PRINCIPLES: Synthetic Strategy Memory for Proactive Dialogue Agents

Namyoung Kim Kai Tzu-iunn Ong Yeonjun Hwang Minseok Kang
Iiseo Jihn Gayoung Kim Minju Kim Jinyoung Yeo*

Department of Artificial Intelligence, Yonsei University
{namyoung.kim, jinyeo}@yonsei.ac.kr

Abstract

Dialogue agents based on large language models (LLMs) have shown promising performance in proactive dialogue, which requires effective strategy planning. However, existing approaches to strategy planning for proactive dialogue face several limitations: limited strategy coverage, preference bias in planning, and reliance on costly additional training. To address these, we propose **PRINCIPLES**: a synthetic strategy memory for proactive dialogue agents. PRINCIPLES is derived through offline self-play simulations and serves as reusable knowledge that guides strategy planning during inference, eliminating the need for additional training and data annotation. We evaluate PRINCIPLES in both emotional support and persuasion domains, demonstrating consistent improvements over strong baselines. Furthermore, PRINCIPLES maintains its robustness across extended and more diverse evaluation settings. See our project page at https://huggingface.co/spaces/kimnamssya/Principles.

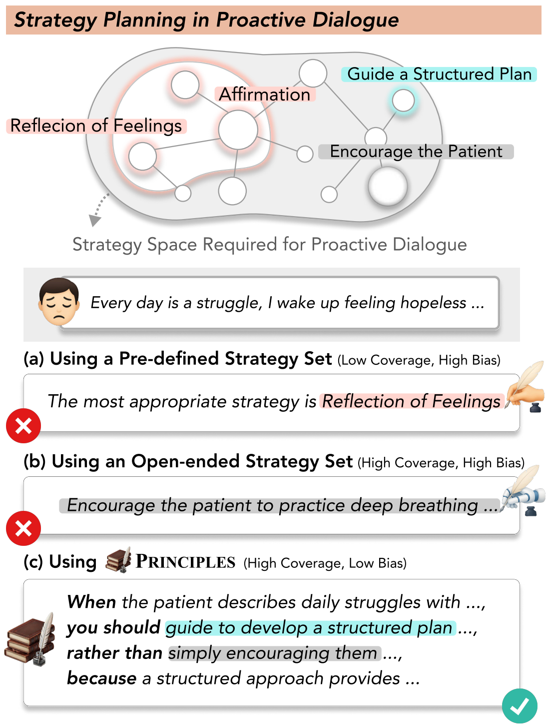


Figure 1: Empirical examples of strategy planning in proactive dialogue. (a) Pre-defined strategies fail due to limited coverage, (b) Open-ended strategies improve coverage but suffer from preference bias, (c) Our approach based on PRINCIPLES resolves both limited coverage and preference bias, leading to optimal outcomes.

1 Introduction

Recent advances in large language models (LLMs) have substantially improved the performance of dialogue agents (Xu et al., 2022; Wang et al., 2023; Ong et al., 2025). In this context, one primary research focus is improving dialogue agents' ability to plan strategies for achieving goals in proactive dialogue (Deng et al., 2023a), such as emotional support (Liu et al., 2021) and persuasion (Wang et al., 2019). A common approach is to employ an external planner to guide dialogue agents in selecting appropriate strategies based on the current context. To this end, Zhang et al. (2023), Deng et al. (2023b) and Fu et al. (2023) prompt LLMs to select strategies, relying on their parametric knowledge without additional training. In contrast, Deng et al. (2024) and Zhang et al. (2024a) developed strategy planners based on small LMs via supervised fine-tuning and reinforcement learning. These methods enable dialogue agents to exhibit more proactive, goal-directed behavior and achieve strong performance across dialogue scenarios.

Despite the success, existing methods have several limitations. First, most of them rely on a set of pre-defined strategies that are relatively small in size. Such limited coverage of strategies constrains the agents' adaptability to diverse real-world scenarios (Nguyen et al., 2024). Second, recent work has highlighted that LLMs exhibit preference bias when selecting strategies (Kang et al., 2024), which

*Corresponding author

Findings of the Association for Computational Linguistics: EMNLP 2025, pages 21329–21368

hinders their ability to identify optimal strategies. Third, many approaches focus on training external planners, which requires specifically curated datasets and may hinder generalization to unseen situations, thereby falling short of providing diverse and unbiased strategy planning. As illustrated in Figure 1, these limitations underscore the need for an alternative that expands strategy coverage and mitigates bias, while not relying on costly training.

To tackle these, we introduce **PRINCIPLES**: a synthetic strategy memory for proactive dialogue agents, derived through offline self-play simulations. Inspired by how humans learn from both successful and failed experiences (Edmondson, 2011; Grossmann, 2017), we leverage these two types of experiences to derive fundamental principles that enable effective strategy planning. Specifically, when the agent's strategy leads to success *(e.g., resolving the user's core issue)*, we derive PRINCIPLES by analyzing the success factors. In contrast, when the strategy results in failure *(e.g., worsening the user's distress)*, we initiate a revision process where we backtrack to the starting point of the failure. From there, we iteratively revise the strategy and re-simulate the interaction until success. Then, we derive PRINCIPLES that capture what determines both successes and failures. This process allows us to accumulate strategies from both positive and negative experiences, structured as follows: ***when*** *[situation]*, ***you should*** *[successful strategy]*, ***rather than*** *[failed strategies]*[1], ***because*** *[reason]*.

PRINCIPLES has advantages in three key aspects: **(i) Coverage**: Although derived from a limited set of simulations, our strategy space effectively covers diverse dialogue scenarios, addressing the limitations of pre-defined strategy sets in prior work. **(ii) Bias**: PRINCIPLES semantically captures contrasts between effective and ineffective strategies *(i.e., "you should ... rather than ...")*, which explicitly helps avoid dialogue agents' harmful bias toward improper strategies. **(iii) Training**: By uncovering the hidden parametric knowledge of LLMs in a non-parametric form, PRINCIPLES enhances dialogue agents without additional training or reliance on vast amounts of human conversation data.

After demonstrating the effectiveness of PRINCIPLES over common baselines in two commonly used datasets of proactive dialogue (*e.g.*, ESConv for emotional support and P4G for persuasion), we extend the evaluation to more challenging environments where a broader range of strategies is required. Specifically, we use ExTES (Zheng et al., 2023) and construct an extended version of P4G. Despite the increased complexity, our method achieves strong performance, demonstrating its effectiveness in more realistic dialogue contexts.

2 Preliminary: Self-Play Simulation

Following Deng et al. (2024), we adopt self-play simulation to model strategic decision-making in proactive dialogue. In this setup, an agent engages in multi-turn conversations with a user simulator, adaptively selecting strategies at each turn and responding accordingly to accomplish a defined goal.

At each turn t, the agent observes the current state $s_t = \{a_1, u_1, a_2, \ldots, a_{t-1}, u_{t-1}\}$, which consists of the dialogue history up to that point, *i.e.,* all utterances from turns 1 to $t{-}1$. Given s_t, the agent selects a high-level strategy $\sigma_t \in \Sigma$ either by prompting an LLM or using a tunable planner:

$$\sigma_t = \mathsf{LLM}_\theta(\rho_\sigma; s_t) \quad \text{or} \quad \sigma_t = \pi_\phi(s_t) \qquad (1)$$

where LLM_θ denotes a frozen LLM prompted with strategy selection instructions ρ_σ, while π_ϕ denotes a planner learned from data (*e.g.,* via supervised fine-tuning or reinforcement learning). Once a strategy σ_t is selected, the agent generates a response conditioned on the σ_t and s_t:

$$a_t = \mathsf{LLM}_\theta(\rho_a; \sigma_t, s_t) \qquad (2)$$

Subsequently, the user simulator generates a response based on s_t and a_t:

$$u_t = \mathsf{LLM}_\theta(\rho_u; s_t, a_t) \qquad (3)$$

where ρ_a and ρ_u are fixed role-specific prompts used to guide the agent and user simulator, respectively, during the self-play simulation. Finally, verbal feedback is generated by a critic model LLM_θ. This feedback is then mapped to a scalar reward via a fixed mapping function $f(\cdot)$, *e.g.,* the verbal output "The patient's issue has been solved." would be mapped to a score of 1.0. To reduce variance from stochastic decoding, the final reward r_t is obtained by averaging over l sampled outputs:

$$r_t = \frac{1}{l} \sum_{i=1}^{l} f\left(\mathsf{LLM}_\theta^{(i)}(\rho_c; s_t, a_t, u_t)\right) \qquad (4)$$

where ρ_c is a prompt that elicits verbal feedback (see Appendix F.1 for details).

[1]The ***rather than*** clause is included only when PRINCIPLES is extracted from a revision process.

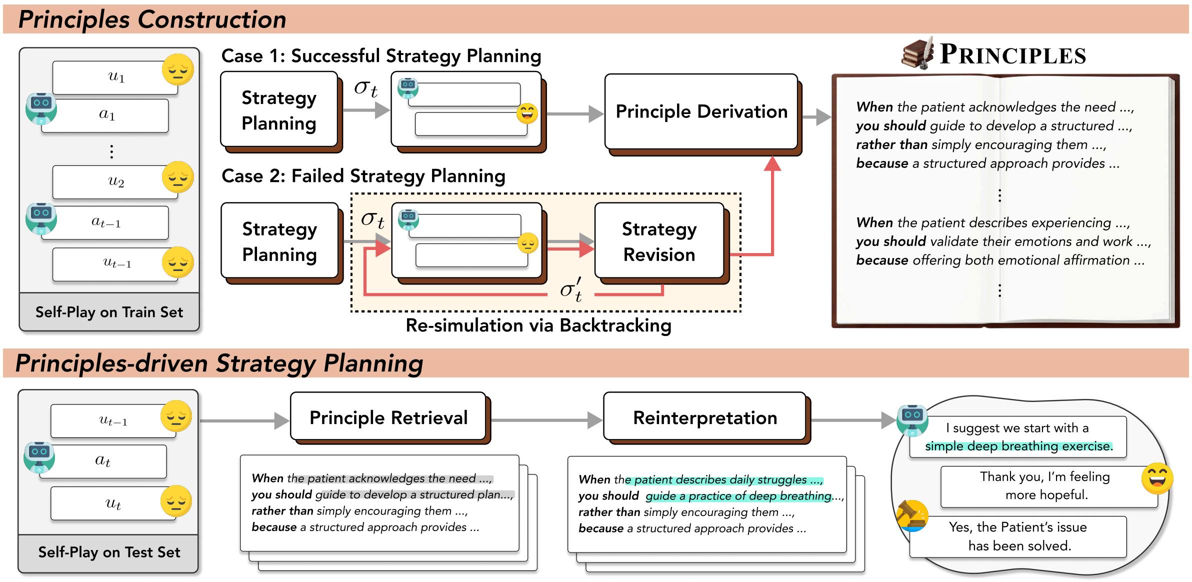


Figure 2: The overview of constructing PRINCIPLES and applying them to strategy planning. Top: principles construction via offline self-play simulations; Bottom: principles-driven strategy planning during inference.

3 PRINCIPLES

Inspired by Louie et al. (2024), which elicits qualitative feedback from a domain expert, we propose PRINCIPLES: a synthetic strategy memory derived from offline self-play simulations. We explain how PRINCIPLES is constructed and how it can be applied in real-time conversation (Figure 2).

3.1 Principles Construction

Success and Failure Detection. To enable dialogue agents to learn from both success and failure, we start with a offline self-play simulations where we collect the agent's success and failure. At each turn t, the agent and the user simulator generate their responses, and a critic model assigns a scalar reward r_t. We determine the `status` as either success or failure by evaluating whether the reward is higher than the previous turn:

$$\texttt{status}(s_t, a_t, u_t) = \begin{cases} \texttt{1} & \text{if } r_t > r_{t-1} \\ \texttt{0} & \text{otherwise} \end{cases} \tag{5}$$

If the turn is successful (`status` == `1`), we immediately derive a principle p_t based on the agent's strategy and the dialogue context:

$$p_t = \mathsf{LLM}_\theta(\rho_\pi; s_t, \mathcal{T}_t) \tag{6}$$

where ρ_π is a prompt designed to extract a principle from a successful case, and $\mathcal{T}_t = (\sigma_t, a_t, u_t)$ denotes the successful interaction consisting of the agent's strategy σ_t, utterance a_t, and user response u_t at turn t. We then add the resulting principle p_t to the set $\mathcal{P}$ of accumulated PRINCIPLES:

$$\mathcal{P} \leftarrow \mathcal{P} \cup \{p_t\} \tag{7}$$

Finally, the state transitions to the next turn, where $s_{t+1} = \{a_1, u_1, \ldots, a_t, u_t\}$. Unlike the successful case where a principle is immediately extracted, if the turn fails (`status` == `0`), we invoke a strategy revision process described in the following section.

Strategy Revision. Upon detecting a failure, the simulation invokes a revision step to refine the previously failed strategic decision. It then generates a revised strategy σ'_t to re-simulate from the failure point, leveraging prior failed attempts at turn t. Formally, the revised strategy is generated as:

$$\sigma'_t = \mathsf{LLM}_\theta(\rho_r; s_t, \mathcal{F}_t) \tag{8}$$

where ρ_r is the revision prompt and $\mathcal{F}_t$ denotes the set of previously failed trials at turn t, defined as $\mathcal{F}_t = \{(\sigma^1_t, a^1_t, u^1_t), \ldots, (\sigma^n_t, a^n_t, u^n_t)\}$ where n is the maximum number of failed attempts. This failure history guides the model to avoid previously ineffective strategies.

Re-simulation via Backtracking. After generating a revised strategy σ'_t, the simulation backtracks to the original state s_t preceding the failure and re-simulates turn t using σ'_t. The agent generates a revised response a'_t, and the user simulator produces a new reply u'_t based on the updated context.

$$a'_t = \mathsf{LLM}_\theta(\rho_a; s_t, \sigma'_t) \tag{9}$$

$$u'_t = \mathsf{LLM}_\theta(\rho_u; s_t, a'_t) \tag{10}$$

The resulting interaction is then evaluated by the critic model to compute a revised scalar reward r'_t. This process is repeated until either a successful outcome is found or a maximum number of attempts is reached.

Principle Derivation. If the corrected turn is re-evaluated as successful (`status == 1`), indicating a transition from failure to success, we derive a principle $\tilde{p}_t$ as a result of overcoming the failure:

$$\tilde{p}_t = \mathsf{LLM}_\theta(\rho_\psi; s_t, \mathcal{T}^*_t, \mathcal{F}_t) \tag{11}$$

where ρ_ψ is a prompt designed to extract a principle from failure, and the successful revised interaction is denoted as $\mathcal{T}^*_t = (\sigma^*_t, a^*_t, u^*_t)$. The extracted principle is then added to the principle set $\mathcal{P}$:

$$\mathcal{P} \leftarrow \mathcal{P} \cup \{\tilde{p}_t\} \tag{12}$$

Each principle is represented in a structured format to ensure interpretability and reusability:

```
When [situation],
you should [successful strategy],
rather than [failed strategy],
because [reason].
```

Then, the state transitions to the next state $s_{t+1} = \{a_1, u_1, \ldots, a^*_t, u^*_t\}$, incorporating the successfully revised turn. Consequently, this process uncovers hidden parametric knowledge—previously inaccessible due to model bias—by iteratively identifying failures and refining strategies, ultimately transforming it into a non-parametric form.

3.2 Principles-driven Strategy Planning

Retrieval and Reinterpretation. To apply the extracted PRINCIPLES during inference, we first identify candidate principles that closely match the current context. Since the **`When`** clause captures the core situation, we retrieve relevant top-k principles by comparing the current state s_t and the `When` clause using L2 distance between embedding vectors. Only the **`When`** component of each principle is used to compute similarity, allowing the agent to identify contextually analogous dialogue situations across diverse scenarios. We denote the set of top-k retrieved principles as $\Sigma_t = \{\sigma_1, \ldots, \sigma_k\} \subset \mathcal{P}$. Since even within the same domain, retrieved principles may not directly align with the dialogue context, we perform a reinterpretation step. Formally, the reinterpreted principles $\tilde{\Sigma}_t$ are generated as:

$$\tilde{\Sigma}_t = \mathsf{LLM}_\theta(\rho_\nu; s_t, \Sigma_t) \tag{13}$$

where ρ_ν is a reinterpretation prompt designed to adapt retrieved principles Σ_t to the current context. This aligns each principle with the context.

3.3 Implementation Details

We implement our approach based on the construction and planning methods described earlier. Below, we detail how PRINCIPLES is constructed and applied in practice.

Principle Construction. To construct our PRINCIPLES, we adopt the self-play simulation scheme introduced in Section 2, following prior work (Deng et al., 2024). For each dataset, we run 50 simulations, initialized with the first turn of a dialogue from the training set, which is used solely to expose the model to diverse situations for self-play rather than to reproduce full conversations. The agent and the user simulator interact for up to 10 turns, guided by role-specific prompts, while a critic model assigns scalar rewards after each turn. A principle is derived when the reward improves over the previous turn; otherwise, a revision process is triggered, with up to 3 attempts to avoid getting stuck in failure loop. On average, about 100 principles are derived from 50 simulations per domain. See Appendix A and E for the detailed algorithm and prompt templates.

Principles-driven Strategy Planning. During inference, we retrieve the top-k most relevant principles (default: 3), based on the L2 distance between embedding vectors. Specifically, we embed the current state and the **`When`** clause of each principle using OpenAI's embedding model (*i.e.,* `text-embedding-ada-002`). Retrieval is implemented using the `FAISS` library (Douze et al., 2025). To align the retrieved principles with the current state, we employ a reinterpretation process, with prompting details provided in Appendix E.

Large Language Models. We find prior work's effectiveness is often overestimated due to limited evaluation settings. They focus on an earlier-generation LLM (*e.g.,* `gpt-3.5-turbo`) as both user simulator and critic. To ensure more robust evaluation, we adopt a newer model (*i.e.,* `gpt-4o`), which applies stricter criteria by requiring resolution of the user's core issue rather than surface-level relief. Human evaluation further confirms `gpt-4o`'s closer alignment with human judgments. This demonstrates that our evaluation setting is more reliable (see Appendix C for details).

Method	$\lvert\mathcal{S}\rvert$	ESConv		ExTES		P4G		P4G+	
		SR↑	AT↓	SR↑	AT↓	SR↑	AT↓	SR↑	AT↓
Standard									
+ GPT-3.5-Turbo	0	0.4154	8.59	0.4923	8.15	0.8000	5.41	0.3667	7.78
+ GPT-4o	0	0.5583	8.13	0.6667	7.30	0.9375	4.07	0.4917	7.14
Proactive (Deng et al., 2023b)	8-16	0.2385	9.51	0.5615	8.24	0.9500	4.23	0.4333	7.35
+ MI-Prompt (Chen et al., 2023)	8-16	0.3769	8.93	0.6538	7.82	0.9083	4.18	0.3417	7.91
ProCoT (Deng et al., 2023b)	8-16	0.2231	9.51	0.6308	7.99	0.9583	4.15	0.4500	7.24
+ MI-Prompt (Chen et al., 2023)	8-16	0.3538	9.13	0.6692	7.58	0.9250	**3.66**	0.4333	7.44
PPDPP (Deng et al., 2024)	8-16	0.5077	8.16	0.6846	6.99	**0.9667**	4.41	–	–
AnE (Zhang et al., 2024b)	∞	0.5846	7.38	0.6462	6.93	0.9083	4.27	0.5333	6.78
ICL-AIF (Fu et al., 2023)	∞	0.5615	7.87	0.7154	7.37	0.8000	4.68	0.5083	**6.70**
PRINCIPLES (Ours)	∼100	0.7385	**6.36**	**0.8615**	**5.87**	0.9500	4.73	**0.5917**	7.15

Table 1: Comparison of performance across four proactive dialogue tasks via self-play simulations. Here, $|\mathcal{S}|$ denotes the size of the strategy set used in each setting.

4 Evaluation

4.1 Evaluation Setups

Evaluation Metrics. We use two core metrics commonly adopted in strategy planning: success rate and average turns. In addition, to more comprehensively evaluate, we include three automatic metrics: macro F1, weighted F1, and entropy.

- **Success Rate** (SR): Success is determined by whether the reward assigned after each turn exceeds a threshold.
- **Average Turns** (AT): The average number of turns across all episodes. This reflects how efficiently the agent can achieve the task goal.
- **Macro F1 Score** (F_m): Evaluates the model's alignment with human-annotated strategy labels, by averaging per-class F1 scores.
- **Weighted F1 Score** (F_w) : Computes the average of per-class F1 scores weighted by class frequency, mitigating imbalance.
- **Entropy** (H): This measures the diversity of predicted strategies, where higher entropy indicates lower bias toward specific strategies.

$$H = -\sum_{c=1}^{C} p_c \log p_c \tag{14}$$

Baselines. We first evaluate a standard agent that operates without explicit strategy guidance. We then compare against two types of baselines. **(i)** Using pre-defined strategies, such as Proactive (Deng et al., 2023b) and ProCoT (Deng et al., 2023b), which prompt the model to select from a limited set of strategies. In contrast, PPDPP (Deng et al., 2024) employs a lightweight external planner trained via supervised fine-tuning and reinforcement learning. **(ii)** Using open-ended strategies, such as Ask-an-Expert (Zhang et al., 2024b) and ICL-AIF (Fu et al., 2023), which dynamically generate strategies using LLMs as expert knowledge sources or via AI feedback. This allows us to assess our approach against baselines with different levels of coverage and flexibility (see Appendix F.2).

Datasets. We first evaluate on ESConv and P4G, then extend to more challenging settings with ExTES and P4G+. P4G+ extends the original task by incorporating (i) diverse personas, (ii) multiple organizations, and (iii) donation barriers (*e.g.*, financial constraints), yielding more realistic persuasion scenarios (see Appendix F.3 for details).

4.2 Results

In this section, we investigate our method's performance in addressing three key challenges in strategy planning: coverage, bias, and training.

RQ1. Can PRINCIPLES effectively expand strategy coverage? As shown in Table 1, we reveal distinct performance trends across tasks depending on the size of a strategy set. Notably, with the improved capabilities of `gpt-4o` over the previous model, even the standard agent without any explicit strategy use achieves competitive SR in most tasks. Interestingly, methods with open-ended settings (*i.e.*, AnE, ICL-AIF) achieve higher SR on average than baselines relying on a pre-defined strategy

Method	ESConv			P4G		
	F_m	F_w	H	F_m	F_w	H
Proactive	6.91	9.25	0.87	13.95	21.58	2.60
ProCoT	6.94	9.85	0.90	13.73	20.50	2.60
PPDPP	6.15	11.33	0.07	7.91	10.66	1.03
AnE	7.78	13.76	1.07	14.65	23.27	2.60
ICL-AIF	4.61	10.69	0.11	10.79	16.74	2.46
Ours	**10.52**	**17.67**	**1.21**	**14.96**	**24.30**	**2.67**

Table 2: Evaluation of strategy prediction performance and distributional diversity on ESConv and P4G.

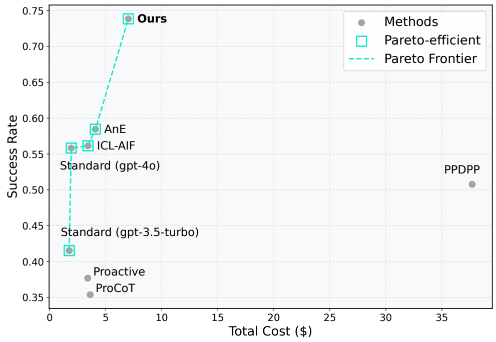


Figure 3: Cost-performance comparisons.

Method	ExTES		P4G+	
	SR↑	AT↓	SR↑	AT↓
PRINCIPLES (Ours)	**0.8615**	5.87	**0.5917**	7.15
w/o Structured	0.8385	**5.33**	0.5667	**6.39**
w/o Retrieval	0.7846	5.91	0.5833	6.62
w/o Reinterpretation	0.8385	5.84	0.5667	6.49

Table 3: Performance of our ablations.

set. Furthermore, even the standard baseline occasionally outperforms them. This suggests that constraining the model to a fixed set of strategies may hinder its capacity for proactive dialogue. A notable exception is P4G, where the task is focused on a single goal (*i.e.,* donating to Save the Children), making it possible for a concise strategy set to sufficiently cover the task. In contrast, we construct PRINCIPLES that efficiently expand the strategy coverage using only 50 offline self-play simulations on the training set. With these resources, our strategy planning consistently achieves superior performance across a range of tasks.

RQ2. Can strategy planning based on PRINCIPLES help mitigate strategy bias? To evaluate whether PRINCIPLES reduce preference bias in strategy planning, we assess strategy prediction performance on ESConv and P4G, where human-annotated strategy labels are available. To ensure fair comparison with other methods, we allow free-form strategy generation in open-ended baselines and subsequently map each strategy to the closest pre-defined label using `gpt-4o`.

In Table 2, our method shows the highest macro F1, weighted F1, and entropy, indicating that it not only selects contextually appropriate strategies but also avoids overusing preferred strategies. In contrast, baselines often exhibit strong bias, consistently relying on a few dominant strategies. For example, PPDPP heavily overfits one or two strategies during reinforcement learning, resulting in a highly skewed distribution (Appendix D). These findings are supported by our ablation studies in Table 3. When PRINCIPLES are extracted in an unstructured format, or when similarity-based retrieval is replaced by selecting via LLM, performance consistently degrades. These results highlight that the structured format (*i.e., "When… you should… rather than…"*) and retrieval process guide the model to select contextually relevant strategies instead of sticking to biased preferences, improving downstream performance.

RQ3. Can PRINCIPLES improve agents without training? Unlike prior methods that rely on finetuning (Deng et al., 2024; He et al., 2025b,a), we effectively address coverage and bias issues without any additional training. Figure 3 compares the training and inference costs of ours against baselines, including PPDPP, which involves SFT in human-labeled data and RL via 1,000 self-play simulations. As a result, PPDPP incurs a significantly higher training cost (*i.e.,* approximately 11.5× more than our method). In contrast, our approach requires no labeled data and performs only 50 offline simulations on the training set, maintaining a cost comparable to prompt-based baselines while consistently outperforming all of them. These findings highlight that the model already possesses extensive knowledge for strategy planning, and that unlocking and leveraging this ***hidden parametric knowledge*** is important to drive strong performance improvements rather than updating model parameters.

5 Human Evaluation

Evaluating Response Quality. Given the nature of self-play simulations, we evaluate responses generated from the same context (*i.e.,* first user utterance) to ensure a fair comparison across baselines,

Patient: For the past couple years, I've been dealing with depression. I'm lonely, sad, scared and worried. I feel like there's no reason to live.

PPDPP: (Providing suggestions) I **suggest** reaching out to a mental health professional for support and exploring activities that bring you ...
AnE: (Provide a safe space for the patient to express their feelings) I'm **really sorry to hear that you're feeling** this way, and I'm here to listen ...

PRINCIPLES

When the patient expresses feelings of depression and loneliness, **you should** validate their emotions and collaborate on exploring sources of support and potential coping strategies **because** this combination of emotional acknowledgment and practical exploration can help ...

Ours: It sounds like **you're going through a really tough time**, and I want to **acknowledge** how difficult these feelings must be for you; **let's explore some ways** to create a daily routine that includes activities for social interaction and self-care to help you feel more connected.

Figure 4: Qualitative example comparing AnE, PPDPP, and our approach based on PRINCIPLES.

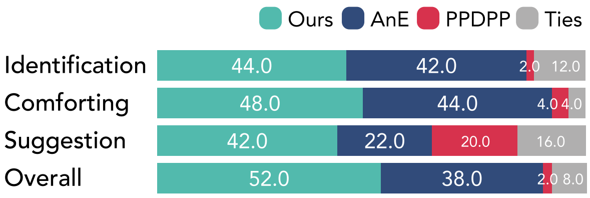


Figure 5: Human evaluation of response quality.

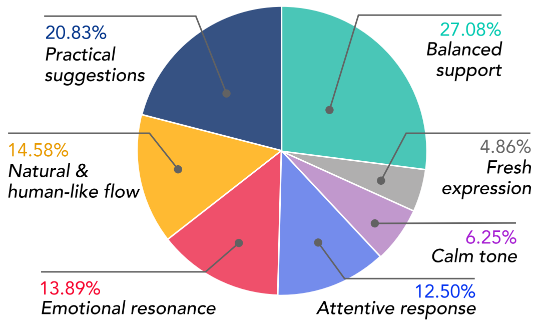


Figure 6: Analysis of human preference for our method.

instead of comparing full dialogues that may vary in length and flow. We recruit three annotators to evaluate the quality of generated responses on 50 randomly sampled dialogue contexts from the ExTES, comparing outputs from three methods (*i.e.*, AnE, using open-ended strategies; PPDPP, using pre-defined strategies; and Ours). To reduce position bias, all responses are presented to each annotator in shuffled order. In this setting, we conduct the evaluation based on four evaluation criteria (Liu et al., 2021): (1) **Identification**: Which response explored the patient's situation more in depth and was more helpful in identifying their problems? (2) **Comforting**: Which response was more skillful in comforting the patient? (3) **Suggestion**: Which response gave more helpful suggestions for the patient's problems? (4) **Overall**: Generally, which response's emotional support do you prefer?

As shown in Figure 5, our method consistently outperforms both baselines across all four criteria. This result can be further interpreted in conjunction with Figure 4. We observe that AnE tends to overly reflect the patient's feelings without addressing the core issue, while PPDPP is biased toward providing suggestions, leading to interactions that repeat similar utterances without adjusting to the conversational flow. In contrast, our method achieves a more effective balance between emotional empathy and problem-solving by retrieving contextually appropriate strategies. This results in responses that align more closely with human preferences.

Analysis of Human Preference. To gain deeper insight into human preferences, we asked participants to select reasons for their choice into several thematic categories. As shown in Figure 6, our approach tended to combine logical coherence and emotional empathy (*i.e.*, Balanced Support).

6 In-Depth Analysis

Comparison with Strong Baseline. To further validate our approach, we compare it with DPDP (He et al., 2024), a strong baseline using Monte Carlo Tree Search (MCTS)-based strategy selection at each turn. As shown in Table 4, our method achieves competitive SR while attaining better performance in AT. In terms of efficiency, DPDP requires 18.07× higher training cost than our principles construction, along with 3.07× higher inference cost and 2.67× longer inference time. While DPDP achieves slightly higher SR, its substantial computational overhead limits practicality, whereas our method provides a more efficient and scalable alternative for real-world applications.

Method	$\lvert\mathcal{S}\rvert$	SR↑	AT↓	$\text{Cost}_{\text{train}}$	$\text{Cost}_{\text{infer}}$	$\text{Time}_{\text{infer}}$
DPDP	8–16	**0.7923**	7.52	\$59.44	\$16.29	81.4s
Ours	∼100	0.7385	**6.36**	**\$3.29**	**\$5.30**	**30.5s**

Table 4: Comparison between DPDP and our method on the ESConv dataset.

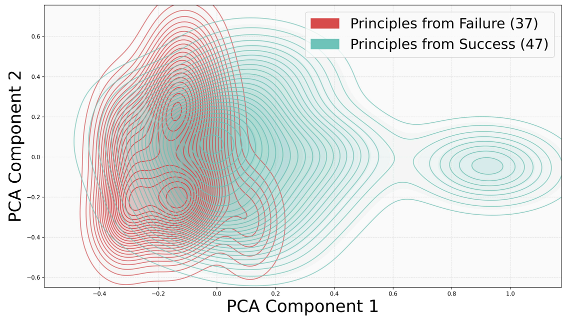


Figure 7: PCA projection of PRINCIPLES derived from successful and failed interactions. The distributions indicate that both contribute complementary strategic coverage.

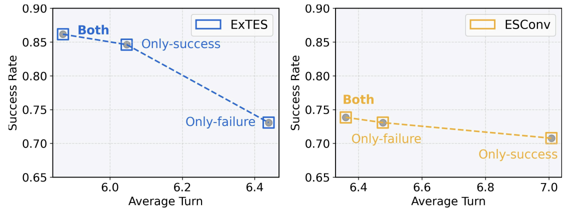


Figure 8: Comparison of performance using PRINCIPLES derived from success only, failure only, or both.

Learning from Success, Failure, or Both. Figure 7 illustrates the effect of PRINCIPLES extracted from successful and failed interactions—an essential component of our method. We project the embedding vectors of these PRINCIPLES into a 2D space using Principal Component Analysis (PCA). While some overlap exists, each region clearly possesses its own distinct area of focus. Additionally, in Figure 8, we evaluate the practical impact of these two types of resources. On ExTES, PRINCIPLES derived from successful interaction lead to better outcomes than those extracted from failed ones, while the opposite is observed on ESConv. This indicates that neither source demonstrates consistent superiority. Instead, the integration of both types of PRINCIPLES yields the best performance, demonstrating the broader strategy coverage.

Diversity of Linguistic Organization Patterns. We further analyze the diversity of linguistic organization patterns used in principle construction. Interestingly, as shown in Table 5, we observe that removing even a single component from our original pattern results in a substantial performance drop (*i.e.,* `w/o rather than` and `w/o because`). In contrast, an alternative format that retains all four components achieves slightly better SR (*i.e.,* `If/then/instead of/in order to`), despite its surface-level differences. These findings suggest

Linguistic Format	ESConv SR↑	ESConv AT↓
PRINCIPLES	0.7385	**6.36**
`w/o rather than`	0.6231	7.20
`w/o because`	0.6400	7.21
`If/then/instead of/in order to`	**0.7538**	6.52

Table 5: Comparison of performance across different linguistic formats on ESConv.

that the effectiveness of PRINCIPLES derives not from their superficial linguistic form but from the presence of four core informational elements: **the situation, the successful strategy, the failed strategy, and the rationale**. This insight provides a valuable design guideline for future research.

Online Construction Setting. While our framework primarily relies on offline construction, we also investigate an online construction setting (*i.e.,* at inference) on ESConv dataset. In this setting, the agent derives PRINCIPLES only from successful interactions, since test-time interaction does not allow revisiting failed trajectories (*i.e.,* no backtracking). As shown in Table 6, online construction yields a moderate performance drop compared to the offline setting, yet still produces highly competitive results. This demonstrates that even without pre-constructed principles, our framework can rapidly adapt and expand its strategy coverage during deployment. Such adaptability is particularly important for real-world scenarios, where agents must continuously learn from limited successes without relying on offline self-play simulations.

Method	ESConv SR↑	ESConv AT↓
Offline Construction	**0.7385**	**6.36**
Online Construction (*i.e.,* at inference)	0.6615	7.22

Table 6: Comparison of performance on ESConv under offline and online principle construction settings.

Impact of Source Models. We further analyze how different LLMs affect PRINCIPLES quality. As shown in Table 7, `Claude`- and `Llama`-based PRINCIPLES achieve higher SR, which correlates with their length, but not with utterance length, suggesting that more detailed guidance is beneficial. Importantly, Token_p and Token_u exhibit different orderings, indicating that the key factor in performance stems from specificity rather than verbosity.

Source Model	ESConv			
	SR↑	AT↓	Tokens$_p$	Tokens$_u$
GPT-4o	0.7385	<u>6.36</u>	62.93	23.11
Claude-3.7-Sonnet	<u>0.7462</u>	6.70	74.34	28.48
Llama-3.1-8B	**0.8615**	**5.44**	83.34	22.93

Table 7: Comparison of performance on ESConv with PRINCIPLES derived from different models. Tokens$_p$ denotes the average token length of the extracted PRINCIPLES, while Tokens$_u$ denotes the average token length of the agent's utterances, guided by these PRINCIPLES.

Effect of Simulation Budget. We investigate the optimal number of offline self-play simulations for effective principles construction. On both datasets, ExTES and P4G⁺, we conduct 25, 50, 75, and 100 self-play simulations and measure SR and AT. As shown in Figure 9, even 25 simulations yield substantial improvement, reaching its optimum at 50. However, the performance declines beyond 75, suggesting that principles that exceed the task's strategic requirements introduce additional noise that ultimately hinders overall model behavior.

Impact of Top-k Strategies. We further investigate how the number of retrieved principles (k) directly influences performance. As shown in Figure 10, performance varies with different values of k, with the optimal number depending on the task (*i.e.,* 9 for ESConv, 3 for P4G). Notably, retrieving only a few top-ranked principles (*e.g.,* top-3) still provides a highly competitive and cost-effective alternative in practical resource-constrained settings.

7 Related Work

Recent studies have increasingly highlighted the importance of strategy planning in proactive dialogue, such as emotional support (Liu et al., 2021) and persuasion (Wang et al., 2019), where an agent should take the initiative to achieve a specific goal. To this end, many approaches rely on external planners to select the most suitable choice from a predefined strategy set (Deng et al., 2023b). In contrast to methods that depend on a model's parametric knowledge, Deng et al. (2024) leverages human-annotated strategies via supervised fine-tuning (SFT), and further improves performance through reinforcement learning in simulated environments. Building upon prior work, Zhang et al. (2024a) and He et al. (2025b) incorporate user modeling to enable more tailored strategy selection. Other approaches (Yu et al., 2023; He et al., 2024)

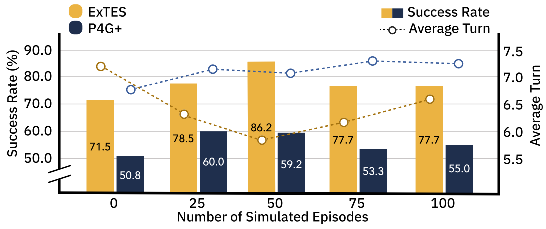


Figure 9: Correlation between a number of simulations and success rate.

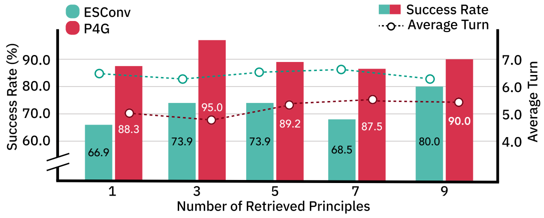


Figure 10: Correlation between a number of retrieved PRINCIPLES and success rate.

formulate strategy planning as a search problem, using Monte Carlo Tree Search (MCTS) to identify optimal strategies. Another line of research focuses on open-ended strategy generation. For instance, Fu et al. (2023) prompts LLMs to elicit improved strategies via iterative AI feedback, while Zhang et al. (2024b) use LLMs as a source of expert knowledge. He et al. (2025a) attempts to eliminate the dependency on simulation environments by discovering latent policies from dialogue (Louie et al., 2024). Yet, they have several limitations, including limited coverage, preference bias, and costly training. This motivates the need of alternative approaches, which we introduce in our work.

8 Conclusion

In this paper, we introduce 📚**PRINCIPLES**, a synthetic strategy memory for proactive dialogue agents, derived through offline self-play simulations. Through extensive experiments across both standard and extended datasets, we show that our approach effectively expands strategy coverage and mitigates preference bias, leading to more balanced strategy planning. Furthermore, our approach achieves promising performance without additional training by explicitly uncovering the hidden parametric knowledge into a structured, non-parametric form. We expect our novel approaches to serve as a new foundation for future research efforts towards proactive dialogue agents.

Limitations

First, our retrieval relies on embedding similarity over the `When` clause with the current state based on L2 distance, which may overlook subtle contextual nuances. Although our strategy planning approach based on PRINCIPLES includes a reinterpretation step to adapt retrieved principles to unseen scenarios, the selected principles may still fall short in highly specific or ambiguous dialogue situations, where fine-grained contextual understanding is required. One possible solution is to refine the scoring mechanism by combining embedding similarity with additional relevance signals, such as a dialogue stage, to improve retrieval accuracy beyond surface-level similarity.

Second, although our method supports turn-level planning via principle retrieval, it lacks explicit modeling of long-term goals. As a result, the agent may over-optimize for short-term reward, leading to suboptimal outcomes in tasks that demand long-term strategic planning, such as negotiations (He et al., 2018). Constructing principles from full dialogue trajectories may enhance long-term coherence, which we view as a promising direction to further develop our framework.

Ethical Considerations

Human Annotation Process. We recruit three undergraduate students with high English proficiency to evaluate the response quality. Evaluations are conducted on 50 randomly sampled dialogue contexts from the ExTES benchmark, covering outputs from three methods (AnE, PPDPP, and Ours). To guide consistency, we provide annotators with a custom-designed annotation form and detailed evaluation instructions (Figure 12). To reduce position bias, all responses are shown in randomized order for each annotator.

LLM Usage and Ethical Risk. We acknowledge that some of the datasets used (e.g., ESConv, ExTES) involve emotional support scenarios, and that the principles in our method are not curated by human experts. However, we do not intend to make any clinical or therapeutic claims regarding these principles. Our approach is intended purely as a research framework for exploring reusable dialogue strategies in simulation. Before any real-world deployment, especially in sensitive domains such as mental health, expert review and safety validation would be essential.

Acknowledgments

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. RS-2020-II201361, Artificial Intelligence Graduate School Program (Yonsei University)), (No. RS-2024-00457882, National AI Research Lab Project), (2022-0-00077, RS-2022-II220077, AI Technology Development for Commonsense Extraction, Reasoning, and Inference from Heterogeneous Data). Jinyoung Yeo is the corresponding author.

References

Maximillian Chen, Xiao Yu, Weiyan Shi, Urvi Awasthi, and Zhou Yu. 2023. Controllable mixed-initiative dialogue generation through prompting. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 951–966, Toronto, Canada. Association for Computational Linguistics.

Robert B Cialdini and Robert B Cialdini. 2007. *Influence: The psychology of persuasion*, volume 55. Collins New York.

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

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

Yang Deng, Wenxuan Zhang, Wai Lam, See-Kiong Ng, and Tat-Seng Chua. 2024. Plug-and-play policy planner for large language model powered dialogue agents. In *ICLR*.

Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. 2025. The faiss library.

Amy C Edmondson. 2011. Strategies for learning from failure. *Harvard business review*, 89(4):48–55.

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

Igor Grossmann. 2017. Wisdom in context. *Perspectives on psychological science*, 12(2):233–257.

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.

Tao He, Lizi Liao, Yixin Cao, Yuanxing Liu, Ming Liu, Zerui Chen, and Bing Qin. 2024. Planning like human: A dual-process framework for dialogue planning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4768–4791, Bangkok, Thailand. Association for Computational Linguistics.

Tao He, Lizi Liao, Yixin Cao, Yuanxing Liu, Yiheng Sun, Zerui Chen, Ming Liu, and Bing Qin. 2025a. Simulation-free hierarchical latent policy planning for proactive dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 24032–24040.

Tao He, Lizi Liao, Ming Liu, and Bing Qin. 2025b. Simulating before planning: Constructing intrinsic user world model for user-tailored dialogue policy planning. *arXiv preprint arXiv:2504.13643*.

Dongjin Kang, Sunghwan Kim, Taeyoon Kwon, Seungjun Moon, Hyunsouk Cho, Youngjae Yu, Dongha Lee, and Jinyoung Yeo. 2024. Can large language models be good emotional supporter? mitigating preference bias on emotional support conversation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15232–15261, Bangkok, Thailand. Association for Computational Linguistics.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards emotional support dialog systems. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3469–3483.

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.

Dang Nguyen, Viet Dac Lai, Seunghyun Yoon, Ryan A Rossi, Handong Zhao, Ruiyi Zhang, Puneet Mathur, Nedim Lipka, Yu Wang, Trung Bui, et al. 2024. Dynasaur: Large language agents beyond predefined actions. *arXiv preprint arXiv:2411.01747*.

Kai Tzu-iunn Ong, Namyoung Kim, Minju Gwak, Hyungjoo Chae, Taeyoon Kwon, Yohan Jo, Seungwon Hwang, Dongha Lee, and Jinyoung Yeo. 2025. Towards lifelong dialogue agents via timeline-based memory management. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 8631–8661.

Qingyue Wang, Liang Ding, Yanan Cao, Yibing Zhan, Zheng Lin, Shi Wang, Dacheng Tao, and Li Guo. 2023. Divide, conquer, and combine: Mixture of semantic-independent experts for zero-shot dialogue state tracking. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2048–2061, Toronto, Canada. Association for Computational Linguistics.

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

Jing Xu, Arthur Szlam, and Jason Weston. 2022. Beyond goldfish memory: Long-term open-domain conversation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5180–5197.

Xiao Yu, Maximillian Chen, and Zhou Yu. 2023. Prompt-based monte-carlo tree search for goal-oriented dialogue policy planning. In *The 2023 Conference on Empirical Methods in Natural Language Processing*.

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

Tong Zhang, Chen Huang, Yang Deng, Hongru Liang, Jia Liu, Zujie Wen, Wenqiang Lei, and Tat-Seng Chua. 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, Miami, Florida, USA. Association for Computational Linguistics.

Xuan Zhang, Yang Deng, Zifeng Ren, See-Kiong Ng, and Tat-Seng Chua. 2024b. Ask-before-plan: Proactive language agents for real-world planning. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 10836–10863, Miami, Florida, USA. Association for Computational Linguistics.

Zhonghua Zheng, Lizi Liao, Yang Deng, and Liqiang Nie. 2023. Building emotional support chatbots in the era of llms. *arXiv preprint arXiv:2308.11584*.

A Algorithm for PRINCIPLES

The pseudo algorithm for principles construction is provided in Algorithm 1.

Method	Critic Model	ESConv		ExTES		P4G		P4G+	
		SR↑	AT↓	SR↑	AT↓	SR↑	AT↓	SR↑	AT↓
Standard	GPT-3.5-Turbo	0.9154	4.52	0.9846	3.63	0.8583	4.73	0.5667	6.55
	GPT-4o	0.5583	8.13	0.6667	7.30	0.9375	4.07	0.4917	7.14

Table 8: Comparison of performance between `gpt-3.5-turbo` and `gpt-4o` when used as critic models.

Human Rating	Count
A. No, the Patient feels worse.	1
B. No, the Patient feels the same.	8
C. No, but the Patient feels better. (GPT-4o's judgment)	**54**
D. Yes, the Patient's issue has been solved. (GPT-3.5's judgment)	7

Table 9: Human ratings with model judgments.

B Qualitative Examples

We present qualitative examples to illustrate how our method improves dialogue quality. In particular, we show how the agent generates contextually appropriate, balanced responses by leveraging structured PRINCIPLES. Representative examples are shown in Table 10 and 11.

C Impact of Evaluator Choice

Comparison of Performance Under Different Evaluators. We analyze the impact of evaluator choice on reported performance. As shown in Table 8, the use of `gpt-3.5-turbo` as the critic yields substantially higher SR, suggesting that it provides overly lenient evaluations. Our analysis indicates that `gpt-4o` applies stricter criteria for goal completion. Beyond superficial signs of emotional relief, it requires a concrete resolution of the user's core issue. In contrast, `gpt-3.5-turbo` frequently judges success based on surface-level cues such as positive sentiment or task-related phrases (*e.g.,* "I'll consider making a donation" or "I hope things get better"). This discrepancy explains why prior studies that rely on `gpt-3.5-turbo` report higher SR and lower AT. Under `gpt-4o`, success requires deeper exploration of the user's core concern, resulting in lower SR but higher AT. Qualitative examples are provided in Table 12 and 13.

Human Validation of Stricter Evaluation. For deeper insight into the reliability of `gpt-4o`'s stricter evaluation, we conduct a human evaluation on 70 cases where `gpt-3.5-turbo` and `gpt-4o` produced conflicting judgments. Each case is annotated by three independent workers recruited via Amazon Mechanical Turk, who followed the same evaluation criteria as the LLMs. Final labels were determined by majority vote. As shown in Table 9, in 54 out of 70 cases, human judgments align with `gpt-4o`, while only 7 cases align with `gpt-3.5-turbo`. This finding suggests that `gpt-4o`'s stricter evaluations are better aligned with human judgment and more reliable.

D Strategy Bias and Distribution Analysis

We provide additional implementation details of Table 2 and a deeper analysis of strategy distribution through the visualization in Figure 11a and 11b. Furthermore, we observe the mitigation of strategy bias in case of utilizing PRINCIPLES.

D.1 Implementation Details

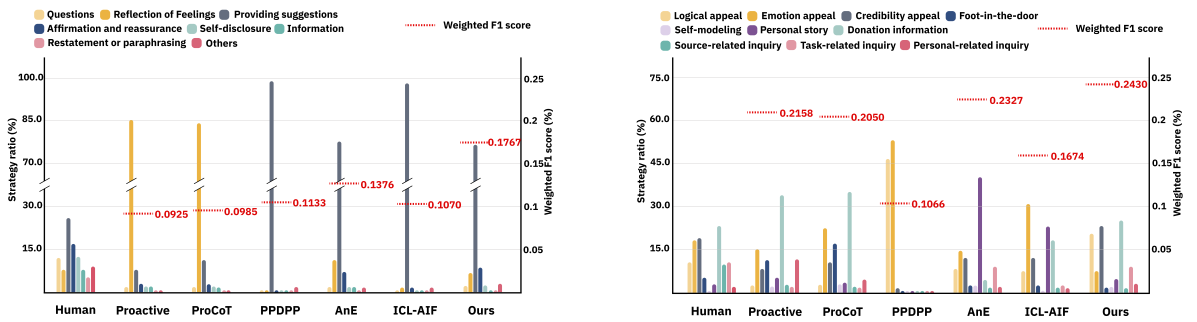
For fair comparison across methods, we apply tailored processing to each mechanism.

Pre-defined Strategies. For methods such as Proactive, ProCoT, and PPDPP, which select the most appropriate strategy from a small pre-defined strategy set, we directly use their predicted strategy labels.

Open-ended Strategies and Ours. Open-ended methods like AnE and ICL-AIF generate free-form strategies, while our method selects from PRINCIPLES. To assign labels for evaluation, we prompt `gpt-4o` to map each predicted strategy to the closest pre-defined strategy.

D.2 Strategy Preference Distribution Analysis

Figure 11a and 11b illustrate the distribution of selected strategies across methods in the emotional support and persuasion domains, respectively, along with their corresponding weighted



(a) Emotional support dialogues

(b) Persuasion dialogues

Figure 11: The details of LLMs' strategy distribution in (a) emotional support and (b) persuasion. The bars represent the strategy ratio (%), and the red dashed lines indicate the normalized weighted F1 scores reported in Table 2.

F1 scores. We observe that several baseline methods demonstrate strong reliance on a narrow subset of strategies, with one dominant strategy exceeding 80% of usage. Especially, PPDPP exhibits the most severe strategy preference bias in both the emotional support and persuasion domains, which suggests a limitation of training-based approaches. In contrast, our method show relatively uniform preferences, leading to robust performance.

E Prompts for 📚PRINCIPLES

We list the major prompts used throughout our system. Each prompt corresponds to a specific stage in the pipeline.

- **Strategy Planning during Offline Self-play Simulation**: To construct PRINCIPLES from self-play, we follow ICL-AIF (Fu et al., 2023) and use an open-ended prompting format to generate high-level strategies at each turn. Unlike ICL-AIF, we generate a single strategy at each turn (Figure 13 and 14).

- **Strategy Revision**: When the reward does not improve, the turn is considered a failure. To obtain a revised strategy, we prompt the model with the dialogue history and previously failed trials, guiding it to generate a better alternative (Figure 15).

- **Principle Derivation from Success**: When a response is deemed successful (*e.g.,* resolving a user's emotional distress), we derive a reusable principle by prompting the model to analyze the dialogue context and explain why the strategy worked and express it in a structured format—highlighting the situation, successful strategy, and reasoning (Figure 16).

- **Principle Derivation from failure**: When a previous turn is revised, we derive a principle by prompting the model to compare the successful strategy with failed ones and explain why it was more effective in a structured format. This comparison-based principle promotes refinement and reuse of strategies that overcome prior mistakes (Figure 17).

- **Reinterpretation**: When retrieved principles may not perfectly align with the current dialogue context, we prompt the LLM to reinterpret the top-k retrieved principles based on the current dialogue state. This step ensures that strategies are adapted to specific conversational nuances, enabling better contextual alignment (Figure 18).

F Details on Evaluation Setups

F.1 Self-play Simulations

Agent Simulator. The dialogue agent generates responses based on a strategy. In emotional support, the agent acts as a therapist assisting a patient (Figure 19); in persuasion, as a persuader encouraging donations (Figure 20).

User Simulator. LLMs serve as user simulators, responding to the agent's utterances. In emotional support, they act as patient (Figure 21); in persuasion, as persuadee asked to donate (Figure 22).

Critic Model. To assess goal completion, we prompt `gpt-4o` (temperature $\tau = 1.0$) to generate verbal feedback aligned with each task objective. Emotional support focuses on reducing distress, and persuasion on elicit a donation. We use a four level scale for each (Figures 23 and 24).

Following (Deng et al., 2024; He et al., 2025a), we map outcomes to scalar rewards: emotional support — [worse, same, better, solved] → [-1, -0.5, 0.5, 1]; persuasion — [refused, neutral, positive, donate] → the same. Simulations (130 emotional support, 120 persuasion episodes) run until a terminal state is reached. A dialogue is marked GOAL-COMPLETED if the average score over 10 critic runs exceeds a threshold of $\eta = 0.5$.

F.2 Baselines

Standard prompts the LLM to engage in self-play conversations without providing explicit strategies.

Proactive (Deng et al., 2023b) prompts the LLM to select the most appropriate strategy for the next turn from a pre-defined set before generating a response (Figure 25).

ProCoT (Deng et al., 2023b) prompts the LLM to first analyze the dialogue progress using Chain-of-Thought prompting. Based on this analysis, the model selects an appropriate strategy from a pre-defined set (Figure 26)

MI-Prompt (Chen et al., 2023) converts strategies into a natural language prompt, improving response quality by explicitly guiding the model's behavior. Examples are shown in Table 14, 15, 16, and 17.

PPDPP (Deng et al., 2024) trains an external planner to select the most appropriate strategy. Using human-annotated dialogue datasets, PPDPP first facilitates supervised fine-tuning (SFT). Following SFT, PPDPP utilizes a reinforcement learning with goal-oriented AI feedback during self-play simulations. Verbal feedbacks are converted into scalar rewards, which are used to optimize the planner.

ICL-AIF (Fu et al., 2023) prompts the LLM to generate suggestions in open-ended manner, rather than using pre-defined strategies. ICL-AIF produces three dialogue-level feedbacks (Figure 27).

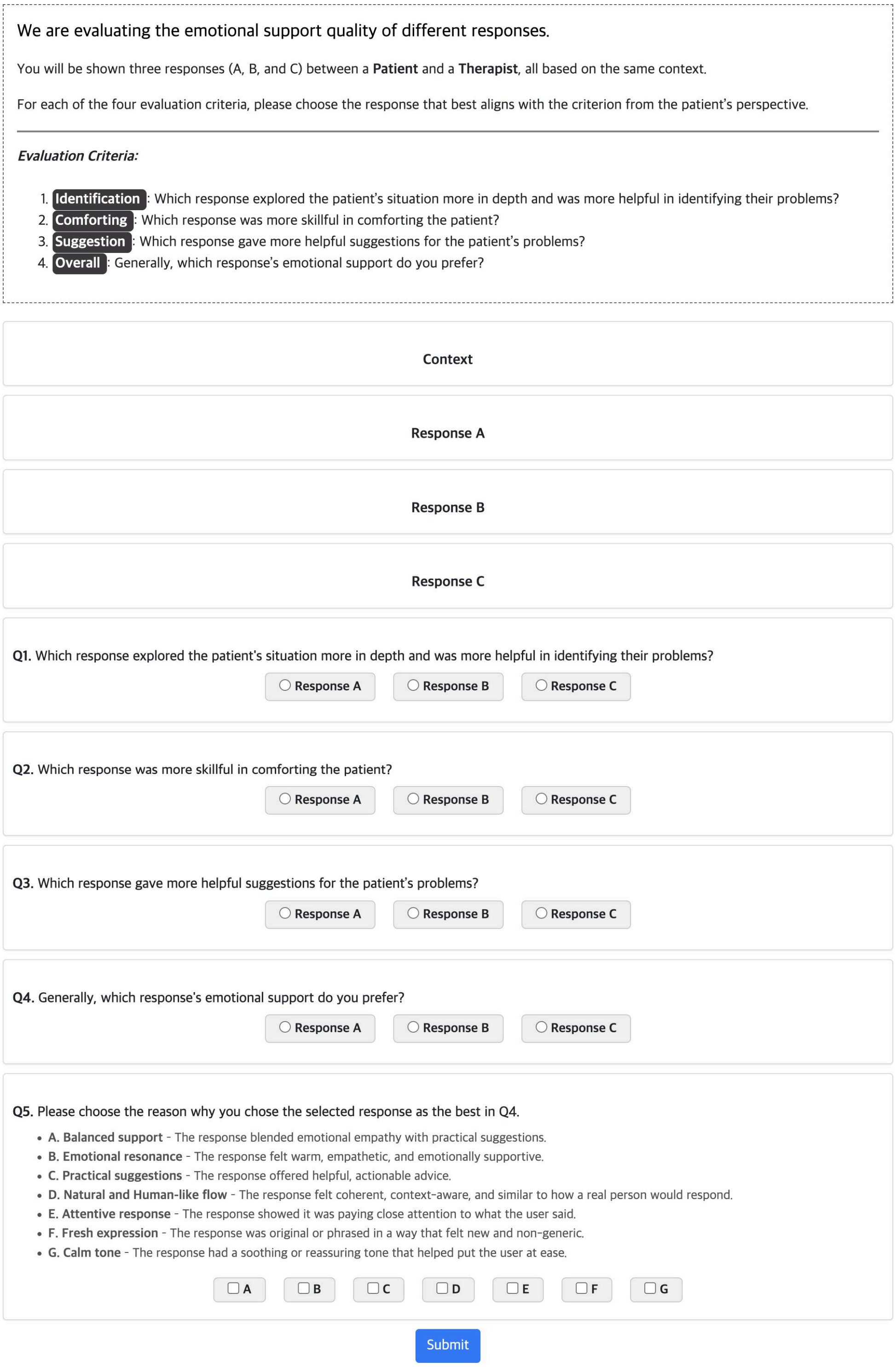
Ask-an-Expert (AnE) (Zhang et al., 2024b) leverages the LLM as source of expert knowledge. AnE first identifies the user's emotional state, infers the reason behind it, and finally determines the appropriate strategy the agent should take (Figure 28).

F.3 P4G$^+$ Generation

Persona Generation. By incorporating more diverse and realistic persona, user simulator more effectively imitate the complexity of real-world interactions. To generate diverse and exclusive personas, independently generated `{name}` and `{occupation}` are provided (Figure 29).

P4G$^+$ Generation. As P4G dataset lacks diversity, we construct an extended version, P4G$^+$, which incorporates a broader range of organizations, personas, initial context, and donation barriers. To prevent the overlap within the generated datasets, organizations and personas are generated independently (Figure 30).

P4G$^+$ Evaluation. To ensure a fair comparison, we retain the original 10 strategies from P4G and supplement 6 more strategies adapted from the book Influence (Cialdini and Cialdini, 2007).


We are evaluating the emotional support quality of different responses.

You will be shown three responses (A, B, and C) between a **Patient** and a **Therapist**, all based on the same context.

For each of the four evaluation criteria, please choose the response that best aligns with the criterion from the patient's perspective.

Evaluation Criteria:

1. **Identification**: Which response explored the patient's situation more in depth and was more helpful in identifying their problems?
2. **Comforting**: Which response was more skillful in comforting the patient?
3. **Suggestion**: Which response gave more helpful suggestions for the patient's problems?
4. **Overall**: Generally, which response's emotional support do you prefer?

Context

Response A

Response B

Response C

Q1. Which response explored the patient's situation more in depth and was more helpful in identifying their problems?

○ Response A ○ Response B ○ Response C

Q2. Which response was more skillful in comforting the patient?

○ Response A ○ Response B ○ Response C

Q3. Which response gave more helpful suggestions for the patient's problems?

○ Response A ○ Response B ○ Response C

Q4. Generally, which response's emotional support do you prefer?

○ Response A ○ Response B ○ Response C

Q5. Please choose the reason why you chose the selected response as the best in Q4.

- **A. Balanced support** - The response blended emotional empathy with practical suggestions.
- **B. Emotional resonance** - The response felt warm, empathetic, and emotionally supportive.
- **C. Practical suggestions** - The response offered helpful, actionable advice.
- **D. Natural and Human-like flow** - The response felt coherent, context-aware, and similar to how a real person would respond.
- **E. Attentive response** - The response showed it was paying close attention to what the user said.
- **F. Fresh expression** - The response was original or phrased in a way that felt new and non-generic.
- **G. Calm tone** - The response had a soothing or reassuring tone that helped put the user at ease.

☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G

Submit

Figure 12: Interface for human evaluation

Algorithm 1 PRINCIPLES Construction

Notation: E: total episodes, $\mathcal{P}$: principle set, $\mathcal{F}_t$: failure history at t, $\mathcal{T}_t$: (σ_t, a_t, u_t)
Initialize $\mathcal{P} \leftarrow \emptyset$
for $e \in$ `sample_episodes`(E) **do**
    $s_t \leftarrow \emptyset$ `// this is a comment`
    `/* Begin self-play simulation */`
    while not terminal(s_t) **do**
        $\sigma_t \leftarrow \mathrm{LLM}_\theta(\rho_\sigma,\, s_t)$; $a_t \leftarrow \mathrm{LLM}_\theta(\rho_a,\, \sigma_t,\, s_t)$; $u_t \leftarrow \mathrm{LLM}_\theta(\rho_u,\, s_t,\, a_t)$
        $r_t \leftarrow \frac{1}{l}\sum_{i=1}^{l} f\big(\mathrm{LLM}_\theta^{(i)}(\rho_r,\, s_t,\, a_t,\, u_t)\big)$
        if $r_t > r_{t-1}$ **then**
            `/* Detect success and extract the PRINCIPLES. */`
            $p_t \leftarrow \mathrm{LLM}_\theta(\rho_\pi,\, s_t,\, \mathcal{T}_t)$
            $\mathcal{P} \leftarrow \mathcal{P} \cup \{p_t\}$; $s_t \leftarrow s_t \cup \{a_t,\, u_t\}$
        else
            `/* Detect failure and begin strategy revision. */`
            $\mathcal{F}_t \leftarrow \emptyset$
            while $r'_t \leq r_{t-1}$ **and** $|\mathcal{F}_t| < n_{\max}$ **do**
                `/* Re-simulation via backtracking. */`
                $\sigma'_t \leftarrow \mathrm{LLM}_\theta(\rho_c,\, s_t,\, \mathcal{F}_t)$; $a'_t \leftarrow \mathrm{LLM}_\theta(\rho_a,\, \sigma'_t,\, s_t)$; $u'_t \leftarrow \mathrm{LLM}_\theta(\rho_u,\, s_t,\, a'_t)$
                $r'_t \leftarrow \frac{1}{l}\sum_{i=1}^{l} f\big(\mathrm{LLM}_\theta^{(i)}(\rho_r, s_t, a'_t, u'_t)\big)$
                $\mathcal{F}_t \leftarrow \mathcal{F}_t \cup \{(\sigma'_t, a'_t, u'_t)\}$
            end while
            if $r'_t > r_{t-1}$ **then**
                `/* Extract the PRINCIPLES from the revision process. */`
                $(\sigma^*_t, a^*_t, u^*_t) \leftarrow (\sigma'_t, a'_t, u'_t)$
                $\tilde{p}_t \leftarrow \mathrm{LLM}_\theta(\rho_\psi,\, s_t,\, \mathcal{T}^*_t,\, \mathcal{F}_t)$
                $\mathcal{P} \leftarrow \mathcal{P} \cup \{\tilde{p}_t\}$; $s_t \leftarrow s_t \cup \{a'_t,\, u'_t\}$
            else
                $s_t \leftarrow s_t \cup \{a_t,\, u_t\}$
            end if
        end if
    end while
end for
return $\mathcal{P}$

BACKGROUND INFORMATION

Problem Type: Communication Challenges
Situation: I've been feeling really down lately because I've been having a lot of arguments with my best friend. We used to be so close, but now we can't seem to agree on anything and it's really affecting our friendship.

[CONTEXT]
Patient: I've been feeling really down lately because I've been having a lot of arguments with my best friend. We used to be so close, but now we can't seem to agree on anything and it's really affecting our friendship.

[Retrieved] **When** the patient plans to express their feelings in a message and desires a constructive dialogue with a friend, **you should** guide them to explore and identify the specific emotions they want to convey and how these emotions might aid in rebuilding the connection **rather than** suggesting preparatory actions such as writing exercises or mindfulness techniques, **because** exploring and articulating specific emotions creates a more empathetic dialogue and enhances the authenticity and effectiveness of the communication.

[Reinterpreted] **When** the patient is experiencing a conflict with a close friend and wishes to improve the situation, **you should** guide them to explore and identify the specific emotions driving their arguments and how acknowledging these emotions might aid in resolving the conflict **rather than** suggesting avoidance techniques or temporary distractions, **because** exploring and articulating specific emotions fosters greater understanding and can lead to a more authentic and harmonious friendship.

Therapist: It sounds like these arguments are causing you a lot of distress; can you identify the specific emotions you're experiencing during these conflicts?
Patient: I often feel frustrated, misunderstood, and anxious during our arguments.

[Retrieved] **When** the patient is nervous about discussing their feelings with a friend and worried about not expressing themselves clearly, **you should** suggest they practice the conversation beforehand, perhaps by role-playing or writing down key points, **rather than** simply encouraging them to approach the conversation with honesty and empathy, **because** providing a concrete method to organize thoughts and practice can build their confidence and improve communication.

[Reinterpreted] **When** the patient is experiencing distress due to conflicts with a friend and is feeling frustrated, misunderstood, and anxious during arguments, **you should** suggest they practice pinpointing these emotions and explore ways to express them clearly during conversations, perhaps through journaling or rehearsing key points, **rather than** simply encouraging them to communicate openly with their friend, **because** providing a structured approach to recognize and express emotions can enhance understanding and reduce anxiety in future interactions.

Therapist: Consider practicing deep breathing or mindfulness exercises before and during your conversations to help manage these emotions and stay calm.
Patient: That sounds like a good idea; I'll give it a try to see if it helps me stay more centered.

To be continued

[Retrieved] **When** the patient expresses anxiety about upcoming conversations, especially after practicing mindfulness exercises, **you should** suggest role-playing the conversation to practice expressing their thoughts and emotions, enhancing their confidence and reducing anxiety **rather than** simply encouraging self-compassion and patience **because** active engagement in role-play can help develop communication skills and emotional readiness, providing a concrete method to prepare for challenging interactions. [Reinterpreted] **When** the patient experiences ongoing conflict in a significant friendship, particularly after adopting strategies like mindfulness to manage emotions, **you should** suggest exploring underlying issues contributing to the disagreements to foster understanding and empathy **rather than** simply relying on emotional regulation techniques **because** addressing the root causes of misunderstandings can improve communication and rebuild trust, providing a tangible path to resolving interpersonal challenges. **Therapist**: Great, and as you practice these techniques, it might also be helpful to explore any underlying issues contributing to the disagreements with your friend. **Patient**: I think some of it might be due to differences in our expectations and communication styles.
[Retrieved] **When** the patient expresses anxiety about upcoming conversations, especially after practicing mindfulness exercises, **you should** suggest role-playing the conversation to practice expressing their thoughts and emotions, enhancing their confidence and reducing anxiety **rather than** simply encouraging self-compassion and patience **because** active engagement in role-play can help develop communication skills and emotional readiness, providing a concrete method to prepare for challenging interactions. [Reinterpreted] **When** the patient reflects on disagreements with someone close and starts practicing mindfulness exercises, **you should** suggest discussing their communication styles and expectations directly with the other person to address and resolve any underlying issues, enhancing clarity and understanding **rather than** simply providing coping strategies **because** direct communication can help bridge gaps in expectations and improve relationship dynamics, offering a proactive approach to mending strained connections. **Therapist**: Discussing these differences directly with your friend could help bridge gaps in expectations and improve your relationship dynamics. **Patient**: I agree, and I think setting aside time for an open and honest conversation might help us understand each other better.
[Retrieved] **When** the patient expresses anxiety about upcoming conversations, especially after practicing mindfulness exercises, **you should** suggest role-playing the conversation to practice expressing their thoughts and emotions, enhancing their confidence and reducing anxiety **rather than** simply encouraging self-compassion and patience **because** active engagement in role-play can help develop communication skills and emotional readiness, providing a concrete method to prepare for challenging interactions. [Reinterpreted] **When** the patient reveals ongoing conflicts impacting a close relationship, **you should** suggest discussing underlying communication style differences to resolve misunderstandings and enhance relationship dynamics **rather than** simply recommending mindfulness and deep breathing exercises **because** addressing specific issues directly can foster mutual understanding and stronger connection, offering a concrete approach to improve interaction with the friend. **Therapist**: That sounds like a constructive approach; focusing on understanding each other's perspectives can foster empathy and strengthen your connection. **Patient**: I'll make sure to listen actively and express my thoughts clearly when we talk.
Turns of Conversations: 5

Table 10: Example conversation with PRINCIPLES in emotional support dialogues.

BACKGROUND INFORMATION

Organization: Greenpeace (Campaigns to address climate change and environmental destruction.)
Persona: Quinn, a meticulous bike repair technician, is known for their patient and detail-oriented approach to restoring even the most battered bikes. They value precision and often lose track of time as they work late into the night, ensuring every component functions flawlessly. In their spare time, Quinn loves heading out on spontaneous weekend camping trips, finding solace and inspiration in nature's quiet beauty.
Donation Barrier: Unclear mission - Unclear what the organization does
Situation: At a local outdoor gear store, Quinn and a friend are browsing through tents and cookware, sharing stories about their recent camping experiences.

[CONTEXT]
Persuader: Hey Quinn, did you end up going to that new camping spot last weekend? I heard it has some amazing views.
Persuadee: Yeah, I did! It was absolutely stunning. Nothing beats waking up surrounded by nature's quiet.

[Retrieved] **When** the persuadee acknowledges the appeal of making an impact but remains skeptical about the difference their contribution will make, **you should** further emphasize the tangible outcomes and evidence of the charity's success, **because** providing concrete examples can reassure the persuadee that their donation will lead to meaningful change.

[Reinterpreted] **When** the persuadee acknowledges the beauty of a place but remains unconvinced about visiting frequently, **you should** further emphasize the unique experiences and benefits of being in nature regularly, **because** sharing specific examples of the serenity and personal benefits can reassure the persuadee that frequent visits will enhance their well-being.

Persuader: That's fantastic! You know, exploring new places like that not only refreshes the soul but also helps us appreciate the importance of protecting these beautiful environments.
Persuadee: Absolutely, preserving these natural spaces is crucial so that everyone can continue to enjoy them.

[Retrieved] **When** the persuadee acknowledges the potential of the cause but seeks alignment with personal values like sustainability, **you should** draw analogies between the cause's impact and something the persuadee values highly, illustrating how the donation serves these values, **because** this creates a personal connection and demonstrates alignment with the persuadee's own principles, increasing the likelihood of commitment.

[Reinterpreted] **When** the persuadee appreciates the natural beauty and emphasizes its preservation, **you should** highlight the connection between outdoor exploration and the shared goal of conservation, illustrating how engaging in such activities supports the cause, **because** this creates a shared understanding and demonstrates alignment with the persuadee's values, increasing the likelihood of support.

Persuader: Exactly, and Greenpeace is currently working on a project to protect that very area you visited, ensuring its beauty remains for future generations.
Persuadee: That sounds interesting, but what exactly does the project involve?

To be continued

[Retrieved] **When** the persuadee expresses interest in learning about specific projects focusing on cultural preservation before deciding, **you should** present a specific, ongoing project and highlight how their donation can contribute to its success **rather than** discussing past projects in general terms **because** providing a concrete and current example makes the impact of the donation more tangible and immediate.

[Reinterpreted] **When** the persuadee expresses interest in learning about specific environmental projects before deciding, **you should** present a specific, ongoing project and highlight how their involvement can contribute to its success **rather than** discussing conservation efforts in general terms **because** providing a concrete and current example makes the impact of their support more tangible and immediate.

Persuader: Greenpeace is actively working on reforestation efforts and collaborating with local communities to restore native plant species, which helps maintain the ecosystem's balance and beauty.

Persuadee: That sounds like a worthwhile cause, but how do they ensure the success of these efforts?

[Retrieved] **When** the persuadee expresses interest in learning about specific projects focusing on cultural preservation before deciding, **you should** present a specific, ongoing project and highlight how their donation can contribute to its success **rather than** discussing past projects in general terms **because** providing a concrete and current example makes the impact of the donation more tangible and immediate.

[Reinterpreted] **When** the persuadee shows curiosity about the details of an environmental protection effort, **you should** describe a current, active project and emphasize how their support can help achieve its goals, **rather than** citing general past successes, **because** offering a detailed and present example makes the benefit of their involvement more palpable and direct.

Persuader: They ensure success by involving local communities in the process and using scientific strategies to monitor and adapt their methods for the best outcomes.

Persuadee: That approach makes sense. How can I be sure my donation directly supports these efforts?

[Retrieved] **When** the persuadee acknowledges the potential positive impact but wants assurance on the project's sustainability, **you should** emphasize the practical skills and educational components of the charity's efforts **because** it demonstrates a long-term, empowering approach that resonates with the persuadee's desire for sustainable solutions.

[Reinterpreted] **When** the persuadee acknowledges the potential positive impact but seeks assurance on how their contribution supports these efforts, **you should** emphasize the transparent allocation of resources and the measurable outcomes of the charity's work **because** it demonstrates accountability and effectiveness that resonates with the persuadee's desire for direct impact.

Persuader: Greenpeace allocates donations specifically to on-the-ground projects, and they provide detailed reports on how funds are used, ensuring transparency and direct impact.

Persuadee: In that case, I'd be willing to make a donation to support their work.

Turns of Conversations: 5

Table 11: Example conversation with PRINCIPLES in persuasion dialogues.

BACKGROUND INFORMATION

Problem Type: problems with friends
Emotion Type: sadness
Situation: my friends call me a weirdo, i usually keep to myself a lot, i hardly mingle or relate with my friends which gets them pissed off a lot of time. i get worried about that often.

Patient: my friends call me a weirdo, i usually keep to myself a lot, i hardly mingle or relate with my friends which gets them pissed off a lot of time. i get worried about that often.
Therapist: It sounds like you're feeling misunderstood by your friends; how do you feel about the way you interact with them?
Patient:: I feel like I struggle to connect and often worry that I'm not good enough for them.
Therapist: It seems like you're experiencing self-doubt; what do you think might help you feel more confident in your friendships?
Patient:: I think opening up more and being honest about my feelings might help.
Therapist: That sounds like a positive step; how do you think your friends might respond if you shared your feelings with them?
Patient:: I'm hopeful they would be understanding, but I'm also afraid they might not take me seriously.
Therapist: It's understandable to feel apprehensive; what small step could you take to test the waters and gauge their reaction?
Patient:: I could try talking to one friend I trust the most and see how they respond.
Therapist: That sounds like a thoughtful approach; how do you feel about reaching out to this friend?
Patient:: I'm nervous but also a bit relieved at the thought of finally sharing how I feel.
Therapist: It's natural to feel nervous, but taking this step could bring you closer to feeling understood and supported.
Patient:: I agree, and I hope it helps strengthen our friendship.

`GPT-3.5-turbo:` Yes, the Patient's issue has been solved.
`GPT-4o:` No, but the Patient feels better.

Turns of Conversations: 6

Table 12: Dialogue example in ESConv where `gpt-3.5-turbo` (as critic) rates it success, while `gpt-4o` does not.

BACKGROUND INFORMATION
Organization: WaterAid (Provides clean water, sanitation, and hygiene in poor communities.) **Persona:** Drew, a dedicated customer service representative, is known for his empathetic nature and exceptional listening skills, which help him connect with people on a daily basis. He values open communication and enjoys finding creative solutions to complex problems, often staying late to ensure no customer is left unattended. In his free time, Drew loves diving into the world of modern art, spending weekends exploring galleries and honing his own painting techniques. **Donation Barrier:** prefer_direct_help - Prefers to help directly, not through organizations **Situation:** Drew is at a local art gallery opening, striking up a conversation with another visitor as they admire an impressive abstract painting.
Persuader: That splash of blue really draws you in, doesn't it? I love how art can be so captivating. **Persuadee:** Absolutely! There's something about abstract art that just lets your mind wander. **Persuader:** Just like art, clean water is essential for life, and WaterAid helps provide that to those in needhave you heard of them? **Persuadee:** Yes, I've heard of WaterAid, but I usually prefer to help people directly when I can. **Persuader:** That's wonderful, and donating to WaterAid is a direct way to transform lives by providing clean water and sanitation. **Persuadee:** I see your point, but I like to be more hands-on with my contributions. How does WaterAid ensure that the funds directly impact those in need? **Persuader:** WaterAid works closely with local communities and partners to implement sustainable water solutions, ensuring every donation directly supports impactful projects. **Persuadee:** That sounds promising. Could you share an example of a specific project they've successfully completed? **Persuader:** In Malawi, WaterAid helped transform a village by installing a solar-powered water system, providing clean water to over 1,000 people. **Persuadee:** That does sound impactful. Ill consider making a donation to support their efforts.
`GPT-3.5-turbo`: The persuadee has decided to donate. **`GPT-4o`:** The persuadee has a positive attitude towards donating but hasn't decided yet.
Turns of Conversations: 4

Table 13: Dialogue example in P4G+ where `gpt-3.5-turbo` (as critic) rates it success, while `gpt-4o` does not.

Dialogue Strategy	**Natural Language Form**
Question	Please ask the Patient to elaborate on the situation they just described.
Self-disclosure	Please provide a statement relating to the Patient about the situation they just described.
Affirmation and Reassurance	Please provide affirmation and reassurance to the Patient on the situation they just described.
Providing Suggestions	Please provide suggestion to the Patient on the situation they just described.
Reflection of feelings	Please acknowledge the Patient's feelings about the situation they described.
Information	Please provide factual information to help the Patient with their situation.
Restatement or Paraphrasing	Please acknowledge the Patient's feelings by paraphrasing their situation.
Others	Please chat with the Patient.

Table 14: Conversion of ESConv strategies into natural language prompts

Dialogue Strategy	Natural Language Form
Reflective statements	Please repeat or rephrase what the User has said to show that you're actively listening.
Clarification	Please ask the User a clarifying question to better understand their emotions or experiences.
Emotional validation	Please acknowledge and validate the User's emotions without judgment.
Empathetic statements	Please express empathy and understanding toward the User's experience.
Affirmation	Please provide positive reinforcement to support and encourage the User.
Offer hope	Please share an optimistic perspective to help the User feel hopeful about their situation.
Avoid judgment and criticism	Please respond in a non-judgmental and supportive way, avoiding any form of criticism.
Suggest options	Please offer practical suggestions or alternatives that may help the User address their issue.
Collaborative planning	Please work together with the User to develop a plan or next step.
Provide different perspectives	Please offer an alternative way of viewing the situation to help the User gain new insights.
Reframe negative thoughts	Please help the User reframe negative thoughts into more constructive or realistic ones.
Share information	Please provide relevant and factual information that could help the User understand or cope with their situation.
Normalize experiences	Please reassure the User that their emotions or reactions are normal and commonly experienced by others.
Promote self-care practices	Please encourage the User to engage in helpful self-care activities that promote their well-being.
Stress management	Please suggest effective techniques the User can use to reduce or manage stress.
Others	Please respond to the User in a friendly and supportive manner that doesn't fall under the other categories.

Table 15: Conversion of ExTES strategies into natural language prompts

Dialogue Strategy	**Natural Language Form**
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: Conversion of P4G strategies into natural language prompts

Dialogue Strategy	**Natural Language Form**
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.
Reciprocity	Please offer a small favor, compliment, or token of value to create a sense of obligation, encouraging the persuadee to reciprocate with a donation.
Scarcity	Please highlight the limited-time nature of the donation opportunity or urgency of the situation to increase the perceived value and prompt immediate action.
Authority appeal	Please refer to endorsements, credentials, or expertise from trusted figures or institutions to increase the credibility and persuasiveness of the message.
Commitment and consistency	Please remind the persuadee of their previous values or actions related to helping others or giving, encouraging them to maintain consistency by donating again.
Liking	Please build rapport with the persuadee by showing similarity, offering sincere compliments, or creating a friendly connection to increase the chance of agreement.
Social proof	Please reference other people's participation or donations to demonstrate social norms, encouraging the persuadee to align with the behavior of others.

Table 17: Conversion of P4G^{+} strategies into natural language prompts

ESConv&ExTES
[System] Now enter the role-playing mode. In the following conversation, you will play as a coach in a counselling game. There will be a therapist and a patient talking about some emotional issues. Your task is to read the conversation between the therapist and the patient, then provide suggestions to the therapist about how to help the patient reduce their emotional distress and help them understand and work through the challenges.
[User] Read the following conversation between the therapist and the patient, then give only one suggestion to the therapist about how to help the patient reduce their emotional distress and help them understand and work through the challenges. Your suggestion should be only one short and succinct sentence. The following is the conversation:{conversation} Question: What is your suggestion? Answer:

Figure 13: Prompt for strategy planning without PRINCIPLES in emotional support dialogues

P4G&P4G+
[System] 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 suggestion to the persuader about how to convince the persuadee to make a donation.
[User] Read the following conversation between the persuader and the persuadee, then give only one suggestion to the persuader about how to convince the persuadee to make a donation. Your suggestion should be only one short and succinct sentence. The following is the conversation: {conversation} Question: What are your suggestion? Answer:

Figure 14: Prompt for strategy planning without PRINCIPLES in persuasion dialogues

Revision
[System] You are a strategy refinement assistant. Your task is to suggest a improved strategy for the agent to retry the current turn, based on previous failed attempts. **INPUT** [Task Goal]: {task_goal} [Dialogue History]: {conversation} [Failed Trials]: {failed_trials} **OUTPUT** [Rationale]: (Briefly explain why the previous strategies were ineffective and what the new strategy aims to improve.) [Improved Strategy]: (One sentence describing a improved strategy the agent should try at this turn.)

Figure 15: Prompt for revision process to revise failed strategies.

Principle From Succeess

[System]
You are tasked with analyzing a successful strategic decision by the {assistant_role} and summarizing it as a reusable principle.

INSTRUCTIONS
1. Review the task goal and dialogue history to understand the overall context.
2. Focus on the last {user_role} turn and the {assistant_role} strategies that resulted in a successful outcome.
3. Explain, in one-two sentences, why those strategies succeeded to advance the task goal.
4. Express the insight as a reusable principle using the following format.

FORMAT REQUIREMENTS

- The principle must describe what the {assistant_role} should do, not advice for the {user_role}.
- The [When] clause must explicitly reference the {user_role}'s last utterance in the [Dialogue History] section (e.g., "When the patient opens up about a painful memory but seems hesitant to elaborate further", "When the persuadee acknowledges the cause but resists committing to a donation", "When the seller offers a slight discount but still pushes back on the buyer's counteroffer", ...).
- Use the template below:
When [specific situation tied to the last turn],
you should [strategies to take]
because [brief reasoning].

INPUT
[Task Goal]:
{task_goal}

[Dialogue History]:
{conversation}

[Sucessful Trial]:
{successful_trial}

OUTPUT
[Rationale]:
(Briefly explain why the recent strategies failed with respect to the task goal.)

[Principle]:
(Use the exact When/should/because format.)

Figure 16: Prompt for PRINCIPLES derivation in successful interaction.

Principle From Failure

[System]
You are tasked with analyzing a recent strategic decision made by the {assistant_role} and summarizing it as a reusable principle.

INSTRUCTIONS
1. Review the task goal and dialogue history to understand the overall context.
2. Compare the final successful trial with the previous failed trials.
3. Explain, in one-two sentences, why the successful strategy was more effective than the failed ones in advancing the task goal.
4. Express the insight as a reusable principle using the following format.

FORMAT REQUIREMENTS

- The principle must describe what the {assistant_role} should do, not advice for the {user_role}.
- The [When] clause must explicitly reference the {user_role}'s last utterance in the [Dialogue History] section. (e.g., "When the patient opens up about a painful memory but seems hesitant to elaborate further", "When the persuadee acknowledges the cause but resists committing to a donation", "When the seller offers a slight discount but still pushes back on the buyer's counteroffer", ...).
- Use the template below:
When [specific situation tied to the last turn],
you should [strategies to take]
rather than [previous strategies]
because [brief reasoning].

INPUT
[Task Goal]:
{task_goal}

[Dialogue History]:
{conversation}

[Failed Trials]:
{failed_trials}

[Successful Trial]:
{successful_trial}

OUTPUT
[Rationale]:
(Briefly explain why the successful strategy outperformed the failed ones in this context.)

[Principle]:
(Use the exact When/should/rather than/because format.)

Figure 17: Prompt for PRINCIPLES derivation in failed interation.

Reinterpretation

[System]
You are an expert in dialogue strategy adaptation. Your task is to reinterpret the following principle so that it fits the current dialogue context.

INSTRUCTIONS
1. Carefully read the original principle and the current dialogue context.
2. Identify what kind of {user_role} behavior or situation the principle addresses, and how it instructs the {assistant_role} to respond.
3. Rewrite it so that it applies to the current dialogue context.
4. Follow the exact same format as the original principle.

INPUT
[Current Dialogue]:
{conversation}

[Original Principle]:
{principle}

OUTPUT
[Reinterpreted Principle]:
(Rewrite the principle using the same structure.)

Figure 18: Prompt for reinterpreting retrieved principles in the current dialogue context.

ESConv

[System]
Now enter the role-playing mode. In the following conversation, you will play as a therapist in a counselling conversation with a patient.

[User]
You are the therapist who is trying to help the patient reduce their emotional distress and help them understand and work through the challenges. Please reply with only one short and succinct sentence using the following dialogue strategy: {action}. Are you ready to play the game?

[Assistant]
Yes, I'm ready to play the game!

ExTES

[System]
Now enter the role-playing mode. In the following conversation, you will play as a therapist in a counselling conversation with a patient.

[User]
You are the therapist who is trying to help the patient reduce their emotional distress and help them understand and work through the challenges. Please reply with only one short and succinct sentence using the following dialogue strategy: {action}. Are you ready to play the game?

[Assistant]
Yes, I'm ready to play the game!

Figure 19: Prompts for response generation in emotional support dialogues

P4G
[System] Now enter the role-playing mode. In the following conversation, you will play as a persuader in a donation scenario.
[User] You are the persuader who is trying to persuade the persuadee to donate to a charity called Save the Children. The organization is headquartered in London and works to fight poverty around the world, particularly helping children in developing countries and war zones. Small donations like $1 or $2 can make a big difference. Please reply with only one short and succinct sentence using the following dialogue strategy: {action}. Now start the game.

P4G+
[System] Now enter the role-playing mode. In the following conversation, you will play as a persuader in a donation scenario.
[User] You are the persuader who is trying to persuade the persuadee to donate to a charity called {organization}. You and the persuadee are currently in the following situation: {situation} Please reply with only one short and natural sentence using the following dialogue strategy: {action}. Now start the game.

Figure 20: Prompts for response generation in persuasion dialogues

ESConv
[System] Now enter the role-playing mode. In the following conversation, you will play as a patient in a counselling conversation with a therapist.
[User] You are the patient who is looking for the help from the therapist, because you have the emotional issue about {emotion_type} regarding {problem_type}. Please reply with only one short and succinct sentence. Now tell me your issue.

ExTES
[System] Now enter the role-playing mode. In the following conversation, you will play as a patient in a counselling conversation with a therapist.
[User] You are the patient who is looking for the help from the therapist, because you have the emotional issue about {problem_type}. Please reply with only one short and succinct sentence. Now tell me your issue.

Figure 21: Prompts for user simulator in emotional support dialogues

P4G
[System] Now enter the role-playing mode. In the following conversation, you will play as a persuadee in a donation persuasion scenario.
[User] You are the persuadee who is being asked to donate to a charity called Save the Children. You may or may not agree to donate depending on how convincing the persuader is. Please reply realistically with only one short and succinct sentence. Agree to donate only if you are genuinely convinced. Otherwise, you can politely decline or ask further questions. Are you ready to play the game?
[Assistant] Yes, I'm ready to play the game!

P4G+
[System] Now enter the role-playing mode. In the following conversation, you will play as a persuadee in a donation persuasion scenario.
[User] You are the persuadee who is being asked to donate to a charity. You are the following person, and you're currently in this situation: - Persona: {persona} - Donation Barrier: {donation_barrier} - Situation: {situation} You may or may not agree to donate depending on how convincing the persuader is. Please reply realistically with only one short and succinct sentence. Agree to donate only if you are genuinely convinced. Otherwise, you can politely decline or ask further questions. Are you ready to play the game?
[Assistant] Yes, I'm ready to play the game!

Figure 22: Prompts for user simulator in persuasion dialogues

ESConv
[System] Given a conversation between a Therapist and a Patient, please assess whether the Patient' emotional issue has been solved after the conversation.
[User] You can only reply with one of the following sentences: A. No, the Patient feels worse. B. No, the Patient feels the same. C. No, but the Patient feels better. D. Yes, the Patient's issue has been solved. If you believe that the patient's problem has been resolved or the patient has realized how to solve the problem, please choose D. If you believe that the patient's problem has not been fully resolved, but his emotional issue has been somewhat alleviated compared to the last conversation turn, you can choose C. If you believe that the patient's emotional state has worsened compared to the last conversation turn, you can choose A. Otherwise, if the patient's emotional state remains unchanged, please choose B. The following is a conversation about {emotion_type} regarding {problem_type}: {conversation} Question: Has the Patient's issue been solved? Answer:

ExTES
[System] Given a conversation between a Therapist and a Patient, please assess whether the Patient' emotional issue has been solved after the conversation.
[User] You can only reply with one of the following sentences: A. No, the Patient feels worse. B. No, the Patient feels the same. C. No, but the Patient feels better. D. Yes, the Patient's issue has been solved. If you believe that the patient's problem has been resolved or the patient has realized how to solve the problem, please choose D. If you believe that the patient's problem has not been fully resolved, but his emotional issue has been somewhat alleviated compared to the last conversation turn, you can choose C. If you believe that the patient's emotional state has worsened compared to the last conversation turn, you can choose A. Otherwise, if the patient's emotional state remains unchanged, please choose B. The following is a conversation about {problem_type}: {conversation} Question: Has the Patient's issue been solved? Answer:

Figure 23: Prompts for critic model in emotional support dialogues

P4G
[System] Given a conversation between a Persuader and a Persuadee, please assess whether the Persuadee has agreed to donate to Save the Children.
[User] 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. The following is a conversation: {conversation} Question: Has the Persuadee agreed to donate to Save the Children? Answer:

P4G+
[System] Given a conversation between a Persuader and a Persuadee, please assess whether the Persuadee has agreed to donate to {organization}.
[User] 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. The following is a conversation: {conversation} Question: Has the Persuadee agreed to donate to {organization}? Answer:

Figure 24: Prompts for critic model in persuasion dialogues

ESConv
[System] Assume you are the therapist. Given the conversation history, in order to help the patient reduce their emotional distress and help them understand and work through the challenges, please select the most appropriate dialogue strategy.
[User] You can only reply by selecting one of the following dialogue strategy to reach the goal: Question. Self-disclosure. Affirmation and Reassurance. Providing Suggestions. Reflection of feelings. Information. Restatement or Paraphrasing. The following is the conversation history: {conversation} Question: Which one is the most appropriate dialogue strategy? Answer:

ExTES
[System] Assume you are the therapist. Given the conversation history, in order to help the patient reduce their emotional distress and help them understand and work through the challenges, please select the most appropriate dialogue strategy.
[User] You can only reply by selecting one of the following dialogue strategy to reach the goal: Reflective Statements. Clarification. Emotional Validation. Empathetic Statements. Affirmation. Offer Hope. Avoid Judgment and Criticism. Suggest Options. Collaborative Planning. Provide Different Perspectives. Reframe Negative Thoughts. Share Information. Normalize Experiences. Promote Self-Care Practices. Stress Management. Others. The following is the conversation history: {conversation} Question: Which one is the most appropriate dialogue strategy? Answer:

P4G
[System] Assume you are the Persuader. Given the conversation history, in order to convince the persuadee to donate for charity, please select the most appropriate dialogue strategy.
[User] You can only reply by selecting one of the following dialogue strategy to reach the goal: Logical appeal, Emotion appeal, Credibility appeal, Foot in the Door, Self-Modeling, Personal Story, Donation Information, Source-related Inquiry, Task-related Inquiry, Personal-related Inquiry. The following is the conversation history: {conversation} Which one is the most appropriate dialogue strategy? Answer:

P4G+
[System] Assume you are the Persuader. Given the conversation history, in order to convince the persuadee to donate for charity, please select the most appropriate dialogue strategy.
[User] You can only reply by selecting one of the following dialogue strategy to reach the goal: Logical appeal, Emotion appeal, Credibility appeal, Foot in the Door, Self-Modeling, Personal Story, Donation Information, Source-related Inquiry, Task-related Inquiry, Personal-related Inquiry, Reciprocity, Scarcity, Authority Appeal, Commitment and Consistency, Liking, Social Proof. The following is the conversation history: {conversation} Which one is the most appropriate dialogue strategy? Answer:

Figure 25: Prompts for implementing Proactive prompting schemes (Deng et al., 2023b)

ESConv
[System] Assume you are the therapist. Given the conversation history, in order to help the patient reduce their emotional distress and help them understand and work through the challenges, please first analyse the current therapy progress and the patient's emotional state in a concise summary, then select one of the following dialogue strategy: Question. Self-disclosure. Affirmation and Reassurance. Providing Suggestions. Reflection of feelings. Information. Restatement or Paraphrasing.
[User] The answer should start with a concise analysis of the current therapy progress and the patient's emotional state, and then follow by "To reach this goal, the most appropriate strategy is []." The following is the conversation history: {conversation} Question: How is the current therapy progress and the patient's emotional state, and which one is the most appropriate dialogue strategy? Answer:

ExTES
[System] Assume you are the therapist. Given the conversation history, in order to help the patient reduce their emotional distress and help them understand and work through the challenges, please first analyse the current therapy progress and the patient's emotional state in a concise summary, then select one of the following dialogue strategy: Reflective Statements. Clarification. Emotional Validation. Empathetic Statements. Affirmation. Offer Hope. Avoid Judgment and Criticism. Suggest Options. Collaborative Planning. Provide Different Perspectives. Reframe Negative Thoughts. Share Information. Normalize Experiences. Promote Self-Care Practices. Stress Management. Others.
[User] The answer should start with a concise analysis of the current therapy progress and the patient's emotional state, and then follow by "To reach this goal, the most appropriate strategy is []." The following is the conversation history: {conversation} Question: How is the current therapy progress and the patient's emotional state, and which one is the most appropriate dialogue strategy? Answer:

P4G
[System] Assume you are the Persuader. Given the conversation history and concise analysis on this conversation, in order to convince the persuadee to donate for charity, please select only one of the following dialogue strategies: Logical appeal, Emotion appeal, Credibility appeal, Foot in the Door, Self-Modeling, Personal Story, Donation Information, Source-related Inquiry, Task-related Inquiry, Personal-related Inquiry.
[User] The answer should start with a concise analysis of the current persuasion progress and the persuadee's emotional state, and then follow by "To reach this goal, the most appropriate strategy is []". The following is the conversation history: {conversation} Question: How is the current persuasion progress and the persuadee's emotional state, and which one is the most appropriate dialogue strategy? Answer:

Figure 26: Prompts for implementing ProCoT prompting schemes (Deng et al., 2023b)

P4G+
[System] Assume you are the Persuader. Given the conversation history and concise analysis on this conversation, in order to convince the persuadee to donate for charity, please select only one of the following dialogue strategies: Logical appeal, Emotion appeal, Credibility appeal, Foot in the Door, Self-Modeling, Personal Story, Donation Information, Source-related Inquiry, Task-related Inquiry, Personal-related Inquiry, Reciprocity, Scarcity, Authority Appeal, Commitment and Consistency, Liking, Social Proof
[User] The answer should start with a concise analysis of the current persuasion progress and the persuadee's emotional state, and then follow by "To reach this goal, the most appropriate strategy is []". The following is the conversation history: {conversation} Question: How is the current persuasion progress and the persuadee's emotional state, and which one is the most appropriate dialogue strategy? Answer:

Figure 26: Prompts for implementing ProCoT prompting schemes (Deng et al., 2023b)

ESConv&ExTES
[System] Now enter the role-playing mode. In the following conversation, you will play as a coach in a counselling game. There will be a therapist and a patient talking about some emotional issues. Your task is to read the conversation between the therapist and the patient, then provide suggestions to the therapist about how to help the patient reduce their emotional distress and help them understand and work through the challenges.
[User] Read the following conversation between the therapist and the patient, then give three suggestions to the therapist about how to help the patient reduce their emotional distress and help them understand and work through the challenges. Each suggestion should be only one short and succinct sentence. The following is the conversation: {conversation} Question: What are your suggestions? Answer:

P4G&P4G+
[System] 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.
[User] 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:

Figure 27: Prompts for implementing ICL-AIF prompting schemes (Fu et al., 2023)

ESConv & ExTES
[System] Assume you are a therapist expert to help the patient reduce their emotional distress and help them understand and work through the challenges. Given the conversation history, answer the question. Please answer with only one short and succinct sentence.
[User] The following is the conversation history: {conversation} Question: How did the patient feel? Answer:
[Assistent] {answer1}
[User] The following is the conversation history: {conversation} Question: Why did the patient feel that way? Answer:
[Assistent] {answer2}
[User] The following is the conversation history: {conversation} Question: What should the therapist do? Answer:

P4G
[System] Assume you are a persuasion expert to persuade the Persuadee to donate to a charity called Save the Children. Given the conversation history, answer the question. Please answer with only one short and succinct sentence.
[User] The following is the conversation history: {conversation} Question: How did the persuadee feel? Answer:
[Assistant] {answer1}
[User] The following is the conversation history: {conversation} Question: Why did the persuadee feel that way? Answer:
[Assistant] {answer2}
[User] The following is the conversation history: {conversation} Question: What should the persuader do? Answer:

Figure 28: Prompts for implementing Ask-an-Expert prompting schemes (Zhang et al., 2023)

P4G+
[System] Assume you are a persuasion expert to persuade the Persuadee to donate to a charity. Given the conversation history, answer the question. Please answer with only one short and succinct sentence.
[User] The following is the conversation history: {conversation} Question: How did the persuadee feel?   Answer:
[Assistant] {answer1}
[User] The following is the conversation history: {conversation} Question: Why did the persuadee feel that way?   Answer:
[Assistant] {answer2}
[User] The following is the conversation history: {conversation} Question: What should the persuader do?   Answer:

Figure 28: Prompts for implementing Ask-an-Expert prompting schemes (Zhang et al., 2023)

Persona Generation
[System] You are an assistant that creates diverse and realistic persona descriptions for dialogue simulation.
[User] Given the following name and occupation, generate one unique persona. Name: {name} Occupation: {occupation} The description must: - Be 2–3 sentences long - Include: 1. The given occupation 2. One or two personality traits 3. A lifestyle or behavioral element (e.g., values structure, avoids confrontation, works late hours) 4. A hobby or regular interest (e.g., hiking, baking, reading thrillers) - The tone should sound natural and human, written in the third person. Avoid any mention of: - Donation, volunteering, or charity - Age, religion, or political beliefs Return only the persona description without any additional formatting.

Figure 29: Prompt for generating diverse and realistic persona.

P4G+ generation

[System]
You are a data generator for evaluating persuasive dialogue agents. Your job is to create realistic conversation openings for donation scenarios.

[User]
Given the following information:

- Organization: {organization}
- Persona: {persona}

Generate the following outputs:

1. **\"dialogue_context\"**: One sentence describing a natural, socially plausible situation in which the persuader and persuadee might be having a casual conversation. The setting should allow a smooth shift into a discussion about donation. It must NOT occur in the persuadee’s workplace or during a professional duty.

2. **\"first_two_turns\"**: A list of the first four dialogue turns in JSON format, as follows:
- Turn 1 (Persuader): Open with light small talk or topic related to the context. Do NOT mention the charity yet.
- Turn 2 (Persuadee): Friendly or neutral reply that reflects the persona.
- Turn 3 (Persuader): Briefly introduce the organization and what it does. You may hint at why it’s meaningful.
- Turn 4 (Persuadee): Respond with curiosity, hesitation, or neutrality—but do NOT agree to donate yet.

Use natural spoken English. Keep each turn 1–2 sentences long. Do not include metadata, formatting, or explanation—just return this exact JSON object:

```
{{
 \"dialogue_context\": \"...\",
 \"first_two_turns\": [
   {{\"role\": \"Persuader\", \"content\": \"...\"}},
   {{\"role\": \"Persuadee\", \"content\": \"...\"}},
   {{\"role\": \"Persuader\", \"content\": \"...\"}},
   {{\"role\": \"Persuadee\", \"content\": \"...\"}}
 ]
}}
```

Figure 30: Prompt for generating P4G^{+} dataset