitable for tasks that prioritize efficiency.
2. **Chain-of-Thought (CoT)** [17]: In this method, the agent performs step-by-step reasoning before making a decision and provides a detailed explanation of its reasoning process. This approach is particularly effective for complex decision-making tasks and mimics the human process of breaking down problems into sequential steps.
3. **Thought Experiment (Thought)** [4]: This method involves counterfactual reasoning, where the agent considers various (often hypothetical) scenarios and carefully analyzes the potential outcomes of these imagined situations, supporting more comprehensive decision-making.
4. **Self-Consistency** [18]: Instead of relying on a single greedy reasoning path, this method samples multiple reasoning paths. The final answer is determined by marginalizing over the sampled reasoning paths to select the most consistent solution.
5. **Self-Refine** [6]: This method is based on large language models (LLMs) and focuses on iterative self-improvement. The agent generates an initial output and then provides feedback on its own output, iteratively refining it to produce a more accurate result.

4.2. Settings

Due to resource limitations, we selected "gpt-3.5-turbo-0613" as the backbone model for all experiments. In our RR-MP framework, we designed an interaction framework between the reactive agent and the reflection agent. The reactive agent receives inputs from datasets, including College Physics, College Mathematics, and Moral Scenarios, makes initial decisions, and passes them to the reflection agent. The reflection agent further refines and optimizes these initial decisions through collaboration and debate, ensuring their accuracy and rationale. The two agents act as the "initial decision-maker" and the "decision optimizer," respectively, working together to complete tasks.

Method	Zero-shot			Few-shot			Average
	Moral Scenarios	College Physics	College Math	Moral Scenarios	College Physics	College Math	
Standard [7]	37.65	40.19	40	46.25	46.09	41	41.86
CoT [17]	48.49	57.84	39	52.29	63.72	38	48.22
Thought [4]	41.45	-	-	49.5	-	-	45.48
Self-Consistency [18]	<u>63.24</u>	65.68	53	**68.49**	62.75	53	61.03
Self-Refine [6]	59.66	61.76	50	67.01	66.67	45	58.35
Same-Domain Collaboration	**70.39**	85.29	<u>71</u>	63.91	86.27	**75**	<u>75.15</u>
Same-Domain Debate	48.71	<u>87.25</u>	70	62.12	<u>87.25</u>	**74**	71.55
Different-Domain Collaboration	60.78	**89.21**	**74**	<u>65.47</u>	**91.18**	**75**	**75.94**
Different-Domain Debate	59.77	85.29	74	56.76	86.27	70	72.02

Table 1: Main Results (Accuracy, %). "Same-Domain Collaboration" indicates that the reactive agent and reflection agent collaborate within the same domain to perform scientific reasoning, while "Different-Domain Debate" means they engage in debate across different domains. In the averages column, bold denotes the best result, and underline denotes the second-best result.

We tested the system in zero-shot and few-shot settings. In the zero-shot setting, the model relies entirely on its reasoning ability to make decisions without any prior examples. In the few-shot setting, five learning examples were provided for each agent to help them better understand the task context and decision-making logic. To further enrich the reasoning paths, we adopted a role-playing approach. For example, in the College Physics experiments, roles such as physicists and mathematicians were defined (based on simple prompt engineering). These roles, following design principles [33], explore diverse reasoning paths [35], with each role assuming specific tasks during the reasoning process and contributing to decision-making. The details of role definitions, task assignments, and the implementation of prompt engineering are thoroughly described in Appendix B.

4.3. Main Results

In the proposed RR-MP framework, we designed four interaction paradigms to investigate the interplay between reactive agents and reflection agents: the first is collaborative interaction between the reactive agent and reflection agent in a same-domain context; the second is debate interaction in a same-domain context; the third is collaborative interaction between the two agents in a different-domain context; and the fourth is debate interaction in a different-domain context. The comparison results with baseline methods are shown in Table 1, demonstrating that our approach achieves significant performance improvements in few-shot scenarios.

From the results in Table 1, it can be observed that the RR-MP framework exhibits significant performance improvements under the human-machine collaboration paradigm across complex datasets in both zero-shot and few-shot scenarios, including College Physics, College Mathematics, and Moral Scenarios. Notably, the collaboration between reactive agents and reflection agents from different domains (Different-Domain Collaboration) achieves the best performance in the majority of tasks, with an average accuracy of 75.94%, outperforming other baseline methods.

Furthermore, Table 1 reveals additional insights. For instance, collaboration modes generally outperform debate modes regardless of whether the agents are within the same domain or across different domains. This trend is consistently observed across multiple tasks in both zero-shot and few-shot settings. The collaboration mode aims to solve problems or reach consensus, enabling the integration of diverse perspectives, fostering a more comprehensive understanding, identifying blind spots, and preventing cognitive rigidity caused by debates. The study also finds that using reactive agents and reflection agents of the same type may lead to decreased performance when performing tasks. This is because when multiple agents of the same role collaborate, their thinking patterns and methodologies tend to converge, reducing diversity and innovation, thereby limiting performance on complex tasks.

In summary, the RR-MP framework significantly enhances the performance of complex reasoning tasks by designing flexible collaboration and debate modes and leveraging the diverse roles of reactive agents and reflection agents. The collaboration mode performs better in most scenarios, especially when integrating knowledge from different domains. Additionally, collaboration within the same domain can effectively facilitate task completion in specific tasks. These results validate the importance of multi-agent interaction design and provide strong support for the optimization of future multi-domain collaboration systems.

5. Ablation Study

5.1. Is It Necessary for reflection to Exist?

In our proposed method, the reflection agent serves as a core component of the RR-MP Framework. We posit that the reflection agent plays a crucial role in exploring reasoning pathways during the reflection phase, particularly when the reactive agent exhibits hallucinations or overconfidence in its

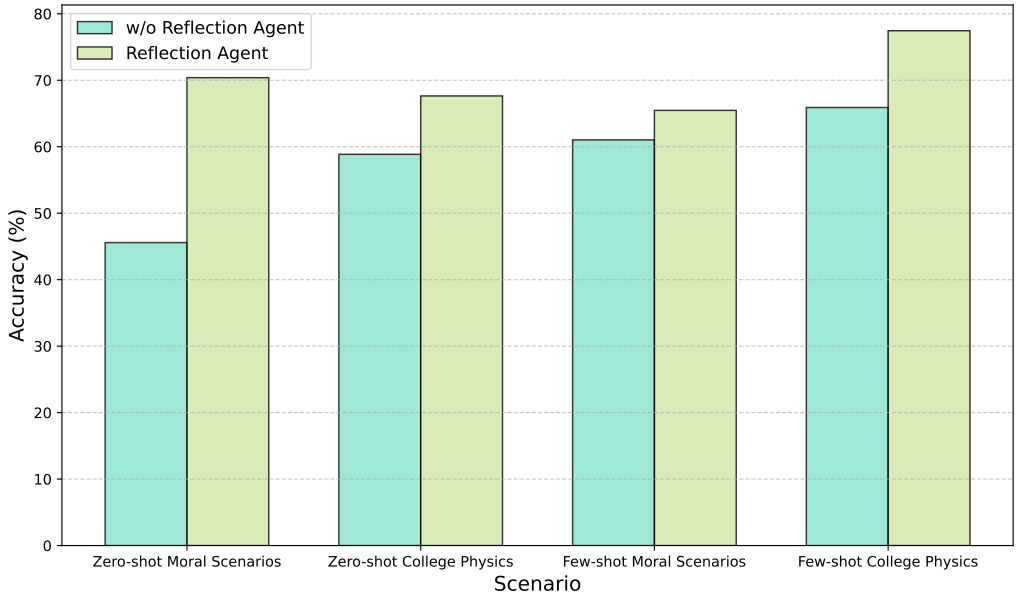

Figure 4: Accuracy (%) with and without stimulation roles.

Mode	College Physics (0-shot)		College Math (0-shot)		College Physics (few-shot)	
	Single	Multiple	Single	Multiple	Single	Multiple
Same-Domain Collaboration	78.43	**85.29**	69	**71**	79.41	**86.27**
Same-Domain Debate	86.27	**87.25**	67	**70**	89.11	87.25
Different-Domain Collaboration	85.29	**89.21**	71	**74**	85.29	**91.18**
Different-Domain Debate	83.30	**85.29**	70	**74**	84.31	**86.27**

Table 2: Performance comparison between single and multiple instances across different collaboration and debate modes.

reasoning. In such scenarios, the reflection agent facilitates further cognitive optimization, analogous to how humans rely on external stimuli to refine their thought processes after encountering overconfident errors. To validate this hypothesis, we designed comparative experiments to assess the difference in reasoning performance between models with and without the reflection agent. Under both zero-shot and few-shot prompting settings, we conducted reasoning tasks on the Moral Scenarios and College Physics datasets, respectively.

The experimental results are presented in Figure 4. Specifically, under the zero-shot prompting setting, the reasoning accuracy on the Moral Scenarios and College Physics datasets improved by 24.81% and 8.78%, respectively. Under the few-shot prompting setting, the accuracy increased by 4.44% and

Method	Zero-shot Accuracy (%)	Few-shot Accuracy (%)	Average Accuracy (%)
Linear	59.00	53.90	56.45
Hierarchical	63.72	57.80	60.76
Network	50.98	58.80	54.89
Ours	**89.21**	**91.18**	**90.20**

Table 3: Comparison of accuracy across three agents interaction methods and our proposed RR-MP framework Results are evaluated on zero-shot, few-shot, and average accuracy (%).

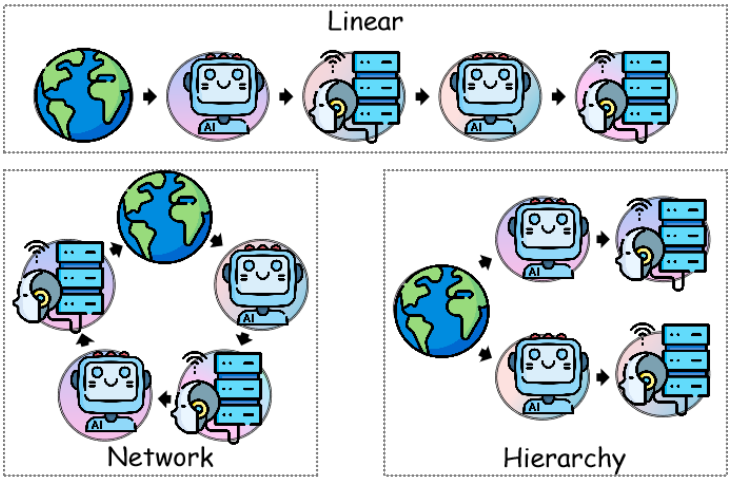

Figure 5: Three typical interaction paradigms in human-agent collaboration frameworks.

11.55%, respectively. These results indicate that incorporating the reflection agent significantly enhances the model's ability to handle complex reasoning tasks within these datasets. By introducing external stimulation to optimize reasoning pathways, the reflection agent can correct and augment cognitive processes, thereby ultimately achieving superior decision-making performance.

5.2. Are Multiple Instances Necessary?

In our proposed method, the reactive agent and reflection agent are both based on the same type of large language models (LLMs), such as ChatGPT-3.5, but operate as independent dialogue instances without interference. This design ensures that each agent can perform its designated tasks independently, avoiding reasoning biases caused by shared context or cross-agent in-

terference. To investigate whether it is possible to achieve multi-path reasoning by dynamically switching agents within a single-instance LLM through prompt engineering, we devised an alternative approach. This method simulates different agents within the same dialogue instance using prompt engineering and conducts four types of interaction experiments on the College Physics and Moral Scenarios datasets, as shown in Table 2.

The experimental results show that, regardless of zero-shot or few-shot prompting settings, the reasoning performance of single-instance dialogues decreases. This decline can be attributed to context conflicts or inconsistencies arising from frequent switching between agent modes, which negatively impact prediction accuracy. In contrast, multi-instance dialogues maintain consistency and independence among agents, significantly enhancing collaboration and improving reasoning performance. Moreover, single-instance dialogues are more costly, as they require frequent input of role-specific information, whereas multi-instance setups only require a single input for each agent.Our findings align with the studies of Xu et al. [38], Chen et al. [33], which emphasize the importance of clear definitions and independent task boundaries for each agent. Well-defined agent roles not only help maintain the self-consistency of LLMs but also effectively prevent cognitive confusion, thereby improving response quality and the reasoning ability to address complex scientific problems.

5.3. Exploring the Impact of Interaction Methods on agents.

In our study, we introduced three typical topological structures to explore interaction strategies in multi-agent systems: Linear Interaction, Network Interaction, and Hierarchical Interaction, as shown in Figure 5. These interaction methods are inspired by common patterns of human team collaboration. Linear Interaction is a sequential approach where agents process and transfer tasks along a fixed linear path, resembling workflows in assembly lines or hierarchical organizations. Network Interaction allows agents to establish arbitrary dependencies within a networked structure, reflecting the flexibility and dynamic adjustments often observed in team-based collaboration. Hierarchical Interaction adopts a layered structure where agents work in parallel across different branches, similar to team collaboration based on roles or functional hierarchies.

We conducted tests on the College Physics datasets. Experimental results (Table 3) demonstrate that, while Hierarchical Interaction exhibits relatively

well performance, our proposed RR-MP framework achieves significantly better results due to its reflection capability. reflection enables agents to dynamically adjust reasoning paths during interactions, effectively enhancing their ability to self-correct and optimize when addressing complex scientific problems. By combining reflection with multi-path reasoning, our method exhibits superior flexibility and efficiency across all scenarios, further validating the importance of reflection and dynamic interactions in the design of multi-agent systems.

6. Conclusion

In this work, we propose a framework named Reactive and Reflection agents with Multi-Path Reasoning. This framework aims to address the issues of decreased accuracy and Degeneration-of-Thought in multi-agent systems during complex scientific reasoning, which are caused by fixed single responses and insufficient execution of intermediate feedback. By doing so, it enhances the reasoning capabilities of LLMs in solving complex scientific problems. Our approach consists of two core components: first, the diversity of multi-path reasoning methods significantly improves the accuracy of LLMs; second, the interaction between multiple agents effectively mitigates hallucinations and Degeneration-of-Thought issues. We have demonstrated the effectiveness of the framework through both theoretical analysis and experimental validation.

Although the proposed framework serves as an effective post-hoc error correction method, significantly improving the decision-making capabilities of agents in complex tasks, it still has certain limitations. Specifically, the framework requires task-specific design of roles and reasoning examples, which is a common challenge in the field of prompt engineering [17, 4, 3, 6]. Future work will focus on exploring how to implement automated prompt design within the framework to further enhance the method's generalizability and adaptability.

References

[1] J. Feng, Q. Wang, H. Qiu, L. Liu, Retrieval in decoder benefits generative models for explainable complex question answering, Neural Networks 181 (2025) 106833.

[2] X. Zhang, F. Zeng, C. Gu, Simignore: Exploring and enhancing multimodal large model complex reasoning via similarity computation, Neural Networks (2024) 107059.

[3] T. Liang, Z. He, W. Jiao, X. Wang, Y. Wang, R. Wang, Y. Yang, Z. Tu, S. Shi, Encouraging divergent thinking in large language models through multi-agent debate, arXiv preprint arXiv:2305.19118 (2023).

[4] X. Ma, S. Mishra, A. Beirami, A. Beutel, J. Chen, Let's do a thought experiment: Using counterfactuals to improve moral reasoning, arXiv preprint arXiv:2306.14308 (2023).

[5] B. Sel, P. Shanmugasundaram, M. Kachuee, K. Zhou, R. Jia, M. Jin, Skin-in-the-game: Decision making via multi-stakeholder alignment in llms, arXiv preprint arXiv:2405.12933 (2024).

[6] A. Madaan, N. Tandon, P. Gupta, S. Hallinan, L. Gao, S. Wiegreffe, U. Alon, N. Dziri, S. Prabhumoye, Y. Yang, et al., Self-refine: Iterative refinement with self-feedback, Advances in Neural Information Processing Systems 36 (2024).

[7] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Advances in neural information processing systems 33 (2020) 1877–1901.

[8] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, et al., A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions, ACM Transactions on Information Systems (2023).

[9] L. Pan, M. Saxon, W. Xu, D. Nathani, X. Wang, W. Y. Wang, Automatically correcting large language models: Surveying the landscape of diverse automated correction strategies, Transactions of the Association for Computational Linguistics 12 (2024) 484–506.

[10] J. Lu, W. Zhong, W. Huang, Y. Wang, F. Mi, B. Wang, W. Wang, L. Shang, Q. Liu, Self: Language-driven self-evolution for large language model, arXiv preprint arXiv:2310.00533 (2023).

[11] Z. Jiang, H. Peng, S. Feng, F. Li, D. Li, Llms can find mathematical reasoning mistakes by pedagogical chain-of-thought, arXiv preprint arXiv:2405.06705 (2024).

[12] N. Shinn, F. Cassano, A. Gopinath, K. Narasimhan, S. Yao, Reflexion: Language agents with verbal reinforcement learning, Advances in Neural Information Processing Systems 36 (2024).

[13] K. E. Stanovich, R. F. West, Advancing the rationality debate, Behavioral and brain sciences 23 (2000) 701–717.

[14] D. Kahneman, Thinking, fast and slow, Farrar, Straus and Giroux (2011).

[15] L. S. Vygotsky, M. Cole, Mind in society: Development of higher psychological processes, Harvard university press, 1978.

[16] K. Christakopoulou, S. Mourad, M. Matarić, Agents thinking fast and slow: A talker-reasoner architecture, arXiv preprint arXiv:2410.08328 (2024).

[17] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al., Chain-of-thought prompting elicits reasoning in large language models, Advances in neural information processing systems 35 (2022) 24824–24837.

[18] X. Wang, J. Wei, D. Schuurmans, Q. Le, E. Chi, S. Narang, A. Chowdhery, D. Zhou, Self-consistency improves chain of thought reasoning in language models, arXiv preprint arXiv:2203.11171 (2022).

[19] W. Saunders, C. Yeh, J. Wu, S. Bills, L. Ouyang, J. Ward, J. Leike, Self-critiquing models for assisting human evaluators, arXiv preprint arXiv:2206.05802 (2022).

[20] P. Chen, Z. Guo, B. Haddow, K. Heafield, Iterative translation refinement with large language models, arXiv preprint arXiv:2306.03856 (2023).

[21] X. Feng, Z.-Y. Chen, Y. Qin, Y. Lin, X. Chen, Z. Liu, J.-R. Wen, Large language model-based human-agent collaboration for complex task solving, arXiv preprint arXiv:2402.12914 (2024).

[22] Y. Wang, W. Zhong, L. Li, F. Mi, X. Zeng, W. Huang, L. Shang, X. Jiang, Q. Liu, Aligning large language models with human: A survey, arXiv preprint arXiv:2307.12966 (2023).

[23] N. Yax, H. Anlló, S. Palminteri, Studying and improving reasoning in humans and machines, Communications Psychology 2 (2024) 51.

[24] S. Hao, Y. Gu, H. Ma, J. J. Hong, Z. Wang, D. Z. Wang, Z. Hu, Reasoning with language model is planning with world model, arXiv preprint arXiv:2305.14992 (2023).

[25] D. Zhou, N. Schärli, L. Hou, J. Wei, N. Scales, X. Wang, D. Schuurmans, C. Cui, O. Bousquet, Q. Le, et al., Least-to-most prompting enables complex reasoning in large language models, arXiv preprint arXiv:2205.10625 (2022).

[26] T. Khot, D. Khashabi, K. Richardson, P. Clark, A. Sabharwal, Text modular networks: Learning to decompose tasks in the language of existing models, arXiv preprint arXiv:2009.00751 (2020).

[27] T. Guo, X. Chen, Y. Wang, R. Chang, S. Pei, N. V. Chawla, O. Wiest, X. Zhang, Large language model based multi-agents: A survey of progress and challenges, arXiv preprint arXiv:2402.01680 (2024).

[28] X. Zhu, J. Li, Y. Liu, C. Ma, W. Wang, Distilling mathematical reasoning capabilities into small language models, Neural Networks 179 (2024) 106594.

[29] S. Rasal, Llm harmony: Multi-agent communication for problem solving, arXiv preprint arXiv:2401.01312 (2024).

[30] X. Zhu, J. Wang, L. Zhang, Y. Zhang, Y. Huang, R. Gan, J. Zhang, Y. Yang, Solving math word problems via cooperative reasoning induced language models, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 4471–4485. URL: https://aclanthology.org/2023.acl-long.245. doi:10.18653/v1/2023.acl-long.245.

[31] L. Pan, Y. Li, C. Yu, Y. Shi, A human-computer collaborative tool for training a single large language model agent into a network through few examples, arXiv preprint arXiv:2404.15974 (2024).

[32] J. He-Yueya, G. Poesia, R. E. Wang, N. D. Goodman, Solving math word problems by combining language models with symbolic solvers, arXiv preprint arXiv:2304.09102 (2023).

[33] P. Chen, B. Han, S. Zhang, Comm: Collaborative multi-agent, multi-reasoning-path prompting for complex problem solving, arXiv preprint arXiv:2404.17729 (2024).

[34] Y. Du, S. Li, A. Torralba, J. B. Tenenbaum, I. Mordatch, Improving factuality and reasoning in language models through multiagent debate, arXiv preprint arXiv:2305.14325 (2023).

[35] Z. M. Wang, Z. Peng, H. Que, J. Liu, W. Zhou, Y. Wu, H. Guo, R. Gan, Z. Ni, J. Yang, et al., Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models, arXiv preprint arXiv:2310.00746 (2023).

[36] E. Hubinger, C. van Merwijk, V. Mikulik, J. Skalse, S. Garrabrant, Risks from learned optimization in advanced machine learning systems, arXiv preprint arXiv:1906.01820 (2019).

[37] D. Hendrycks, C. Burns, S. Basart, A. Zou, M. Mazeika, D. Song, J. Steinhardt, Measuring massive multitask language understanding, arXiv preprint arXiv:2009.03300 (2020).

[38] B. Xu, A. Yang, J. Lin, Q. Wang, C. Zhou, Y. Zhang, Z. Mao, Expertprompting: Instructing large language models to be distinguished experts, arXiv preprint arXiv:2305.14688 (2023).