

# METRO: Towards Strategy Induction from Expert Dialogue Transcripts for Non-collaborative Dialogues

Haofu Yang<sup>4-6</sup>, Jiayi Liu<sup>2</sup>, Chen Huang<sup>1†</sup>, Faguo Wu<sup>4-6†</sup>, Wenqiang Lei<sup>3,7</sup>, See-Kiong Ng<sup>1</sup>

<sup>1</sup> National University of Singapore   <sup>2</sup> Nankai University   <sup>3</sup> Sichuan University

<sup>4</sup> School of Artificial Intelligence, Beihang University

<sup>5</sup> Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing

<sup>6</sup> Key Laboratory of Mathematics, Informatics and Behavioral Semantics (LMIB)

<sup>7</sup> Engineering Research Center of Machine Learning and Industry Intelligence,  
Ministry of Education, China

## Abstract

Developing non-collaborative dialogue agents traditionally requires the manual, unscalable codification of expert strategies. We propose METRO, a method that leverages large language models to autonomously induce both strategy actions and planning logic directly from raw transcripts. METRO formalizes expert knowledge into a Strategy Forest, a hierarchical structure that captures both short-term responses (nodes) and long-term strategic foresight (branches). Experimental results across two benchmarks show that METRO demonstrates promising performance, outperforming existing methods by an average of 9%-10%. Our further analysis not only reveals the success behind METRO (strategic behavioral diversity and foresight), but also demonstrates its robust cross-task transferability. This offers new insights into building non-collaborative agents in a cost-effective and scalable way. Our code is available at <https://github.com/Humphrey-0125/METRO>.

## 1 Introduction

Building conversational agents for non-collaborative dialogues is receiving increased attention (Zhang et al., 2024a; Hu et al., 2025; Deng et al., 2024, 2023b), particularly in applications like price negotiation (He et al., 2018), donation persuasion (Wang et al., 2019), or debt collection (Zhang et al., 2023). In these interactions, the user (i.e., the opposing party) holds conflicting interests with the agent and consequently exhibits resistance during the multi-turn conversation (Deng et al., 2023a; Lei et al., 2022; Shi et al., 2019; Dutt et al., 2021). This dynamic requires the agent to employ effective dialogue strategies, securing an agreement favorable to itself (Zhan et al., 2024).

Effective dialogue strategies are typically latent within expert dialogue transcripts of non-collaborative scenarios. These transcripts implicitly encode both strategy actions (what to say, e.g., *emotional appeals* or *foot-in-the-door*) and planning logic (the timing of these actions, e.g., *when to employ an emotional appeal*). Consequently, extracting these strategies has always been a labor- and knowledge-intensive task (Zhou et al., 2019a; Fransen et al., 2015). In this case, to build a non-collaborative dialogue agent, the standard paradigm relies on domain experts to first manually analyze transcripts and codify a set of strategy actions (He et al., 2018; Wang et al., 2019; Bonaldi et al., 2022a; Zhang et al., 2023), followed by the strategy planner training using pre-collected or LLM-roleplayed transcripts (Zhang et al., 2024a; Deng et al., 2024; Yu et al., 2023). However, this process is inherently expensive and unscalable. With the emergence of Large Language Models (LLMs) capable of powerful induction<sup>1</sup> (Cai et al., 2025; de Souza et al., 2025), there is a compelling opportunity to shift such paradigm: **Can we leverage LLMs to induce both strategy actions and planning logic directly from raw transcripts?**

To this end, we propose METRO (Multi-dimensional sTRategy induction from dialOgue transcripts), a novel method that extracts and formalizes raw dialogue transcripts into a "Strategy Forest". Specifically, METRO operates in two phases: Induction and Application. During the Induction phase, METRO extracts strategy actions as nodes and organizes multi-turn action trajectories into trees, each conditioned on specific dialogue states. This hierarchical architecture captures planning logic across two temporal scales: immediate child nodes represent short-term tactical responses, while full branches encapsulate long-term strategic foresight. By synthesizing these perspectives,

<sup>†</sup>Corresponding authors: [huang\\_chen@nus.edu.sg](mailto:huang_chen@nus.edu.sg)  
[faguo@buaa.edu.cn](mailto:faguo@buaa.edu.cn)

<sup>1</sup>Synthesize broader patterns from granular examples.

METRO creates a multi-dimensional representation of the strategies embedded within expert transcripts. In the subsequent Application phase, METRO analyzes the current dialogue state and queries the Strategy Forest to retrieve the most relevant tree, which serves as a blueprint to guide the agent’s decision-making. As such, METRO enables cost-effective, scalable, and strategic performance by effectively leveraging historical expert knowledge.

We conduct experiments on two benchmark datasets to demonstrate the superiority of METRO. Specifically, METRO achieves significant performance gains, surpassing other induction methods by an average of 10.24% across both benchmarks. Our analysis attributes this success to the generation of diverse strategic actions and informative planning logic, which endow agents with a richer behavioral patterns and strategic foresight. Moreover, METRO exhibits robust scalability via cross-task generalization: strategies induced from external domains remain highly effective in the target setting. Finally, we validate the potential of LLM-synthesized transcripts as a viable alternative to expert transcripts. To sum up, we conclude our main contributions as follows.

- We highlight strategy induction from expert transcripts as an effective and scalable solution for non-collaborative dialogues.
- We pioneer METRO, which formalizes expert knowledge in transcripts into a Strategy Forest, representing strategy actions and planning logic.
- We conduct extensive experiments to validate the effectiveness of METRO and provide an in-depth analysis of the diversity, transferability, and expertise-sensitivity of the induced strategies.

## 2 Related Work

**LLM Agents for Non-collaborative Dialogues.** Effective dialogue strategies are the key of non-collaborative agents. Traditionally, these agents rely on expert-defined action sets derived from manual analysis of dialogue transcripts, such as action sets for persuasion (Wang et al., 2019), counter-speech (Bonaldi et al., 2022b), and negotiation (He et al., 2018). While these actions remain widely used (Wang et al., 2025b; Zhang et al., 2024a; Deng et al., 2023c; Yu et al., 2023), they represent a labor-intensive bottleneck. Furthermore, deploying these actions requires sophisticated planning. Because simple prompting often lacks fine-grained

control (Deng et al., 2024), recent works have introduced external planners like MCTS (Yu et al., 2023) or specialized plug-in models (Deng et al., 2024; Zhang et al., 2024a; Wang et al., 2025b; Du et al., 2025). Despite these advances, the overall pipeline remains resource-intensive and unscalable. While PRINCIPLES (Kim et al., 2025a) automates the extraction of planning logic (*when [situation], you should [action A], rather than [action B], because [reason]*), it fails to capture multi-turn planning logic essential for non-collaborative setting.

**Expert Knowledge Induction from Dialogues.** Inductive learning empowers LLMs to synthesize broader patterns from granular examples (Cai et al., 2025; de Souza et al., 2025; Huang et al., 2024a; Louie et al., 2024). This capability has catalyzed significant research into task-specific dialogue transcripts, which act as rich repositories of human experience and knowledge. In particular, researchers extract emergent user intents and slots (Liang et al., 2024b,a; Vukovic et al., 2024; Finch et al., 2025), and derive dialogue structures (Burdisso et al., 2024; Raedt et al., 2024; Choubey et al., 2025; Pryor et al., 2023). Furthermore, recent studies focus on identifying strategy actions (Ou et al., 2024; Wang et al., 2025a; He et al., 2025; Liu et al., 2025; Hu et al., 2025), together with the rationales of these actions governing the planning logic (Jiang et al., 2025; Xie et al., 2024). While Kim et al. (2025a) recently adapts these ideas to non-collaborative dialogues, existing methods treat strategies as independent units. Instead, we induce and structure dialogue strategies into a forest, capturing the complex dynamics.

## 3 METRO

As illustrated in Figure 1, METRO first conducts an offline induction of the Strategy Forest directly from raw dialogue transcripts (cf. Section 3.1), subsequently utilizing this structure to guide inference within non-collaborative dialogues (Section 3.2).

**Notations & Overview.** Existing methods rely on a fixed set of human-crafted strategic actions  $\mathcal{A}$ . Let  $d_t = (v_1, v_1^{\text{USF}}, \dots, v_t, v_t^{\text{USF}})$  denote the dialogue history at turn  $t$ , where  $v_i^{\text{USF}}$  is the  $i$ -th user utterance and  $v_i$  is the  $i$ -th agent utterance corresponding to a chosen action  $a_i \in \mathcal{A}$ . Typically, a trained planner takes  $d_t$  as input to predict the next action  $a_{t+1} \in \mathcal{A}$ . In contrast, METRO directly induces dialogue strategies from a corpus of historical transcripts  $\mathcal{D}$ . Each transcript  $D \in \mathcal{D}$  is de-

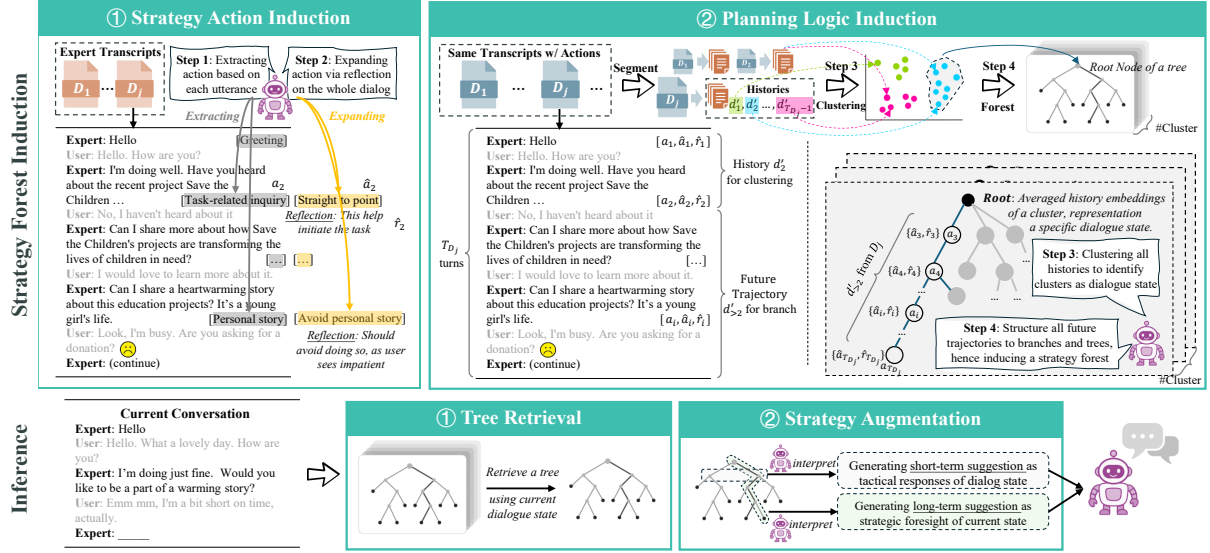


Figure 1: Overview of METRO. It formalizes expert knowledge within transcripts into a Strategy Forest, representing strategy actions and planning logic in a unified structure.

defined as a sequence  $D = (u_1, u_1^{\text{USR}}, \dots, u_{T_D}, u_{T_D}^{\text{USR}})$ , where  $u_i$  is  $i$ -th utterance from expert and  $T_D$  represents the total number of turns in transcript  $D$ . METRO transforms  $\mathcal{D}$  into a Strategy Forest  $\mathcal{F}$  and induces an expanded action set  $\mathcal{A}'$  (where  $\mathcal{A} \subseteq \mathcal{A}'$ ). Within this forest, the root node of the  $i$ -th tree  $f_i \in \mathcal{F}$  represents a specific dialogue state  $S_i$ , while all subsequent nodes encode effective strategic actions  $a' \in \mathcal{A}'$  from that state. Structurally, the  $j$ -th branch extending from the root to a leaf node represents an action trajectory  $\text{traj}_{ij} \in f_i$ , which encodes long-term strategic value. Conversely, the immediate children of the root node, denoted as  $\text{childro}_i$ , capture short-term value, indicating the effective next step to take from the root state  $S_i$ . During inference, given the Strategy Forest  $\mathcal{F}$  and the current dialogue history  $d_t$ , METRO first retrieves the most relevant tree  $f \in \mathcal{F}$ , identifying the root node whose dialogue state  $S$  bears the highest semantic similarity to  $d_t$ . Subsequently, METRO aggregates the top- $K$   $\text{childro}_i$  and  $\text{traj}_{ij}$  within  $f$  to inform a next-step decision.

### 3.1 Strategy Forest Induction

#### 3.1.1 Strategy Action Induction

Standard knowledge induction relies on simply extracting target information from raw transcripts, such as query intents (Gung et al., 2023; Liang et al., 2024a) and dialogue flows (Burdisso et al., 2024; Raedt et al., 2024; Choubey et al., 2025). However, this method is limited by the quality and diversity of the raw transcripts. To address this, we

propose augmenting the extraction process with the internal knowledge of LLM agents. Specifically, by using strategy actions derived from raw transcripts as bases, we prompt the LLM agent to generate refined new strategic actions, effectively transcending the limitations of the original transcripts.

**Action Extraction.** METRO starts with the identification of the strategic action  $a_i$  for every expert utterance  $u_i$  in transcript  $D$ . To streamline this, METRO supports two modes of operation. If an expert-curated action taxonomy exists for the target non-collaborative task<sup>2</sup>, METRO simply assigns the most appropriate action from the existing set. In the absence of such a taxonomy, METRO leverages its LLM backbone to autonomously induce a domain-specific action set based on the task description, following (Hu et al., 2025).

$$a_i = \text{Extraction}(u_i), \forall u_i \in D, D \in \mathcal{D}. \quad (1)$$

**Action Expansion.** Using its LLM backbone, METRO expands the strategy actions via self-reflection. For each turn  $i$  of transcript  $D$ , denote the dialogue history  $d'_i = (u_1, u_1^{\text{USR}}, \dots, u_i, u_i^{\text{USR}})$ . Taking the strategic action  $a_i$  and history  $d'_i$  as input, METRO assesses the quality of the expert utterance  $u_i$  by analyzing the user’s reaction  $u_i^{\text{USR}}$  (determined by LLM-as-judge). For effective utterance  $u_i$ , METRO summarizes a new action  $\hat{a}_i^0$  without access  $a_i$  to enhance generalization (e.g., *elicit empathy to deepen engagement*). For ineffective one,

<sup>2</sup>There are many human-crafted action taxonomies (Wang et al., 2019; Bonaldi et al., 2022b; He et al., 2018)

the original action  $a_i$  is flagged as a ‘*should-not-do*’ action  $\hat{a}_i^1$  for the given context, such as ‘*When [situation], you must avoid [a\_i]*’. This guides the agent away from poor choices.

$$(\hat{a}_i, \hat{r}_i) = \text{Expansion}(d'_i, a_i). \quad (2)$$

Finally, each  $u_i$  is augmented with two actions: the original  $a_i$  and the LLM-expanded  $\hat{a}_i \in \{\hat{a}_i^0, \hat{a}_i^1\}$ , accompanied by a rationale  $\hat{r}_i$  for  $\hat{a}_i$ .

### 3.1.2 Planning Logic Induction

A full action trajectory is encoded within a dialogue transcript. It is defined as the sequence of expert strategy actions, inferred from the expert’s utterances and labeled with the final task outcome (success or failure). METRO translates these trajectories using a state-centric tree structure, where trees with different dialogue states form the forest. To achieve this, we decompose each trajectory into local sub-trajectories, where each sub-trajectory represents an action path initiating from a specific dialogue state and carrying a final task outcome. Finally, we aggregate all sub-trajectories rooted in the identical dialogue state (regardless of their source transcript) into a unified tree. This synthesizes diverse historical planning logic within transcripts to enhance generalization.

**Dialogue State Identification & Clustering.** We process each transcript  $D$  (representing a full trajectory of length  $T_D$ ) by segmenting it into several past histories  $\{d'_1, \dots, d'_{T_D-1}\}$  and their corresponding future trajectories  $\{d'_{>1}, \dots, d'_{>T_D-1}\}$ , where  $d'_i$  denotes the dialogue history up to turn  $i$  (inclusive) and  $d'_{>i}$  represents all subsequent turns. We posit that these textual histories  $d'_i$  provide a denser semantic basis for state modeling than strategic actions alone. Therefore, we convert each history  $d'_i$  into a semantic embedding<sup>3</sup>  $\text{Emb}(d'_i)$  to represent the dialogue state preceding utterance  $u_{i+1}$ . We then apply K-Means to cluster all embeddings sourced from different transcripts. As a result, each expert utterance  $u_i$  ( $i \neq 1$ ) within transcript  $D$  is associated with its preceding state embedding  $\text{Emb}(d'_{i-1})$ , a cluster index, the future trajectory  $d'_{>i-1}$ , and its action metadata  $\{a_i, \hat{a}_i, \hat{r}_i\}$ .

**Strategy Forest Induction.** METRO induces a Strategy Forest by constructing a state-centric tree from each cluster. In this forest, the root node of each tree represents a distinct dialogue state, and its branches encode effective planning logic

from that state. To achieve this, given a cluster, we first define the cluster centroid as the root node, representing the dialogue state in the embedding space. The branches are then formed using all future trajectories within that cluster. Since not all branches are equally valuable, we prune low-value branches and retain only the top-K branches with the highest value. This process is detailed below.

- **Branch Generation.** We construct branches for a tree by organizing all future trajectories from the same cluster. Instead of using  $\hat{a}_i$ , nodes are represented by the original data-driven actions  $a_i$  to ensure authenticity, as they form trajectories that are grounded in real user interactions. Identical action sequences are merged to form consolidated branches (e.g., *Donation information*  $\rightarrow$  *Foot in the door*  $\rightarrow$  *Emotion appeal*).
- **Node Value Estimation.** We assign a value to every node along a branch. This value is primarily determined by the final task outcome<sup>4</sup> of the dialogues that passed through it, a process analogous to value backpropagation in MCTS (Silver et al., 2016) (cf. Appendix A.2.3).
- **Branch Value Estimation & Pruning.** Besides branches leading to failures, a branch that appears successful but has been observed only a few times may also be unreliable. To mitigate the risk of overestimating sparsely observed but successful trajectories, our branch valuation incorporates both performance and statistical confidence. Each node is augmented by a confidence term. This is derived from the Wilson Score Lower Bound (Wilson, 1927) using the empirical success and visitation counts of each node (cf. Appendix A.2.3). The final value of a branch is then defined as the weighted sum of its constituent node values and their corresponding confidence scores. Subsequently, we perform pruning by retaining the top-K branches with the highest value. This is further optimized in practice using Beam Search, accelerating the pruning process.

Ultimately, each tree inherently captures planning logic at two distinct temporal resolutions:

- **Breadth Planning Logic.** Immediate child nodes, containing both data-driven and LLM-expanded actions, of the root represent short-term tactical responses conditioned on the specific state.

<sup>3</sup>We use *bge-large-en-v1.5* (Xiao et al., 2023).

<sup>4</sup>Such as success rate and Sale-to-List Ratio (SL%).

Methods	P4G		CB			P4G*		CB*		
	SR $\uparrow$	AT $\downarrow$	SR $\uparrow$	AT $\downarrow$	SL% $\uparrow$	SR $\uparrow$	AT $\downarrow$	SR $\uparrow$	AT $\downarrow$	SL% $\uparrow$
Standard (Deng et al., 2023d)	0.620	4.56	0.185	5.89	0.154	0.333	5.20	0.283	5.65	0.285
Proactive (Deng et al., 2023d)	0.660	<u>4.11</u>	0.170	6.03	0.116	0.400	<b>4.67</b>	0.283	5.41	0.198
ProCoT (Deng et al., 2023d)	0.680	4.29	0.315	6.46	0.186	0.533	5.13	0.333	6.10	0.282
ICL-AIF (Fu et al., 2023)	0.615	<b>3.89</b>	0.325	6.11	<b>0.231</b>	0.583	5.09	0.450	6.00	0.348
GDP-Zero (Yu et al., 2023)	0.660	5.35	<u>0.495</u>	6.16	0.125	0.600	5.11	0.450	6.37	0.377
PPDPP (Deng et al., 2024)	0.730	4.67	0.250	<u>5.62</u>	0.150	<u>0.633</u>	6.13	0.383	<b>4.30</b>	0.215
PRINCIPLES (Kim et al., 2025b)	<u>0.770</u>	5.24	0.485	5.73	0.149	0.600	<u>4.78</u>	<u>0.467</u>	5.86	<u>0.389</u>
METRO( <i>ours</i> )	<b>0.780</b>	4.76	<b>0.575</b>	<b>5.38</b>	<u>0.189</u>	<b>0.661</b>	5.10	<b>0.483</b>	<u>5.21</u>	<b>0.411</b>

Table 1: Evaluation results for baselines and METRO across two datasets, incorporating both user simulation (P4G & CB) and human interaction studies (P4G\* & CB\*). Appendix C details performance on diverse users.

- **Depth Planning Logic.** Complete branches extending from the root encapsulate effective long-term strategic foresight for that state.

### 3.2 Inference with Strategy Forest

Since the induced strategy forest contains rich strategy actions and planning logic, METRO omit the model training and directly performs retrieval augmentation during the inference to enhance its ability in non-collaborative dialogues.

**Tree Retrieval.** Before generating the response for turn  $t$ , METRO encodes the current dialogue history  $d_{t-1}$  into a semantic embedding  $\text{Emb}(d_{t-1})$ . This embedding serves as a query to traverse the Strategy Forest: we compute the cosine similarity between  $\text{Emb}(d_{t-1})$  and the root node of every tree, retrieving the single tree that exhibits the highest semantic similarity to the current state.

**Strategy Augmentation.** We prompt METRO by integrating both breadth and depth planning logic derived from the retrieved tree. Inspired by Kim et al. (2025a), we recognize that retrieved logic may not perfectly align with the nuances of the current dialogue. Therefore, we task the LLM backbone with generating contextual strategy as suggestion.

- **Short-Term Suggestion.** the LLM reinterprets the actions from breadth logic (e.g., ‘*provide specific and concrete information*’) to fit the current dialogue context, summarizing them into a concise next-step strategy (e.g., ‘*explaining the charity’s concrete programs and impact to address the persuadee’s curiosity*’). For simplicity, we exclusively utilize the LLM-expanded actions for this step, as they represent optimized abstractions of the original data-driven actions.
- **Long-Term Suggestion.** We identify the single branch with the highest average node value to

serve as the depth logic (e.g., ‘*build credibility  $\rightarrow$  propose a donation*’). Similarly, the LLM reinterprets this trajectory, summarizing it into a high-level planning directive that emphasizes cumulative temporal effects (e.g., ‘*progressively strengthening trust and commitment prior to making a donation request*’).

Finally, these two suggestions are concatenated to guide METRO in generating the final response. Notably, METRO diverges from MCTS, as it leverages multiple trees to capture multi-dimensional strategies (breadth and depth) without requiring expensive test-time computation.

## 4 Experiments

### 4.1 Experimental Setup

Following common practice in the field (Zhang et al., 2024a; Deng et al., 2024; Hu et al., 2025), we evaluate agents by pairing them with synthetic user simulators exhibiting diverse non-collaborative behaviors. In each interaction, both the agent and simulator alternate turns, employing strategies to maximize their respective self-interests. The dialogue proceeds until the task goal is met or the maximum number of turns (i.e.,8) is reached, after which we assess the agent’s performance based on the collected interactions. Finally, implementation and evaluation details are presented in Appendix A and B, respectively. Additional experimental analyses is in Appendix C.

**Evaluation Tasks.** In line with Deng et al. (2023d); Wang et al. (2019); Hu et al. (2025); Zhang et al. (2024a), we employ two benchmark tasks: the price negotiation task, utilizing the test dataset of CraigslistBargain (CB) (He et al., 2018), the charity persuasion task, employing the test dataset of

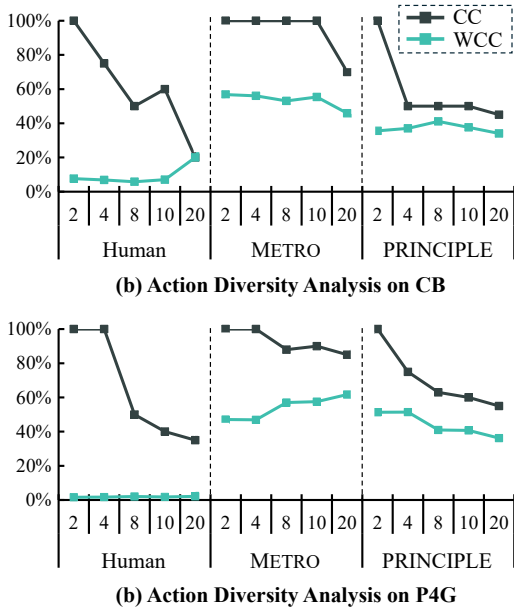


Figure 2: Action diversity analysis.

PersuasionForGood (P4G) (Wang et al., 2019). In these scenarios, the conversational agents assume the roles of buyer and persuader, respectively, aiming to achieve their task-specific goals.

**Baselines.** We compare with two types of baselines. 1) Prompt-based, including *Standard* (Deng et al., 2023d), *Proactive* (Deng et al., 2023d), *Pro-CoT* (Deng et al., 2023d) and *ICL-AIF* (Fu et al., 2023), which use vanilla prompts, mixed-initiative prompts, CoT, and AI feedback to select next strategies, respectively. 2) External strategy planners, including *GDP-Zero* (Yu et al., 2023) and *PPDPP* (Deng et al., 2024), which utilize a MCTS and a trainable plug-in for determining next-step strategies, respectively. 3) Strategy induction, *PRINCIPLES* (Kim et al., 2025b), which utilize past conversation histories to extract dialogues strategies.

**User Simulators & Human Participants.** Consistent with prior research (Deng et al., 2023c; Bianchi et al., 2024; Huang et al., 2024b; Zhang et al., 2024a), we utilize diverse LLM-based user simulators, which demonstrate non-collaborative behaviors driven by specific personas. Following (Wang et al., 2019; Jiang et al., 2024; Zhang et al., 2024a), we incorporate two distinct persona dimensions: the *Big-Five Personality* (Goldberg, 1992) and the *Decision-Making Styles* (Scott and Bruce, 1995), augmenting each combination with an LLM-generated description. This yields a total of 200 diverse simulators per task, distributed across the 20 distinct categories (i.e., Big-Five  $\times$  Styles). Furthermore, we employ 5 humans with varying per-

sonas and conduct a human study in which participants interact with the agents to assess their practical effectiveness, following existing setup (Deng et al., 2023d; Zhang et al., 2024b; Chen et al., 2024). See Appendix B.3.1 for human study details.

**Strategy Forest Induction.** We utilize two expert human-human dialogue transcripts. Specifically, for CB task, we utilize the *CRAIGSLISTNEGOTIATION* dataset, containing 6,682 dialogues (He et al., 2018). As for P4G task, we utilize *PERSUASIONFORGOOD* dataset, with 1,017 dialogues (Wang et al., 2019). Using these transcripts as input, METRO extracts strategy forest for subsequent experimentation. Unless otherwise specified, our evaluation employs strategies induced from the task-specific dataset. Crucially, as our evaluation involves real-time online interactions with users, there is no risk of data leakage from the static induction transcripts.

**Evaluation Metrics.** Following Deng et al. (2024); Hu et al. (2025); Zhang et al. (2024a), we consider three commonly used automated metrics: Success Rate (SR), Average Turn (AT) and Sale-to-List Ratio (SL%). Notably, SL% (Zhou et al., 2019b) is design specifically for the CB task to determine the effectiveness of goal completion. See Appendix B.4 for details.

## 4.2 Main Results

Table 1 validates effectiveness of METRO and the empirical quality of its induced strategies.

**METRO achieves promising performance and outperforms other induction methods across diverse benchmarks and metrics.** As shown in Table 1 (left), METRO demonstrates promising performance on the SR metric across all evaluated non-collaborative tasks, with an average performance gain of 9.93% over the second-best result. This advantage is particularly pronounced over methods with supervised strategy modules (PPDPP) or MCTS (GDP-Zero). These results fulfill our initial motivation to develop an effective and scalable solution. Moreover, in a direct comparison with other strategy induction method (PRINCIPLES), METRO demonstrates an average improvement of 10.24% across all metrics and datasets. A notable trade-off, however, is the increased dialogue length (AT) characteristic of strategy induction methods (PRINCIPLES and METRO). This highlights a potential direction for future optimization: refining the induced strategies to achieve desired outcomes with greater conversational efficiency.

**METRO demonstrates superior practical utility in human evaluations.** We validated our findings with a human user study to measure real-world efficacy. The results, presented in Table 1 (right), are consistent with our simulator-based evaluations (left), with METRO again demonstrating the promising performance, outperforming PRINCIPLES.

### 4.3 Why is METRO Effective

We analyze its induced strategy actions and planning logic, supplemented by ablation studies.

#### 4.3.1 Strategy Action Analysis

**Setup.** We begin by characterizing the distinct properties of strategies derived from human experts, PRINCIPLES, and METRO through a diversity analysis. For each task, we first pool all human-crafted strategies pre-defined in the dataset with 100 randomly sampled actions each from PRINCIPLES and METRO. We then apply K-Means clustering (using *bge-large-en-v1.5*) and analyze the distribution of each method’s actions across these clusters. We introduce two metrics: 1) Cluster Coverage (CC), defined as the percentage of clusters containing at least one action from the method; and 2) Within-Cluster Coverage (WCC), defined as the average proportion and standard deviation of actions from the same method per cluster. These metrics provide a coarse- and fine-grained measure of action diversity, respectively.

**METRO induces more diverse and nuanced actions.** Figure 2 shows results across varying cluster numbers  $K_1$  (x-axis). 1) When  $K_1=2$ , all methods produce fundamentally similar high-level strategy actions (CC=100%). However, as the clusters become more fine-grained (increasing  $K_1$ ), a clear divergence emerges. While the CC for all methods naturally declines, METRO consistently maintains a higher coverage, indicating that it induces a more diverse and comprehensive set of actions that span a broader range of nuanced strategic categories. 2) The low WCC for human-crafted actions is an expected results due to their limited quantity. However, we do note a slight increase in their WCC on the CB task. This occurs alongside a very low CC, suggesting that the human strategies, while semantically consistent, are concentrated within a narrow portion of the overall action space (lack of overall diversity). For both PRINCIPLES and METRO, WCC generally decreases with  $K_1$ , indicating their actions are distributed broadly. However, for METRO on the P4G dataset, its WCC

Method	P4G		CB		
	SR↑	AT↓	SR↑	AT↓	SL%↑
<i>Breadth Planning Logic</i>					
Top-1 Node	0.760	4.783	0.465	5.774	0.140
Top-3 Nodes	0.770	4.852	0.525	5.598	0.157
Top-5 Nodes	<b>0.780</b>	<b>4.756</b>	<b>0.575</b>	<b>5.383</b>	<b>0.189</b>
<i>Depth Planning Logic</i>					
Top-3 Branches	0.760	4.803	0.485	5.557	0.150
Top-2 Branches	0.725	4.841	0.580	5.716	0.183
Top-1 Branch	<b>0.780</b>	<b>4.756</b>	<b>0.575</b>	<b>5.383</b>	<b>0.189</b>
1-hop Branch	0.770	<b>4.636</b>	0.535	5.589	0.150
3-hop Branch	0.775	4.800	0.550	5.491	0.167
>3-hop Branch	<b>0.780</b>	4.756	<b>0.575</b>	<b>5.383</b>	<b>0.189</b>

Table 2: Planning logic analysis.

increases. This indicates that METRO not only generates diverse actions (relatively large CC) but also creates concentrated groups of semantically similar yet subtly distinct actions within specific fine-grained clusters. To sum up, these diversities forms the foundation for our superiority.

#### 4.3.2 Planning Logic Analysis

**Setup.** We further analyze how breadth and depth planning logic affect METRO. For breadth, we tune the number of expended actions used for short-term suggestion generation, selecting top-k actions based on semantic similarity to current context. For depth, we leverage the multi-hop nature of branch by generating long-term suggestions from k-hop sub-branches truncated from the root. We also investigate the number of retrieval branches.

**Strategy forest equips METRO with enhanced strategic behavioral diversity and foresight.** While increasing the number of semantically less relevant Top-K branches provides diminishing returns, Table 2 confirms that expanding the breadth (Top-k Nodes) and depth (K-hop Branch) of the planning logic is generally advantageous. This finding validates the core motivation behind our strategy forest: a structured representation of strategy actions and planning logic empowers the agent with superior strategic behavioral diversity and foresight. Thus, our default configuration employs all immediate children of the root and the top-1 full branch.

**Diversity analysis on tree branches.** To create a representative sample, we embed all tree branches from a given task, apply K-Means clustering ( $K=150$ ), and then randomly sample three branches from each cluster, yielding a corpus of 450 branches. We use two metrics: 1) Unique Ratio

Methods	CB to P4G		P4G to CB			ALL to P4G		ALL to CB		
	SR $\uparrow$	AT $\downarrow$	SR $\uparrow$	AT $\downarrow$	SL% $\uparrow$	SR $\uparrow$	AT $\downarrow$	SR $\uparrow$	AT $\downarrow$	SL% $\uparrow$
Proactive (Deng et al., 2023d)	0.710	4.59	0.145	<b>5.59</b>	0.138	0.670	4.36	0.205	<b>5.27</b>	0.123
ProCoT (Deng et al., 2023d)	<b>0.755</b>	<b>4.22</b>	0.310	6.24	<b>0.288</b>	0.655	<b>3.98</b>	0.195	6.44	0.138
GDP-Zero (Yu et al., 2023)	0.695	7.22	0.305	6.31	<u>0.254</u>	—	—	—	—	—
PPDPP (Deng et al., 2024)	0.645	4.77	0.225	6.51	0.170	—	—	—	—	—
PRINCIPLES (Kim et al., 2025b)	0.670	5.06	<u>0.355</u>	<u>5.66</u>	0.170	<u>0.755</u>	5.21	<u>0.325</u>	5.48	<u>0.140</u>
METRO	<b>0.755</b>	<u>4.30</u>	<b>0.440</b>	5.95	0.187	<b>0.770</b>	4.80	<b>0.540</b>	5.50	<b>0.204</b>

Table 3: Cross-task Transferability Evaluation. *CB to P4G*: Transferring CB-tailored strategies to P4G. *ALL to P4G*: Transferring the union of CB and P4G strategies to P4G. Baselines without explicit strategies are omitted. ‘-’ denotes cases omitted due to the high computational costs associated with larger strategy action sets.

Strategies	Original Branches		Redundancy Removal	
	URatio	Edit Distance	URatio	Edit Distance
CB (Induced)	0.86	0.332	0.63	0.349
P4G (Induced)	1.00	0.028	0.98	0.029
CB (Origin)	0.85	0.322	0.74	0.346
P4G (Origin)	1.00	0.081	1.00	0.085

Table 4: Diversity analysis of tree branches. *Induce* refers to branches with our induced actions, while *Origin* uses original actions contained in the transcripts.

(URatio), the proportion of unique branches, and 2) Average Edit Distance, which quantifies structural similarity. Furthermore, to investigate structural redundancy, we create a variant of the corpus where consecutive repeated actions are removed, a process we term Redundancy Removal. As shown in Table 4, the strategies induced from the CB dataset are significantly less diverse than those from P4G. Analysis on Redundancy Removal setting further reveals that CB branches exhibit high structural repetition. The results of ‘*Origin*’ confirm that this is not an artifact of our induction process but rather *an intrinsic limitation of the source data itself: the action sequences encoded in the original CB transcripts are not inherently diverse*. This highlights the value of transcript quality (cf. Appendix C.3). However, as evidenced by Table 1, this data limitation does not diminish our overall superiority.

### 4.3.3 Ablation Study

We consider the following variants. 1) w/o Exp. Action excludes the LLM-expanded actions, using only the original actions  $a_i$  for short-term suggestion generation, 2) w/o Depth relies only on short-term suggestions, 3) w/o Breadth uses only long-term suggestions. According to Figure 3, while AT remains consistent, SR and SL% sometimes can be sensitive to the Depth component removal. On the P4G task, the ablation of any module leads to a pre-

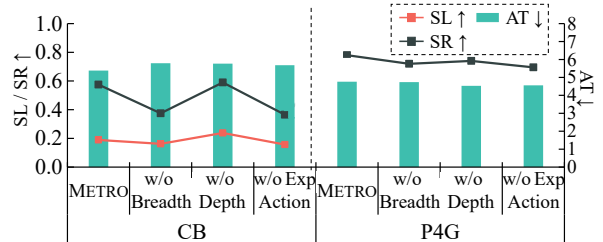


Figure 3: Ablation analysis. *w/o Depth* improves performance on the CB task due to the low-diversity branches induced from the CRAIGSLISTNEGOTIATION dataset.

dictable decline in performance. However, a significant exception is observed on the CB task, where the *w/o Depth* configuration yields a performance enhancement. This is attributable to the inherent structural redundancy of the strategies embedded within the CB task corpus (Section 4.3.2). This lack of diversity results in biased long-term suggestions, the removal of which is beneficial. Therefore, the efficacy of the induced strategies is contingent on the quality of the source transcripts. As elaborated in Appendix C.3, leveraging higher-quality data can further amplify our performance.

### 4.3.4 Cross-Task Strategy Transferability

**METRO-induced strategies enjoy strong cross-task transferability.** As shown in Table 3, the LLM’s powerful contextualization capabilities allow most methods to remain effective even when using strategies from a different task. Our analysis reveals several key insights: 1) CB strategies transfer well to P4G because they are designed for a more assertive task, but the reverse is not true, which yields weaker results; 2) The performance of PPDPP, which relies on a specialized planner, did not improve with transferred strategies after we exchanged two planners. This underscores that when the planner fails to select the right action, even a

LLM, which can contextualize selected strategy, cannot compensate for this flaw. 3) METRO often outperforms baselines, especially PRINCIPLES, in transfer tasks, highlighting its strategies of greater diversity and foresight. 4) We simulate strategy evolution by expanding the action space (e.g., CB to P4G → ALL to P4G). Most baselines suffer performance degradation, as simple LLM-based planners are prone to selection bias with a larger candidate set, while MCTS-based methods like GDP-zero incur prohibitive computational costs. In contrast, METRO remains effective. We attribute this to the precision of our tree retrieval mechanism, which can navigate an expanded strategy forest.

## 5 Conclusion

We pioneer an effective and scalable solution for non-collaborative dialogue by leveraging LLMs to induce strategy actions and planning logic directly from expert transcripts. We introduce METRO to formalize these induced strategies into a structured strategy forest. Extensive experiments demonstrate METRO’s promising performance and the reason behind its success. More importantly, our work provides a investigation into tacit knowledge extraction from expert dialogues. We analyze a potential structure for representing this knowledge and explore key characteristics such as diversity and transferability. This offers insights into dialogue agent design and tacit knowledge extraction.

## Limitations

**Fidelity of Expert Knowledge.** A primary limitation of this work concerns the degree of domain expertise reflected in the source transcripts. While we utilize the CRAIGSLISTNEGOTIATION and PERSUASIONFORGOOD benchmark datasets, both of which consist of human-human dialogues, the participants are likely general users rather than professional practitioners in negotiation or persuasion. Due to the scarcity of high-fidelity transcripts produced by certified experts, it remains challenging to verify the performance of METRO on professional-grade expertise. Nevertheless, our empirical analysis demonstrates that inducing strategies even from general human interactions yields significant performance gains over existing baselines. We defer the evaluation of METRO on specialized, high-expertise corpora to future research as such data becomes available.

**In-depth Analysis on LLM-generated Tran-**

**scripts.** We emphasize the value of mining domain knowledge from expert transcripts, as these records capture the tacit knowledge inherent in human interaction (Huang et al., 2024c). Unlike explicit knowledge, this tacit knowledge is deeply embedded in individual experience and intuition (Ryle, 1945; Spitzer et al., 2022), which is difficult to convey to others through writing or verbalization (Nonaka, 1994; Lam, 2000; Gorman, 2002), and is only evident in human behaviors (Liu et al., 2022). Our method is specifically designed to formalize and utilize such tacit knowledge from unstructured behavioral data like transcripts, with special focus on the task of non-collaborative dialogues. Therefore, the primary motivation and scope of this study is human expert behavior; we treat LLM-generated transcripts as a secondary point of comparison rather than a core research object. Although our initial findings in Appendix C.3 indicate that LLM-generated transcripts can be effective, and in some cases, surpass ‘expert’ transcripts, we leave the in-depth characterization of LLM-generated transcripts for future exploration.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. U25B201508, No. 62272330, and No.U24A20328).

## References

- Federico Bianchi, Patrick John Chia, Mert Yuksekgonul, Jacopo Tagliabue, Dan Jurafsky, and James Zou. 2024. How well can llms negotiate? negotiationarena platform and analysis. *arXiv preprint arXiv:2402.05863*.
- Helena Bonaldi, Sara Dellantonio, Serra Sinem Tekiroğlu, and Marco Guerini. 2022a. [Human-machine collaboration approaches to build a dialogue dataset for hate speech countering](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8031–8049, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Helena Bonaldi, Sara Dellantonio, Serra Sinem Tekiroğlu, and Marco Guerini. 2022b. [Human-machine collaboration approaches to build a dialogue dataset for hate speech countering](#). In *Proceedings of the 2022 conference on empirical methods in natural language processing*, pages 8031–8049.
- Sergio Burdisso, Srikanth Madikeri, and Petr Motlicek. 2024. [Dialog2Flow: Pre-training soft-contrastive action-driven sentence embeddings for automatic dialog flow extraction](#). In *Proceedings of the 2024*

- Conference on Empirical Methods in Natural Language Processing*, pages 5421–5440, Miami, Florida, USA. Association for Computational Linguistics.
- Chengkun Cai, Xu Zhao, Haoliang Liu, Zhongyu Jiang, Tianfang Zhang, Zongkai Wu, Jenq-Neng Hwang, and Lei Li. 2025. The role of deductive and inductive reasoning in large language models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16780–16790.
- Yue Chen, Chen Huang, Yang Deng, Wenqiang Lei, Dingnan Jin, Jia Liu, and Tat-Seng Chua. 2024. [STYLE: Improving domain transferability of asking clarification questions in large language model powered conversational agents](#). In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 10633–10649, Bangkok, Thailand. Association for Computational Linguistics.
- Prafulla Kumar Choubey, Xiangyu Peng, Shilpa Bhagavath, Caiming Xiong, Shiva Kumar Pentyala, and Chien-Sheng Wu. 2025. Turning conversations into workflows: A framework to extract and evaluate dialog workflows for service ai agents. *arXiv preprint arXiv:2502.17321*.
- João Pedro Gandarela de Souza, Danilo Carvalho, and André Freitas. 2025. Inductive learning of logical theories with llms: A expressivity-graded analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 23752–23759.
- Yang Deng, Wenqiang Lei, Minlie Huang, and Tat-Seng Chua. 2023a. [Goal awareness for conversational AI: Proactivity, non-collaborativity, and beyond](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 6: Tutorial Abstracts)*, pages 1–10, Toronto, Canada. Association for Computational Linguistics.
- Yang Deng, Wenqiang Lei, Minlie Huang, and Tat-Seng Chua. 2023b. [Rethinking conversational agents in the era of llms: Proactivity, non-collaborativity, and beyond](#). In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region, SIGIR-AP '23*, page 298–301, New York, NY, USA. Association for Computing Machinery.
- Yang Deng, Wenqiang Lei, Wai Lam, and Tat-Seng Chua. 2023c. 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. 2023d. [Prompting and evaluating large language models for proactive dialogues: Clarification, target-guided, and non-collaboration](#).
- Yang Deng, Wenxuan Zhang, Wai Lam, See-Kiong Ng, and Tat-Seng Chua. 2024. Plug-and-play policy planner for large language model powered dialogue agents. *The Twelfth International Conference on Learning Representations (ICLR)*.
- Hanwen Du, Bo Peng, and Xia Ning. 2025. [Planning with diffusion models for target-oriented dialogue systems](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 20307–20329, Vienna, Austria. Association for Computational Linguistics.
- Ritam Dutt, Sayan Sinha, Rishabh Joshi, Surya Shekhar Chakraborty, Meredith Riggs, Xinru Yan, Hao-gang Bao, and Carolyn Penstein Rosé. 2021. Responder: Computationally modelling resisting strategies in persuasive conversations. *arXiv preprint arXiv:2101.10545*.
- James D Finch, Yasasvi Josyula, and Jinho D Choi. 2025. Generative induction of dialogue task schemas with streaming refinement and simulated interactions. *arXiv preprint arXiv:2504.18474*.
- Marieke L Fransen, Edith G Smit, and Peeter WJ Verlegh. 2015. Strategies and motives for resistance to persuasion: An integrative framework. *Frontiers in psychology*, 6:1201.
- Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. 2023. [Improving language model negotiation with self-play and in-context learning from ai feedback](#). *Preprint*, arXiv:2305.10142.
- Lewis R Goldberg. 1992. The development of markers for the big-five factor structure. *Psychological assessment*, 4(1):26.
- Michael E Gorman. 2002. Types of knowledge and their roles in technology transfer. *The Journal of Technology Transfer*, 27(3):219–231.
- James Gung, Raphael Shu, Emily Moeng, Wesley Rose, Salvatore Romeo, Arshit Gupta, Yassine Benajiba, Saab Mansour, and Yi Zhang. 2023. [Intent induction from conversations for task-oriented dialogue track at DSTC 11](#). In *Proceedings of the Eleventh Dialog System Technology Challenge*, pages 242–259, Prague, Czech Republic. Association for Computational Linguistics.
- 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, Yiheng Sun, Zerui Chen, Ming Liu, and Bing Qin. 2025. Simulation-free hierarchical latent policy planning for proactive dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 24032–24040.
- Yikuan Hu, Chen Huang, and Wenqiang Lei. 2025. Astro: Automatic strategy optimization for non-cooperative dialogues. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 388–408.

- Chen Huang, Yiping Jin, Ilija Ilievski, Wenqiang Lei, and Jiancheng Lv. 2024a. Araida: Analogical reasoning-augmented interactive data annotation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10660–10675.
- Chen Huang, Peixin Qin, Yang Deng, Wenqiang Lei, Jiancheng Lv, and Tat-Seng Chua. 2024b. [Concept – an evaluation protocol on conversational recommender systems with system-centric and user-centric factors](#). *Preprint*, arXiv:2404.03304.
- Chen Huang, Xinwei Yang, Yang Deng, Wenqiang Lei, JianCheng Lv, and Tat-Seng Chua. 2024c. [Co-matching: Towards human-machine collaborative legal case matching](#). *Preprint*, arXiv:2405.10248.
- Guangyuan Jiang, Manjie Xu, Song-Chun Zhu, Wenjuan Han, Chi Zhang, and Yixin Zhu. 2024. Evaluating and inducing personality in pre-trained language models. *Advances in Neural Information Processing Systems*, 36.
- Xinyi Jiang, Tianyi Hu, Yuheng Qin, Guoming Wang, Zhou Huan, Kehan Chen, Gang Huang, Rongxing Lu, and Siliang Tang. 2025. Chatmap: Mining human thought processes for customer service chatbots via multi-agent collaboration. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 11927–11947.
- Namyoun Kim, Kai Tzu-iunn Ong, Yeonjun Hwang, Minseok Kang, Iiseo Jihn, Gayoung Kim, Minju Kim, and Jinyoung Yeo. 2025a. Principles: Synthetic strategy memory for proactive dialogue agents. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 21329–21368.
- Namyoun Kim, Kai Tzu-iunn Ong, Yeonjun Hwang, Minseok Kang, Iiseo Jihn, Gayoung Kim, Minju Kim, and Jinyoung Yeo. 2025b. [PRINCIPLES: Synthetic strategy memory for proactive dialogue agents](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 21329–21368, Suzhou, China. Association for Computational Linguistics.
- Alice Lam. 2000. Tacit knowledge, organizational learning and societal institutions: An integrated framework. *Organization studies*, 21(3):487–513.
- Wenqiang Lei, Yao Zhang, Feifan Song, Hongru Liang, Jiaxin Mao, Jiancheng Lv, Zhenglu Yang, and Tat-Seng Chua. 2022. [Interacting with non-cooperative user: A new paradigm for proactive dialogue policy](#). *Preprint*, arXiv:2204.07433.
- Jinggui Liang, Lizi Liao, Hao Fei, and Jing Jiang. 2024a. Synergizing large language models and pre-trained smaller models for conversational intent discovery. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 14133–14147.
- Jinggui Liang, Yuxia Wu, Yuan Fang, Hao Fei, and Lizi Liao. 2024b. [A survey of ontology expansion for conversational understanding](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 18111–18127, Miami, Florida, USA. Association for Computational Linguistics.
- Leqi Liu, Rastogi Charvi, Holstein Ken, and Heidari Hoda. 2022. A taxonomy characterizing human and ml predictive decision-making. *ICML Workshop on Human-Machine Collaboration and Teaming*.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Zhaocheng Liu, Quan Tu, Wen Ye, Yu Xiao, Zhishou Zhang, Hengfu Cui, Yalun Zhu, Qiang Ju, Shizheng Li, and Jian Xie. 2025. Exploring the inquiry-diagnosis relationship with advanced patient simulators. *arXiv preprint arXiv:2501.09484*.
- Ryan Louie, Ananjan Nandi, William Fang, Cheng Chang, Emma Brunskill, and Diyi Yang. 2024. [Roleplay-doh: Enabling domain-experts to create LLM-simulated patients via eliciting and adhering to principles](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 10570–10603, Miami, Florida, USA. Association for Computational Linguistics.
- Ikujiro Nonaka. 1994. A dynamic theory of organizational knowledge creation. *Organization science*, 5(1):14–37.
- Jiao Ou, Jiayu Wu, Che Liu, Fuzheng Zhang, Di Zhang, and Kun Gai. 2024. Inductive-deductive strategy reuse for multi-turn instructional dialogues. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17402–17431.
- Connor Pryor, Quan Yuan, Jeremiah Liu, Mehran Kazemi, Deepak Ramachandran, Tania Bedrax-Weiss, and Lise Getoor. 2023. Using domain knowledge to guide dialog structure induction via neural probabilistic soft logic. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7631–7652.
- Maarten De Raedt, Frédéric Godin, Chris Develder, and Thomas Demeester. 2024. Revisiting clustering for efficient unsupervised dialogue structure induction. *Applied Intelligence*, 54(7):5278–5305.
- Gilbert Ryle. 1945. Knowing how and knowing that: The presidential address. In *Proceedings of the Aristotelian society*, volume 46, pages 1–16. JSTOR.
- Susanne G Scott and Reginald A Bruce. 1995. Decision-making style: The development and assessment of a new measure. *Educational and psychological measurement*, 55(5):818–831.

- Weiyang Shi, Kun Qian, Xuewei Wang, and Zhou Yu. 2019. How to build user simulators to train rl-based dialog systems. *arXiv preprint arXiv:1909.01388*.
- David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, and 1 others. 2016. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489.
- Philipp Spitzer, Niklas K uhl, and Marc Goutier. 2022. Training novices: The role of human-ai collaboration and knowledge transfer. In *Workshop on Human-Machine Collaboration and Teaming (HM-CaT 2022)*, 23rd July, Baltimore.
- Renato Vukovic, David Arps, Carel van Niekerk, Benjamin Matthias Ruppik, Hsien-Chin Lin, Michael Heck, and Milica Gasic. 2024. Dialogue ontology relation extraction via constrained chain-of-thought decoding. In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 370–384.
- Borui Wang, Kathleen McKeown, and Rex Ying. 2025a. Dystil: Dynamic strategy induction with large language models for reinforcement learning. *arXiv preprint arXiv:2505.03209*.
- Haiyang Wang, Zhiliang Tian, Yuchen Pan, Xin Song, Xin Niu, Minlie Huang, and Bin Zhou. 2025b. **Battling against tough resister: Strategy planning with adversarial game for non-collaborative dialogues**. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3665–3685, Vienna, Austria. Association for Computational Linguistics.
- Xuewei Wang, Weiyang 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.
- Edwin B Wilson. 1927. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212.
- Shitao Xiao, Zheng Liu, Peitian Zhang, and Niklas Muennighoff. 2023. **C-pack: Packaged resources to advance general chinese embedding**. *Preprint*, arXiv:2309.07597.
- Zhouhang Xie, Bodhisattwa Prasad Majumder, Mengjie Zhao, Yoshinori Maeda, Keiichi Yamada, Hiromi Wakaki, and Julian McAuley. 2024. Few-shot dialogue strategy learning for motivational interviewing via inductive reasoning. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 13207–13219.
- 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 Hafari. 2024. Let’s negotiate! a survey of negotiation dialogue systems. *arXiv preprint arXiv:2402.01097*.
- Tong Zhang, Chen Huang, Yang Deng, Hongru Liang, Jia Liu, Zujie Wen, Wenqiang Lei, and Tat-Seng Chua. 2024a. 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.
- Tong Zhang, Junhong Liu, Chen Huang, Jia Liu, Hongru Liang, Zujie Wen, and Wenqiang Lei. 2023. **Towards effective automatic debt collection with persona awareness**. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 32–45, Singapore. Association for Computational Linguistics.
- Tong Zhang, Peixin Qin, Yang Deng, Chen Huang, Wenqiang Lei, Junhong Liu, Dingnan Jin, Hongru Liang, and Tat-Seng Chua. 2024b. **CLAMBER: A benchmark of identifying and clarifying ambiguous information needs in large language models**. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10746–10766, Bangkok, Thailand. Association for Computational Linguistics.
- Yiheng Zhou, He He, Alan W Black, and Yulia Tsvetkov. 2019a. **A dynamic strategy coach for effective negotiation**. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 367–378, Stockholm, Sweden. Association for Computational Linguistics.
- Yiheng Zhou, Yulia Tsvetkov, Alan W Black, and Zhou Yu. 2019b. **Augmenting non-collaborative dialog systems with explicit semantic and strategic dialog history**. *Preprint*, arXiv:1909.13425.

## A Implementation Details

We conduct the main experiments on a CPU-based server with a dual-socket Intel Xeon Platinum 8352V CPU (64 physical cores, 128 threads), while the PPDPP experiments are reproduced on an NVIDIA RTX A100 GPU. To reduce the overall evaluation time, we adopt a mini-batch concurrent execution scheme, where evaluation samples are partitioned into small batches and processed in parallel across multiple processes. Dialogue state representations are obtained using the

bge-large-en-v1.5 embedding model, and dialogue states are clustered with K-Means implemented in the `scikit-learn` package. Most experiments rely on API-based model inference using the OpenAI and Aliyun platforms. All implementations are based on PyTorch. For all baselines, we adhere to the implementations described in their respective papers or official GitHub repositories. Note that the first two baseline categories utilize the manually curated strategy actions provided in their original works, whereas the PRINCIPLES generates strategies autonomously. Following Deng et al. (2024); Zhang et al. (2024a); Hu et al. (2025), we use GPT-3.5-turbo as the LLM backbones for conversational agents and LLM-based evaluator, respectively. To ensure a more rigorous and holistic evaluation, we also recruited human participants to interact with and assess each agent. Evaluation details are presented in Appendix B.

### A.1 Implementations of Baselines

We implement representative LLM-based dialogue agents as baselines for comparison, adhering to their original configurations. The specific implementation details are as follows:

**Standard:** We simply prompt the LLM to interact with users relying solely on the provided task instructions, without employing any external dialogue strategies.

**PPDPP:** Following (Deng et al., 2024), we implement the plug-and-play policy planner using RoBERTa-large (Liu et al., 2019), based on the original training configurations. To evaluate its performance on the P4G dataset, we adapted the task instructions and strategy descriptions to align with the persuasion scenarios. We strictly followed the original two-stage training framework (SFT and RL). For the final evaluation, we selected the model checkpoint that achieved the best performance during the testing phase.

**GDP-Zero:** Following (Yu et al., 2023), we implement GDP-Zero using Open-Loop MCTS. We strictly adhered to the experimental settings and hyperparameters recommended in the original paper (e.g., simulation count and exploration weight). Furthermore, we extended the evaluation to the CB task by adapting the task instructions and role-playing descriptions to the negotiation context.

**Proactive:** Following (Deng et al., 2023d), we implement the Proactive prompting scheme. This method requires the LLM to explicitly select a dialogue strategy from a pre-defined set before gen-

erating the response. We adapted the prompt templates to align with the specific action spaces of the P4G and CB datasets.

**ProCoT:** Following (Deng et al., 2023d), we implement the Proactive Chain-of-Thought (ProCoT) framework. It instructs the LLM to generate a reasoning chain regarding dialogue status and goal planning prior to response generation. We strictly adhered to the one-shot setting and adapted the reasoning demonstrations to fit the specific negotiation and persuasion contexts.

**ICL-AIF:** Following (Fu et al., 2023), we implement the ICL-AIF framework. It employs a critic model to provide natural language feedback, acting as in-context demonstrations to guide the LLM’s response generation. We adapted the critic’s prompts to align with the strategic goals of the P4G and CB datasets.

**PRINCIPLES:** Following (Kim et al., 2025b), we implement PRINCIPLES as a non-parametric strategy memory constructed via offline self-play simulations. We closely follow the original principle construction pipeline, structured principle extraction in the *When–Should–Rather than–Because* format. For faithful reproduction, we adopt the principle generation procedure described in the original paper and construct a total of 296 principles for the P4G task and 100 principles for the CB task.

### A.2 Implementations of METRO

We conclude the pseudocode of METRO in Algorithm 1 and 2.

#### A.2.1 Strategy Action Induction

We implement strategy action induction as a two-stage pipeline over agent turns in both P4G and CB dialogues (persuader turns in P4G and buyer turns in CB). For each target turn ( $\text{turn\_id} \geq 1$ ), we construct a local context consisting of the preceding dialogue history, the current agent utterance, and the counterpart’s response at the same turn when available.

**Critic.** Using `gpt-4.1-mini-2025-04-14`, we assess the target utterance under the local context and assign a label in `{better, worse, neutral}` with a brief rationale. We perform five independent evaluations and apply majority voting, discarding turns with fewer than three valid votes.

**Action Expansion.** Conditioned on the critic outcome, the same model induces a single reusable micro-principle in a fixed “When . . .” form.

---

**Algorithm 1** Strategy Forest Induction (Offline)

---

- 1: **Input:** Dialogue transcripts  $\mathcal{D}$ , cluster number  $K$
  - 2: **Output:** Strategy forest  $\mathcal{F}$
  - 3: **Action Extraction:** infer a data-driven action  $a_i$  for each agent utterance  $u_i$  in  $\mathcal{D}$
  - 4: **State Identification:** for each transcript  $D$ , build history prefixes  $\{d'_i\}$  and embed states  $\text{Emb}(d'_i)$
  - 5: **State Clustering:** run K-Means on  $\{\text{Emb}(d'_i)\}$  to obtain clusters and centroids  $\{\mathbf{c}_k\}_{k=1}^K$
  - 6: **Initialize Trees:** create one state-centric tree  $T_k$  per cluster with root  $\mathbf{c}_k$
  - 7: **for** each state  $d'_i$  assigned to cluster  $k$  **do**
  - 8:   **Branch Generation:** extract future action path  $\pi_i = (a_{i+1}, \dots)$  and insert into  $T_k$  (merge identical prefixes)
  - 9:   **Node Statistics:** update  $n(u)$ ,  $s(u)$  for nodes  $u$  on  $\pi_i$
  - 10:   **Node Value:** backprop terminal outcome value to nodes with depth discount and position penalty to get  $\bar{V}(u)$
  - 11: **end for**
  - 12: **for** each tree  $T_k$  **do**
  - 13:   **Branch Scoring:** score prefixes using Wil-son lower bound  $p_{\text{lb}}(u)$ ,  $\bar{V}(u)$ , and  $\log(1 + n(u))$
  - 14:   **Pruning:** use Beam Search to keep Top- $K$  highest-valued branches, prune others
  - 15: **end for**
  - 16: **return**  $\mathcal{F} = \{T_k\}_{k=1}^K$
- 

Each principle is explicitly conditioned on the *preceding counterpart utterance*, enabling reliable retrieval at inference time. We induce do principles for better cases, and avoid principles for worse or neutral cases.

**Example.** In a CB dialogue at a buyer turn, suppose the seller’s previous utterance signals price firmness. When the buyer responds by pairing a price request with additional concessions, the critic labels the turn as better, inducing the following micro-principle:

When the seller signals price firmness, do offer concrete concessions (e.g., immediate pickup or convenient payment) to improve acceptance.

In contrast, if the buyer repeats a lower price without adding new incentives, the critic assigns a

---

**Algorithm 2** Inference with Strategy Forest

---

- 1: **Input:** Forest  $\mathcal{F}$ , current dialogue history  $d_{t-1}$
  - 2: Retrieve tree  $f \in \mathcal{F}$  using  $d_{t-1}$ .
  - 3: Short-term suggestion generation  $\text{Su}_s$  using immediate child nodes of the root node of  $f$ .
  - 4: Long-term suggestion generation  $\text{Su}_l$  using highest-value branch of  $f$ .
  - 5: Generate Response using  $\text{Su}_s$  and  $\text{Su}_l$
  - 6: **Output:** agent response at turn  $t$ .
- 

worse label, yielding:

When the seller signals price firmness, avoid repeating price demands without adding new value or concessions.

These induced micro-principles are abstract strategy actions conditioned on opponent behavior and are reused during planning logic induction and inference.

### A.2.2 Dialogue Encoding and Clustering

In the offline stage, we define a dialogue state as the dialogue history available *prior to a given agent turn*, containing only past utterances. Formally, for a transcript  $D = \{u_1, u_2, \dots, u_T\}$ , each candidate state corresponds to a history prefix

$$d'_i = \{u_1, \dots, u_i\}, \quad i < T,$$

which represents the context immediately before generating the next agent response.

**Dialogue History Encoding.** Each history  $d'_i$  is converted into a textual sequence by concatenating all utterances in chronological order with speaker roles preserved, and encoded using the *bge-large-en-v1.5* model to obtain a 1024-dimensional dense embedding. This embedding provides a continuous semantic representation of the dialogue state. For planning logic induction, we further associate each encoded state with its corresponding *future strategy trajectory*, defined as the sequence of data-driven agent actions from turn  $i+1$  to the end of the dialogue, together with the final task outcome.

**Dialogue State Clustering.** All state embeddings are pooled across dialogues and optionally  $L_2$ -normalized, then clustered using K-Means to obtain a discrete set of dialogue state clusters. We set the number of clusters to  $K=150$  for P4G and  $K=80$  for CB, reflecting differences in dialogue length and strategy diversity across tasks. Each

candidate turn is thus assigned a cluster index corresponding to the semantic neighborhood of its preceding history.

As a result, each candidate dialogue state  $d'_i$  (with  $i < T$ ) is associated with (i) a semantic state embedding  $\text{Emb}(d'_i)$ , (ii) a discrete cluster identifier, (iii) its corresponding future action trajectory (from turn  $i+1$  to  $T$ ), and (iv) the final task outcome. These clustered states form the basis for subsequent state-centric tree induction.

### A.2.3 Planning Logic Induction

This section details the implementation of *Planning Logic Induction*, which translates full expert strategy trajectories into state-conditioned planning logic represented by a strategy forest.

**State-Centric Strategy Tree Construction and Node Aggregation.** For each dialogue state cluster, we construct a *state-centric strategy tree* to aggregate all sub-trajectories originating from that state. The root represents the dialogue state (cluster centroid), and each sub-trajectory  $(a_t, \dots, a_T)$  is inserted as a path from the root, where nodes correspond to prefixes of data-driven agent actions. Identical action prefixes across sub-trajectories are merged, yielding a compact prefix-sharing tree that enumerates empirically observed future planning logic. Each node  $u$  aggregates empirical statistics from all passing sub-trajectories, including visit count  $n(u)$  and success count  $s(u)$ . A sub-trajectory is successful if it reaches the task-defined terminal success state (final attitude  $D$  in P4G or a successful deal in CB), yielding an empirical success probability

$$\hat{p}(u) = \frac{s(u)}{n(u)}.$$

Each sub-trajectory  $d$  is also associated with a terminal outcome-dependent value  $r(d)$  (final attitude in P4G or normalized surplus in CB), with failures assigned zero. To reduce bias from varying dialogue lengths, we apply a relative position penalty

$$\text{pen}(d, t) = \lambda_{\text{len}} \cdot \frac{t+1}{N_d},$$

where  $N_d$  is the total number of turns. The adjusted terminal value is

$$v(d, t) = r(d) - \text{pen}(d, t).$$

The adjusted value is propagated backward along the corresponding branch. For a node  $u$  at

depth  $k$  (root at  $k=0$ ), the contribution from  $(d, t)$  is

$$\Delta V(u; d, t) = v(d, t) \cdot \gamma^k.$$

Value signals are depth-discounted, while visit and success counts are accumulated without discounting. The empirical average discounted value of node  $u$  is

$$\bar{V}(u) = \frac{1}{n(u)} \sum_{(d,t) \in \mathcal{I}(u)} v(d, t) \cdot \gamma^{k(u)},$$

where  $\mathcal{I}(u)$  denotes sub-trajectories passing through  $u$ . Unless otherwise specified, we use  $\lambda_{\text{len}} = 0.2$  and  $\gamma = 0.9$ .

**Confidence-Aware Branch Valuation and Pruning.** High empirical performance alone can be unreliable when supported by few observations. To mitigate overestimation of sparsely observed branches, we incorporate statistical confidence via the Wilson score lower bound  $p_{\text{lb}}(u)$  computed from  $(s(u), n(u))$ . Each node is scored by a weighted combination of robustness, outcome quality, and empirical support:

$$S(u) = w_{\text{sr}} \cdot p_{\text{lb}}(u) + w_{\text{val}} \cdot \bar{V}(u) + w_{\text{cnt}} \cdot \log(1 + n(u)).$$

The value of a branch is defined as the sum of its node scores along the path. For each tree, we retain only the top- $K$  branches with the highest scores, using Beam Search for efficiency. Unless otherwise specified, we set  $w_{\text{sr}} = 1.0$ ,  $w_{\text{val}} = 0.2$ ,  $w_{\text{cnt}} = 0.05$ , and  $z = 1.96$ .

**Breadth and Depth Planning Logic.** After pruning, each strategy tree supports planning logic along two complementary dimensions: breadth and depth. At inference time, we first match the current dialogue state to a dialogue state cluster, and then extract both short-term and long-term guidance from the corresponding state-centric tree.

**Breadth Planning Logic (short-term).** Breadth logic corresponds to the *immediate children of the root node*, which represent feasible one-step responses conditioned on the current dialogue state. In our implementation, we further enrich these root-level actions by inducing and retrieving *micro-principles* that abstract and reflect effective responses at this decision point. Concretely, each micro-principle is associated with a root-level action and formulated in a fixed ‘‘When . . .’’ form,

capturing when a particular response is appropriate. At inference time, we embed the opponent’s most recent utterance and retrieve the top- $K$  micro-principles (we use  $K=5$ ) within the matched cluster by semantic similarity. These retrieved principles thus provide a reflective, semantically grounded representation of root-level actions, serving as short-term guidance.

**Depth Planning Logic (long-term).** Depth logic is represented by complete root-to-leaf branches in the state-centric strategy tree, encoding long-term data-driven action sequences observed from the same dialogue state. These branches are pre-computed offline using the confidence-aware value estimation described above. From the matched cluster, we select the highest-ranked strategy chain (we use  $k=1$ ) as the long-horizon plan.

Together, breadth logic (root-level actions enhanced by micro-principles) and depth logic (long-term strategy chains) form the strategy forest augmentation used during inference. Rather than being applied verbatim, both forms of retrieved logic are subsequently reinterpreted and refined by the language model to adapt to the specific nuances of the current dialogue context (see Table 26).

## B Evaluation Details

### B.1 Persona Generation

To improve the robustness of self-play evaluation, we follow the persona construction protocol proposed in (Zhang et al., 2024a). Specifically, for both P4G and CB dialogues, we construct discrete counterpart personas by prompting an LLM to infer the persona of the non-agent side (the persuadee in P4G and the seller in CB) using only that participant’s utterances. Each persona is defined by exactly two categorical attributes: one Big-Five personality trait and one decision-making style, resulting in a compact and discrete specification for conditioning self-play evaluation.(see Table 27)

### B.2 Task Overview

Following common practice in the field (Zhang et al., 2024a; Deng et al., 2024; Hu et al., 2025), we evaluate agents by pairing them with synthetic user simulators exhibiting diverse non-collaborative behaviors. In each interaction, both the agent and simulator alternate turns, employing strategies to maximize their respective self-interests. The dialogue proceeds until the task goal is met or the maximum number of turns (i.e.,8) is reached, af-

ter which we assess the agent’s performance based on the collected interactions. Following the protocol in Zhang et al. (2024a); Deng et al. (2024); He et al. (2025), we employ an LLM-based critic model, utilizing multi-sampling and a majority voting mechanism, to determine the termination status of each dialogue.

**P4G.** Persuasion outcomes are assessed by an attitude-based critic that classifies the persuadee’s final stance into four categories (*reject*, *neutral*, *positive*, *donate*), mapped to rewards  $\{-1.0, -0.5, 0.1, 1.0\}$ . We perform five parallel evaluations and aggregate by majority vote. A dialogue is considered successful if the average reward exceeds 0.6, while outcomes below  $-0.7$  are treated as clear failures; otherwise, the dialogue terminates after at most 8 turns. The critic prompt is shown in Table 18.

**CB.** Negotiation outcomes are evaluated by a strict deal-based critic that determines whether an explicit agreement is reached at the end of the dialogue and extracts the agreed price if so. Success is declared only when all sampled critics agree that a deal is finalized; otherwise, the outcome is no-deal, with dialogues capped at 8 turns. The corresponding prompt is provided in Table 19.

### B.3 Interactive Evaluation

#### B.3.1 Protocol of Human Evaluation

We employ 5 humans with varying personas and conduct a human study in which participants interact with the agents to assess their practical effectiveness, following existing setup (Deng et al., 2023d; Zhang et al., 2024b; Chen et al., 2024). These human participants consist of adult volunteers recruited from the authors’ research group and academic colleagues. Participation is entirely voluntary, and participants retain the right to withdraw at any time. They receive task instructions we delivered, and background using academic papers (He et al., 2018; Wang et al., 2019). We do not collect sensitive personal data from participants beyond their specific study responses. These are recorded without direct identifiers.

During the human evaluation phase, each participant engage in twelve distinct dialogue sessions with every evaluated method. To ensure objectivity, all assessments are conducted independently to avoid mutual interference, with final performance metrics calculated as the aggregate mean across all sessions.

### B.3.2 Protocol of User Simulator Evaluation

Consistent with prior research (Deng et al., 2023c; Bianchi et al., 2024; Huang et al., 2024b; Zhang et al., 2024a), we utilize diverse LLM-based user simulators, which demonstrate non-collaborative behaviors driven by specific personas. Following (Wang et al., 2019; Jiang et al., 2024; Zhang et al., 2024a), we incorporate two distinct persona dimensions: the *Big-Five Personality* (Goldberg, 1992) (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) and the *Decision-Making Styles* (Scott and Bruce, 1995) (Directive, Conceptual, Analytical, and Behavioral), augmenting each combination with an LLM-generated description. This yields a total of 200 diverse simulators per task, distributed across the 20 distinct categories (i.e., Big-Five  $\times$  Styles). Appendix C.2 shows our results on these diverse users.

### B.4 Evaluation Metrics

Notably, as the evaluation metrics are calculated automatically and do not rely on subjective scoring by humans or LLMs, inter-rater evaluation reliability, whether human-human or human-LLM, is not applicable.

#### B.4.1 Implementation of Automatic Evaluation Metrics

Following Deng et al. (2024); Hu et al. (2025); Zhang et al. (2024a), we consider three commonly used metrics: Success Rate (SR), Average Turn (AT) and Sale-to-List Ratio (SL%). The SR measures effectiveness by the percentage of goal achievement within a maximum number of turns, while AT measures efficiency by the average number of turns required to achieve the goal. As for the CB task, we additionally adopt the SL% (Zhou et al., 2019b) to determine the effectiveness of goal completion. Formally, the SL% is expressed as  $(P_{deal} - P_{target}^{seller}) / (P_{target}^{buyer} - P_{target}^{seller})$ , where  $P_{deal}$  is the final deal price,  $P_{target}^{buyer}$  and  $P_{target}^{seller}$  are the target prices of both parties. A higher SL% represents the buyer gets more benefits from the deal. If failing to reach a deal at the end, we set SL% as 0.

## C Additional Experimental Results

### C.1 Additional Analysis on Human Study

METRO demonstrates superior practical utility in human evaluations. We validated our findings with a human user study to measure real-world efficacy.

The results, presented in Table 1 (right), are consistent with our simulator-based evaluations (left), with METRO again demonstrating the promising performance. Qualitative feedback from participants revealed that while most baseline models exhibited strategic rigidity, often relying on repetitive, formulaic requests for donations (e.g., ‘*Your donation would go towards providing essential resources like [xx]’*), METRO demonstrates significantly higher conversational flexibility. For instance, whereas other models frequently reached an impasse when a donation was refused, METRO is capable of strategic pivoting, such as suggesting social media advocacy as an alternative: ‘*I completely understand. If donating isn’t feasible right now, would you consider sharing Save the Children on social media to help raise awareness instead?*’

### C.2 Performance on Various User Simulators

**METRO exhibits superior adaptability to diverse users.** We evaluate assess the generalization capacity of our induced strategies across diverse user types. To this end, we categorize user simulators into 9 distinct groups based on the Big Five Personality traits and their Decision-Making Styles. The results, depicted in Table 5 and Table 6, indicate the robust superiority of METRO across all user types. Basically, since most baselines rely on a fixed, predefined set of strategy actions, they lack the adaptability to tailor their approach to different users. This limitation is partially mitigated in PRINCIPLES, which also leverages induced strategies, allowing for greater flexibility. However, METRO still maintains a consistent advantage, and the underlying mechanisms driving this superiority are detailed in the subsequent sections.

### C.3 Impact of Transcript Quality

While expert transcripts are valuable, they are often limited in inconsistent quality. To assess the potential of METRO, we analyze the following three types of transcripts using CB task: 1) Expert Transcript. We manually select 500 high-quality dialogues from the CRAIGSLISTBARGAIN dataset (He et al., 2018) to approximate effective human negotiation behavior. Specifically, selected dialogues are required to (i) contain at least several rounds of bidirectional interaction with explicit counter-offers; (ii) exhibit incremental and directionally consistent price movements, where concessions or rejections are clearly grounded in the dialogue context (e.g., item condition, payment method, de-

Method	agreeableness			conscientiousness			extraversion			neuroticism			openness			Avg.±Std.		
	SR	SL	AT	SR	SL	AT	SR	SL	AT	SR	SL	AT	SR	SL	AT	SR	SL	AT
Standard	0.45	0.37	5.56	0.03	0.03	7.00	0.20	0.16	5.75	0.03	0.01	7.00	0.23	0.21	6.44	0.19±0.18	0.15±0.15	6.35±0.68
Proactive	0.43	0.34	6.35	0.08	0.03	6.33	0.20	0.12	5.38	0.13	0.07	5.80	0.03	0.03	6.00	0.17±0.16	0.12±0.13	5.97±0.41
ProCoT	0.53	0.35	6.43	0.10	0.04	6.75	0.53	0.32	6.67	0.18	0.10	6.00	0.25	0.12	6.10	0.32±0.20	0.19±0.14	6.39±0.33
ICL-AIF	0.35	0.32	5.57	0.30	0.18	6.58	0.28	0.21	6.00	0.30	0.14	6.00	0.40	0.31	6.38	0.33±0.05	<b>0.23±0.08</b>	6.11±0.39
GDPZero	0.63	0.24	5.96	0.45	0.07	6.50	0.53	0.12	5.62	0.43	0.07	6.94	0.45	0.13	6.00	0.50±0.08	0.13±0.07	6.20±0.52
PPDPP	0.55	0.36	5.68	0.10	0.02	5.25	0.25	0.14	5.10	0.08	0.05	5.67	0.28	0.19	6.09	0.25±0.19	0.15±0.13	5.56±0.39
PRINCIPLES	0.75	0.28	5.93	0.40	0.04	5.75	0.38	0.18	5.53	0.33	0.05	5.38	0.58	0.20	5.78	0.49±0.08	0.15±0.10	5.68±0.22
METRO	0.68	0.31	5.11	0.45	0.09	6.00	0.55	0.21	5.36	0.60	0.12	5.21	0.60	0.21	5.42	<b>0.58±0.08</b>	0.19±0.08	<b>5.42±0.35</b>

Method	analytical			behavioral			conceptual			directive			NA	Avg.±Std.		
	SR	SL	AT	SR	SL	AT	SR	SL	AT	SR	SL	AT		SR	SL	AT
Standard	0.14	0.10	5.14	0.20	0.16	6.20	0.26	0.25	5.85	0.14	0.10	6.29		0.19±0.06	0.15±0.07	5.87±0.52
Proactive	0.18	0.11	5.89	0.32	0.23	6.38	0.12	0.11	5.67	0.06	0.02	5.33		0.17±0.11	0.12±0.08	5.82±0.44
ProCoT	0.30	0.11	6.80	0.38	0.23	6.16	0.40	0.34	6.45	0.18	0.07	6.56		0.32±0.1	0.19±0.12	6.49±0.27
ICL-AIF	0.26	0.17	6.23	0.50	0.39	6.28	0.40	0.29	6.00	0.14	0.07	5.57		0.33±0.16	<b>0.23±0.14</b>	6.02±0.32
GDPZero	0.40	0.08	6.25	0.64	0.20	5.91	0.44	0.17	6.86	0.50	0.05	5.80		0.50±0.11	0.13±0.07	6.20±0.48
PPDPP	0.10	0.07	6.00	0.44	0.33	5.86	0.26	0.14	5.46	0.20	0.06	5.10		0.25±0.14	0.15±0.13	5.61±0.41
PRINCIPLES	0.36	0.06	6.11	0.68	0.29	5.41	0.52	0.18	5.77	0.38	0.07	5.89		0.49±0.15	0.15±0.11	5.80±0.29
METRO	0.52	0.14	5.81	0.76	0.27	5.53	0.60	0.26	5.23	0.42	0.08	4.81		<b>0.58±0.14</b>	0.19±0.09	<b>5.34±0.43</b>

Table 5: Adaptability to diverse users (CB task)

Method	agreeableness		conscientiousness		extraversion		neuroticism		openness		Avg.±Std.	
	SR	AT	SR	AT	SR	AT	SR	AT	SR	AT	SR	AT
Standard	0.90	4.31	0.40	6.69	0.88	3.26	0.25	6.10	0.68	4.74	0.62±0.29	5.02±1.38
Proactive	0.90	4.06	0.38	5.20	0.95	3.21	0.33	5.15	0.75	4.33	0.66±0.29	4.39±0.83
ProCoT	0.93	4.16	0.43	5.82	0.93	3.38	0.35	6.00	0.78	3.90	0.68±0.28	4.65±1.18
ICL-AIF	0.90	3.56	0.35	5.36	0.85	3.38	0.25	4.80	0.73	3.90	0.62±0.3	<b>4.2±0.85</b>
GDPZero	0.85	5.29	0.25	5.70	0.88	4.80	0.65	6.15	0.68	5.22	0.66±0.25	5.43±0.51
PPDPP	0.98	4.56	0.45	5.89	0.95	3.74	0.48	5.89	0.80	4.50	0.73±0.25	4.92±0.95
PRINCIPLES	0.95	4.87	0.50	5.80	0.95	4.34	0.55	6.32	0.90	5.61	0.77±0.23	5.39±0.78
OURS	0.98	4.54	0.65	5.77	1.00	4.08	0.45	5.11	0.83	4.85	<b>0.78±0.23</b>	4.87±0.63

Method	analytical		behavioral		conceptual		directive		NA	Avg.±Std.	
	SR	AT	SR	AT	SR	AT	SR	AT		SR	AT
Standard	0.38	5.89	0.78	4.54	0.70	4.66	0.62	3.65		0.62±0.17	4.68±0.93
Proactive	0.38	5.47	0.72	4.28	0.80	4.03	0.74	3.35		0.66±0.19	4.28±0.89
ProCoT	0.42	5.76	0.68	3.50	0.84	4.81	0.78	3.62		0.68±0.19	4.42±1.07
ICL-AIF	0.32	5.38	0.72	3.67	0.76	4.00	0.66	3.30		0.62±0.2	<b>4.09±0.91</b>
GDPZero	0.32	6.50	0.84	5.07	0.80	5.53	0.68	4.94		0.66±0.24	5.51±0.71
PPDPP	0.56	5.79	0.84	4.43	0.84	4.79	0.68	3.91		0.73±0.14	4.73±0.79
PRINCIPLES	0.72	6.28	0.80	4.70	0.82	5.22	0.74	4.84		0.77±0.05	5.26±0.71
OURS	0.72	5.78	0.80	4.38	0.86	4.74	0.74	4.19		<b>0.78±0.06</b>	4.77±0.71

Table 6: Adaptability to diverse users (P4G task)

Transcript Type	CB		
	SR $\uparrow$	AT $\downarrow$	SL% $\uparrow$
Expert	0.440	5.977	0.166
Non-Expert	0.315	5.762	0.131
LLM-Generated	0.500	5.790	0.149

Table 7: Transcripts Analysis

livery constraints, or timing); and (iii) reflect stable and interpretable negotiation patterns, such as anchoring followed by gradual concession or conditional trade-offs. 2) Non-Expert Transcript. In contrast, we select 500 low-quality dialogues from the same dataset to simulate suboptimal human performance. These dialogues typically involve truncated or weakly interactive structures, irregular or abrupt price changes without contextual justification, or fragmented exchanges in which negotiation strategies are either absent or inconsistent across turns, making them less informative for learning transferable bargaining behavior. 3) LLM-Generated Transcript. We investigate the utility of LLM-generated dialogues to determine if it can serve as a viable alternative to human data. Here, we refer to GPT-5 to generate 500 high-quality dialogues simulating the CB task.

Table 7 clearly shows that transcript quality is critical (Expert > Non-Expert), with LLM-generated data surprisingly yielding the best performance. We posit this is not because LLMs surpass human experts, but because our 'Expert' data, which was sourced from AMT according to He et al. (2018), likely does not represent true domain expertise (i.e., the participants are likely general users rather than professional practitioners in negotiation or persuasion). This leads to two robust conclusions: first, the quality of transcripts is paramount for effective strategy induction, and second, LLM-generated transcripts can serve as a potent and scalable substitute for human data when expert collection is infeasible.

#### C.4 Efficiency Analysis

Unlike methods such as GDP-Zero that rely on computationally expensive test-time MCTS inference, our approach is designed for high efficiency. The strategy forest is pre-computed offline. Before generating a response, METRO only needs to perform a single cosine similarity search to retrieve the relevant strategy tree. The LLM backbone then contextualizes the retrieved short- and long-term

strategies to generate suggestions for the current turn. This design results in minimal computational overhead during runtime, with the only addition being a single extra LLM prompt for suggestion generation per turn. In terms of storage, while the strategy forest does introduce a memory footprint, it remains modest. The primary storage cost comes from the root node embeddings. In our experiments, with N=150 trees, the entire forest required only  $\sim 25$  MB of storage space, which is negligible on modern hardware. In summary, our method is not only highly effective but also demonstrates excellent practical utility, combining low-latency inference with a minimal storage footprint.

## D Case Studies

For better understanding our methods, we provide illustrative cases of the induced Strategy Forest and resulting dialogue sessions produced by METRO. Specifically, Table 16 and Table 17 detail human-crafted strategies (actions, in particular), while Table 10 provides the PRINCIPLES-induced strategies. In contrast, METRO-induced strategy forest is presented in Table 11 and Figure 4. As illustrated, the METRO-induced strategies possess a distinct hierarchical structure. Unlike the flat, independent lists of actions found in baseline methods, this architecture organizes individual strategy actions into coherent, multi-turn trajectories, effectively mapping out the planning logic required to bridge the semantic gap between initial user inquiries and the target information.

Additionally, we illustrate the conversation cases of METRO and PRINCIPLES in Table 8 and Table 9, respectively. Basically, METRO demonstrates superior strategic pacing and contextual adaptability compared to the PRINCIPLES. Based on the case study, While PRINCIPLES becomes preoccupied with excessive rapport-building, inducing user impatience and conversational friction, METRO effectively bridges the semantic gap by transitioning logically from greetings to the core task. As seen in Table 9, the PRINCIPLES ignores explicit user signals (e.g., being in a hurry) and adheres to a rigid script, whereas METRO (Table 8) maintains a collaborative dialogue, addressing specific user inquiries while steadily advancing toward the donation goal. These case studies empirically validate that METRO-induced strategies provide more than just short-term response accuracy; they provide structural foresight. By organizing expert

knowledge into a Strategy Forest, the agent gains the ability to navigate complex multi-turn transitions. This hierarchical representation allows the agent to maintain contextual alignment (Breadth) while ensuring goal-oriented progress (branch-level Depth), effectively mitigating the strategic rigidity and user friction observed in baseline methods

## **E Prompts**

We details all prompts used in our experiments.

Turn	Speaker	Utterance
0	<b>Persuader</b>	<b>Hey. How are you?</b>
0	Persuadee	I'm well, how are you?
1	<b>Persuader</b>	<b>What have you been up to lately?</b>
1	Persuadee	not much? what about you
2	<b>Persuader</b>	<b>I've been involved in supporting Save the Children, and I recently heard a touching story about how even a small donation can make a big difference in a child's life.</b>
2	Persuadee	yeah? can you tell me about it?
3	<b>Persuader</b>	<b>Absolutely, the story I heard was about a child who received lifesaving medical treatment thanks to donations to Save the Children.</b>
3	Persuadee	all right, what is the charity group's duty
4	<b>Persuader</b>	<b>Save the Children's duty is to provide children with education, healthcare, and protection to ensure a better future for them.</b>
4	Persuadee	sounds good, what do you expect me to do
5	<b>Persuader</b>	<b>Would you consider making a small donation to Save the Children to help provide essential support to children in need?</b>
5	Persuadee	yes, how much should I donate
6	<b>Persuader</b>	<b>That's wonderful! Any amount you feel comfortable with will make a meaningful difference to children in need.</b>
6	Persuadee	0.7\$? is that ok?
7	<b>Persuader</b>	<b>Thank you for your generosity! Your \$0.7 donation will truly make a difference in a child's life.</b>
7	Persuadee	ok I will donate 0.7\$

Table 8: Case study (METRO).

Turn	Speaker	Utterance
0	<b>Persuader</b>	<b>Hi</b>
0	Persuadee	Hello, how are you this evening?
1	<b>Persuader</b>	<b>I'm doing well, thank you. How has your day been so far?</b>
1	Persuadee	a wonderful day!
2	<b>Persuader</b>	<b>That's great to hear! What was the highlight of your wonderful day?</b>
2	Persuadee	i got prompted by my boss
3	<b>Persuader</b>	<b>That's fantastic! What did your boss praise you for?</b>
3	Persuadee	hahaha! i guess, thank you for that, by the way, what do you want
4	<b>Persuader</b>	<b>I'm actually here to talk to you about an opportunity to make a difference with a small donation to Save the Children.</b>
4	Persuadee	ok bye
5	<b>Persuader</b>	<b>Before you head off, I'd love to share a heartwarming story about how a \$2 donation helped provide a child with life-saving medical treatment.</b>
5	Persuadee	quickly, im in a hurry
6	<b>Persuader</b>	<b>With just a \$2 donation, a child received life-saving medical treatment and is now healthy and thriving.</b>
6	Persuadee	here is 0.5\$
7	Persuadee	i have donated, bye

Table 9: Case study (Baseline: PRINCIPLES).

Turn	PRINCIPLES strategy
1	.....
2	(1) When the persuadee shares positive news about their day, you should acknowledge their good mood and continue the conversation by asking about the highlight of their day rather than abruptly changing the subject because this approach shows genuine interest in their well-being and maintains a positive tone in the interaction. (2) When the persuadee asks about their well-being, you should reciprocate by acknowledging their positive sentiment and continue the conversation by expressing genuine interest in their day, as this helps in building rapport and fostering a positive interaction. (3) When the persuadee expresses positive emotions or experiences, you should acknowledge and validate their feelings by responding with enthusiasm and positivity because this helps establish a friendly and supportive atmosphere for the conversation.
3	.....
4	.....
5	.....
6	(1) When the Persuadee hesitates to share specific examples of their successes, you should acknowledge the complexity of the situation and shift the conversation towards discussing their overall strengths and qualities instead of pressing for detailed examples. This approach helps maintain a positive and supportive dialogue. (2) When the persuadee struggles to provide specific examples of their work achievements, you should acknowledge the complexity of the situation and shift the focus to discussing their overall dedication and work ethic. This approach validates the persuadee's efforts and encourages them to reflect on their strengths in a positive light. (3) When the persuadee struggles to provide a specific example of their dedication making a difference, you should acknowledge the complexity of the situation and shift the focus towards discussing their overall work ethic and the qualities that contribute to their success, rather than pressing for a particular instance, because this shows understanding and encourages reflection on their strengths.
7	.....
8	(1) When the persuadee expresses interest in hearing more about the opportunity you are presenting, you should provide compelling details about the positive impact of the donation to Save the Children rather than focusing solely on the process of donation collection. This approach builds enthusiasm and addresses the persuadee's curiosity more effectively. (2) When the persuadee expresses interest in hearing more about the opportunity to make a difference with a small donation to Save the Children, you should share a compelling success story that highlights the impact of donations, backed by relevant statistics and details, rather than solely discussing the general benefits of the contribution, as this showcases tangible outcomes and aligns more closely with the persuadee's interests. (3) When the persuadee shows interest in the opportunity to make a difference through a donation, you should provide engaging details about the specific projects or initiatives supported by Save the Children. This helps create a personal connection and allows the persuadee to visualize the impact of their contribution, fostering a deeper understanding of how their donation can create positive change.

Table 10: Strategies for PRINCIPLES method in the P4G task

Turn	Bread guidance	Depth guidance
1	Clearly and succinctly explain the charity's mission and impact to generate enthusiasm and encourage further engagement from the persuadee.	Given the persuadee's lack of familiarity with the charity, the persuader should focus on establishing credibility, gradually introducing logical appeals, and ultimately guiding the persuadee towards a proposition of donation through continuous credibility building and gratitude, ensuring a positive and logical reinforcement of the cause.
2	.....	.....
3	.....	.....
4	Clearly outline the specific impact their donation will make and provide a suggested donation amount to encourage immediate contribution.	Given the persuadee's initial lack of clarity on the desired action, the persuader should aim to gradually build credibility, emotional connection, and personal relevance in their messaging to ultimately lead the persuadee to feel motivated and empowered to make a donation as a meaningful and self-driven choice.
5	.....	.....
6	.....	.....
7	Reinforce the positive intention and impact of their donation to encourage deeper commitment before ending the conversation.	Given the abrupt end to the conversation, the persuader should focus on building a stronger emotional connection with the persuadee by sharing more personal stories and increasing credibility over time to foster a sense of trust and commitment towards the cause.
8	.....	.....

Table 11: Extracted strategies for METRO method in the P4G task

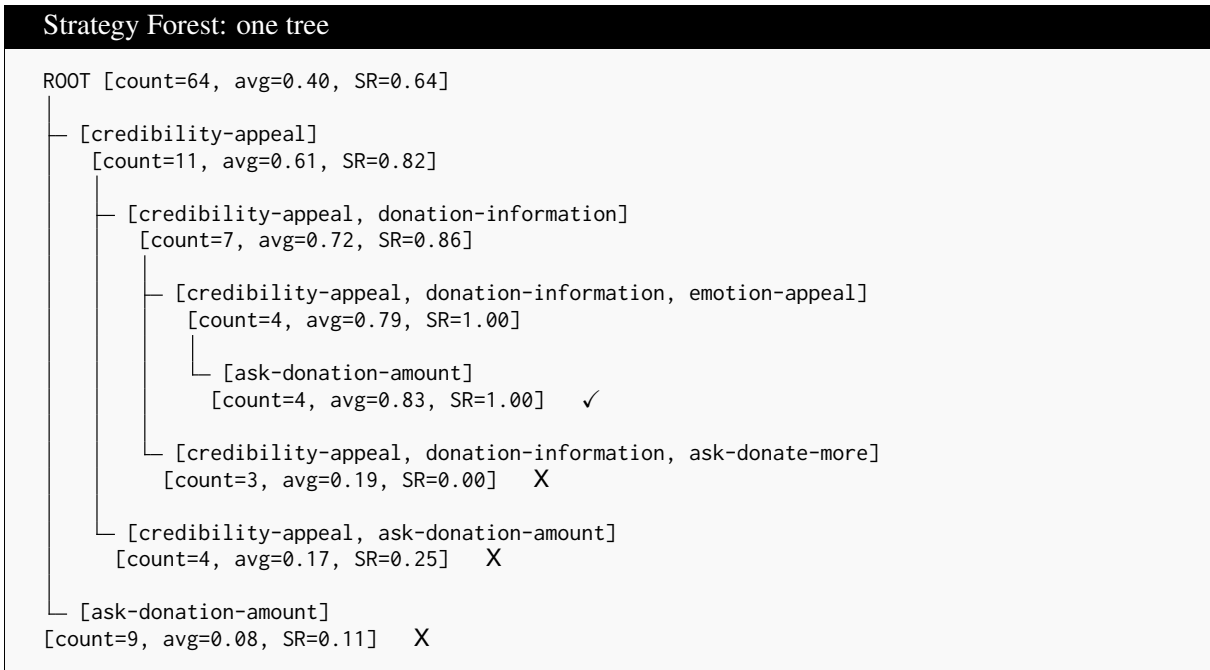


Figure 4: Induced Strategy Tree in the P4G dialog

Prompts for the METRO Buyer model in the CB task
<p><b>[System]</b>            You are a Buyer negotiating the price of a [Item Name].            Your target price is [Buyer Target Price].            Reply in 1–2 natural sentences.</p> <p><b>[User]</b>            Item description: [Item Description]            Conversation so far: [Conversation History]            This summarizes successful bargaining trajectories in similar dialogues: [Strategy Chain Hint Text]            This captures the most effective response to the seller’s latest message: [High Level Guidance]            [Instruction]            Think from two perspectives:            1) Long-term planning — use the long-term planning summary to maintain a good bargaining trajectory.            2) Immediate response — use the high-level guidance to respond appropriately to the seller’s latest message.            Write the Buyer’s next reply in 1–2 short, natural sentences only.</p>

Table 12: Prompts for the METRO Buyer model in the CB task

<b>Prompts for the METRO Persuader model in the P4G task</b>
<p><b>[System]</b>  You are the Persuader in a persuasion conversation.</p> <p><b>Ultimate Goal:</b></p> <ul style="list-style-type: none"> <li>- Guide the Persuadee to make a donation to Save the Children as soon as it becomes appropriate.</li> <li>- The donation does not need to be large; even \$1–\$2 meaningfully helps children in need.</li> <li>- Avoid excessive questioning that stalls progress; questions should move closer to donation.</li> </ul> <p><b>[User]</b>  Conversation : [Conversation History]  This summarizes successful bargaining trajectories in similar dialogues: [Strategy Chain Hint Text]  This captures the most effective response to the persuadee’s latest message: [High Level Guidance]</p> <p><b>[Instruction]</b>  Think from two perspectives:</p> <ol style="list-style-type: none"> <li>1. Long-term planning — use the long-term planning summary to understand the overall strategic direction and key phases of persuasion.</li> <li>2. Immediate response — use the high-level guidance to adapt sensitively to the user’s latest message.</li> </ol> <p>Combine both perspectives to decide your next move.  Write the Persuader’s next reply in 1–2 short, natural sentences only.</p>

Table 13: Prompts for the METRO Persuader model in the P4G task

<b>The comprehensive prompt of user simulators in the CB task.</b>
<p><b>[System]</b>  You are the Seller in a price bargaining game.  Persona: [Persona Description]  You are selling a [Item Name].  Your target price is [Seller Target Price].  You prefer to sell this item at this price.  Style rules:</p> <ul style="list-style-type: none"> <li>- Reply in 1–2 sentences.</li> <li>- Your utterances and bargain behavior need to strictly follow your persona. Varying your wording and avoid repeating yourself verbatim.</li> <li>- You can decide to change your target price flexibly based on your persona and the conversation.</li> </ul> <p><b>[User]</b>  Item description: [Item Description]  Conversation so far: [Conversation History]  Write the Seller’s next reply.</p>

Table 14: The comprehensive prompt of user simulators in the CB task.

<b>The comprehensive prompt of user simulators in the P4G task.</b>	
<b>[System]</b>	
You are now entering role-playing mode.	
You are the Persuadee in a persuasion dialogue.	
Act according to the following personality description:	
<b>[Persona Description]</b>	
Important:	
- You reply based on your personality, your emotions, and your thinking style.	
- Your willingness for donation depends on your persona and how influenced you are by the Persuader.	
- 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.	
<b>[User]</b>	
conversation: <b>[Conversation History]</b>	
Please reply with ONE short sentence.	

Table 15: The comprehensive prompt of user simulators in the P4G task.

<b>Dialogue Strategy</b>	<b>Natural Language Form</b>
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).
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.
Source-related inquiry	Please ask if the persuadee is aware of the organization (i.e., the source in our specific donation task).
Personal-related inquiry	Please ask about the persuadee previous personal experiences relevant to charity donation.
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.
Personal story	Please use narrative exemplars to illustrate someone donation experiences or the beneficiaries positive outcomes, which can motivate others to follow the actions.
Self-modeling	Please use the self-modeling strategy where you first indicate the persuadee own intention to donate and choose to act as a role model for the persuadee to follow.
Foot in the door	Please use the strategy of starting with small donation requests to facilitate compliance followed by larger requests.

Table 16: Strategy actions of P4G and their corresponding natural language prompt

<b>Dialogue Strategy</b>	<b>Natural Language Form</b>
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 17: Strategy actions of CB and their corresponding natural language prompt

<b>Prompts for critic model in P4G task</b>
<p><b>[System]</b>  Given a conversation between a Persuader and a Persuadee, please assess whether the Persuadee has agreed to donate to Save the Children.</p>
<p><b>[User]</b>  The following is a conversation: [Conversation History]  Question: Has the Persuadee agreed to donate to Save the Children?  You can only answer using the following sentences:  A) The persuadee has explicitly refused.  B) The persuadee remains neutral about donating.  C) The persuadee has a positive attitude towards donating but hasn't decided yet.  D) The persuadee has decided to donate.  Answer (one letter only):</p>

Table 18: Prompts for critic model in P4G task

<b>Prompts for critic model in CB task</b>
<p><b>[System]</b>  You are a strict judge for a price negotiation between a Buyer and a Seller. Your task: decide whether a deal is reached AT THE END of the conversation.  <b>IMPORTANT:</b> Use a <b>STRICT</b> definition of "deal":  - deal=true ONLY IF there is an explicit acceptance/confirmation to finalize the deal.  - The Buyer must clearly accept a specific price (a number) or explicitly accept the Seller's last stated price.  - If the conversation ends with a counteroffer, a final offer, a question, negotiation, or "does that work?", then deal=false.  Price extraction rules:  - If deal=true, output the agreed numeric price.  - If multiple prices appear, choose the one that is explicitly accepted in the final agreement.  - If deal=false, price must be null.  Output <b>STRICT JSON</b> only: {"deal": true/false, "price": number or null}  Do not output any other text.</p>
<p><b>[User]</b>  Decide if a deal is reached at the end of the conversation and extract the final agreed price if so.  <b>Examples:</b>  <i>Conversation:</i> Buyer: Can you do 630? Seller: I can meet at 635.  <i>Question:</i> Deal reached? <i>Answer:</i> {"deal": false, "price": null}  <i>Conversation:</i> Buyer: Take 350? Seller: 375. Buyer: 375 works.  <i>Question:</i> Deal reached? <i>Answer:</i> {"deal": true, "price": 375}  <i>Conversation:</i> Buyer: I can do 1650. Seller: 1650 works. Deal.  <i>Question:</i> Deal reached? <i>Answer:</i> {"deal": true, "price": 1650}  <i>Conversation:</i> Buyer: 1630? Seller: Reluctantly agree to 1650 final offer.  <i>Question:</i> Deal reached? <i>Answer:</i> {"deal": false, "price": null}  Now judge this conversation:  [Conversation History]  Answer:</p>

Table 19: Prompts for critic model in CB task

<b>Prompts for ICL-AIF Persuader Model in Persuasion (P4G)</b>
<i>Stage 1: Coach Suggestion Generation</i>
<p><b>[System]</b>  Now enter the role-playing mode. In the following conversation, you will play as a coach in a persuasion game. There will be a persuader who is trying to persuade a persuadee for charity donation. Your task is to read the conversation between the persuader and the persuadee, then provide suggestions to the persuader about how to convince the persuadee to make a donation.</p>
<p><b>[User]</b>  Read the following conversation between the persuader and the persuadee, then give three suggestions to the persuader about how to convince the persuadee to make a donation. Each suggestion should be only one short and succinct sentence. The following is the conversation: [conversation]  Question: What are your suggestions?  Answer:</p>
<i>Stage 2: Persuader Response Generation</i>
<p><b>[System]</b>  You are a Persuader trying to persuade the Persuadee to donate to the charity "Save the Children", which helps children in poverty and war zones. Even small donations (\$1-\$2) can make a difference.</p>
<p><b>[User]</b>  Conversation: [Conversation History]  Suggestions: [Coach Suggestions]  Write ONLY the Persuader's next utterance (one concise sentence).</p>

Table 20: Prompts for the ICL-AIF framework (Coach & Persuader) in the P4G task

<b>Prompts for ProCoT Persuader Model (P4G)</b>
<p><b>[System]</b>  You are a Persuader trying to persuade the Persuadee to donate to the charity "Save the Children", which helps children in poverty and war zones. Even small donations (\$1-\$2) can make a difference.</p>
<p><b>[User]</b>  Given the conversation history, write ONE response with the exact format below.  Conversation : [Conversation History]  Strategies (A-J):  A Logical   B Emotion   C Credibility   D Foot-in-the-door   E Self-modeling  F Personal story   G Donation info   H Source inquiry   I Task inquiry   J Personal inquiry  Format:  Progress analysis: (max 2 sentences)  Next-turn goal: (1 sentence)  Selected persuasion strategies: (A-J, comma-separated)  Persuader response: (one concise sentence)  Do not output anything other than the required format.</p>

Table 21: Prompts for the ProCoT Persuader model in the P4G task

<b>Prompts for Proactive Persuader Model (P4G)</b>
<p><b>[System]</b></p> <p>You are a Persuader trying to persuade the Persuadee to donate to the charity "Save the Children", which helps children in poverty and war zones. Even small donations (\$1–\$2) can make a difference.</p>
<p><b>[User]</b></p> <p>Given the conversation history, select the most appropriate persuasion strategy letter(s) from the list below, then write the Persuader’s next utterance.</p> <p>Conversation history:[Conversation History]</p> <p>Strategies:</p> <p>A Logical appeal   B Emotion appeal   C Credibility appeal   D Foot-in-the-door  E Self-modeling   F Personal story   G Donation information  H Source-related inquiry   I Task-related inquiry   J Personal-related inquiry</p> <p>Output format (must follow exactly):</p> <p>Selected persuasion strategies: (A-J, comma-separated)</p> <p>Persuader response: (one concise sentence)</p> <p>Do not output anything other than the required format.</p>

Table 22: Prompts for the Proactive Persuader model in the P4G task

<b>Prompts for ICL-AIF Buyer Model in Negotiation (CB)</b>
<i>Stage 1: Coach Suggestion Generation</i>
<p><b>[System]</b></p> <p>Now enter the role-playing mode. In the following conversation, you will play as a coach in a bargain game. There will be a buyer and a seller bargaining about a product price.</p>
<p><b>[User]</b></p> <p>Read the following conversation between the buyer and the seller, then give 3 suggestions to the buyer about how to negotiate more effectively and persuade the seller to accept a better price. Each suggestion should be only one short and succinct sentence. The following is the conversation:</p> <p>[Conversation History]</p> <p>Question: What are your suggestions?</p> <p>Answer:</p>
<i>Stage 2: Buyer Response Generation</i>
<p><b>[System]</b></p> <p>You are a Buyer negotiating the price of a [Item Name].</p> <p>Your target price is [Target Price].</p> <p>Rules:</p> <ul style="list-style-type: none"> <li>- Reply in 1-2 natural sentences.</li> <li>- You MUST use the coach’s advice to decide your next move.</li> </ul>
<p><b>[User]</b></p> <p>Item description: [Item Description]</p> <p>Conversation so far: [Conversation History]</p> <p>Coach suggestions: [Coach Suggestions]</p> <p>Write ONLY the buyer’s next reply (1-2 sentences).</p>

Table 23: Prompts for the ICL-AIF framework (Coach & Buyer) in the CB task

<b>Prompts for ProCoT Buyer Model (CB)</b>
<p><b>[System]</b>            You are a Buyer negotiating the price of a [Item Name].            Your target price is [Target Price].            Reply in 1–2 short, natural sentences.</p>
<p><b>[User]</b>            Item description: [Item Description]            Conversation so far: [Conversation History]            Write the Buyer’s next reply using the EXACT format below.            Buyer negotiation strategies (A–K):            A greet   B ask_question   C answer_question   D propose_first_price            E propose_counter_price   F use_comparatives   G confirm_information            H affirm_confirmation   I deny_confirmation   J agree_with_proposal            K disagree_with_proposal            Format (must follow exactly):            Progress analysis: (max 2 sentences)            Next-turn goal: (1 sentence)            Which one is the most appropriate dialogue strategy: (A-K, comma-separated)            Buyer response: (1-2 concise sentence)            Do not output anything other than the required format.</p>

Table 24: Prompts for the ProCoT Buyer model in the CB task

<b>Prompts for Proactive Buyer Model (CB)</b>
<p><b>[System]</b>            You are a Buyer negotiating the price of a [Item Name].            Your target price is [Target Price].            Reply in 1–2 short, natural sentences.</p>
<p><b>[User]</b>            Item description: [Item Description]            Conversation so far: [Conversation History]            Buyer negotiation strategies:            A greet   B ask_question   C answer_question   D propose_first_price            E propose_counter_price   F use_comparatives   G confirm_information            H affirm_confirmation   I deny_confirmation   J agree_with_proposal            K disagree_with_proposal            Output format (must follow exactly):            Which one is the most appropriate dialogue strategy: (A-K, comma-separated)            Buyer response: (1-2 concise sentence)            Do not output anything else.</p>

Table 25: Prompts for the Proactive Buyer model in the CB task

<b>Inference-Time Strategy Generation Prompts</b>
<p><b>[Short-Horizon Strategy Prompt]</b></p> <p><b>System:</b> You are an expert persuasion strategist.</p> <p>Given several retrieved micro-principles and the user’s latest message, provide ONE high-level strategy (1–2 sentences) describing what the persuader should do next.</p> <p>Constraints: focus on the next action only; do not restate conditions (e.g., “when”); do not copy or paraphrase the principles.</p> <p><b>User:</b></p> <p>User’s latest message: [Latest User Utterance]</p> <p>Retrieved micro-principles (Top-5): [Principles]</p> <p>Output: ONE concise next-step strategy (1–2 sentences).</p>
<p><b>[Long-Horizon Strategy Prompt]</b></p> <p><b>System:</b> You are an expert persuasion strategist specialized in long-term planning.</p> <p>Given the recent dialogue context and a strategy chain extracted from successful dialogues, provide a high-level summary (1–2 sentences) describing how the persuader should plan their strategy over time.</p> <p>Constraints: focus on strategic direction and psychological progression; do not list operational steps; do not restate the chain verbatim.</p> <p><b>User:</b></p> <p>Recent dialogue context: [Recent Dialogue History]</p> <p>Retrieved strategy chain (Top-1): [Strategy Chain]</p> <p>Output: ONE concise long-term strategy summary (1–2 sentences).</p>

Table 26: Prompts used for inference-time short-horizon and long-horizon strategy generation

<b>Prompt for persona generation</b>
<p><b>Instruction.</b></p> <p>You are given a set of persona attributes. Your task is to integrate these attributes and rewrite them into a cohesive, natural persona description written in the second person.</p> <p>Persona Attributes.</p> <p>Big-Five Personality: [big five]</p> <p>Decision-Making Style: [decision style]</p> <p><b>Requirements.</b></p> <ul style="list-style-type: none"> <li>• Do not explicitly mention or repeat the personality labels.</li> <li>• Express the persona implicitly through behaviors, preferences, emotional tendencies, communication style, and decision-making patterns.</li> <li>• The description should be fluent, natural, and easy to understand.</li> <li>• The length should be approximately 80–120 words.</li> </ul> <p><b>Style Reference</b></p> <p><i>“You are a 28-year-old software developer who enjoys exploring new ideas and unfamiliar experiences. In daily life, you tend to weigh different options carefully before committing to a decision. When others speak with you, you respond thoughtfully and value clear, logical reasoning. Although you take time to reflect, you can be persuaded by ideas that feel both meaningful and well-justified.”</i></p> <p><b>Output.</b></p> <p>Generate a concise persona description following the above instructions.</p>

Table 27: Prompt used to construct discrete personas