A Dual-Mind Framework for Strategic and Expressive Negotiation Agent

Yutong Liu[1], **Lida Shi**[2], **Rui Song**[3], **Hao Xu**[3]*
[1]College of Software , Jilin University, Changchun, China
[2]School of Artificial Intelligence, Jilin University, Changchun, China
[3]College of Computer Science and Technology, Jilin University, Changchun, China
{yutong23,shild21}@mails.jlu.edu.cn
{songrui,xuhao}@jlu.edu.cn

Abstract

Negotiation agents need to influence the attitudes or intentions of users to reach a consensus. Strategy planning and expressive optimization are crucial aspects of effective negotiations. However, previous studies have typically focused on only one of these aspects, neglecting the fact that their combined synergistic effect can lead to better performance. Inspired by the dual-process theory in human cognition, we propose a Dual-Mind Negotiation Agent (DMNA) framework. This framework integrates an intuitive module for rapid, experience-based response and a deliberative module for slow, expression optimization. The intuitive module is trained using Monte Carlo Tree Search (MCTS) and Direct Preference Optimization (DPO), enabling it to make suitable strategic planning and expression. The deliberative module employs a multifaceted reflexion mechanism to enhance the quality of expression. Experiments conducted on negotiation datasets confirm that DMNA achieves state-of-the-art results, demonstrating an enhancement in the negotiation ability of agents [1].

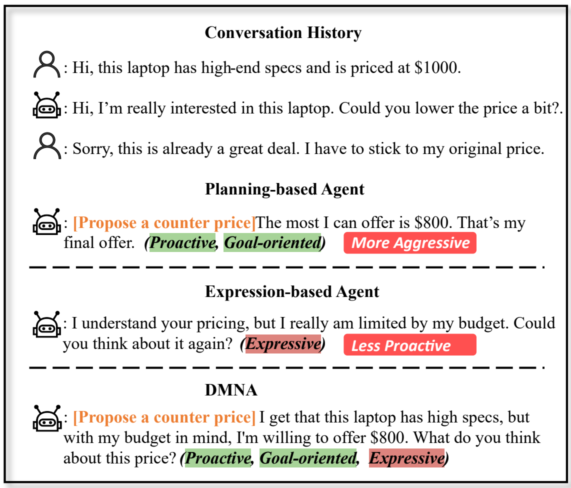


Figure 1: Examples of the response of planning-based Agent, expression-based Agent, and DMNA respectively.

1 Introduction

Negotiation dialogue involves psychology, politics, and communication, playing a crucial role in daily life (Deng et al., 2024a). Characterized as a mixed-initiative dialogue system, negotiation dialogue reflects real-world situations where users might be unwilling to cooperate with the system to reach goals. It aims the agents to mediate conflicts and facilitate mutual agreement among individuals (Zhan et al., 2024). To achieve it, the agents need to possess negotiation capabilities, such as strategic planning and expression skills (Deng et al., 2023a).

Recently, Large Language Models (LLMs) possess the remarkable ability to generate highly convincing content that can rival, even surpass human-crafted negotiation, which significantly empowers negotiation agents. Existing approaches can be broadly divided into two categories: planning-based and expression-based methods. Planning-based methods (e.g., GDP-ZERO (Yu et al., 2023), TRIP (Zhang et al., 2024), and DPDP (He et al., 2024)) operate by selecting suitable dialogue strategy from the predefined set based on current conversational state. These methods treat strategy planning as a decision-making problem and achieve efficiency through elaborate algorithms. However, their reliance on fixed strategies limits adaptability in real negotiations. Moreover, the quality of response is only dependent on the language backbone, neglecting fine-grained expressive constraints. In contrast, expression-based methods (e.g., AnE (Zhang et al., 2023), and ICL-AIF (Fu et al., 2023)) focus on the expression optimizing in negotiation. These methods excel in user perception and adaptive responses, but they lack quantifiable mechanisms for achieving goals, resulting in

*Corresponding author

[1]Code available at: https://github.com/i-ytt/DMNA.git

Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 23840–23860

suboptimal outcomes.

As illustrated in Figure 1, it presents a common negotiation scenario where a buyer expresses interest in a laptop and wants a price reduction. But the seller insists original price. In this situation, three types of agents offer their negotiation response. The planning-based agent employs a 'propose a counter price' strategy, proposing a lower price and stating it as the final offer. This response proactively guides the conversation and is goal-oriented, aiming to purchase the laptop at the lowest price. However, there are evident shortcomings, such as the aggressive expression potentially leading to the failure of the negotiation. Conversely, the expression-based agent adopts a more empathetic and persuasive expression. It acknowledges the price of the seller and cites budget limitation as the reason for price reduction, and politely asks the seller to reconsider. This response is expressive and considerate, which can build rapport and understanding with the seller. However, it sacrifices some proactivity in negotiation, which might lead to a suboptimal outcome. Outstanding negotiation agents require both capabilities synergistically: strategic planning and expression skills, such as the response of DMNA in Figure 1. Our work addresses this integration gap by developing a unified framework that combines strategic planning with expressive optimization.

Inspired by the dual-process theory (Kahneman, 2003a), we propose a Dual-Mind Negotiation Agent. This agent uses a response model trained with strategy and expression experiences as the intuitive module and employs a multifaceted reflexion mechanism for expression as the deliberative module. During the negotiation process, the two cognitive systems are connected and interact with each other, achieving both strategy planning and expression optimization. Specifically, the intuitive module is trained on past negotiation experiences, which involve negotiation strategies and expression. We employ the MCTS process to sample strategies and expressions from negotiation dialogues and form preference data pairs. These data pairs are used to train the small model (e.g., LLaMA-8B (Touvron et al., 2023)) with Direct Preference Optimization (DPO)(Rafailov et al., 2023), equipping it with basic intuitive strategic and expressive capabilities across general dialogue scenarios. The deliberative module utilizes reflextion based on multi-critics and moderator, ensuring high-quality expression even in complex and unfamiliar dialogue states. The two modules are linked and interact through memory storage, influencing each other and working together to enhance the negotiation performance. In summary, our key contributions are concluded below:

- We develop a novel framework DMNA that combines strategic planning with expression optimization, enabling existing planning-based methods and expression-based methods to complement each other.
- We utilize MCTS to obtain preference data and fine-tune the model with DPO, endowing it with planning capability. We also propose a multifaceted reflexion mechanism that is more suited to negotiation than reflexion.
- Experimental results on two datasets suggest that DMNA outperforms planning-based and expression-based baselines, and it effectively enhances the negotiation ability of agents.

2 Related Work

2.1 Dialogue Policy Planning in Negotiation

To achieve a successful negotiation, some studies have focused on the planning of negotiation dialogue strategies. Early methods mostly used neural-focused, algorithm-focused, and reinforcement learning approaches, but these rely on annotated data and depend on the elaborate algorithm design (Zhang et al., 2022; Cheng et al., 2022; Gao et al., 2021).

LLMs present both opportunities and challenges for dialogue planning. For example, Deng et al. (2023a) adopts the prompts to select the strategy proactively, but nontrainable parameters limit effectiveness. Yu et al. (2023) integrates Monte Carlo Tree Search (MCTS) with LLMs prompts to optimize strategy selection, achieving promising performance. However, this method suffers from inefficiency and high costs. Additionally, there is a paradigm that employs a trainable policy model as plugins to assist LLMs, such as Yu et al. (2023) and Zhang et al. (2024). While these can reduce costs, they still fall short in simulating the cognitive processes of future dialogue behaviors. He et al. (2024) is a dual system based on Yu et al. (2023) and Deng et al. (2024b), which can mitigate the limitations of both approaches. Nevertheless, in practical negotiation scenarios, relying solely on

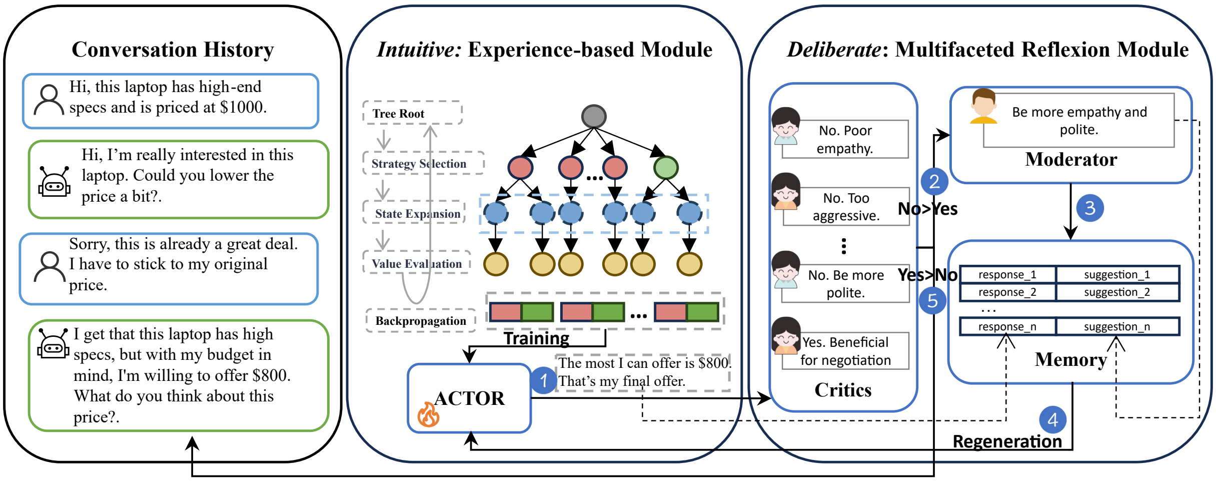


Figure 2: The overview of DMNA. This method includes an intuitive module and a deliberate module. The intuitive module consists of an experience-based module that provides intuitive responses by preference learning based on strategy-expression pairs. The deliberate module is composed of multifaceted reflexion that adjusts the suboptimal expression from the intuitive module.

strategy planning is not sufficient. It is also crucial to have the ability to finely perceive complex situations and generate appropriate responses.

2.2 Expression Quality in Negotiation

Existing studies emphasize the significance of negotiation expression. Recent work employs reinforcement learning to enhance specific dimensions: politeness (Mishra et al., 2022), empathy (Samad et al., 2022), and linguistic richness through human-guided demonstrations (Shi et al., 2021).

The aforementioned methods enhance expression quality from certain fine-grained aspects. In some recent research, LLMs are utilized to improve the expression quality of negotiation dialogues. Fu et al. (2023) uses self-play simulations to iteratively refine negotiation expression based on feedback from other LLMs regarding the current dialogue state. And Zhang et al. (2023) enhances the persuasiveness of responses by prompting another LLM as an expert to act for verbal suggestion. However, these methods lack strategic guidance in the dialogue process, which results in insufficient look-forward capability of the negotiation agents.

2.3 Dual-Mind of LLMs

Dual-process thinking is a cognitive pattern unique to humans, characterized by both intuitive and rational processes (Kahneman, 2003b; Bengio, 2019). Inspired by the theory of dual-process thinking, many fields have applied its principles to the workflow of LLMs (Yang et al., 2024; Xiao et al., 2024; Lin et al., 2023). For example, Cheng et al. (2025) propose *HaluSearch*, which treats text generation as a stepwise reasoning process (fast system) and tree search algorithms (slow system) effectively mitigate hallucinations during inference. In dialogue planning, He et al. (2024) incorporates a policy LM model for fast responses and a Monte Carlo Tree Search (MCTS) mechanism for slow planning. They also use a two-stage training regimen to enhance planning in proactive dialogue together. Similarly, Tian et al. (2023) introduces *DUMA*, which embodies a dual mind mechanism through two LLMs for fast and slow thinking. This allows seamless transitions between intuitive responses and deliberate problem-solving in conversational scenarios. In our work, we apply the theory of dual-process to the construction of a negotiation agent. During the negotiation process, we aim to enhance negotiation capabilities by leveraging two types of cognitive patterns.

3 Dual-Mind Negotiation Agent

The proposed DMNA framework is shown in Figure 2. It includes two modules that mimic human cognitive patterns: an intuitive module based on experience and a deliberate module for multifaceted reflexion. These modules together improve the quality of negotiation statements and significantly enhance the agent's negotiation ability.

3.1 Experience-based Response Module

In this module, we aim to enable the agent to provide intuitive responses that are suitable for the current dialogue state. To achieve this, we intend to train a model as the actor to learn from the look-forward strategy chosen and corresponding expression generation, allowing it to make suitable strategic planning and expression. There are two challenges we faced: (1) *The lack of high-quality datasets that reflect the look-forward strategy and appropriate expression*. and (2) *How to learn the look-forward capability*?

To address these challenges, drawing inspiration from the work of GDP-ZERO (Yu et al., 2023) and DPO (Rafailov et al., 2023), we design the following processes:

Data Generation through MCTS. To obtain the look-forward strategy and the corresponding expression data, we build upon the work of GDP-ZERO (Yu et al., 2023) and collect strategy-expression data from the MCTS process, structured in a '[strategy] expression' format. Specifically, we treat dialogue states as nodes in the search tree. To identify the next negotiation strategy, MCTS iteratively performs **selection**, **expansion**, **evaluation**, and **backpropagation**. Initially, we set the number of iterations to K. In each iteration, the search tree is constructed with values for Q values, state values V, and visit counts N being updated accordingly. After K iterations, MCTS predicts the best negotiation strategy for the current dialogue state based on tree statistics. For details of MCTS, please see Appendix A.

Turn-Level Strategy-Expression Preference Data Pairs Construction. To endow the model with look-forward ability in strategy and expression, we construct turn-level strategy-expression preference data pairs for subsequent training.

We utilize the Q values to reflect the preference for the next strategy at every turn. Specifically, the strategy with the highest Q value is selected as the optimal choice for the current dialogue turn. Correspondingly, the expression associated with this strategy is chosen with the highest value function score. Together, the selected strategy and its corresponding expression form the positive sample. This pair represents the most favorable outcome given the current dialogue state. Alongside the positive sample, we also construct negative samples to provide a contrastive learning signal. These samples are derived from strategies that are explored during the search but ultimately not selected due to lower Q values. For each unselected strategy, we identify its corresponding expression with the highest value function score. These pairs of unselected strategies and their associated expression serve as negative samples, indicating less favorable outcomes compared to the positive sample.

Preference Learning. Given the turn-level strategy-expression preference data collected via MCTS, we fine-tune the model (e.g., LLaMA) using DPO. The dataset $\mathcal{D} = (x, y_w, y_l)$ is a collection of data items, where each item is represented as a triplet (x, y_w, y_l). The triplet consists of an input prompt x, a preferred response y_w, and a dispreferred response y_l.

$$\mathcal{L}_{\text{dpo}}(\theta) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}} \log \sigma\left(\beta h_{\pi_{\text{ref}}}^{\pi_\theta}(x, y_w, y_l)\right) \quad (1)$$

The term $h_{\pi_{\text{ref}}}^{\pi_\theta}(x, y_w, y_l)$ quantifies the reward differential between the preferred response and the dispreferred response, defined as:

$$h_{\pi_{\text{ref}}}^{\pi_\theta}(x, y_w, y_l) = \log \frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)} - \log \frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)} \quad (2)$$

This process does not necessarily involve feeding the model with strictly optimal strategies or the most suitable expression. Instead, it aims to learn from past dialogue experiences and develop a capacity for look-forward planning, allowing it to make more informed decisions in new dialogue states. Additionally, compared to the MCTS method, this approach significantly reduces complex search processes and value calculations.

3.2 Multifaceted Reflexion Module.

Due to the complexity of negotiation scenarios that come from diverse dialogue states and different users, expression needs to be optimized from multiple perspectives. Additionally, even after fine-tuning, the model may still produce unstable outputs, making it insufficient to rely solely on the experience-based response module. To address this, we employ a reflective approach to make fine-grained adjustments to expression derived from the experience-based response module when necessary. Drawing on the Reflexion framework (Shinn et al., 2024), which features dynamic memory and reflective capabilities, we design a multi-critics and moderator reflexion mechanism named Multifaceted Reflexion. This mechanism is more adaptable for enhancing negotiation expression in multiple aspects. As shown in Figure 2, multifaceted reflexion comprises four key components:

Multi-Critics. The quality of negotiation expression can be evaluated from multiple perspectives, such as repetitiveness, continuity, richness, empathy, etc. Therefore, using a single-critic approach to assess sentence quality is insufficient. We employ a multi-critics approach to evaluate the current expression. On one hand, this method enhances the exploratory capabilities of critics, allowing for a more comprehensive identification of the shortcomings in the current expression and providing richer feedback. On the other hand, it ensures the quality and stability of the reflexion process, thereby reducing the occurrence of hallucinations and errors.

The reflexion module is only activated to rewrite suboptimal expressions, determined by multi-critics. If the majority of critics deem the expression quality inadequate, reflexion is triggered.

Moderator. To organize and synthesize the diverse evaluations from multi-critics, we introduce a moderator role. The primary function of a moderator is to maintain an objective stance and provide a comprehensive summary that captures the essence of feedback from the critics. This involves: (1) *Aggregating and Balancing Opinions*. The moderator consolidates the evaluations, highlighting common themes while reconciling conflicting views to ensure a balanced perspective. (2) *Enhancing Clarity*. By refining complex feedback into clear and actionable insights, the moderator ensures that the summarized evaluations are easily understood and implemented. (3) *Facilitating Improvement*. The moderator's summary serves as a guide for iterative refinement, bridging the gap between evaluations and the improvement of negotiation sentences.

Memory. In order to integrate dynamic memory and iterative reflexion, we introduce the memory structure. It functions as a repository that stores suboptimal negotiation sentences along with their corresponding feedback from the moderator in the current turn. DMNA leverages this memory to enhance the generation of new responses through specific instructions. After reflexion in a certain dialogue turn, the memory is updated and cleared to ensure that it contains only the most relevant and recent information for the current dialogue state. For more details regarding the regeneration of the actor component, please refer to Tables 8 and Table 9 in the Appendix.

Actor. The actor is based on the intuitive response model discussed in Section 3.1. Through experience-based learning, this model develops a certain level of negotiation capability. Due to the limitations of experience learning and inherent instability within the model, it is necessary to impose fine-grained quality constraints on the expression by the actor regenerating.

In general negotiation scenarios, the actor can provide a direct response. For complex dialogue states where the initial expression quality is poor, the actor employs the multifaceted reflexion to regenerate response based on refined experiences extracted from the memory. This iterative process improves response performance and better adapts to complex dialogue states.

4 Experiments

4.1 Experimental Setups

Dataset. To evaluate our framework, we conduct our experiments on the negotiation datasets, including PersuasionForGood (P4G; Wang et al. (2019)) and CraigslistBargain (CB; He et al. (2018)). These negotiation datasets involve two roles with distinct goals and pre-defined negotiation strategies, aiming to reach consensus through conversation. For details of predefined negotiation strategies, please see Appendix D.3. P4G is set in a persuading donation scenario, where the persuader attempts to convince the persuadee to donate to an organization called Save the Children. In contrast, CB is based on a bargaining scenario, the buyer tries to persuade the Seller to accept a lower price, while the seller aims to reach a consensus at a higher price.

Baselines. We provide comparisons with two types of negotiation agents: 1) Enhance pre-defined strategies planning (Planning-based), including *Pro-CoT* (Deng et al., 2023b), *TRIP* (Zhang et al., 2024) and *GDP-Zero* (Yu et al., 2023) 2) Optimize negotiation expression (Expression-optimized), including *DialoGPT* (Zhang et al., 2020), *ICL-AIF* (Fu et al., 2023) and *AnE* (Zhang et al., 2023). For detailed implementation of the above methods, please refer to the Appendix B.1.

Evaluation Metrics. We utilize two types of evaluation methodologies: goal-based metrics and quality-based metrics. In line with Deng et al. (2024b), we employ average turn (AT) and success rate (SR) as goal-based metrics. AT evaluates the number of dialogue turns needed to achieve the negotiation goal, while SR measures the proportion of dialogues that successfully reach the negotiation goal. For the CB, we use the Sale-to-List Ratio (SL) to evaluate the deal of the buyer. A higher SL indicates the buyer gets more benefits from deals and we set SL to 0 if the deal fails. Following Shi et al.

Agents	Price Negotiation						
	AT↓	SR↑	SL↑	N-Rep↑	Coh↑	Emp↑	Pers↑
ProCoT (Deng et al., 2023b)	7.62	0.60	0.2307	2.73	4.32	3.34	2.67
TRIP (Zhang et al., 2024)	6.34	0.68	0.4096	2.74	4.76	3.69	2.72
GDP-ZERO (Yu et al., 2023)	7.63	0.44	0.2401	4.24	4.78	3.51	3.21
DialoGPT (Zhang et al., 2020)	6.73	0.32	0.2012	2.70	4.01	3.32	2.62
LLaMA3.3-70b (Grattafiori et al., 2024)	10.26	0.57	0.2734	3.05	4.86	3.58	3.29
GPT-4o-mini (OpenAI et al., 2024)	10.02	0.48	0.2097	2.66	4.71	3.65	3.10
ICL-AIF (Fu et al., 2023)	8.42	0.34	0.2503	4.25	4.86	3.46	2.94
AnE (Zhang et al., 2023)	**5.60**	0.34	0.1742	4.37	**4.89**	3.46	3.03
DMNA (Ours)	6.80	**0.84**	**0.5359**	**4.52**	4.82	**3.69**	**4.04**

Agents	Persuasion for Good					
	AT↓	SR↑	N-Rep↑	Coh↑	Emp↑	Pers↑
ProCoT (Deng et al., 2023b)	9.90	0.18	2.89	4.27	3.56	3.13
TRIP (Zhang et al., 2024)	8.51	0.55	3.61	4.82	4.02	3.91
GDP-ZERO (Yu et al., 2023)	9.74	0.25	3.92	4.16	3.71	3.77
DialoGPT (Zhang et al., 2020)	9.73	0.22	2.67	4.51	3.24	2.92
LLaMA3.3-70b (Grattafiori et al., 2024)	12.13	0.52	4.36	**4.80**	**4.50**	4.12
GPT-4o-mini (OpenAI et al., 2024)	12.86	0.47	4.09	4.65	4.41	3.78
ICL-AIF (Fu et al., 2023)	10.54	0.43	3.27	4.66	3.78	3.89
AnE (Zhang et al., 2023)	10.32	0.46	3.40	4.79	4.07	3.84
DMNA (Ours)	**8.14**	**0.76**	**4.43**	4.74	4.13	**4.14**

Table 1: Evaluation results on Price Negotiation and Persuasion for Good. Compared to the planning-based and expression-optimized baselines, DMNA enhances the negotiation ability of the agent.

(2021) and Samad et al. (2022), we assess expression quality based on Non-repetitiveness (***N-Rep***), Coherence (***Coh***), Empathy (***Emp***), and Persuasiveness (***Pers***) as quality-based metrics. For this purpose, we utilize an evaluation method powered by LLMs, supplemented by human evaluation for validation. Detailed definitions of the evaluation metrics see the Appendix C.1.

Experimental Details. To enhance the realism of the negotiation environment, we employ the user simulators in TRIP as comprehensive user simulators. These simulators enable the LLMs to exhibit diverse personas and incorporate resistance strategies to counteract the persuasion attempts of agents. Our implementation of MCTS is based on the GDP-Zero framework, refer to its code and parameters. Moreover, we adopt LLaMA-3-8B-Instruct (Touvron et al., 2023) as the backbone for the experience model and GPT-3.5-turbo-1106 as the backbone for the multifaceted reflexion module.

4.2 Main Results & Human Evaluation

Tables 1 presents the evaluation results of our framework compared with selected baselines on the CB and P4G datasets respectively. Among the baseline methods, planning-based baselines achieve the highest SR and SL, while expression-optimized baselines demonstrate superior performance in expression metrics. This observation aligns with our expectations: planning-based baselines exhibit stronger look-forward capability for the goal, whereas expression-optimized baselines generate higher-quality expression.

DMNA effectively improves the negotiation ability of conversation agent, enabled by its dual cognitive mechanism comprising intuitive module and deliberate module. As illustrated in Table 1, DMNA significantly outperforms all baselines across both tasks. Specifically, DMNA achieves a higher SR and fewer AT in task completion, while also attaining superior performance in quality-based metrics. This comprehensive improvement makes DMNA better suited for human-centric conversational agents (Deng et al., 2024a), as it not only efficiently accomplishes tasks but also emphasizes human needs and expectations.

The proposed evaluation approach exhibits significant reliability and aligns closely with outcomes from human judgment. In Section 4.1, it is noted that we employ evaluation powered by LLMs. Given the generation of LLMs can be unstable, we further assess the reliability of the evaluation results by comparing them with human annotators. Initially, we sample 50 dialogues generated by DMNA on both CB and P4G datasets. These dialogues are then annotated by 3 human annota-

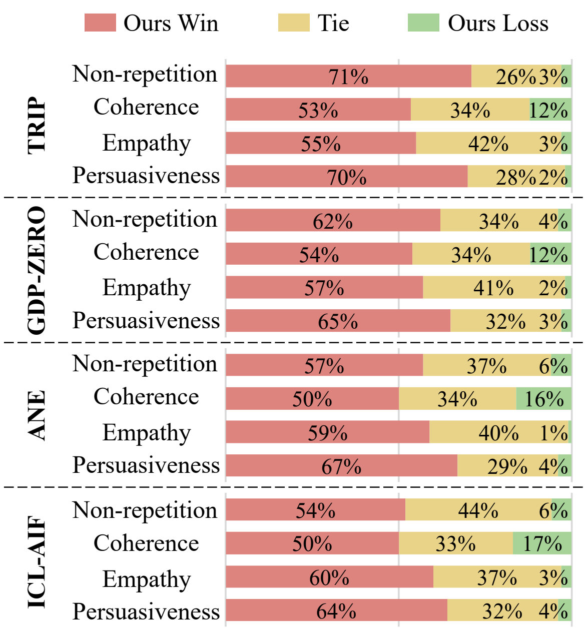


Figure 3: The results of human A/B test. Each bar shows the ratios for "Ours Wins," "Tie," and "Ours Loss" from left to right.

tors using the same standard as those applied by the LLM-based evaluator. Subsequently, we calculate the Spearman correlations between the average value of human annotations and the annotations of the LLMs evaluator. The results, as presented in Table 2, demonstrate a significant consistency, suggesting that LLM-powered evaluators can serve as a viable alternative to human annotators.

Moreover, our evaluation results exhibit a high degree of consistency with human judgment. Specifically, we conduct an A/B test comparing expressions generated by DMNA with those from baseline methods, evaluating them based on quality-based metrics. As illustrated in Figure 3, DMNA consistently outperforms other baseline methods, further confirming the effectiveness of our approach.

Dataset	N-Rep	Coh	Emp	Pers
CB (He et al., 2018)	0.61	0.52	0.56	0.67
P4G (Wang et al., 2019)	0.68	0.62	0.64	0.70

Table 2: The result of Spearman correlation statistics. The Spearman correlations between human evaluation results and our method's evaluation results indicate a strong correlation.

Agents	**Price Negotiation**							
	AT	SR	SL	N-Rep	Coh	Emp	Pers	Ref
DMNA	**6.80**	**0.84**	**0.5359**	**4.52**	**4.82**	**3.69**	**4.04**	**22.06**
w/o DPO	10.68	0.42	0.2524	3.57	4.11	3.46	3.50	24
w/o Ref	10.72	0.40	0.2303	3.00	3.53	2.96	2.69	–
w/o MC	9.80	0.58	0.4048	3.48	3.98	3.59	3.40	30.3
Agents	**Persuasion for Good**							
	AT	SR		N-Rep	Coh	Emp	Pers	Ref
DMNA	**8.14**	**0.76**		**4.43**	**4.74**	**4.13**	4.14	**3.24**
w/o DPO	9.51	0.66		3.87	4.73	4.01	**4.15**	5.32
w/o Ref	8.72	0.67		4.31	4.76	3.92	3.92	–
w/o MC	9.20	0.70		4.36	4.77	3.95	4.06	5.66

Table 3: The evaluation results of ablation study. The experience-based response module, multifaceted reflexion module, and multi-critics and moderator reflexion mechanism are effective in improving agents and complement each other.

4.3 Ablation Study

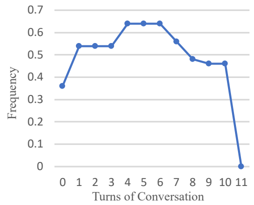
To explore the effects of each component in DMNA, we devise several variants as follows:

- **w/o DPO** represents removing the experience-based response module, the backbone of the intuitive module being LLaMA-3-8B-Instruct.
- **w/o Ref** represents removing the multifaceted reflexion module, where the model only provides intuitive responses.
- **w/o MC** represents removing the multi-critics and moderator reflexion mechanism, using single-critic reflexion instead.

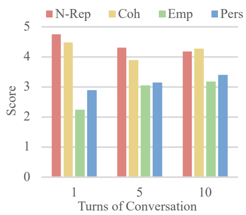
We summarize the performance of each model variation. Based on the results in Table 3, we obtain the following observations:

The intuitive module, after experience learning, exhibits look-forward planning capability. Specifically, the result illustrates that DMNA outperforms w/o DPO by improving the SR and SL. This suggests that the intuitive module, following experience learning, possesses look-forward proactive and goal-oriented ability. In addition, DMNA reduces the number of reflexion iterations, indicating that the expression of DMNA also has a certain degree of negotiation capability, enabling it to provide intuitive responses.

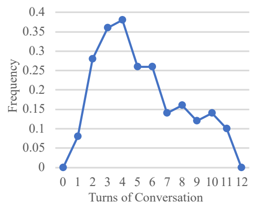
The deliberate module can enhance the quality of expression and further improve negotiation outcomes. In the investigation of DMNA and w/o Ref, we note that DMNA achieves greater improvements in the expression metrics (N-Rep, Coh, Emp, and Pers) of negotiation expression. It indicates that the deliberate module constrains and



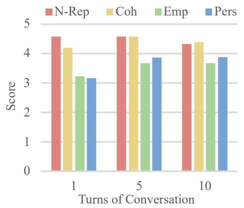
(a) Frequency of triggering reflexion w.r.t conversation turns in CB.



(b) Expression quality w.r.t conversation turns in CB.

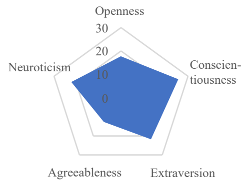


(c) Frequency of triggering reflexion w.r.t conversation turns in P4G.

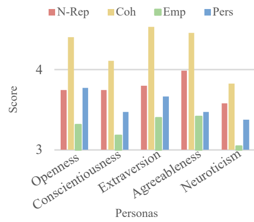


(d) Expression quality w.r.t conversation turns in P4G.

Figure 4: Frequency of triggering reflexion and expression quality of DMNA in different conversation turns.



(a) The number of reflexion iterations in DMNA corresponding to different personas.



(b) The performance of expression quality in DMNA across different personas.

Figure 5: The number of reflexion iterations and expression quality in different persons.

optimizes the quality of expression, leading to better negotiation performance (SR and SL).

Multifaceted reflexion outperforms reflexion and reduces the number of reflexion iterations. By comparing DMNA and w/o MC, we find that multifaceted reflexion demonstrates higher performances than single-critic reflexion, especially in expression evaluations. It suggests that multifaceted reflexion provides a more comprehensive feedback for expression quality than single-critic. Moreover, multifaceted reflexion reduces the number of reflexion iterations proving that this approach offers greater stability. These findings collectively prove that the multifaceted reflexion not only offers stable feedback but is also better suited for the complex dynamics of dialogue states.

4.4 In-depth Analysis

Frequency of triggering Reflexion and Quality of Expression w.r.t Conversation Turns. In this part, we analyze the variations in the frequency of triggering reflexion and the quality of expression as the conversation progresses. As shown in Figure 4(a) and 4(c), we note that there is an initial gradual increase in the frequency of reflexion, which peaks during the middle turns of the conversation, followed by a significant decline in the later stages. The findings suggest that reflexion plays a crucial role during the early and middle stages of the dialogue, while its importance tends to diminish in the later stages.

To further analyze the relationship between reflexion frequency and expression quality, we examine different intervals of dialogue turns (0-1, 2-5, 6-10). As illustrated in Figure 4(b) and 4(d), expression quality generally improves with an increase in the number of turns. This suggests that DMNA tends to optimize its expression through reflexion on iterations. For instance, comparing the intervals of dialogue turns 0–1 and 2–5, the average reflexion frequency increase. Correspondingly, we observe that DMNA yields positive improvements across expression metrics, such as a significant increase in Coh, Emp, and Pers. It is worth noting that, as the conversation proceeds, the N-Rep metric exhibits a declining trend. This may be attributed to the agent's increasing focus on topics and expression patterns that may interest the user in order to more effectively advance the negotiation process. These results demonstrate that multiple reflexions can significantly enhance expression quality.

Analysis the Adaptability of DMNA across Different Personas. As shown in Figure 5(a) and 5(b), we conduct an in-depth analysis of the quality-based performance of DMNA across different Big-Five Personality (Goldberg, 1992). Specifically, we assess the average iterations number of reflexion and the average value of expression metrics (N-Rep, Coh, Emp, and Pers) for every persona in Big-Five personality types. For example, individuals with high conscientiousness typically demand detailed and well-planned information. This leads to more frequent adjustments and optimizations in DMNA's expression, resulting in the highest number of reflexion among all persona types. However, DMNA's performance in Emp and Pers is relatively weaker for this group. This may be because individuals with high conscientiousness prioritize task completion and accuracy over emotional resonance. The evaluation results indicate that DMNA exhibits the flexibility of the iterations number of reflexion and varying expression depending on the user's persona. We also analyze the adaptability adjustments in the other four persona types. For more details, please see Appendix C.2.

5 Conclusion

In this study, we propose the Dual-Mind Negotiation Agent (DMNA) framework, which integrates strategic planning and expressive optimization to enhance the negotiation ability of agents. Inspired by the dual-process theory in human cognition, DMNA comprises two modules: an intuitive module trained using MCTS and DPO, and a deliberative module that employs a multifaceted reflexion mechanism to optimize expression quality. Our experimental results on two negotiation datasets demonstrate that DMNA significantly outperforms existing state-of-the-art methods. These findings indicate that DMNA effectively bridges the gap between planning-based and expression-based methods, offering a more human-centric approach to negotiation dialogue agents.

Limitation

Although the DMNA framework has shown promising results on existing datasets, its performance may be limited in more complex scenarios such as multi-party negotiations, cross-cultural interactions, and multimodal negotiation domains. For instance, in multi-party negotiations, the intertwined interests and diverse strategies of multiple participants require the agent to coordinate and optimize negotiation goals in real time. In cross-cultural negotiations, differences in cultural backgrounds can lead to varying interpretations of the same strategy, demanding stronger cultural adaptability and expressive flexibility from the negotiation agent. Moreover, as negotiations increasingly involve multimodal interactions, DMNA needs to enhance its ability to process non-textual information such as visual and auditory cues to better meet the demands of complex negotiation scenarios.

Therefore, our future work will focus on addressing these underexplored challenges by incorporating richer training data, optimizing strategy generation mechanisms, and enhancing multimodal interaction capabilities, thereby further improving DMNA's performance in diverse and complex real-world negotiation settings.

Ethics Statement

Our Dual-Mind Negotiation Agent (DMNA) to enhance the effectiveness of dialogue systems in assisting users or systems in accomplishing tasks and goals. We explicitly reject the use of DMNA for unethical purposes such as manipulation or fraud. We are committed to ensuring our work benefits users and society. The risks involved are as follows:

Automation bias and user dependence: Systems may have automation bias. Users may overly rely on DMNA's negotiation strategies and expressions. To address this, we clarify that DMNA is designed to assist users in negotiating, not to replace human judgment. Users should critically assess the system's output and make decisions based on their own understanding.

Non-zero-sum negotiation dynamics: Negotiation scenarios are not always zero-sum games. The DMNA aims to promote mutual agreements and cooperation among parties. It does not prioritize one party's interests over another's. DMNA focuses on finding common ground and achieving mutually beneficial outcomes.

Personality trait measurement bias: Section 4.4 shows performance variations among the 'Big Five' personality types. While the analysis highlights adaptive differences, we do not explicitly address fairness or mitigate potential drawbacks of specific traits. This remains an open challenge.

Human annotator conditions: Human annotators are involved in verifying the quality of expressions. They are compensated at a rate of 15 dollars per hour, with tasks limited to 2 hours to prevent fatigue.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant No.62476111 and Grant No.72091315, the Changchun Science and Technology Bureau under Grant No. 24GNYZ11, the Department of Science and Technology of Jilin Province, China under Grant No.20230201086GX and Industry University Research Innovation Fund of the Ministry of Education under Grant No.2022XF017.

References

Yoshua Bengio. 2019. The consciousness prior.

Xiaoxue Cheng, Junyi Li, Wayne Xin Zhao, and Ji-Rong Wen. 2025. Think more, hallucinate less: Mitigating hallucinations via dual process of fast and slow thinking.

Yi Cheng, Wenge Liu, Wenjie Li, Jiashuo Wang, Ruihui Zhao, Bang Liu, Xiaodan Liang, and Yefeng Zheng. 2022. Improving multi-turn emotional support dialogue generation with lookahead strategy planning. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 3014–3026, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Yang Deng, Wenqiang Lei, Wai Lam, and Tat-Seng Chua. 2023a. A survey on proactive dialogue systems: Problems, methods, and prospects. *arXiv preprint arXiv:2305.02750*.

Yang Deng, Lizi Liao, Liang Chen, Hongru Wang, Wenqiang Lei, and Tat-Seng Chua. 2023b. Prompting and evaluating large language models for proactive dialogues: Clarification, target-guided, and non-collaboration. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 10602–10621, Singapore. Association for Computational Linguistics.

Yang Deng, Lizi Liao, Zhonghua Zheng, Grace Hui Yang, and Tat-Seng Chua. 2024a. Towards human-centered proactive conversational agents. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 807–818.

Yang Deng, Wenxuan Zhang, Wai Lam, See-Kiong Ng, and Tat-Seng Chua. 2024b. Plug-and-play policy planner for large language model powered dialogue agents. In *The Twelfth International Conference on Learning Representations*.

Ritam Dutt, Sayan Sinha, Rishabh Joshi, Surya Shekhar Chakraborty, Meredith Riggs, Xinru Yan, Haogang Bao, and Carolyn Rose. 2021. ResPer: Computationally modelling resisting strategies in persuasive conversations. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 78–90, Online. Association for Computational Linguistics.

Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. 2023. Improving language model negotiation with self-play and in-context learning from ai feedback. *arXiv preprint arXiv:2305.10142*.

Xiaoyang Gao, Siqi Chen, Yan Zheng, and Jianye Hao. 2021. A deep reinforcement learning-based agent for negotiation with multiple communication channels. In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 868–872.

Lewis R. Goldberg. 1992. The development of markers for the big-five factor structure. *Psychological Assessment*, 4:26–42.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, Danny Wyatt, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Francisco Guzmán, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Govind Thattai, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jack Zhang, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Karthik Prasad, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Kushal Lakhotia, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Maria Tsimpoukelli, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Ning Zhang, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohan Maheswari, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten

Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vítor Albiero, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaofang Wang, Xiaoqing Ellen Tan, Xide Xia, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aayushi Srivastava, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Amos Teo, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Dong, Annie Franco, Anuj Goyal, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Ce Liu, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Cynthia Gao, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Eric-Tuan Le, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Filippos Kokkinos, Firat Ozgenel, Francesco Caggioni, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hakan Inan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Hongyuan Zhan, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Ilias Leontiadis, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Janice Lam, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kiran Jagadeesh, Kun Huang, Kunal Chawla, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Miao Liu, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikhil Mehta, Nikolay Pavlovich Laptev, Ning Dong, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Rangaprabhu Parthasarathy, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Russ Howes, Ruty Rinott, Sachin Mehta, Sachin Siby, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Mahajan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shishir Patil, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Summer Deng, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Koehler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaojian Wu, Xiaolan Wang, Xilun Wu, Xinbo Gao, Yaniv Kleinman, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yu Zhao, Yuchen Hao, Yundi Qian, Yunlu Li, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, Zhiwei Zhao, and Zhiyu Ma. 2024. The llama 3 herd of models.

He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2333–2343, Brussels, Belgium. Association for Computational Linguistics.

Tao He, Lizi Liao, Yixin Cao, Yuanxing Liu, Ming Liu, Zerui Chen, and Bing Qin. 2024. Planning like human: A dual-process framework for dialogue planning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Vol-*

ume 1: Long Papers), pages 4768–4791, Bangkok, Thailand. Association for Computational Linguistics.

Daniel Kahneman. 2003a. Maps of bounded rationality: Psychology for behavioral economics. *American economic review*, 93(5):1449–1475.

Daniel Kahneman. 2003b. Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review*, 93(5):1449–1475.

Bill Yuchen Lin, Yicheng Fu, Karina Yang, Faeze Brahman, Shiyu Huang, Chandra Bhagavatula, Prithviraj Ammanabrolu, Yejin Choi, and Xiang Ren. 2023. Swiftsage: A generative agent with fast and slow thinking for complex interactive tasks. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Kshitij Mishra, Azlaan Mustafa Samad, Palak Totala, and Asif Ekbal. 2022. PEPDS: A polite and empathetic persuasive dialogue system for charity donation. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 424–440, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2024. Gpt-4 technical report.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Azlaan Mustafa Samad, Kshitij Mishra, Mauajama Firdaus, and Asif Ekbal. 2022. Empathetic persuasion: Reinforcing empathy and persuasiveness in dialogue systems. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 844–856, Seattle, United States. Association for Computational Linguistics.

Weiyan Shi, Yu Li, Saurav Sahay, and Zhou Yu. 2021. Refine and imitate: Reducing repetition and inconsistency in persuasion dialogues via reinforcement learning and human demonstration. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3478–3492, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2024. Reflexion: language agents with verbal reinforcement learning. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA. Curran Associates Inc.

Xiaoyu Tian, Liangyu Chen, Na Liu, Yaxuan Liu, Wei Zou, Kaijiang Chen, and Ming Cui. 2023. Duma: a dual-mind conversational agent with fast and slow thinking.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5635–5649, Florence, Italy. Association for Computational Linguistics.

Changrong Xiao, Wenxing Ma, Qingping Song, Sean Xin Xu, Kunpeng Zhang, Yufang Wang, and Qi Fu. 2024. Human-ai collaborative essay scoring: A dual-process framework with llms.

Cheng Yang, Chufan Shi, Siheng Li, Bo Shui, Yujiu Yang, and Wai Lam. 2024. Llm2: Let large language models harness system 2 reasoning.

Fu-Hao Yu, Ke-Han Lu, Yi-Wei Wang, Wei-Zhe Chang, Wei-Kai Huang, and Kuan-Yu Chen. 2021. 2020 (a preliminary study of Formosa speech recognition challenge 2020 – Taiwanese ASR). In *International Journal of Computational Linguistics & Chinese Language Processing, Volume 26, Number 1, June 2021*, Taipei, Taiwan. Association for Computational Linguistics and Chinese Language Processing.

Xiao Yu, Maximillian Chen, and Zhou Yu. 2023. Prompt-based Monte-Carlo tree search for goal-oriented dialogue policy planning. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7101–7125, Singapore. Association for Computational Linguistics.

Haolan Zhan, Yufei Wang, Tao Feng, Yuncheng Hua, Suraj Sharma, Zhuang Li, Lizhen Qu, Zhaleh Semnani Azad, Ingrid Zukerman, and Gholamreza Haffari. 2024. Let's negotiate! a survey of negotiation dialogue systems. *arXiv preprint arXiv:2402.01097*.

Haodi Zhang, Zhichao Zeng, Keting Lu, Kaishun Wu, and Shiqi Zhang. 2022. Efficient dialog policy learning by reasoning with contextual knowledge. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(10):11667–11675.

Qiang Zhang, Jason Naradowsky, and Yusuke Miyao. 2023. Ask an expert: Leveraging language models to improve strategic reasoning in goal-oriented dialogue models. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 6665–6694, Toronto, Canada. Association for Computational Linguistics.

Tong Zhang, Chen Huang, Yang Deng, Hongru Liang, Jia Liu, Zujie Wen, Wenqiang Lei, and Tat-Seng Chua. 2024. Strength lies in differences! improving strategy planning for non-collaborative dialogues via diversified user simulation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 424–444, Miami, Florida, USA. Association for Computational Linguistics.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2020. DIALOGPT : Large-scale generative pre-training for conversational response generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 270–278, Online. Association for Computational Linguistics.

A Details of MCTS

In the process of constructing the strategy-expression preference dataset, we refer to the implementation of GDP-ZERO (Yu et al., 2021). As a supplement to the main body, we describe the implementation details of the MCTS algorithm here. We implement the MCTS algorithm on two datasets respectively. We treat the current dialogue state in turn i look as a sequence of dialogue actions $s_i^{tr} = (a_0, \dots, a_i)$ and the chosen dialogue strategy as a^*. By iteratively executing the four phases of MCTS, we continuously update the relevant variables: Q-values, state values V, and visit counts N. After K iterations, MCTS selects the optimal strategy based on these variables. Specifically:

Selection. For the current node s^{tr}, this phase aims to choose a dialogue strategy from a^* to reach a child node. To balance exploration and expansion, we employ the PUCB function to select the dialogue strategy:

$$\mathrm{PUCT}\left(s^{tr}, a\right) = Q\left(s^{tr}, a\right) + \mathrm{c}_p \frac{\sqrt{\sum_a N\left(s^{tr}, a\right)}}{1 + N\left(s^{tr}, a\right)}$$

The algorithm will keep selecting a child node in sequence until a leaf node is reached.

Expansion. Upon reaching a leaf node, we use the LLM as a policy network to prompt it to generate the next dialogue action distribution. This is achieved by sampling the LLM at a temperature of 1.0 for m times and then converting the sampled dialogue acts into a probability distribution. Finally, each strategy is set with $Q\left(s^{tr}, \cdot\right) = Q_0$, which is a hyperparameter that influences exploration.

Evaluation. We determine the value of a state a based on the likelihood of its dialogue context leading to task success. In the Persuasion for Good task, it is evaluated at convincing a user to donate to a charity, this can be done by adding the utterance "Would you like to make a donation?" to the context. In the CraigslistBargain task, the goal is to negotiate with the Seller to reach a mutually agreeable price, which is evaluated by identifying the negotiated price to context.

Backpropagation. After each search iteration concludes, we perform the following updates for the above variables:

$$\begin{aligned} N\left(s^{tr}, a\right) &\leftarrow N\left(s^{tr}, a\right) + 1 \\ Q\left(s^{tr}, a\right) &\leftarrow Q\left(s^{tr}, a\right) + \Delta Q\left(s^{tr}, a\right) \\ \Delta Q\left(s^{tr}, a\right) &= \frac{v\left(s^{tr} - Q\left(s^{tr}, a\right)\right)}{N\left(s^{tr}, a\right)} \end{aligned}$$

After all simulations, we select the optimal strategy for the current state s^{tr} using a formula:

$$a^* = \arg\max_a N\left(s_0^{tr}, a\right)$$

B More Implementation Details

B.1 Implemention of Baselines

We follow the original design of the baseline methods and categorize them into planning-based agents and expression-based agents. To compare the two types of agents with DMNA, we adapt these agents to the applications in our experiments for two tasks:

ProCoT (Planning-based): Following Deng et al. (2023b), we instruct the LLMs(e.g., gpt-3.5-turbo-1106) to analyze the current dialogue context, choose the next strategy, and generate a response accordingly.

TRIP (Planning-based): Following Zhang et al. (2024), We implement TRIP based on the details provided in the paper, utilizing a user-aware strategic planning module and a population-based training paradigm to enhance the adaptability of dialogue agents to diverse users.

GDP-ZERO (Planning-based): Following Yu et al. (2023), we leverage the open-MCTS method to enable strategic planning by LLMs. Specifically, we utilize a large language model (e.g., ChatGPT) to serve as a policy prior, value function, user simulator, and system model during the tree search process.

DialoGPT (Expression-based): DialoGPT is a widely used model in the field of dialogue systems, known for its strong performance in generating coherent and contextually relevant responses(Zhang et al., 2020). We instruct DialoGPT-large as a negotiation agent to chat with the user in two tasks.

ICL-AIF (Expression-based): Following Fu et al. (2023), we prompt GPT3.5 to provide verbal feedback, offering suggestions to the dialogue agent at the end of each interaction. Specifically, our implementation includes presenting three suggestions after each interaction, ensuring that only the most recent 20 suggestions are retained to prevent excessive accumulation.

AnE (Expression-based): Following Zhang et al. (2023), we prompt GPT3.5 to act as a negotiation expert by posing M-part questions that guide reasoning about the next response suggestion.

B.2 Implemention of Training

We use LLaMA-3-8B-Instruct as our base pre-train model. The DPO experiment is conducted with a 24G GPU(NVIDIA RTX4090). We choose the learning rate 5e-6 for DPO training, with a cosine learning rate scheduler. The training epoch is 3. The maximum sequence length of models is 512. We train the model with a batch size of 1. Follow the DPO paper(Rafailov et al., 2023) to set the KL constraint parameter as 0.1. Each sample in DPO is a set of step-level preference data decomposed by MCTS. We set the number of MCTS iterations as $K = 10$ for two tasks.

C More Details of Evaluation

C.1 Definitions of quality-based metrics

To assess the quality of expression during the negotiation process, we refer to Shi et al. (2021) and Samad et al. (2022), establish four metrics for evaluation. Each of the four metrics is evaluated on a five-point scale:

Non-repetition (N-Rep): Non-repetition measures the diversity and uniqueness of the expression generated by the negotiation agent. It evaluates whether the agent can produce a variety of responses instead of repeating the same phrases or sentences.

Coherence (Coh): Coherence assesses the logical flow and consistency of the agent's responses within the conversation. It ensures that the agent's statements are relevant, connected, and make sense in the context of the ongoing dialogue.

Empathy (Emp): Empathy evaluates the agent's ability to understand and share the feelings of the counterpart. It measures how well the agent can respond in a way that demonstrates emotional intelligence and sensitivity.

Persuasiveness (Pers): Persuasiveness measures the agent's ability to influence the counterpart's opinions or decisions. It evaluates how effectively the agent can use language and arguments to persuade the counterpart to agree with its proposals or suggestions.

C.2 Analysis of the Adaptability of Different Personas

In Section 4.4, we conclude that DMNA demonstrates strong adaptability across different personas. For users with conscientiousness, we provided an analysis in the main body. As a supplement, we offer here an analysis of its adaptability to the other four personality types:

Openness: As shown in Figure 5(a) and 5(b), individuals with high openness require a moderate number of reflexions (17.71) to optimize expression. When interacting with openness individuals, DMNA generates expression with higher Coh and Pers but shows slightly weaker Emp. This may be because openness users are more willing to accept new perspectives and strategies, allowing DMNA to adapt with fewer reflexions.

Extraversion: Extraversion individuals require a moderate number of reflexions (21.78) to optimize expression. When interacting with extraversion individuals, DMNA generates expression with higher Coh and Pers. Extraversion individuals typically value interaction and energy, and DMNA can meet these needs with an appropriate number of reflexions while maintaining Coh and Pers.

Agreeableness: Agreeableness individuals require the fewest reflexions (12.57), indicating that DMNA can quickly generate expressions that meet their expectations. Agreeableness individuals focus more on cooperation and empathy, and DMNA can rapidly produce expression with high N-Rep and Coh. Although the value of Pers is average, agreeableness individuals may prioritize cooperation and emotional resonance.

Neuroticism: Neuroticism individuals require a relatively high number of reflexions (22.22), indicating that DMNA needs to frequently adjust and optimize expression when interacting with them. Neuroticism individuals experience higher emotional volatility, and DMNA needs multiple reflexions to alleviate their anxiety. However, their expression quality is relatively weak across all dimensions, likely because neuroticism users have lower adaptability to stress and challenges and require more support and reassurance.

D Details of Prompting

D.1 User Simulation

Due to the human involving conversations is time-consuming, we resort to role-playing to simulate users with LLMs. To make the negotiation scenarios more realistic, we employ a user simulator with comprehensive prompts which are endowed with different personas, resistance strategies, and task descriptions. We draw upon the prompts of the user simulator from TRIP (Zhang et al., 2024).

Specifically, for persona, we sample one attribute from the *Big Five personality* and one from *Decision-Making Styles*, serving as a set of basic persona for the user. In total, we sample 20 sets of personas and ensure the balance of each attribute. Then, we utilize GPT4 to generate 300 specific task descriptions for sampling into the user role-playing prompt. Regarding resistance strategies, we adopt those from Dutt et al. (2021) as user behaviors and integrate them into the user role-playing prompt.

The comprehensive prompt encompasses several parts: the background of the task, conversation history, user personality, resistance strategy, and specific prompts used in two tasks, which can be found in Tables 6 and 7.

D.2 Assistant Simulation

In the intuitive module, we prompt the actor to respond to the current dialogue state through role-playing. The template content includes the background of the task, conversation history, dialogue strategies, and previous dialogue experience (have if regenerated after reflexion). The prompts used in the two tasks can be seen in Tables 8 and 9.

D.3 Details of Strategy

Here, we present the negotiation strategies used in the two tasks. In the CB task, there are 11 strategies involved. In the P4G task, 10 strategies are involved. For detailed negotiation strategies and their descriptions, see Tables 4 and 5.

D.4 Details of Component in Multifaceted Reflexion Module

This part describes the relevant prompt content used in the multifaceted reflexion module. The multifaceted reflexion module involves roles such as multi-critics, monitor, and actor. The specific roles of these characters be introduced in the main body, and the specific prompt content for each role can be found in Tables 10, 11, and 12 respectively.

Negotiation Strategy	Explanation
Greetings	Please say hello or chat randomly.
Ask a question	Please ask any question about product, year, price, usage, etc.
Answer a question	Please provide information about the product, year, usage, etc.
Propose the first price	Please initiate a price or a price range for the product.
Propose a counter price	Please propose a new price or a new price range.
Use comparatives	Please propose a vague price by using comparatives with existing price.
Confirm information	Please ask a question about the information to be confirmed
Affirm confirmation	Please give an affirmative response to a confirm.
Deny confirmation	Please give a negative response to a confirm.
Agree with the proposal	Please agree with the proposed price.
Disagree with a proposal	Please disagree with the proposed price.

Table 4: The negotiation strategies which DMNA employs in CB.

Negotiation Strategy	Explanation
Logical Appeal	Please use of reasoning and evidence to convince the persuadee.
Emotion Appeal	Please elicit the specific emotions to influence the persuadee.
Credibility Appeal	Please use credentials and cite organizational impacts to establish credibility and earn the user's trust. The information usually comes from an objective source (e.g., the organization's website or other well-established websites).
Foot in the Door	Please use the strategy of starting with small donation requests to facilitate compliance followed by larger requests.
Self-Modeling	Please use the self-modeling strategy where you first indicates the persuadee own intention to donate and chooses to act as a role model for the persuadee to follow.
Personal Story	Please use narrative exemplars to illustrate someone donation experiences or the beneficiaries positive outcomes, which can motivate others to follow the actions.
Donation Information	Please provide specific information about the donation task, such as the donation procedure, donation range, etc. By providing detailed action guidance, this strategy can enhance the persuadee's self-efficacy and facilitates behavior compliance.
Source-related Inquiry	Please ask if the persuadee is aware of the organization (i.e., the source in our specific donation task).
Task-related Inquiry	Please ask about the persuadee opinion and expectation related to the task, such as their interests in knowing more about the organization.
Personal-related Inquiry	Please asks about the persuadee previous personal experiences relevant to charity donation.

Table 5: The negotiation strategies which DMNA employs in P4G.

The user simulator prompt for CB
Now enter the role-playing mode. In the following conversation, you will play as a seller in a price bargaining game. Your persona: %s. You must follow the instructions below during chat. 1. Your utterances and bargain behavior need to strictly follow your persona. Varying your wording and avoid repeating yourself verbatim! 2. You can decide to change your target price flexibly based on your persona and the conversation. Here are some conversation strategies you can follow: 1. "Source Derogation": Attacks the other party or questions the item. 2. "Counter Argument": Provides a non-personal argument/factual response to refute a previous claim or to justify a new claim. 3. "Personal Choice": Provides a personal reason for disagreeing with the current situation or chooses to agree with the situation provided some specific condition is met. 4. "Information Inquiry": Requests for clarification or asks additional information about the item or situation. 5. "Self Pity": Provides a reason (meant to elicit sympathy) for disagreeing with the current terms. 6. "Hesitance": Stalls for time and is hesitant to commit; specifically, they seek to further the conversation and provide a chance for the other party to make a better offer 7. "Self-assertion": Asserts a new claim or refutes a previous claim with an air of finality/ confidence. 8. "Others": Do not explicitly foil the negotiation attempts. You are the seller who is trying to sell the %s with the initial price of %s. Product description: %s. Please reply with only one short and succinct sentence. ******** Conversation History ********

Table 6: The comprehensive prompt of user simulators in CB.

The user simulator prompt for P4G
Now enter the role-playing mode. In the following conversation, you will play as a Persuadee in a persuasion game. Your persona: %s. You must follow the instructions below during chat. 1. Your utterances need to strictly follow your persona. Varying your wording and avoid repeating yourself verbatim! 2. Pretend you have little knowledge about the Save the Children charity. You have little willingness for donation at the beginning of conversation. 3. Your willingness for donation depends on your persona and how influenced you are by the Persuader. 4. You need to determine whether to donate money or not. If you think your are convinced by the Persuader, you should donate money for the charity. Your Response Strategy: 1. "Donate": show your willingness to donate. 2. "Source Derogation": attacks or doubts the organisation's credibility. 3. "Counter Argument": argues that the responsibility is not on them or refutes a previous statement. 4. "Personal Choice": Attempts to saves face by asserting their personal preference such as their choice of charity and their choice of donation. 5. "Information Inquiry": Ask for factual information about the organisation for clarification or as an attempt to stall. 6. "Self Pity": Provides a self-centred reason for not being willing to donate at the moment. 7. "Hesitance": Attempts to stall the conversation by either stating they would donate later or is currently unsure about donating. 8. "Self-assertion": Explicitly refuses to donate without even providing a personal reason. 9. "Others": Do not explicitly foil the persuasion attempts. You are the Persuadee who is being persuaded by a Persuader. Please reply with only one short and succinct sentence. ******** Conversation History ********

Table 7: The comprehensive prompt of user simulators in P4G.

The prompt of DMNA actor in CB
Now enter the role-playing mode. In the following conversation, you will play as a buyer in a price bargaining game. You should choose among the following actions during a conversation to response to the Seller: %s. Please chat with the Seller using short and natural sentence. You are the buyer who is trying to buy the %s with the price of %s. Product description: %s. Choose a strategy from the list provided and provide a corresponding response based on the Conversation Context. Please reply with only one short and succinct sentence. Conversation Context:%s. (If is Actor after multifaced reflexion: In the context of the conversation above, here are some experience poor expression quality, along with corresponding reflections: %s.) Please output the appropriate and high quality response and format your answer as '[strategy]response'. Your Response:

Table 8: The prompt design of the DMNA actor for CB.

The prompt of DMNA actor in P4G
Now enter the role-playing mode. In the following conversation, you will play as a Persuader who is trying to persuade the Persuadee to donate to the charity called Save the Children. Save the Children is head-quartered in London, and they work to help fight poverty around the world. Children need help in developing countries and war zones. Small donations like $1 or $2 go a long way to help. You should choose among the following actions during a conversation to respond to the Persuader: %s. Please chat with the Persuadee using short and natural sentence. As the Persuader, in order to persuade the Persuadee to donate to a charity called Save the Children, choose a strategy from the list provided and provide a corresponding response based on the Conversation Context. Please chat with the Persuadee using short and natural sentence. Conversation Context:%s. (If is Actor after multifaced reflexion: In the context of the conversation above, here are some experience poor expression quality, along with corresponding reflections: %s.) Please output the appropriate and high quality response and format your answer as '[strategy]response'. Your Response:

Table 9: The prompt design of the DMNA actor for P4G.

The prompt of DMNA Multi-Critics in CB
You are the critic for a conversation. Your task is to perform a fine-grained analysis of the Buyer's latest response in current communication. Determine can the Buyer's latest response positively influence the progression of future bargain efforts to negotiate down the Seller's price? You can consider the following example aspects in your analysis: 1. Whether the Buyer maintains a polite and respectful tone throughout the conversation, even when disagreements arise. 2. Does the price given conform to the bargain logic? The Buyer's price should be more and more to reach an agreement with the seller. 3. Whether the Buyer offers different angles or reasons for their request, rather than repeating the same point. Please format your response as: Answer:Yes or No. (If No)Suggestion: your concrete suggestion. The following is the conversation: %s. Question: Does the Buyer's latest response positively influence the progression of future bargain?

Table 10: The prompt design of the Multi-Critics for CB.

The prompt of DMNA Multi-Critics in P4G
You are the critic for a conversation. Your task is to perform a fine-grained analysis of the Persuader's current communication. Determine if these expression can positively influence the progression of future persuasion efforts to persuade Persuadee donate. You can consider the following example aspects in your analysis: 1. Whether address the Persuadee's expressed needs and concerns. 2. Whether lack of empathy and trust with the Persuadee. 3. Whether keep open and respectful in the communication. 4. Whether the Persuader's last response is similar to previous turn, lack of initiative and richness. Please format your response as: Answer:Yes or No. (If No)Suggestion: your concrete suggestion. The following is the conversation: %s. Question: Does the Persuader's latest response positively influence the progression of future persuasion?

Table 11: The prompt design of the Multi-Critics for P4G.

The prompt of DMNA Moderator
You are a reflection craft proposer. Your task is summarize the ideas that have been presented into a draft designed to satisfy the maximum number of agents. Below is the ideas from num agents:reflectionsdraft of reflection:

Table 12: The prompt design of the Moderator.