

Cost-Effective Communication: An Auction-based Method for Language Agent Interaction

Yijia Fan^{1*}, Jusheng Zhang^{1*}, Kaitong Cai¹, Jing Yang¹,
Chengpei Tang^{1†}, Jian Wang², Keze Wang^{1,3†}

¹Sun Yat-sen University

²Snap Inc.

³Guangdong Key Laboratory of Big Data Analysis and Processing

Abstract

Multi-agent systems (MAS) built on large language models (LLMs) often suffer from inefficient “free-for-all” communication, leading to exponential token costs and low signal-to-noise ratios that hinder their practical deployment. We challenge the notion that more communication is always beneficial, hypothesizing instead that the core issue is the absence of resource rationality. We argue that “free” communication, by ignoring the principle of scarcity, inherently breeds inefficiency and unnecessary expenses. To address this, we introduce the **Dynamic Auction-based Language Agent (DALA)**, a novel framework that treats communication bandwidth as a scarce and tradable resource. Specifically, our DALA regards inter-agent communication as a centralized auction, where agents learn to bid for the opportunity to speak based on the predicted *value density* of their messages. Thus, our DALA intrinsically encourages agents to produce concise, informative messages while filtering out low-value communication. Extensive and comprehensive experiments demonstrate that our economically-driven DALA achieves new state-of-the-art performance across seven challenging reasoning benchmarks, including 84.32% on MMLU and a 91.21% pass@1 rate on HumanEval. Note that this is accomplished with remarkable efficiency, i.e., our DALA uses only 6.25 million tokens, a fraction of the resources consumed by current state-of-the-art methods on GSM8K. Further analysis reveals that our DALA cultivates the emergent skill of strategic silence, effectively adapting its communication strategies from verbosity to silence in a dynamical manner via resource constraints.

Introduction

The research for Large Language Models (LLMs)(Zhao and et al. 2025; Zhang et al. 2025f) is rapidly shifting from single-agent(Wang, Ma, and et al. 2024; Yao et al. 2023) solvers to complex Multi-Agent Systems (MAS)(Chen et al. 2024; Zhang et al. 2025b,d; Fan et al. 2025), which hold the promise of tackling problems beyond the reach of any individual model. This collaborative approach has shown initial success in domains like complex reasoning(Li et al. 2025a) and software development(Huang et al. 2024), where

agents can refine solutions by pooling distributed knowledge and engaging in iterative dialogue. However, these advancements often rely on “free-for-all” communication protocols, which permit agents to broadcast information at a low cost. This architectural reliance leads to an exponential increase in token consumption, inundating agents with low-value, verbose exchanges. Existing methods, such as PHP(Zheng et al. 2024) and DyLAN(Liu et al. 2024), despite their performance gains, overlook the resource cost of communication, resulting in high operational overhead and a low signal-to-noise ratio. For instance, on the GSM8K benchmark(Cobbe, Kosaraju, and Bavarian 2021), these methods consume tens of millions of tokens, failing to achieve efficient deployment. In contrast to the current research that has predominantly focused on enhancing individual agent capabilities(Zhou and et al. 2024; Zhang et al. 2025c; Li et al. 2025b; Wen et al. 2025) or simple message pruning, the optimal design of communication protocols, i.e., shaping *how* and *when* agents should communicate, might also be beneficial and promising.

To achieve this objective, we challenge a fundamental question: the bottleneck in MAS performance is not a lack of communication, but the absence of *resource rationality* in its budget management. We hypothesize that by treating communication bandwidth as a scarce, common resource and compelling agents to operate within a simulated market, the MAS can self-organize to prioritize concise, high-value information exchange. The core insight motivating this hypothesis is that “free” communication, by ignoring resource scarcity, breeds inefficiency; introducing incentive mechanisms can cultivate emergent rational intelligence, thereby balancing performance with efficiency of the MAS.

To validate this hypothesis, we introduce the **Dynamic Auction-based Language Agent (DALA)**, a novel framework that integrates principles of scarce resource allocation. Specifically, our DALA redefines inter-agent communication as a centralized auction process, such that agents must “pay to speak” and the entire framework is trained via the MAPPO strategy(Yu et al. 2022). Agents formulate bids based on the predicted value density of their potential messages, which mechanistically incentivizes them to generate concise and informative statements to win the auction. This mechanism intrinsically prunes low-value chatter, fostering

*These authors contributed equally.

†Corresponding author: kezewang@gmail.com and tchengp@mail.sysu.edu.cn

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

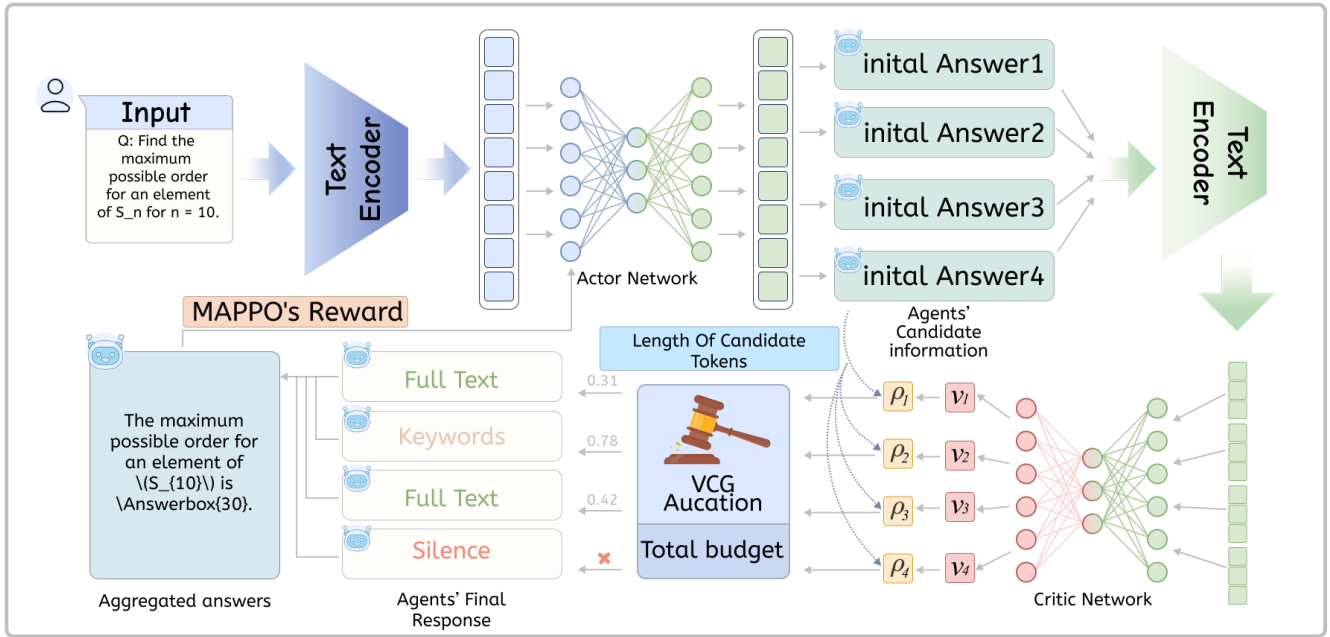


Figure 1: An overview of our DALA. Specifically, an Actor Network generates candidate messages, and a Critic Network computes their value density (ρ_i) based on utility (V_i) and cost. This density serves as a bid in a budget-constrained VCG auction that allocates speaking rights. The entire actor-critic system is trained with MAPPO to maximize a reward signal combining task success and auction costs, thereby learning an efficient communication strategy.

a more deliberate and efficient conversational environment.

In summary, our **main contributions** are: i) We introduce the principle of resource rationality to the field of MAS, identifying the management of available tokens in communication as a scarce resource, rather than its volume, as the key to unlocking both performance and efficiency; ii) We present DALA, a novel framework with a dynamic auction mechanism, compelling agents to evaluate the value density of their messages and bid for the opportunity to speak; iii) We empirically demonstrate that our DALA not only achieves new state-of-the-art performance with a fraction of the token cost but also cultivates the critical emergent skill of strategic silence, effectively teaching a MAS when to be quiet. Extensive and comprehensive experiments demonstrate that our DALA yields significant improvements in both performance and efficiency, further establishing a new state-of-the-art across seven challenging reasoning benchmarks.

Related Work

The Evolution of Multi-Agent Communication. The challenge of efficient communication in Multi-Agent Systems (MAS) is not new, but has been amplified by the token-based economics of modern Large Language Models (LLMs). Early MAS research, guided by Speech Act Theory (Searle 1969; Wang et al. 2024a,b), led to the development of standardized Agent Communication Languages (ACLs) to ensure semantic clarity (Suguri 1999). The subsequent Contract Net Protocol (CNP) (Smith 1980) established a pragmatic negotiation framework for task allocation, but its de-

signers already recognized that unconstrained communication could lead to prohibitive overhead. With the rise of deep reinforcement learning, the focus shifted from fixed protocols to learning emergent communication strategies. However, "free-for-all" communication was found to introduce noise often and hinder collaboration (Foerster et al. 2016). This has prompted research on communication efficiency, focusing on three aspects: topology pruning (Jiang and Lu 2018); gated pruning (Ding, Huang, and Lu 2020; Fan, Zhang, and Wang 2025); and representation learning (Wang et al. 2023b; Lo et al. 2024). While effective, they often lack a unified, principled framework to explicitly weigh the benefits of a message against its quantifiable cost.

Economic Mechanisms and Information-Theoretic Principles. When communication bandwidth is a scarce resource, its allocation becomes an economic problem. Auction theory provides a rigorous framework for such problems, with the Vickrey-Clarke-Groves (VCG) mechanism being particularly notable (Vickrey 1961; Zhang et al. 2025e). VCG is incentive-compatible, encouraging truthful bidding, and maximizes social welfare, ensuring resources are allocated to agents who value them most. Its successful application in complex scenarios, from radio spectrum allocation (Cramton, Shoham, and Steinberg 2006a) to Multi-armed Bandit System (Leon and Etesami 2025), validates its suitability for managing scarce communication resources in MAS. Our work operationalizes this by introducing the concept of *value density* ($\rho = \text{value}/\text{cost}$), which determines an agent's bid. This concept deeply resonates with the

Information Bottleneck (IB) principle (Tishby, Pereira, and Bialek 2000). DALA puts the information bottleneck principle into practice by training agents to compress their private knowledge into high-value density messages that contribute the most to the team’s tasks while having the lowest token cost for themselves.

Communication Efficiency in LLM-based MAS. Recent MAS frameworks built on LLMs have achieved success in complex reasoning and software development (Zheng et al. 2023; Du et al. 2024; Zhang et al. 2024, 2025a). However, their performance often comes at the cost of exponential token consumption due to “free-for-all” communication protocols, resulting in a low signal-to-noise ratio. While some works (Li and et al. 2024; Gandhi et al. 2025) have started to address this, for instance by pruning communication links based on random graph topologies like in AgentPrune-R (Zhang and et al. 2024a), these methods lack an intelligent, value-driven basis for their pruning decisions. By integrating multi-agent reinforcement learning with the VCG auction mechanism, our DALA creates a communication market to force agents to bid for the opportunity to speak, thereby cultivating a resource rationality characterized by “strategic silence” for a balance between performance and cost.

Methodology

Problem Formulation. The learning objective of our DALA is to find the optimal joint policy parameters $\Theta^* = \{\theta_1^*, \dots, \theta_N^*\}$ that solve:

$$\begin{aligned} \Theta^* &= \arg \max_{\Theta} \mathbb{E}_{\tau \sim \Pi_{\Theta}} [\mathcal{R}(\tau)] \\ \text{s.t. } \mathbb{E}_{\tau \sim \Pi_{\Theta}} \left[\sum_{t=1}^T \sum_{a_j \in W_t} L(m_j^{(t)}) \right] &\leq B \end{aligned} \quad (1)$$

Note that the budget in Eqn. (1) is an *expectation constraint* (a soft constraint), whereas the per-round budget in the auction mechanism is a hard constraint, practically enforcing the soft constraint’s optimization objective.

Value-Based Bidding Mechanism

We design a principled bidding mechanism where agents learn to evaluate the marginal utility of their messages via a learned value network and bid accordingly. The value-based bidding process is initiated through a prompt annotation mechanism guided by the **Actor Network**. Upon receiving an input question, it is first tokenized. The **Actor Network** then processes this sequence of tokens and outputs a scalar importance weight for each token, effectively appraising the salience of each word in the input. A new, annotated prompt is then constructed by interleaving the original words with their computed importance weights. For instance, a prompt like “Explain the Pythagorean theorem” might be transformed into “Explain [0.3] the [0.2] Pythagorean [0.9] theorem [0.8]”. This annotated prompt, along with a meta-instruction explaining that the bracketed values signify token importance, is then passed to the agent’s core language model. This directs the LLM’s attention to the most critical

parts of the query, guiding it to generate a more focused and relevant initial candidate message, which forms the basis for the subsequent value evaluation and bidding process.

Guiding Rational Bidding. The core design objective of our DALA is to guide agents to progressively learn an approximately rational bidding strategy through our proposed reward mechanism and learning algorithm. Ideally, an agent’s bid can accurately reflect the expected marginal contribution of its candidate message to the team’s task objective, thereby achieving efficient allocation of communication resources. To evaluate the utility of a specific message, we learn a **Message-Conditioned Value Function** $V_{\theta_i}^V$ for each agent a_i . This function aims to predict the expected cumulative return for the team, given the current observation $o_i^{(t)}$, if a candidate message m is generated and broadcast. It is computed as follows:

$$v_i(m, o_i^{(t)}) = g_V(\phi_M(m) \oplus \phi_O(o_i^{(t)}); \theta_i^V), \quad (2)$$

where ϕ_M and ϕ_O are message and observation encoders, respectively, \oplus is a multi-modal fusion operator (such as an attention mechanism), and g_V is an MLP head. v_i is essentially an action-conditioned value, similar to a Q-value. To balance the absolute value of a message with its cost (length) during bidding, we define the normalized value density $\rho_i(m, o_i^{(t)})$ as:

$$\rho_i(m, o_i^{(t)}) = \frac{v_i(m, o_i^{(t)}) - \bar{v}_t}{\sigma_{v_t} + \epsilon} \cdot \frac{1}{L(m)}, \quad (3)$$

where \bar{v}_t and σ_{v_t} are the mean and standard deviation of the values of all candidate messages in the current auction round, used to dynamically normalize the value to a Z-score, and ϵ is a small stability term.

The agent a_i ’s **Bidding Function** $b_i^{(t)}$ is directly determined by the value density of its message:

$$b_i^{(t)} = \max(0, \rho_i(m_i^{(t)}, o_i^{(t)})). \quad (4)$$

This ensures that only messages with an expected positive marginal utility (after normalization) enter the auction. Based on the magnitude of the value density, we design a tiered content output strategy: when $\rho_i \geq \tau_{full}$, a full, detailed message is generated (Full); when $\tau_{summary} \leq \rho_i < \tau_{full}$, a summary is generated (Summary); when $\tau_{keywords} \leq \rho_i < \tau_{summary}$, only keywords are output (Keywords); and when $\rho_i < \tau_{keywords}$, the agent chooses to remain silent (Silence), where $\tau_{full} > \tau_{summary} > \tau_{keywords} > 0$ are preset value density thresholds.

These thresholds effectively divide the positive value density range, enabling a four-level message generation strategy. After an agent generates a candidate message, our DALA computes its value density ρ_i . Based on where this value falls, a secondary prompt is used to format the final output. A value in the top third of the positive range prompts a Full text; the middle third prompts a Summary; and the lowest positive third prompts KeyWords. If the value density is not positive (i.e., $\rho_i \leq 0$ as per Eqn. (4)), the agent chooses to remain in Silence. For example: “Q: Answer the question about the origin of the Pythagorean

theorem”, A: *Keywords: Multiple origins (China, Babylonian civilization), Geometric relationship (right triangle side length formula), Pythagorean proof (from experience to deduction), Mathematical foundation (wide application and far-reaching influence).*”

The DALA Combinatorial Auction Mechanism

To enable multiple agents to communicate in a single round, we upgrade the auction to a **Combinatorial Auction** and apply the VCG payment rule (Groves 1973). Instead of selecting a single winner, the goal is to determine a winning *set* of agents W_t whose combined messages provide the highest total value to the team, without exceeding the round’s budget. This is known as the **Winner Determination Problem (WDP)** (Cramton, Shoham, and Steinberg 2006b).

Winner Determination Problem (WDP). Given the set of valid bids $\mathcal{B}_{\text{valid}}$, the auctioneer solves for the winning set W_t that maximizes the sum of bids, subject to the round’s budget constraint $B_{\text{max}}^{(t)}$. This is a classic 0/1 Knapsack Problem:

$$W_t = \arg \max_{S \subseteq \mathcal{B}_{\text{valid}}} \left\{ \sum_{a_j \in S} b_j^{(t)} \right\} \quad \text{s.t.} \quad \sum_{a_j \in S} L(m_j^{(t)}) \leq B_{\text{max}}^{(t)} \quad (5)$$

While the WDP is NP-hard in general, it can be solved efficiently for moderately sized agent groups using dynamic programming.

VCG Payment in Combinatorial Auction. The VCG mechanism ensures that bidding truthfully remains the dominant strategy for each agent. For each winner $a_j \in W_t$, its payment $p_j^{(t)}$ is the social cost or “harm” its presence imposes on all other agents. This is calculated as the optimal total value the framework *could have achieved without* agent j , minus the value that all *other winners actually received* in the presence of agent j .

$$p_j^{(t)} = \left(\max_{S \subseteq \mathcal{B}_{\text{valid}} \setminus \{j\}} \sum_{a_k \in S} b_k^{(t)} \right) - \left(\sum_{a_k \in W_t \setminus \{j\}} b_k^{(t)} \right) \quad (6)$$

If a_i does not win (i.e., $a_i \notin W_t$), its payment $p_i^{(t)}$ is 0.

Token Budget Management Strategy remains conceptually the same, but the cost per round $C^{(t)}$ is now the sum of token lengths from all winners in the set W_t . We employ a hierarchical control mechanism: first, a total episode budget B_{episode} ; second, a dynamic round-level budget $B_{\text{round}}^{(t)} = \frac{B_{\text{episode}} - \sum_{k=0}^{t-1} C^{(k)}}{T-t+1}$; and finally, an instantaneous hard cap $B_{\text{max}}^{(t)} = \min(B_{\text{round}}^{(t)}, B_{\text{hard}})$. The winner “pays” not with currency, but with the communication budget consumed by its message length. When the budget is tight or depleted, our DALA automatically compresses messages, forces silence, and issues budget warnings to agents to guide them toward generating shorter content.

Multi-Agent Policy Optimization. The dynamics and reward structure of our DALA form the basis for multi-agent reinforcement learning. We optimize the agent policy parameters $\Theta = \{\theta_i\}_{i=1}^N$ through interaction with the environment. Given the sparsity and team-based nature of the

Algorithm 1: DALA Combinatorial Auction Mechanism

Require: Action set $\{(m_i^{(t)}, b_i^{(t)})\}_{i=1}^N$, current budget $B_{\text{max}}^{(t)}$.

- 1: **Budget Filtering:** Create a set of valid candidates $\mathcal{B}_{\text{valid}} = \{(m_i^{(t)}, b_i^{(t)}) \mid L(m_i^{(t)}) \leq B_{\text{max}}^{(t)}\}$.
- 2: **Winner Determination:** Solve the WDP for $\mathcal{B}_{\text{valid}}$ to find the winning set W_t . ▷ Solves Eqn. (5)
- 3: Initialize payments $p_i^{(t)} \leftarrow 0$ for all agents i .
- 4: Let total value of the winning set be $V(W_t) = \sum_{a_k \in W_t} b_k^{(t)}$.
- 5: **for** each winner $a_j \in W_t$ **do**
- 6: Let $\mathcal{B}'_{\text{valid}} = \mathcal{B}_{\text{valid}} \setminus \{a_j\}$.
- 7: Solve WDP for $\mathcal{B}'_{\text{valid}}$ to find the hypothetical winning set W_t^{-j} .
- 8: Let the optimal value without j be $V(W_t^{-j}) = \sum_{a_k \in W_t^{-j}} b_k^{(t)}$.
- 9: Let value of other winners in the actual outcome be $V(W_t \setminus \{j\}) = V(W_t) - b_j^{(t)}$.
- 10: **VCG Payment** $p_j^{(t)} \leftarrow V(W_t^{-j}) - V(W_t \setminus \{j\})$. ▷ Calculates Eqn. (6)
- 11: **end for**
- 12: Actual budget cost $C^{(t)} \leftarrow \sum_{a_j \in W_t} L(m_j^{(t)})$.
- 13: **Output:** Winning agent set W_t , their messages $\{m_j^{(t)}\}_{j \in W_t}$, total cost $C^{(t)}$, and individual payments $\{p_j^{(t)}\}_{j \in W_t}$.

reward $R(\tau)$, we adopt the **Multi-Agent Proximal Policy Optimization (MAPPO)** algorithm. For agent a_i , we define the policy ratio as:

$$r_i(t)(\theta_i) = \frac{\pi_{\theta_i}(a_i(t) \mid o_i(t))}{\pi_{\theta_i^{\text{old}}}(a_i(t) \mid o_i(t))} \quad (7)$$

where θ_i^{old} are the policy parameters from the last update.

The MAPPO clipped objective function is:

$$L_i^{\text{CLIP}}(\theta_i) = \mathbb{E}_t \left[\min \left(r_i(t)(\theta_i) A_i(t), \text{clip}(r_i(t)(\theta_i), 1 - \epsilon, 1 + \epsilon) A_i(t) \right) \right] \quad (8)$$

where $A_i(t)$ is the advantage function for agent i at time step t , and ϵ is the clipping parameter. The value function loss is:

$$L_i^{\text{VF}}(\theta_i) = \mathbb{E}_t \left[\max \left((V_{\theta_i}(o_i(t)) - R_i(t))^2, (\text{clip}(V_{\theta_i}(o_i(t)), V_{\theta_i^{\text{old}}}(o_i(t)) \pm \epsilon_{vf}) - R_i(t))^2 \right) \right] \quad (9)$$

where $R_i(t)$ is the actual return, and ϵ_{vf} is the clipping parameter for the value function. The final MAPPO objective function combines the policy loss, value loss, and entropy regularization:

$$L_i^{\text{MAPPO}}(\theta_i) = L_i^{\text{CLIP}}(\theta_i) - c_1 L_i^{\text{VF}}(\theta_i) + c_2 S[\pi_{\theta_i}](o_i(t)) \quad (10)$$

where c_1, c_2 are weighting coefficients, and $S[\pi_{\theta_i}]$ is the policy entropy.

Learning Functions and Reward Design

To guide rational bidding in the new combinatorial context, we adapt the immediate reward function.

$$r_i(t) = \alpha \cdot \Delta_{task}(t) - \beta \cdot p_i^{(t)} \cdot \mathbb{I}[a_i \in W_t] \quad (11)$$

where: $\Delta_{task}(t)$ remains the marginal contribution of the team’s joint action to the task progress, e.g., $\Delta_{task}(t) = G_t - G_{t-1}$. $p_i^{(t)} \cdot \mathbb{I}[a_i \in W_t]$ is the VCG payment for agent a_i if it is in the winning set W_t . This penalty, calculated via Eqn. (6), serves as agent i ’s individual communication cost and incentivizes truthful bidding on its marginal value. α and β are hyperparameters to balance task rewards and communication costs.

Gradual Learning Mechanism: Message Value Learning: This part remains unchanged. The message-conditioned value network $V_{\theta_i}^V$ still learns to predict the expected return for a specific message by minimizing the TD error:

$$L_{value} = \mathbb{E} [(v_i(m_i(t), o_i(t)) - R_i(t))^2] \quad (12)$$

This allows v_i to form the basis for the agent’s bid $b_i(t)$. **Bidding Strategy Learning:** Through the updated reward function (11), agents learn to generate messages and bids that are valuable not only in isolation, but also to contribute positively to a high-value *combination*, i.e., winning a spot in the auction at a reasonable cost $p_i^{(t)}$.

Regarding the feedback mechanism, the reward signal guides agents toward forming beneficial coalitions. If an agent is part of a winning set W_t whose collective action yields a high positive contribution (i.e., a high $\Delta_{task}(t)$), the entire team receives a positive task-related reward. This reinforces the behavior of generating messages that are valuable in combination with others. Conversely, if the winning set’s collective action leads to a low or negative contribution, the team reward is negative, which penalizes and suppresses the formation of such ineffective coalitions in the future. Crucially, agents that are not part of the winning set ($a_i \notin W_t$) do not receive a payment penalty $p_i^{(t)}$ and only experience the team-wide task reward. This dynamic teaches agents to remain silent when their potential message is unlikely to be part of a high-value, winning coalition. Through multiple rounds of optimization on this reward signal using the MAPPO algorithm, the agents collectively internalize a more sophisticated rational logic: “my information should be part of a high-value coalition to be worth bidding on,” thus collaboratively achieving the team’s objectives while adhering to budget constraints.

Experiments

Baselines. To comprehensively evaluate the performance of our proposed DALA, we compare it against a suite of strong baselines. These are categorized into two groups: i) single-agent methods, including Vanilla prompting, Chain-of-Thought (CoT)(Wei et al. 2023), Complex CoT(Fu et al. 2023), and Self-Consistency (SC)(Wang et al. 2023a); ii) multi-agent system (MAS) methods, including PHP(Zheng et al. 2024), LLM-Debate(Khan and et al. 2024), DyLAN(Liu et al. 2024), and AgentPrune-R(Zhang and et al.

2024b), which represents the current state-of-the-art communication pruning method based on a random graph topology.

Datasets. Our evaluation spans seven challenging benchmarks across three core reasoning domains. For general reasoning, we use MMLU(Hendrycks et al. 2021a). For mathematical reasoning, we employ a diverse set of five benchmarks: GSM8K(Cobbe, Kosaraju, and Bavarian 2021), MultiArith(Roy and Roth 2016), SVAMP(Patel, Bhattamishra, and Goyal 2021), AQuA(Ling et al. 2017), and the particularly challenging MATH-500(Hendrycks et al. 2021b) subset. For code generation, we use HumanEval (hum 2021) to assess function-level programming capabilities. For benchmarks with established training sets, we utilize them directly. For those without a standard training set, we adopt a data optimization paradigm(Similar to AgentPrune-R(Zhang and et al. 2024b)). Specifically, we designate 2% of the data from each benchmark as an optimization set to train the DALA agent’s bidding strategy, which is based on value density. After this policy is optimized via the MAPPO algorithm and subsequently fixed, we perform the final performance evaluation on the remaining 98% of unseen data.

Experimental Settings. Following (Zhang and et al. 2024b) to ensure a fair comparison of the communication frameworks, all experiments are conducted using gpt-4-1106-preview(OpenAI 2024) as the base large language model for all agents. For all tasks, our DALA uses different total budgets to control the communication cost (see our supplementary materials). For each task instance, essential information, such as distinct mathematical formulas, contextual facts, or required code snippets, is distributed among the different agents. This setup ensures that no single agent possesses sufficient information to solve the problem alone, making effective and efficient communication a prerequisite for success and thereby creating an ideal testbed for our DALA. For specific experimental hyperparameter settings, please refer to Appendix “Implementation Details and Hyperparameters”. For the specific comparison models, benchmarks, and datasets, please refer to Appendix “Benchmarks and Baseline Methods”. All experiments are performed on NVIDIA A100 80GB.

Performance and Cost-Effectiveness

Table 1 demonstrates that our DALA consistently establishes the new state-of-the-art across all seven benchmarks. On the MMLU general reasoning task, DALA achieves an accuracy of 84.32%, outperforming all evaluated baselines. This pattern of superiority is particularly evident in the domain of mathematical reasoning. For instance, on GSM8K, DALA reaches 96.18% accuracy, a 1.35-point improvement over the previous best, AgentPrune-R. Similarly, on the challenging HumanEval and MATH-500 benchmarks, DALA demonstrates its robust capabilities by achieving top scores of 91.21% and 83.42%, respectively. These results underscore DALA’s consistent ability to foster more effective collaboration that translates directly into superior problem-solving accuracy, irrespective of the task domain. Please refer to our supplementary materials for model computational

Method	MMLU	MultiArith	GSM8K	SVAMP	AQuA	HumanEval	MATH-500
<i>Single-Agent Methods</i>							
Vanilla	82.14	93.15	85.40	87.18	70.34	71.68	73.72
CoT	82.65 (+0.51)	94.79 (+1.64)	87.17 (+1.77)	88.32 (+1.14)	73.91 (+3.57)	75.52 (+3.84)	75.18 (+1.46)
ComplexCoT	83.78 (+1.64)	95.86 (+2.71)	87.62 (+2.22)	90.17 (+2.99)	77.58 (+7.24)	74.94 (+3.26)	76.85 (+3.13)
SC	82.66 (+0.52)	96.88 (+3.73)	87.93 (+2.53)	88.69 (+1.51)	75.08 (+4.74)	77.30 (+5.62)	77.02 (+3.30)
<i>Multi-Agent Methods</i>							
PHP	83.45 (+1.31)	96.41 (+3.26)	92.45 (+7.05)	90.62 (+3.44)	76.25 (+5.91)	82.96 (+11.28)	79.24 (+5.52)
LLM-Debate	83.69 (+1.55)	96.27 (+3.12)	90.23 (+4.83)	90.56 (+3.38)	77.52 (+7.18)	83.79 (+12.11)	80.15 (+6.43)
DyLAN	80.16 (-1.98)	94.27 (+1.12)	88.16 (+2.76)	87.40 (+0.22)	74.16 (+3.82)	89.70 (+18.02)	81.66 (+7.94)
AgentPrune-R	83.94 (+1.80)	96.30 (+3.15)	95.83 (+10.43)	91.68 (+4.50)	78.60 (+8.26)	90.30 (+18.62)	82.81 (+9.09)
DALA (Ours)	84.32 (+2.18)	97.87 (+4.72)	96.18 (+10.78)	92.33 (+5.15)	80.92 (+10.58)	91.21 (+19.53)	83.42 (+9.70)

Table 1: Overall performance comparison across seven reasoning benchmarks. We report accuracy (%) for all tasks except HumanEval, for which we report pass@1 (%). Values in parentheses show absolute gain over the Vanilla baseline.

overhead. For a specific QA case analysis, please see the Appendix "Case Study Analysis".

Method	MMLU		GSM8K	
	Accuracy (%)	Token Cons.	Accuracy (%)	Token Cons.
Ours (DALA)	84.3	3.80×10^5	96.2	6.20×10^6
AgentPrune-R	84.0	9.60×10^5	95.8	7.50×10^6
LLM-Debate	83.7	1.50×10^6	90.2	2.20×10^7
PHP	83.4	2.60×10^6	92.5	2.60×10^7
DyLAN	80.2	1.20×10^6	88.2	1.40×10^7
Vanilla	82.1	1.50×10^5	85.4	3.50×10^6

Table 2: Consolidated performance comparison on MMLU and GSM8K datasets. Our DALA, highlighted in gray, demonstrates superior or competitive performance with significantly better token efficiency.

Beyond superior performance, DALA is designed for economic efficiency, achieving top-tier accuracy at a fraction of the communication cost. As illustrated in Figure 2, DALA consistently occupies the high-accuracy, low-cost quadrant across benchmarks. On MMLU, it consumes only 1.81×10^5 tokens, an order of magnitude less than methods like PHP (2.60×10^6). The advantage is even more pronounced on GSM8K, where DALA’s token usage (6.25×10^6) starkly contrasts with costly approaches like DyLAN (1.40×10^7). This efficiency is a direct result of DALA’s “Pay-to-Speak” mechanism, a value-density auction that intrinsically prunes low-value, verbose messages, establishing DALA as a more practical and scalable framework.

Analyses of Emergent Communication Strategies

To validate the core hypothesis of our DALA that an economic-driven communication market can cultivate sophisticated and efficient discursive strategies, we conduct a multi-faceted analysis of the agents’ emergent behaviors. This investigation aims to look beyond aggregate task performance and elucidate the intelligent, fine-grained decision-making processes fostered by our auction mechanism. To analyze strategic adaptation, we established two distinct total budget conditions. These settings are benchmarked against the token consumption of a standard baseline (Vanilla, i.e., gpt-4-1106-preview), which uses approximately 1.5×10^5 tokens for the MMLU task. We there-

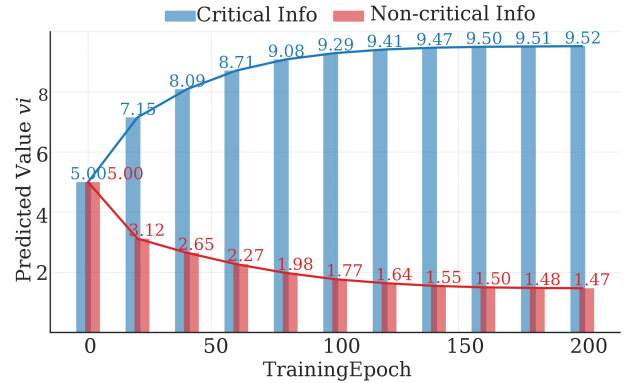


Figure 2: The learning curve of the average predicted value (v_i) on MMLU, showing the agent’s internal value network rapidly learns to assign high value to critical information while devaluing non-critical information, with the valuation gap stabilizing as training converges around epoch 160.

fore set one budget at 1×10^5 tokens to create a resource-scarce scenario that necessitates efficiency, and a second, larger budget of 1×10^6 tokens to represent a resource-abundant environment. Please note that the tokens budget here refers to the budget relative to the entire task design. Under each condition, we recorded the frequency of four key communication strategies. To further evaluate the adaptability of the strategy, we operationally define information types within the MMLU task: **critical information** refers to reasoning steps or facts that are essential for deriving the correct answer, while **non-critical information** includes redundant statements, incorrect reasoning paths, or conversational filler; the specific classification is performed by an external adjudicator model, i.e., openai/o3-2025-04-16.

The empirical results, presented visually, offer compelling validation of our DALA’s underlying mechanics. The foundation of all strategic action is the agent’s ability to appraise information. Figure 2 presents the learning trajectory of this appraisal, showing that the value network rapidly learns to distinguish information types. Starting from near-identical valuations at epoch 0, the curves for critical and non-critical

Domain	Full (%)	Sum. (%)	Key. (%)	Silence (%)
Low Budget (1×10^5 Tokens)				
Humanities	5.1	24.9	30.1	39.9
Social Science	4.7	20.1	35.7	39.5
STEM	2.1	15.6	48.9	33.4
Other (General)	6.9	28.3	29.5	35.2
High Budget (1×10^6 Tokens)				
Humanities	48.7	38.6	9.4	3.2
Social Science	45.9	42.1	8.9	3.0
STEM	30.6	42.1	21.1	6.3
Other (General)	45.0	41.9	10.1	3.0

Table 3: Comparison of agent communication strategies under different budget constraints. The table is split vertically into Low Budget (top) and High Budget (bottom) scenarios.

information diverge significantly, with the value delta stabilizing as the model converges around epoch 160. This learned ability to discern value is the crucial prerequisite that drives behavior in the communication market.

This internal valuation directly informs how agents adapt to resource constraints, as visualized in Figure 3. Under the stringent 1×10^5 token budget (Figure 3, bottom), agents across all domains are forced to be economical. They significantly curb the use of “Full-text” messages (e.g., dropping to just 2.1% in STEM) and increase their reliance on highly compressed “keywords” (peaking at 48.9% in STEM) and “strategic silence” (39.9% in Humanities). Conversely, with the generous 1×10^6 token budget (Figure 3, top), agents become more expressive. “Full-text” communication becomes a dominant strategy, especially in Humanities (48.7%), while “silence” diminishes to a negligible level across all domains (e.g., 3.0% in Social Science). This stark contrast between the two conditions provides clear visual evidence that agents are not following a static policy, but are dynamically modulating their communication strategies based on the learned value of their information and the current state of the market. By analyzing the silence ratio of agents under different costs, the agents in our DALA have truly learned to “speak less or not” when the total budget is small, and have truly learned “when to shut up”.

Module Ablation. To demonstrate the detailed contributions of DALA’s key components, we conduct a series of ablation studies on the MMLU and GSM8K benchmarks. We compare the full DALA model against several ablated variants, where each variant disables a single, critical mechanism. All other settings remain identical to our main experiments. The variants are as follows: **w/o Value Learning:** Replaces the learned value function V_ϕ with a heuristic bid (e.g., proportional to message length), to verify that intelligent, learned valuation is essential for the auction’s success. **w/o Value Density:** Bids are based on raw predicted value v_i , removing the $1/L(m)$ term. This tests the core hypothesis that optimizing for value-per-token, not absolute value, drives efficiency. **w/o Tiered Content:** Collapses the communication strategy to a binary choice (Full message vs. Silence), to assess the importance of adaptive granularity in managing the information-cost trade-off. **w/o Dynamic**

Variant	MMLU (Acc. %)	GSM8K (Acc. %)	Avg. Tokens ($\times 10^6$)
DALA (Full Model)	84.32	96.18	3.22
w/o Value Learning	75.77 ($\downarrow 8.55$)	86.35 ($\downarrow 9.83$)	4.13
w/o Value Density	79.16 ($\downarrow 5.16$)	89.76 ($\downarrow 6.42$)	4.91
w/o Tiered Content	79.83 ($\downarrow 4.49$)	90.95 ($\downarrow 5.23$)	4.49
w/o Dynamic Budget	80.71 ($\downarrow 3.61$)	91.30 ($\downarrow 4.88$)	3.65
w/o Cost Penalty ($\beta = 0$)	77.48 ($\downarrow 6.84$)	88.27 ($\downarrow 7.91$)	6.23

Table 4: Ablation study on MMLU and GSM8K. We report accuracy (%) and the average token consumption (Avg. Tokens) across both tasks. The results demonstrate that each component of our DALA contributes significantly to its overall performance and efficiency. The best performance and lowest token consumption are highlighted in bold.

Budget: Uses a fixed per-round budget instead of an adaptive one, to evaluate the benefits of a forward-looking resource allocation policy. **w/o Cost Penalty ($\beta = 0$):** Removes the communication cost from the reward function (Eq. 9), to demonstrate that an explicit economic penalty is necessary to enforce communicative discipline and prevent verbosity. The results of our ablation study (Table 4) confirm the architectural integrity of our DALA, as removing any component causes a substantial decline in performance. The most pronounced degradation occurs with the ablation of **Value Learning**, causing an accuracy drop of up to 9.83 points. This validates our core hypothesis that an effective multi-agent system requires not just communication, but an intelligent appraisal of information’s worth. Similarly, removing economic principles like **Value Density** and the **Cost Penalty ($\beta = 0$)** leads to verbose, inefficient communication and information overload. This demonstrates the contribution of our DALA’s components, i.e., value-based intelligence, economic discipline, and adaptive mechanisms like **Tiered Content** and a **Dynamic Budget**.

Conclusion

We presented DALA, a simple yet effective framework that recasts communication as a market-driven auction. By incentivizing agents to bid according to the *value density* of their messages, DALA prunes low-value exchanges and encourages more deliberate communication. Experiments show that DALA achieves state-of-the-art performance on seven reasoning benchmarks with only a fraction of the token cost.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62276283, in part by the China Meteorological Administration’s Science and Technology Project under Grant CMA-JBGS202517, in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515012985, in part by Guangdong-Hong Kong-Macao Greater Bay Area Meteorological Technology Collaborative Research Project

under Grant GHMA2024Z04, in part by Fundamental Research Funds for the Central Universities, Sun Yat-sen University under Grant 23hytd006, and in part by Guangdong Provincial High-Level Young Talent Program under Grant RL2024-151-2-11.

References

2021. Evaluating Large Language Models Trained on Code.
- Chen, L.; Davis, J.; Hanin, B.; Bailis, P.; Stoica, I.; Zaharia, M.; and Zou, J. 2024. Are More LM Calls All You Need? Towards Scaling Laws of Compound AI Systems.
- Cobbe, K.; Kosaraju, V.; and Bavarian, M. e. a. 2021. Training Verifiers to Solve Math Word Problems. *arXiv preprint arXiv:2110.14168*.
- Cramton, P.; Shoham, Y.; and Steinberg, R. 2006a. *Combinatorial Auctions*. The MIT Press. ISBN 0262033429.
- Cramton, P.; Shoham, Y.; and Steinberg, R. 2006b. *Combinatorial Auctions*. The MIT Press. ISBN 0262033429.
- Ding, Z.; Huang, T.; and Lu, Z. 2020. Learning individually inferred communication for multi-agent cooperation. In *NeurIPS, NIPS '20*. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781713829546.
- Du, Y.; Li, S.; Torralba, A.; Tenenbaum, J. B.; and Mordatch, I. 2024. Improving factuality and reasoning in language models through multiagent debate. ICML'24. JMLR.org.
- Fan, Y.; Zhang, J.; Cai, K.; Yang, J.; and Wang, K. 2025. CCG: Rare-Label Prediction via Neural SEM-Driven Causal Game. In Christodoulopoulos, C.; Chakraborty, T.; Rose, C.; and Peng, V., eds., *Findings of the Association for Computational Linguistics: EMNLP 2025*, 6243–6256. Suzhou, China: Association for Computational Linguistics. ISBN 979-8-89176-335-7.
- Fan, Y.; Zhang, J.; and Wang, K. 2025. Towards More Efficient Post-training via Fourier Domain Adapter Framework. In Christodoulopoulos, C.; Chakraborty, T.; Rose, C.; and Peng, V., eds., *Findings of the Association for Computational Linguistics: EMNLP 2025*, 6175–6193. Suzhou, China: Association for Computational Linguistics. ISBN 979-8-89176-335-7.
- Foerster, J. N.; Assael, Y. M.; de Freitas, N.; and Whiteson, S. 2016. Learning to communicate with Deep multi-agent reinforcement learning. In *NeurIPS, NIPS'16*, 2145–2153. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781510838819.
- Fu, Y.; Peng, H.; Sabharwal, A.; Clark, P.; and Khot, T. 2023. Complexity-Based Prompting for Multi-Step Reasoning. *arXiv:2210.00720*.
- Gandhi, S.; Patwardhan, M.; Vig, L.; and Shroff, G. 2025. BudgetMLAgent: A Cost-Effective LLM Multi-Agent system for Automating Machine Learning Tasks. *AIMLSys-tems '24*. New York, NY, USA: Association for Computing Machinery. ISBN 9798400711619.
- Groves, T. 1973. Incentives in Teams. *Econometrica*, 41(4): 617–631.
- Hendrycks, D.; Burns, C.; Basart, S.; Zou, A.; Mazeika, M.; Song, D.; and Steinhardt, J. 2021a. Measuring Massive Multitask Language Understanding. *arXiv:2009.03300*.
- Hendrycks, D.; Burns, C.; Kadavath, S.; Arora, A.; Basart, S.; Tang, E.; Song, D.; and Steinhardt, J. 2021b. Measuring Mathematical Problem Solving With the MATH Dataset. In Vanschoren, J.; and Yeung, S., eds., *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1.
- Huang, D.; Zhang, J. M.; Luck, M.; Bu, Q.; Qing, Y.; and Cui, H. 2024. AgentCoder: Multi-Agent-based Code Generation with Iterative Testing and Optimisation. *arXiv:2312.13010*.
- Jiang, J.; and Lu, Z. 2018. Learning attentional communication for multi-agent cooperation. In *NeurIPS, NIPS'18*, 7265–7275. Red Hook, NY, USA: Curran Associates Inc.
- Khan, A.; and et al., J. H. 2024. Debating with More Persuasive LLMs Leads to More Truthful Answers. *arXiv:2402.06782*.
- Leon, V.; and Etesami, S. R. 2025. Online Learning for Dynamic Vickrey-Clarke-Groves Mechanism in Sequential Auctions under Unknown Environments. *arXiv:2506.19038*.
- Li, W.; Hu, B.; Shao, R.; Shen, L.; and Nie, L. 2025a. Lions: Fast & slow video-language thinker as online video assistant. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 3240–3251.
- Li, W.; Zhang, R.; Shao, R.; He, J.; and Nie, L. 2025b. Cogvla: Cognition-aligned vision-language-action model via instruction-driven routing & sparsification. *arXiv preprint arXiv:2508.21046*.
- Li, Y.; and et al., Y. D. 2024. Improving Multi-Agent Debate with Sparse Communication Topology. *arXiv:2406.11776*.
- Ling, W.; Yogatama, D.; Dyer, C.; and Blunsom, P. 2017. Program Induction by Rationale Generation: Learning to Solve and Explain Algebraic Word Problems. In Barzilay, R.; and Kan, M.-Y., eds., *ACL*, 158–167. Vancouver, Canada: Association for Computational Linguistics.
- Liu, Z.; Zhang, Y.; Li, P.; Liu, Y.; and Yang, D. 2024. A Dynamic LLM-Powered Agent Network for Task-Oriented Agent Collaboration. *arXiv:2310.02170*.
- Lo, Y. L.; Sengupta, B.; Foerster, J.; and Noukhovitch, M. 2024. Learning Multi-Agent Communication with Contrastive Learning. *arXiv:2307.01403*.
- OpenAI. 2024. GPT-4 Technical Report. *arXiv:2303.08774*.
- Patel, A.; Bhattamishra, S.; and Goyal, N. 2021. Are NLP Models really able to Solve Simple Math Word Problems? In Toutanova, K.; Rumshisky, A.; Zettlemoyer, L.; Hakkani-Tur, D.; Beltagy, I.; Bethard, S.; Cotterell, R.; Chakraborty, T.; and Zhou, Y., eds., *ACL*, 2080–2094. Online: Association for Computational Linguistics.
- Roy, S.; and Roth, D. 2016. Solving General Arithmetic Word Problems. *arXiv:1608.01413*.
- Searle, J. R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

- Smith. 1980. The Contract Net Protocol: High-Level Communication and Control in a Distributed Problem Solver. *IEEE Transactions on Computers*, C-29(12): 1104–1113.
- Suguri, H. 1999. A standardization effort for agent technologies: The Foundation for Intelligent Physical Agents and its activities. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences*. 1999. HICSS-32. *Abstracts and CD-ROM of Full Papers*, volume Track8, 10 pp.—.
- Tishby, N.; Pereira, F. C.; and Bialek, W. 2000. The information bottleneck method. arXiv:physics/0004057.
- Vickrey, W. 1961. Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance*, 16(1): 8–37.
- Wang, L.; Ma, C.; and et al. 2024. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6).
- Wang, X.; Chen, X.; Ren, W.; Han, Z.; Fan, H.; Tang, Y.; and Liu, L. 2024a. Compensation Atmospheric Scattering Model and Two-Branch Network for Single Image Dehazing. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(4): 2880–2896.
- Wang, X.; Ren, W.; Chen, X.; Fan, H.; Tang, Y.; and Han, Z. 2024b. Uni-YOLO: Vision-Language Model-Guided YOLO for Robust and Fast Universal Detection in the Open World. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, 1991–2000. New York, NY, USA: Association for Computing Machinery. ISBN 9798400706868.
- Wang, X.; Wei, J.; Schuurmans, D.; Le, Q.; Chi, E.; Narang, S.; Chowdhery, A.; and Zhou, D. 2023a. Self-Consistency Improves Chain of Thought Reasoning in Language Models. arXiv:2203.11171.
- Wang, Z.; Du, Y.; Zhang, Y.; Fang, M.; and Huang, B. 2023b. MACCA: Offline Multi-agent Reinforcement Learning with Causal Credit Assignment. arXiv:2312.03644.
- Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.; Le, Q.; and Zhou, D. 2023. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. arXiv:2201.11903.
- Wen, Z.; Hu, H.; Fang, N.; Qian, S.; and Cao, J. 2025. DANet: A RAG-inspired Dual Attention Model for Few-shot Time Series Prediction. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, CIKM '25, 3302–3311. New York, NY, USA: Association for Computing Machinery. ISBN 9798400720406.
- Yao, S.; Zhao, J.; Yu, D.; Du, N.; Shafran, I.; Narasimhan, K.; and Cao, Y. 2023. ReAct: Synergizing Reasoning and Acting in Language Models. In *ICLR*.
- Yu, C.; Velu, A.; Vinitzky, E.; Gao, J.; Wang, Y.; Bayen, A.; and Wu, Y. 2022. The surprising effectiveness of PPO in cooperative multi-agent games. In *NeurIPS*, NIPS '22. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781713871088.
- Zhang, G.; and et al., Y. Y. 2024a. Cut the Crap: An Economical Communication Pipeline for LLM-based Multi-Agent Systems. arXiv:2410.02506.
- Zhang, G.; and et al., Y. Y. 2024b. Cut the Crap: An Economical Communication Pipeline for LLM-based Multi-Agent Systems. arXiv:2410.02506.
- Zhang, H.; Du, W.; Shan, J.; Zhou, Q.; Du, Y.; Tenenbaum, J. B.; Shu, T.; and Gan, C. 2024. Building Cooperative Embodied Agents Modularly with Large Language Models. arXiv:2307.02485.
- Zhang, J.; Cai, K.; Yang, J.; Wang, J.; Tang, C.; and Wang, K. 2025a. Top-Down Semantic Refinement for Image Captioning. arXiv:2510.22391.
- Zhang, J.; Cai, K.; Zeng, Q.; Liu, N.; Fan, S.; Chen, Z.; and Wang, K. 2025b. Failure-Driven Workflow Refinement. arXiv:2510.10035.
- Zhang, J.; Fan, Y.; Cai, K.; Huang, Z.; Sun, X.; Wang, J.; Tang, C.; and Wang, K. 2025c. DrDiff: Dynamic Routing Diffusion with Hierarchical Attention for Breaking the Efficiency-Quality Trade-off. arXiv:2509.02785.
- Zhang, J.; Fan, Y.; Cai, K.; Sun, X.; and Wang, K. 2025d. OSC: Cognitive Orchestration through Dynamic Knowledge Alignment in Multi-Agent LLM Collaboration. arXiv:2509.04876.
- Zhang, J.; Fan, Y.; Cai, K.; and Wang, K. 2025e. Kolmogorov-Arnold Fourier Networks. arXiv:2502.06018.
- Zhang, J.; Fan, Y.; Lin, W.; Chen, R.; Jiang, H.; Chai, W.; Wang, J.; and Wang, K. 2025f. GAM-Agent: Game-Theoretic and Uncertainty-Aware Collaboration for Complex Visual Reasoning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.
- Zhao, W. X.; and et al., K. Z. 2025. A Survey of Large Language Models. arXiv:2303.18223.
- Zheng, C.; Liu, Z.; Xie, E.; Li, Z.; and Li, Y. 2023. Progressive-hint prompting improves reasoning in large language models. *arXiv preprint arXiv:2304.09797*.
- Zheng, C.; Liu, Z.; Xie, E.; Li, Z.; and Li, Y. 2024. Progressive-Hint Prompting Improves Reasoning in Large Language Models. arXiv:2304.09797.
- Zhou, Q.; and et al., Z. Z. 2024. Enhancing the General Agent Capabilities of Low-Parameter LLMs through Tuning and Multi-Branch Reasoning. arXiv:2403.19962.