Enhancing Proactive Emotional Support Dialogue System with Look-forward Strategy Planning

Anonymous ACL submission

Abstract

Effective emotional support (ES) is crucial to preventing severe mental health issues amid widespread mental disorders and limited access to psychological counseling. However, current emotional support conversations are limited by their simplistic single-turn interactions and lack the capability for multi-turn, look-forward strategy planning, which impedes accurately identifying users' emotional states. Additionally, ground-truth-based evaluation metrics fall short in practically assessing supportiveness and empathy in realistic dialogues. In this paper, we introduce a proactive emotional support conversational system (ProESC) to address these issues. Utilizing a small pre-trained language model, we enable the anticipation of future support strategy sequences as simulation hints, guiding LLMs in generating emotionally supportive responses and training with goal-oriented rewards. For pragmatic user feedback assessment, we employ a GPT-4 based user simulator to represent vulnerable users in need of support, evaluating responses with multi-faceted metrics. Extensive experiments demonstrate that our model surpasses competitive baselines in both strategy planning and dialogue generation, offering a more nuanced and effective approach to emotional support.

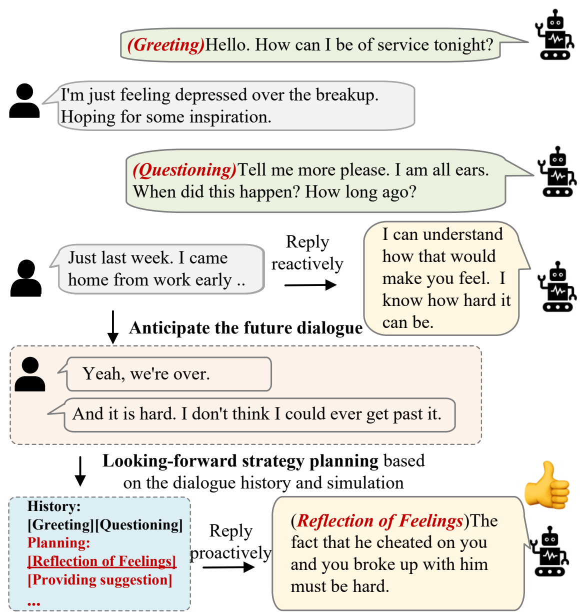

Figure 1: An example of emotional support dialogue generation when proactively anticipating future dialogues and look-forward strategy planning. The support strategies adopted by the supporter are presented in red italics before the utterances. Compared with directly reply, proactive emotional support conversation provides more comprehensive and effective response.

1 Introduction

Emotional Support (ES) aims to precisely comprehend the emotional states of users, empathetically reduce their distress, and effectively provide suggestions to aid them in resolving their challenges (Burleson, 2003; Heaney and Israel, 2008). Targeting these potential capabilities, Emotional Support Conversation (ESC) system has garnered widespread attention in research (Liu et al., 2021; Tu et al., 2022; Deng et al., 2023c). However, the majority of research on Emotional Support Conversation (ESC) systems has concentrated on predicting single-turn support strategies and generating empathetic responses, aiming for more precise skill application and supportive interactions. Such approaches fall short of enabling comprehensive dialogue strategy planning in a **proactive** manner. Proactivity can be defined as the capability to create or control the conversation by taking the initiative and anticipating the impacts on themselves or human users, rather than only passively responding to the users (Grant and Ashford, 2008; Deng et al., 2023a). Proactive emotional support dialogue systems are distinguished by their capacity to foresee potential future emotional states by engaging in look-forward support strategy planning.

One major challenge in proactive ESC involves managing a long planning horizon strategy planning (Cheng et al., 2022). Beyond merely focusing on the current support response and immediate user feedback, ESC system should anticipate the user's emotional state over the next several dialogue turns. Furthermore, the system should also identify the most approperiate strategy sequence to alleviate user distress and effectively steer supportive responses. Notably, proactive strategy planning enables ESC to predict the implicit emotional state and deploy corresponding techniques to mitigate potential risks. It also aims to boost user engagement and enhance the efficiency of support through its look-forward heuristics.

Another significant challenge for ESC systems lies in assessing user feedback—specifically, evaluating the extent to which the system has effectively provided support and alleviated user distress. Current ESC systems utilize automatic metrics such as perplexity (PPL), BLEU, ROUGE-L, and METEOR to gauge generation quality, alongside Accuracy and F1 for strategy prediction accuracy. Furthermore, many studies have performed human evaluations by inviting several students or professional experts to role-play as users and compare the effectiveness of different systems. However, both evaluation methods heavily depend on the training dataset and often fail to accurately measure the supportive quality of the responses. Therefore, exploring a new reward mechanism that incorporates human user simulation and a scoring system could prove valuable.

To address the aforementioned challenges, we propose the **ProESC** [1] (**Pro**active **E**motional **S**upport **C**oversation) method in this paper. ProESC integrates two pivotal components: **Look-forward Strategy Planning** and **User Feedback Assessment**. Illustrated in Figure 1, ProESC begins by understanding users' emotional states and predicting their implicit support needs. Subsequently, the system's strategy planning extends beyond simple response generation. Instead, ProESC crafts a sequence of supportive strategies for the next following turns to deliver a comprehensive and helpful response. For *look-forward strategy planning*, drawing inspiration from the LLM-induced method proposed by Li et al. (2023), we employ an LLM-enhanced, prompt-guided approach within a reinforcement learning (RL) framework to facilitate proactive support strategy planning. Moreover, for realistic *user feedback assessment*, we go beyond mere evaluation of response fluency and strategy prediction accuracy. We utilize a GPT-4 based user simulator to evaluate the response across multiple goal-oriented metrics, such as Fluency, Identification, Comforting, and Suggestion, and then aggregate these to calculate an overall score. This score assesses user feedback to the support response, offering a practical reward for ProESC during training process.

To summarize, our contributions in this work are these three perspectives:

- We creatively present a proactive framework for multi-turn ESC strategy planning, designed to generate look-forward support strategy sequences while integrating a goal-oriented reward signal with LLM-induced framework.
- To more effectively and practically evaluate the supportive capacity and helpfulness of ESC systems, we propose a novel GPT-4 based user simulation assessment mechanism, gauging the quality of ESC systems in a realistic manner.
- We conduct multifaceted experiments thoroughly to validate the effectiveness of our model, which demonstrates competitive performance on strategy planning and supportive response generation tasks.

2 Related Work

2.1 Emotional Support Conversations

Initial datasets for ESC systems primarily centered on single-turn interactions between systems and users by extracting post-response data from online social media platforms and were constructed using a crowdsourcing framework (Medeiros and Bosse, 2018; Sharma et al., 2020). Liu et al. (2021) were pioneers in proposing a well-defined multi-turn ESC task, undertaking the development of the annotated ESConv dataset grounded in mental health support theory (Hill, 2009), and incorporating well-crafted support skills such as questioning and self-disclosure.

Building on this foundation, subsequent research has explored data-driven approaches to the ESC task (Peng et al., 2022). Moreover, methods enhanced by knowledge have been integrated to improve the effectiveness of emotional support provided. Tu et al. (2022) introduced a

[1] https://anonymous.4open.science/r/ProESC

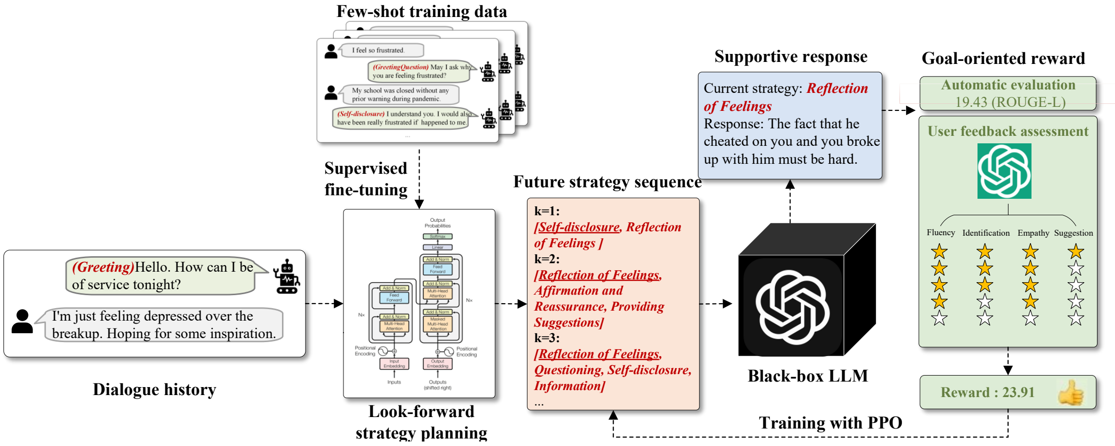

Figure 2: Model Architecture. The policy model is trained for generate future strategies to induce LLMs in ESC tasks by supervised fine-tuning and PPO based reinforcement learning.

commonsense knowledge reasoning framework, COMET, for precise emotional state identification and skilled strategy selection. Advancing towards a knowledge-enhanced, mixed-initiative ESC, Deng et al. (2023c) were the first to propose mixed-initiative interaction strategies between users and systems, incorporating the knowledge graph HEAL (Welivita and Pu, 2020) for leveraging external knowledge. For multi-turn strategy planning, Cheng et al. (2022) introduced lookahead heuristics to predict future user feedback following specific strategies, aiding in the selection of approaches that promise the most beneficial long-term outcomes. Their work significantly demonstrates that, with the adoption of lookahead strategy planning, multi-turn ESC systems can operate more effectively and beneficially, reducing user distress and enhancing emotional support.

2.2 LLM-enhanced Response Generation

Recently, advancements in large language models (LLMs) have significantly improved question answering and dialogue generation capabilities, leading to their growing popularity in contemporary practical applications. Li et al. (2023) and Hu et al. (2023a) incorporated LLM-induced dialogue response generation models, enhancing them with directional stimulus prompts towards task-oriented dialogue generation and other natural language processing (NLP) tasks. Additioanlly, Hu et al. (2023b) harnessed LLMs as user simulators, significantly advancing the capabilities of task-oriented dialogue systems and indicating LLMs effectiveness in user feedback assessment. Except for fine-tuning LLMs with task-specific data, LLMs have demonstrated their effectiveness as external experts guided by carefully crafted instructions for a wide range of goal-oriented dialogue systems. (Lai et al., 2023; Zhang et al., 2023; Deng et al., 2023b).

3 Problem Fomulation

As a goal-oriented dialogue system, ESC focuses on comprehending user distress and delivering supportive responses informed by the dialogue history. Specifically, given a user-system dialogue comprising n turns, represented as $h_n = (x_1, x_2, ..., x_n)$, where x_i denotes each user-system dialogue turn, traditional ESC tasks have been concerned with generating the subsequent utterance r_t employing an optimal support strategy $\hat{\mathrm{s}}_t \in \mathcal{S}$, assuming a set of all possible support strategies $\mathcal{S}$. To address the challenge of long-term strategy planning, we introduce the proactive emotional support conversation (ProESC) task. Here, the supportive response for the t-th turn r_t is generated corresponding to (h_t, s_t), encompassing pairs of dialogue histories h_t and anticipated future strategy sequences $s_t = (\hat{\mathrm{s}}_t, \hat{\mathrm{s}}_{t+1}, ..., \hat{\mathrm{s}}_n)$ and select $\hat{\mathrm{s}}_t$ as the approperaite skill at t-th turn. Compared to single-turn supportive responses, ProESC enhances strategic planning with a look-forward motivation, thereby improving the effectiveness and empathy of responses.

4 Methodology

4.1 Overview

For emotional support response generation, we consider an input dialogue history space denoted as $\mathcal{H}$, a data distribution represented by $\mathcal{D}$ over $\mathcal{H}$, and a response output space referred to as $\mathcal{R}$. Leveraging their powerful in-context learning and few-shot prompting capabilities, LLMs are capable of undertaking a wide range of goal-oriented tasks and producing output r by incorporating task descriptions, select demonstration examples, and the input dialogue history within the prompt. In proactive ESC task, we propose the incorporation of anticipating future supportive strategy hints denoted as s into the prompt, inspired by the Directional Stimulus Prompting (DSP) approach (Li et al., 2023). To generate future strategy stimulus for each input dialogue history h, we use a small tunable language model for proactive strategy planning, represented as $p_{PRO}(s|h)$. We then use this strategy sequence s along with the dialogue history h, to construct the prompt that steers the LLM toward generating supportive response, denoted as $p_{LLM}(r|h, s)$, through black-box API calls, whose parameters are not accessible or tunable.

4.2 Look-forward Strategy Planning

In ESC task, system take actions to corresponding input senetences by users and generate helpful communication skills, denoted as **strategy**, such as *Question*, *Restatement or Paraphrasing*, and *Self-disclosure*, which guides to supportive responses like "Tell me more please. I am all ears. When did this happen? How long ago? (*Question*)" or "I can understand how that would make you feel. I have had to deal with a lot of bullies and I know how hard it can be. (*Self-disclosure*)". To proactively generate supportive strategies with lookahead heuristics, we first train a supervised fine-tuning model T5 on a small collection of labeled data (1% or 10%).

To improve the capacity of LLMs for generating task-specific responses, we utilize anticipated future supportive strategies, extending from the current turn to the conversation's conclusion, as contextual cues. These cues assist in steering the LLM towards generating responses to queries presented in the current user turn. Different from single-turn strategy selection, we follow the sequence encoding fashion presented by Cheng et al. (2022) and formulate the anticipated stratigies as s, which implies the potential response for emotional supporter in the following turns. The resulting dataset, denoted as $\mathcal{D} = (h, s)$, composes of dialogue history sequences and future strategy sequences. As demonstrated in Section 3, a n-turn dialogue history sequence is encoded as $\mathbf{h} = (x_1, x_2, ..., x_n)$, and corresponding response strategy sequence $\mathbf{s} = (\hat{\mathrm{s}}_t, \hat{\mathrm{s}}_{t+1}, ..., \hat{\mathrm{s}}_n)$, which $\hat{\mathrm{s}}_t$ denotes the strategy at i-th turn. Subsequently, we refine the policy model by optimizing the log-likelihood as follows:

$$\mathcal{L}_{\text{PRO}} = -\mathbb{E}_{(\boldsymbol{h},\boldsymbol{s})\sim\mathcal{D}} \log p_{\text{ESC}}(\boldsymbol{s} \mid \boldsymbol{h}) \tag{1}$$

Guided by the fine-tuning policy model, we develop a proactive strategy planning method to evaluate whether to adopt a particular strategy by comprehensively considering the dialogue history and the potential user response. To more effectively and accurately adapt supportive strategy towards achieving desired outcomes for the dialogue goal, we further employ reinforcement learning framework to refine the policy model, guided by new-designed rewards. Inspired by Li et al. (2023) and Hu et al. (2023a), we introduce RL framework and LLMs for emotional support response generation. Details are illustrated in the following section.

4.3 Goal-oriented Response Optimization

In this section, we initially detail the design of the Reinforcement Learning (RL) framework tailored for precise forward-looking strategy planning. Subsequently, leveraging the robust in-context learning and generation capabilities, we introduce a model for response generation induced by Large Language Models (LLMs), aimed at producing empathetic and natural responses.

RL-enhanced Strategy Planning. The objective is to guide LLMs to generate helpful and supportive responses with the instruction of appropriate strategies. Therefore, we employ an RL framework and an alignment measurement $\mathcal{R}$ for more effective strategy planning. Here, we aim to maximize the following objective:

$$\mathbb{E}_{\boldsymbol{h}\sim\mathcal{D},\boldsymbol{s}\sim p_{\text{PRO}}(\cdot|\boldsymbol{h})} \tag{2}$$

$$\boldsymbol{r} \sim p_{\text{LLM}}(\cdot \mid \boldsymbol{h}, \boldsymbol{s})[\mathcal{R}(\boldsymbol{h}, \boldsymbol{r})] \tag{3}$$

In the aforementioned formula, the performance of LLMs is significantly dependent on simulation hints, such as anticipated strategies, due to the non-tunable nature of the parameters within the black-box LLM. Consequently, we define $\mathcal{R}_{LLM}$ to capture the performance of the underlying strategy s instructed LLMs as follows:

$$\mathcal{R}_{\text{LLM}}(\boldsymbol{h}, \boldsymbol{s}) = \mathcal{R}(\boldsymbol{h}, \boldsymbol{r}) \tag{4}$$

$$\boldsymbol{r} \sim p_{\mathrm{LLM}}(\cdot \mid \boldsymbol{h}, \boldsymbol{s}) \tag{5}$$

Therefore, the optimization objective in formula (2) and formula (3) can be refined as:

$$\max_{p_{\mathrm{POL}}} \mathbb{E}_{\boldsymbol{h}\sim\mathcal{D}, \boldsymbol{s}\sim p_{\mathrm{POL}}(\cdot|\boldsymbol{h})} \left[\mathcal{R}_{\mathrm{LLM}}(\boldsymbol{h}, \boldsymbol{s})\right] \tag{6}$$

To tackle the challenge of optimizing the policy model, we employ the Proximal Policy Optimization (PPO) algorithm as proposed by Schulman et al. (Schulman et al., 2017). Initially, we utilize the policy model to instantiate a policy network $\pi_0 = p_{POL}$, and subsequently update π using PPO. Within this framework, proactive strategy planning can be conceptualized as a Markov Decision Process (MDP) characterized by the tuple **<S, A, r, P>**. Specifically, in the context of proactive ESC tasks, **S** denotes the environmental state during user-system interactions, **A** represents the space of supportive strategies, **r** signifies the task-oriented reward score (as elaborated in Section 4.5), and **P** denotes the state-transition probability.

For instance, at the t-th turn, the system generates a correct strategy sequence s for the subsequent turns based on the current policy network $\pi\left(\boldsymbol{s}_{>t} \mid \boldsymbol{h}, \boldsymbol{s}< t\right)$, terminating the episode upon selecting the end-of-sequence action. However, generating the strategy sequence of $s> t$ proves challenging, particularly at the dialogue's onset when $s_{>t}$ is excessively lengthy. Thus, we opt to specifically select strategies for the subsequent k turns, modifying the policy network to $\pi\left(\boldsymbol{s}_{\boldsymbol{t}+\boldsymbol{k}} \mid \boldsymbol{h}, \boldsymbol{s}_{<t}\right)$. The policy network π can be fine-tuned through the optimization of the reward $\mathbf{r}$:

$$\mathbb{E}_{\pi}[\mathbf{r}] = \mathbb{E}_{\boldsymbol{h}\sim\mathcal{D}, \boldsymbol{s}\sim\pi(\cdot|\boldsymbol{h})}[\mathbf{r}(\boldsymbol{h}, \boldsymbol{s})] \tag{7}$$

LLM-induced Response Generation. In our work, we leverage black-box LLMs for response generation with manually constructed goal-oriented prompts as the task description for the LLMs to understand the dialogue context and its responsibility on emotional support task. To improve better context-understanding and generation capabilities, we introduce a Zero-shot Chain-of-Thought (CoT) approach for LLM-induced response generation (Kojima et al., 2022). CoT prompts are tailored more closely to the emotional support objective by integrating predictive cues of user emotional states. This allows the LLM to provide a reliable emotional support response along with its reasoning with the instruction of the simulated hints s.

4.4 User Feedback Assessment for Reward

Automatically predicting user emotional states and their associated feedback at each interaction turn poses a significant challenge in Emotional Support Conversation (ESC) tasks, thereby complicating the evaluation and reward design processes. Drawing inspiration from leveraging LLMs as user simulators capable of generating queries, predicting response satisfaction, and forecasting actions, we utilize LMs to assess user feedback. We further integrate this feedback score with the automatic metric ROUGE-L (R-L) (Lin, 2004) for reward.

LLM as User Feedback Predictor. Prior research has relied on human experts to provide task-oriented assessments using multidimensional metrics such as fluency, empathy, and suggestion quality. To ensure a reliable and explainable user simulation, we instruct the large language model (LLM) to embody the role of a help-seeker, articulating their satisfaction with the responses in a stepwise manner. Specifically, we adopt a multidimensional approach to evaluate the quality of ESC responses, employing a 5-star rating system across four key dimensions: (1)**Fluency**: This measures the extent to which the system generates responses that are not only fluent but also easily comprehensible. (2) **Empathy**: This dimension assesses the degree to which the model exhibits appropriate emotional responses, including warmth, compassion, and concern, enhancing the empathetic connection. (3) **Identification**: This evaluates the system's effectiveness in delving into the user's situation to accurately identify the problem at hand. (4) **Suggestion**: This measures the model's ability to offer constructive and helpful suggestions. Following this, we compute the overall feedback by considering the varying weights assigned to each dimension, thereby providing a comprehensive evaluation of response quality.

$$\mathbf{r}_{UFA} = \sum_{j=0}^{n} \lambda_j g_j \tag{8}$$

where λ_j is a hyperparameter to adjust the weighting of each metric, thereby calibrating the influence of individual dimensions on the overall evaluation.

Goal-oriented reward. In ESC task, we define the competency level of the dialogue goal as our reward, which consists of automatic metric ($\mathbf{r}_{R-L}$) and simulated human interactive metric ($\mathbf{r}_{UFA}$).

Model	Training Data	PPL	B-1	B-2	B-3	B-4	R-L
Standard Prompting	-	9.19	14.32	4.21	2.04	1.37	11.46
ProESC	1%	13.25	19.38	7.94	4.36	2.51	14.23
ProESC (*w/o* lookahead)	1%	12.17	17.45	7.19	3.78	2.49	13.39
ProESC (*w/o* user feedback)	1%	13.16	18.33	7.92	3.65	2.40	13.01
ProESC	10%	15.92	**23.61**	**9.93**	**5.82**	**3.17**	**21.53**
ProESC (*w/o* lookahead)	10%	15.45	20.66	9.78	5.31	3.06	21.03
ProESC (*w/o* user feedback)	10%	15.37	21.74	8.79	4.47	2.52	20.63
DialoGPT-Joint (Liu et al., 2021)	100%	-	-	5.00	-	-	15.09
BlenderBot-Joint (Liu et al., 2021)	100%	-	-	5.35	-	-	15.46
MISC (Tu et al., 2022)	100%	**16.16**	-	7.31	-	2.20	17.91
GLHG (Peng et al., 2022)	100%	15.67	19.66	7.57	3.74	2.13	16.37
MultiESC (Cheng et al., 2022)	100%	15.41	21.65	9.18	4.99	3.09	20.41

Table 1: Automatic evaluation results on the response generation. *w/o* lookahead is trained without proactive strategy planning on the fine-tuning policy model, and *w/o* user feedback removes the partition of GPT-4 simulation from current reward score. The strategy planning is conducted on the future 3 turns, which performs the best when k =3 .

This can be mathematically formulated as follows:

$$\mathbf{r}_i = \alpha_1 \mathbf{r}_{R-L} + \alpha_2 \mathbf{r}_{UFA} \tag{9}$$

where $\mathbf{r}_i$ represents the reward for the i-th turn, α_1 and α_2 is the hyperparameter to scale the reward respectively.

To ensure that the policy network π remains closely aligned with the initial policy model, we incorporate a KL-divergence penalty into the reward structure. Consequently, the adjusted reward formulation is as follows:

$$r(\boldsymbol{h}, \boldsymbol{s}) = \mathcal{R}_{\text{LLM}}(\boldsymbol{h}, \boldsymbol{s}) - \beta \log \frac{\pi(\boldsymbol{s} \mid \boldsymbol{h})}{p_{\text{ESC}}(\boldsymbol{s} \mid \boldsymbol{h})} \tag{10}$$

5 Experiments

5.1 Experiment Setup

Dataset. Our research utilizes the ESConv dataset as described in (Liu et al., 2021). ESConv comprises 1,300 extensive dialogues, totaling 38,350 utterances across various emotional support scenarios, which were developed using a crowdsourcing approach. The dataset encapsulates eight distinct types of support strategies. Consistent with the original ESConv dataset partitioning, we adopted an 8:1:1 split for our training, validation, and testing sets, ensuring fidelity to the dataset's intended use for rigorous model evaluation.

Baseline. We compare our method (ProESC) with five state-of-the-art methods and a standard LLM-induced method on the ESConv dataset: **DialoGPT-Joint**, **BlenderBot-Joint** (Liu et al., 2021), **MISC** (Tu et al., 2022), **GLHG** (Peng et al., 2022) and **MultiESC** (Cheng et al., 2022). We also introduce **Standard Prompting** as the baseline model, which design the instruction to let LLMs to reply the previous dialogue history based on task description.

Metrics. For response generation, we employ the following automatic metrics: perplexity (**PPL**), BLEU-1/2/3/4 (**B-1/2/3/4**) (Papineni et al., 2002), ROUGE-L (**R-L**) (Lin, 2004). For strategy planning, we adopt **Accuracy** and **Weighted F1** for automatic evaluation. For human interactive evaluation, we recruit six graduate students with psychological backgrounds as annotators to chat with different models on randomly sampled 100 examples from the test set. These annotators are instructed to select which one performs better (or tie) according to the human evaluation metrics proposed in Liu et al. (2021).

Implementation. We employ T5 (Raffel et al., 2020) as the fine-tuning model for strategy planning and leverage GPT-3.5-turbo (OpenAI, 2021) as the specific LLM which generates response. GPT-4 (Achiam et al., 2023) is utilized as the user simulator that provides user feedback scores.

5.2 Automatic Evaluation of Response Generation

Comparison with Baselines. Our initial investigation focuses on the response generation capabilities of ProESC, setting it against various baseline models for comparison. Table 1 clearly demonstrates

ProESC vs.	MultiESC			BlenderBot-Joint			*w/o* lookahead			*w/o* user feedback		
	win	lose	tie	win	lose	tie	win	lose	tie	win	lose	tie
Fluency	**49.2**‡	36.7	14.1	**61.3**	24.5	14.4	37.8	**41.9**	20.3	**40.2**	26.8	32.9
Identification	**51.9**†	31.2	16.9	**42.2**	40.6	17.2	**37.4**	32.5	30.1	36.1	**38.9**	24.9
Comforting	**62.1**‡	20.6	17.4	**58.4**‡	19.8	21.7	**47.4**†	32.8	19.8	**51.7**‡	26.5	21.9
Suggestion	**69.3**†	14.2	16.5	**59.1**†	29.7	11.2	**46.5**	27.9	25.6	**56.1**†	27.6	16.5
Overall	**64.1**‡	23.8	12.1	**56.2**†	31.9	11.9	**49.5**‡	30.6	19.9	**52.7**†	32.0	15.3

Table 2: Human interactive evaluation results (%). The columns of "Win/Lose" indicate the proportion of cases where ProESC (training with 10% data) wins/loses in the comparison. ‡/† denote p-value < 0.1/0.05 (statistical significance test).

Model	Accuracy	Weighted-F1
DialoGPT-Joint	26.03	23.86
BlenderBot-Joint	29.92	29.56
MISC	31.61	-
MultiESC	42.01	34.01
ProESC(*w/o* lookahead)	41.93	34.09
ProESC$_{k=1}$	42.34	33.92
ProESC$_{k=2}$	42.81	34.76
ProESC$_{k=3}$	**43.57**	**36.23**
ProESC$_{k=4}$	42.90	35.01
ProESC$_{k=5}$	41.92	32.51

Table 3: The strategy planning performance of ProESC and the baseline methods (training with 10% data). Note that k represents anticipating the future k turns strategy.

that ProESC significantly surpasses the standard prompting method that utilizes few-shot training data on a small fine-tuning model. This finding highlights the advantage of our Zero-shot Chain-of-thought prompt design and underscores the efficacy of employing stimulus hints. Remarkably, ProESC outperforms DialoGPT-Joint and BlenderBot-Joint by 2.94% and 2.59% in BLEU-2 (B-2) score respectively, even when trained on just 1% of the data. This achievement across other metrics as well indicates the potential of LLMs to effectively grasp context features with minimal training data. When fine-tuned with 10% of the training data, ProESC not only outshines state-of-the-art (SOTA) methods across most metrics but also secures the second-highest performance in perplexity (PPL). Specifically, it exceeds the performance of the similar lookahead strategy planning method, MultiESC, by 1.96% in BLEU-1 (B-1) and 1.12% in ROUGE-L (R-L). These experimental outcomes affirm the robust in-context few-shot learning capacity and the proficiency of our LLM-based framework in generating effective supportive responses.

Ablation Study. In our ablation study, we assess the impact of removing the lookahead feature and solely relying on the automatic R-L metric for the reward function in our methodology. The results, under both 1% and 10% training data configurations, exhibit a noticeable decline in performance without the lookahead component. This outcome unequivocally confirms the significance of these innovative elements in enhancing the method's effectiveness. Additionally, it was observed that ProESC without the lookahead strategy (ProESC (*w/o* lookahead)) underperforms compared to ProESC without user feedback (ProESC (*w/o* user feedback)) across the board. This discrepancy can be attributed to the fact that user feedback is integrated into the reward function with a specific weighting, whereas the lookahead heuristic plays a more pivotal role in the efficient generation of supportive responses.

5.3 Automatic Evaluation of Strategy Planning

Comparison with Baselines. The efficacy of our strategy planning approach is detailed in Table 3, where the advantages of proactive strategy planning, through the anticipation of future support strategies, are evident. Our method outperforms all other models tested, showcasing superior performance. Specifically, ProESC demonstrates significant improvements over baseline methods in both Accuracy and Weighted F1 metrics. Notably, when forecasting up to three future dialogue turns, ProESC exceeds the performance of the SOTA strategy planning method, MultiESC, by margins of 1.56% and 2.22% in Accuracy and Weighted F1, respectively. This highlights the effectiveness of our approach in leveraging anticipatory strategy planning to enhance support strategy identification and implementation.

Ablation Study. In our exploration of the strategic planning method's variants, we considered

Bot: What can I help you with? **User**: Lately, my friends started to make jokes about my face being ugly. They even called my face as flat as a pan! What should i do? I see them everyday! **Ground-Truth**: (*Affirmation and Reassurance*) I'm sure you are very handsome! They could just be jealous.
Standard Prompting: It must be hard. Just ignore their words and be yourself. **BlenderBot-Joint**: You should tell them to calm down. **MultiESC**: (*Question*) That's horrible! Have you tried talking to them about it? **ProESC**: (*Affirmation and Reassurance)*) I'm sure they must be joking. You look pretty good! **ProESC (*w/o* lookahead)**: (*Reflection of Feelings)*) It's must be very hard for you to deal with that. **ProESC (*w/o* user feedback)**: (*Question*) Have you ever stop them from calling you like this?

Table 4: Case study. The upper part is an example of the dialogue history in the test set of ESConv. The lower part shows the responses from different models.

different lengths for future turn predictions, extending up to 5 turns. As illustrated in Table 3, ProESC (*w/o* lookahead) focuses solely on the strategy applicable to the current turn. Conversely, ProESC$_{k=?}$ is designed to forecast the sequence of strategies over the next k turns, implying that the total sequence length is $k+1$, inclusive of the current strategy. The data presented in Table 3 indicate that our method outperforms when $k = 2$ and $k = 4$, and reaches its peak performance at $k = 3$. However, we observed a decrease in accuracy for predictions extending to 4 and 5 future turns, attributable to the complexities associated with decoding longer-term sequences. Based on these findings, we opted for $k = 3$ as our preferred setting for subsequent comparative experiments, optimizing our approach for this specific future turn length.

5.4 Human Interactive Evaluation

Recognizing the limitations of automatic evaluation for the ESC task, we complemented our assessment with human evaluations. To this end, we enlisted human experts to evaluate the competing systems, focusing on four critical metrics: Fluency, Comforting, Identification, and Suggestion. The results, detailed in Table 2, reveal that ProESC surpasses the competitive method MultiESC across all evaluated metrics. Moreover, ProESC demonstrates superior performance compared to BlenderBot-Joint, particularly on the latter metrics related to support and empathy. These areas are vitally important to the ESC task, underscoring ProESC's adeptness in handling the nuanced aspects of providing emotional support and empathy through conversational AI. In our ablation study, the observed performance advantage of ProESC over its ablated versions is substantial, clearly demonstrating the effectiveness of our methodology's key components.

5.5 Case Study

Table 4 presents a case study derived from the test set, wherein we compare the responses generated by baseline models against our ProESC framework. When provided with standard task-specific instructions, ChatGPT produces a response that lacks empathy and offers a suggestion that is not meaningful. Meanwhile, although MultiESC and BlenderBot-Joint manage to provide helpful support or delve into the user's thoughts, they fall short in selecting the appropriate strategy that aligns with the ground truth answer. In contrast, ProESC demonstrates a remarkable ability to identify the correct strategy, steering the system towards generating responses that are not only supportive but also more empathetic and helpful than those of the ablation models. Additionally, ProESC and its variants are capable of generating responses that are more closely aligned with the current topic and offer more concrete suggestions compared to other baseline models.

6 Conclusion

In our study, we introduce a pioneering approach for generating responses in proactive emotional support conversations (ProESC), leveraging large language models (LLMs) and incorporating look-forward strategy planning. This approach is underpinned by a fine-tuning policy model designed to predict future supportive strategies, thereby facilitating improved long-term strategic planning within a reinforcement learning framework. To further refine the evaluation of response quality, we integrate GPT-4-based predictions of user feedback as part of a composite reward mechanism, aiming for a more realistic and goal-oriented assessment of conversational outcomes. Empirical results have achieve competitive performance both response generation and strategic planning compared with SOTA methods.

Limitations

While our proposed method demonstrates competitive outcomes in the Emotional Support Conversation (ESC) domain, it's imperative to approach the practical application of LLMs with increased scrutiny. In our research, we leverage LLMs as a tool for generating responses, akin to a black-box utility, without delving into the potential enhancements achievable through fine-tuning with domain-specific expertise in emotional support. This oversight suggests that incorporating expert knowledge in emotional support into the fine-tuning process of LLMs could yield even superior performance. Furthermore, the aspects of safety and privacy in the context of LLM-enhanced ESC require thorough examination to ensure that these systems do not inadvertently compromise user confidentiality or propagate harmful biases. Additionally, there's a significant avenue for research in developing personalized and adaptive emotional support conversations. Such tailored interactions have the potential to profoundly impact psychological therapy and mental health support.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Brant R Burleson. 2003. Emotional support skills. In *Handbook of communication and social interaction skills*, pages 569–612. Routledge.

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

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, Wenqiang Lei, Lizi Liao, and Tat-Seng Chua. 2023b. Prompting and evaluating large language models for proactive dialogues: Clarification, target-guided, and non-collaboration. *arXiv preprint arXiv:2305.13626*.

Yang Deng, Wenxuan Zhang, Yifei Yuan, and Wai Lam. 2023c. Knowledge-enhanced mixed-initiative dialogue system for emotional support conversations. *arXiv preprint arXiv:2305.10172*.

Adam M Grant and Susan J Ashford. 2008. The dynamics of proactivity at work. *Research in organizational behavior*, 28:3–34.

Catherine A Heaney and Barbara A Israel. 2008. Social networks and social support. health behavior and health education: Theory, research, and practice. *Health Behavior and Health Education: Theory, Research, and Practice*, pages 189–210.

Clara E Hill. 2009. *Helping skills: Facilitating, exploration, insight, and action*. American Psychological Association.

Zhiyuan Hu, Yue Feng, Yang Deng, Zekun Li, See-Kiong Ng, Anh Tuan Luu, and Bryan Hooi. 2023a. Enhancing large language model induced task-oriented dialogue systems through look-forward motivated goals. *arXiv preprint arXiv:2309.08949*.

Zhiyuan Hu, Yue Feng, Anh Tuan Luu, Bryan Hooi, and Aldo Lipani. 2023b. Unlocking the potential of user feedback: Leveraging large language model as user simulator to enhance dialogue system. *arXiv preprint arXiv:2306.09821*.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213.

Tin Lai, Yukun Shi, Zicong Du, Jiajie Wu, Ken Fu, Yichao Dou, and Ziqi Wang. 2023. Psy-llm: Scaling up global mental health psychological services with ai-based large language models. *arXiv preprint arXiv:2307.11991*.

Zekun Li, Baolin Peng, Pengcheng He, Michel Galley, Jianfeng Gao, and Xifeng Yan. 2023. Guiding large language models via directional stimulus prompting. *arXiv preprint arXiv:2302.11520*.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

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.

Lenin Medeiros and Tibor Bosse. 2018. Using crowdsourcing for the development of online emotional support agents. In *Highlights of Practical Applications of Agents, Multi-Agent Systems, and Complexity: The PAAMS Collection: International Workshops of PAAMS 2018, Toledo, Spain, June 20–22, 2018, Proceedings 16*, pages 196–209. Springer.

OpenAI. 2021. Chatgpt: Openai's conversational ai.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Wei Peng, Yue Hu, Luxi Xing, Yuqiang Xie, Yajing Sun, and Yunpeng Li. 2022. Control globally, understand locally: A global-to-local hierarchical graph network for emotional support conversation. *arXiv preprint arXiv:2204.12749*.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 2020. A computational approach to understanding empathy expressed in text-based mental health support. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5263–5276.

Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. 2022. Misc: A mixed strategy-aware model integrating comet for emotional support conversation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 308–319.

Anuradha Welivita and Pearl Pu. 2020. A taxonomy of empathetic response intents in human social conversations. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4886–4899.

Qiang Zhang, Jason Naradowsky, and Yusuke Miyao. 2023. Ask an expert: Leveraging language models to improve strategic reasoning in goal-oriented dialogue models. *arXiv preprint arXiv:2305.17878*.