# ESCA: An Emotional Support Conversation Agent for Enhancing Reasonable Strategy Planning and Effective Expression

**Jing Li**[1,2], **Yanxin Luo**[1], **Donghong Han**[1*], **Yimeng Zhan**[1], **Xiaoming Fu**[2*], **Baiyou Qiao**[1], **Gang Wu**[1]

[1]School of Computer Science and Engineering, Northeastern University, 110169, China
[2] Institute of Computer Science, University of Göttingen, 37073, Germany
{2110662, 2401877}@stu.neu.edu.cn, handonghong@cse.neu.edu.cn, 2310754@stu.neu.edu.cn, fu@cs.uni-goettingen.de, {qiaobaiyou, wugang}@mail.neu.edu.cn

## Abstract

Emotional Support Conversation (ESC) aims to alleviate individuals' negative emotions through multi-turn dialogues, where effective strategy planning and response generation are essential. However, existing methods often suffer from limitations in both planning reasonable support strategies and effectively expressing them in responses. To the end, we propose a novel LLM-based Emotional Support Conversation Agent (ESCA) with a plug-in strategy planner and a strategy-aligned prompt generator. The strategy planner cooperates with four aspects of the seeker's state, including emotion intensity, trust degree, dialogue behavior, and stage of change, to enhance the rationality and effectiveness of the strategy prediction. To ensure that predicted strategies are better conveyed, the prompt generator integrates strategy-aligned instructions, knowledge, and context to generate the soft prompt for guiding the LLM to generate supportive responses. In addition to supervised fine-tuning, the prompt generator is further optimized by reinforcement learning. Experimental results demonstrate that ESCA significantly improves both response quality and the success rate of achieving the ESC task goal.

**Code and dataset** — https://github.com/outsider-lj/ESCA

## Introduction

Emotional support aims to alleviate individuals' emotional distress and assist them in understanding and coping with their challenges (Liu et al. 2021). Emotional Support Conversation (ESC) extends this concept to human-machine interaction, enabling dialogue systems to deliver empathetic, supportive responses through multi-turn conversations. Developing effective ESC systems holds great promise for applications such as emotional companionship, psychological counseling, and customer service.

The global goal of ESC is to reduce the seeker's negative emotions through multi-turn interactions. To achieve this, it's crucial for supporters to select appropriate support strategies and generate responses that can effectively express these strategies. **For the strategy planning**, recent research mainly focuses on modeling the context, topic to predict the support strategy (Deng et al. 2024; He et al. 2024) or utilizing the seeker's emotion change as feedback to optimize the strategy planning (Peng et al. 2023; Zhou et al. 2023). However, these studies have overlooked that strategy selection requires considering more aspects than just context and emotion. As illustrated in Figure 1, when a seeker displays intense negative emotion and low trust in the early stage of the conversation, strategies such as affirmation and reassurance are more appropriate for stabilizing emotions and building trust. The strategy, like providing suggestions, may have a counter-effect. As trust grows and the seeker progresses into the stage of preparation for change, directive strategies become more appropriate. Therefore, based on the real conversations and relevant theories about helping skills mentioned in (Hill 1999) (e.g., the seeker's readiness to engage in self-exploration or action determines the level of directiveness in helping), we argue that the strategy plan-

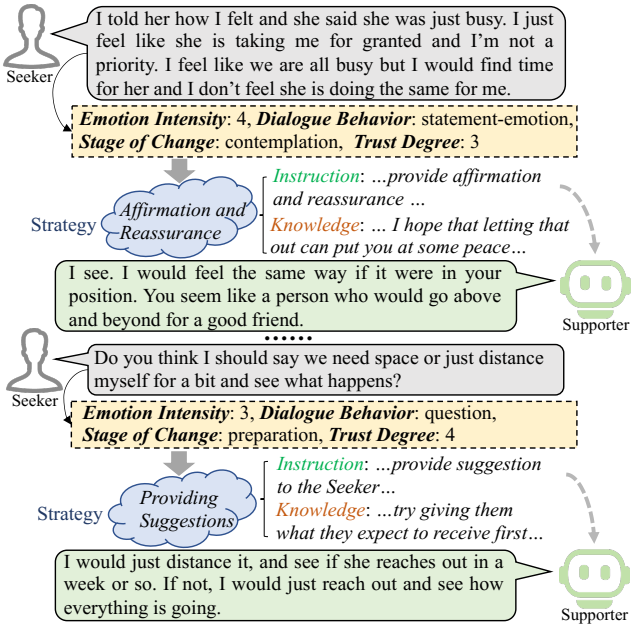


Figure 1: An example of ESC. Yellow boxes show four aspects of the seeker's state. Based on this, the supporter selects strategies (blue boxes) and conducts instructions and knowledge to generate responses.

*Corresponding author

ning requires considering more comprehensive state aspects. Besides the emotion intensity, we introduce novel state aspects, including trust degree, dialogue behavior, and stage of change in ESC. In this way, it also improves the interpretability of strategy planning. But there is no existing ESC dataset containing these labels. To address this gap, we utilize Large Language Models (LLMs) to automatically annotate the common dataset ESConv (Liu et al. 2021) with these labels under the unified rules. Based on the enriched dataset, we propose a novel strategy planner considering multiple state aspects for prediction.

**For the response generation**, it is the directive expression of emotional support. Deng et al. (2024) show that stronger generation capabilities can improve goal achievement with the same strategy planner. Recently, some researches introduce knowledge (Tu et al. 2022; Deng et al. 2023b) and response exemplars (Xu, Meng, and Wang 2022) to improve the response quality including the diversity and relevance. However, few studies focus on how well responses express the selected strategy, i.e., whether they provide informative responses under providing suggestions, or show empathy in affirmation strategies. Despite LLMs' strong generation ability, responses prompted only by strategy instructions often lead to weak strategy expression and limited supportive content. The reason for this is that general LLMs have not learned how to express strategies in the ESC scenario, and the simple prompt is difficult to fully elicit its ability. To address this limitation, we introduce a prompt generator to cooperate the strategy-aligned, context-aware instruction and knowledge to generate the soft prompt for better stimulating the LLM to generate content-rich and strategy-relevant responses. Additionally, we employ the online Reinforcement Learning (RL) to optimize the prompt generator for further enhancing the response performance.

In this paper, we propose a novel LLM-based **E**motional **S**upport **C**onversation **A**gent (**ESCA**) that enhances the capability of both strategy planning and response generation for the ESC tasks. Our key contributions are as follows: 1) We introduce multiple novel aspects of the seeker's state to improve the rationality and effectiveness of strategy planning. To train the strategy planner, we annotate the existing ESConv dataset with these aspect labels using LLMs. 2) We model a prompt generator that incorporates strategy-aligned, context-aware instruction and knowledge to help LLMs generate more strategy-aligned and content-rich responses. The online RL is further used to optimize the prompt generator for enhancing the strategies' expression. 3) We integrate the strategy planner and prompt generator into a base LLM to get a novel ESC Agent. Comprehensive experiments show that our approach improves response quality and increases the success rate of achieving dialogue goals.

## Related Work

### Strategy Planning

Strategy planning is vital for goal-oriented dialogue generation, and its research can be categorized into prompt-based, search-based, and RL-based approaches. Prompt-based methods guide LLMs using designed prompts without extensive training. For instance, mixed-initiative prompting (Chen et al. 2023a), self-reflection (Zhang, Naradowsky, and Miyao 2023), and chain-of-thought prompting (Deng et al. 2023a) aim to improve strategic responses. However, LLMs tend to talk passively, and the capabilities of prompt-based methods are still poor. Search-based approaches address this by selecting strategies from predefined sets. GDP-ZERO (Yu, Chen, and Yu 2023) uses Open-Loop MCTS with LLMs as policy priors and simulators, while Conversational Tree Search (Väth, Vanderlyn, and Vu 2023) integrates FAQ retrieval with task-oriented dialogue. Search-based methods are often complex, requiring consideration of multiple conditions. RL-based approaches optimize dialogue policies by maximizing long-term rewards. PPDPP (Deng et al. 2024) introduces a small plug-in language model trained with supervised and reinforcement learning for better planning. DPDP (He et al. 2024) combines a policy network with MCTS for complex decisions. Although an increasing number of studies are focusing on strategy planning, there is still a lack of attention to interpretability, and performance still needs improvement.

### Emotional Support Response Generation

The ESC task is proposed by Liu et al. (2021), along with the release of the ESConv dataset. Research on ESC has three main directions: understanding seekers' situations, enhancing strategy planning, and improving response quality.

Many studies improve understanding of the seeker's situation by incorporating emotional causes, persona information, or external knowledge. KEMI (Deng et al. 2023b) integrates COMET (Bosselut et al. 2019) and Heal (Welivita and Pu 2022) to enrich context. CauESC (Chen et al. 2024) captures the emotion-causes relations with attention. PAL (Cheng et al. 2023) utilizes co-attention to highlight persona information in context. And DKPE (Hao and Kong 2025) further filters relevant knowledge dynamically to improve contextual understanding.

Strategy planning is beneficial for emotion alleviation in long-term interactions. FADO (Peng et al. 2023) combines turn-level and dialogue-level feedback to predict strategies. SUPPORTER (Zhou et al. 2023) designs emotion support and dialogue coherence rewards to guide the strategy's learning for responding. DSR (Liu et al. 2025) captures seekers' intentions to conduct the strategy prompt for generation. However, these approaches overlook other influential factors, except context and emotion, in long-term planning.

To enhance response quality, some works integrate prophetic commonsense generated by LLMs (Wang et al. 2023) or external knowledge (Chen et al. 2023b) to improve responses. Other studies retrieve exemplars through a one-to-many strategy relationship (Xu, Meng, and Wang 2022) or employ in-context learning (Xu et al. 2024), where exemplars guide generation. COOPER (Cheng et al. 2024) leverages multiple agents with distinct stage goals to guide LLM generation. However, these approaches often overlook whether responses effectively convey support strategies, which is crucial for goal fulfillment. Despite LLMs' strong generative capabilities, prompt engineering alone struggles to elicit strategy-aligned and informative responses in ESC.

# Data Preparation

## State Definition

In this work, we model four key aspects of the seeker's state to dynamically capture deeper interaction features for enhancing strategy planning: emotion intensity, trust degree, stage of change, and dialogue behavior. **Emotion intensity** is commonly used in the ESC task. The ESConv dataset also provides the emotion intensity annotations of the first and last conversation turns. Emotion intensity is measured on a discrete scale from 1 to 5, with higher scores indicating more negative emotions. **Trust degree** refers to the extent to which a user is confident in, and willing to rely on, the recommendations, actions, and decisions of an artificial intelligence-based decision aid (Madsen and Gregor 2000). Following the framework proposed by (Madsen and Gregor 2000), we focus on the cognition-based trust in our work and assess it from three dimensions: the supporter's reliability, the response competence of the supporter, and the seeker's perceived understandability. These dimensions are all on the scale from 1 to 5. The mean of these three scores will be converted to a five-level trust degree. **Stage of change** is based on the Transtheoretical Model (Prochaska, Norcross, and DiClemente 1994, 2005), which provides a structured framework for understanding behavioral change. The five stages (*precontemplation, contemplation, preparation, action, and maintenance*) serve as a useful perspective for interpreting the seeker's readiness to change. **Dialogue behavior** plays an important role in understanding the seeker's interaction style. Inspired by the ESC setting and the categorization proposed by (Saha et al. 2020), we adopt the following behavior categories: *greeting, question, feedback, statement-fact, statement-opinion, statement-emotion, command, acknowledgement, and others.*

## State Annotation

Because of the lack of above four aspects, for training the strategy planner, we annotate these labels in the common dataset ESConv. Recently, there have been many works that utilize LLM to extend the dataset (Zheng et al. 2022) or annotate the dataset (Bagdon et al. 2024). Using LLM for annotation can save labor costs and better ensure the stability of annotation. Therefore, we utilize LLMs to annotate above four aspects. In order to obtain more accurate annotation results, we use one-shot in-context learning to annotate these four state aspects with the fixed rule and example. Additionally, for ensuring the quality of annotation, three LLMs (Qwen2.5-72b (Bai et al. 2023), LLaMA3.3-instruct-70b (Grattafiori et al. 2024), Deepseek-distill-llama-70b (Guo et al. 2025)) are independently used to annotate the dataset, and their outputs are integrated using a voting mechanism. Among the three results, the result with the largest number of occurrences is the final result. If the three results are different, Deepseek-r1 (Guo et al. 2025) is introduced for further annotation as the final result. We evaluate annotation consistency using Fleiss' kappa scores (Fleiss and Cohen 2016), which are 0.60 (emotion intensity), 0.58 (stage of change), 0.55 (behavior), and 0.33 (trust), indicating moderate to fair agreement depending on the aspect. Additionally, we randomly sample 100 annotated instances and conduct human validation by three annotators, with the average of the annotation accuracy about six dimensions (Trust has three dimensions) reported as 89.67%, 87.67%, 94.33%, 94.67%, 94% and 94.33%.

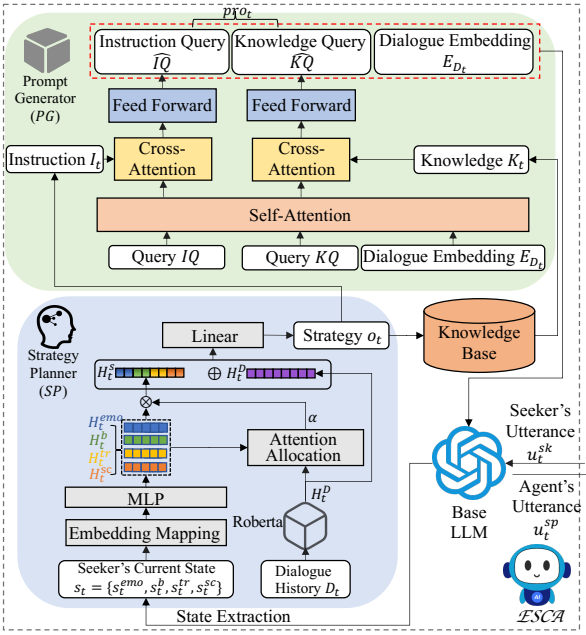


Figure 2: The framework of the ESCA. It includes a strategy planner, a prompt generator and a base LLM.

# Methodology

## Preliminaries

**Problem Formalization** At the $t$ turn, given the dialogue history $D_t = \{u_1^{sk}, u_1^{sp}, \ldots, u_i^{sk}, u_i^{sp}, \ldots, u_t^{sk}\}$, where the $u_i^{sk}$ and $u_i^{sp}$ mean the seeker and supporter utterances of the $i$-th turn, ESCA extracts the current state $s_t = \{s_t^{emo}, s_t^b, s_t^{tr}, s_t^{sc}\}$, where $s_t^{emo}$, $s_t^{tr}$, $s_t^b$, and $s_t^{sc}$ denote the emotion intensity, trust degree, dialogue behavior and stage of change respectively. Based on $s_t$, ESCA selects a strategy $o_t \in O$, where $O$ is a pre-defined set of emotional support strategies. Following $o_t$, ESCA generates an appropriate emotional support response $u_t^{sp}$. In return, the seeker will reply with the next utterance $u_{t+1}^{sk}$. This interaction process continues until either the dialogue goal is achieved or the maximum number of turns $T_d$ is reached.

**Overview** In this paper, the ESCA contains three main parts: a plug-in strategy planner ($SP$), a prompt generator ($PG$) and a base LLM. The framework of ESCA is shown in Figure 2. Firstly, the base LLM extracts the seeker's current state $s_t$ based on the dialogue history $D_t$. Then, the $SP$ selects a strategy $o_t$ condition on $D_t$ and $s_t$. To better express the strategy $o_t$, we retrieve relevant external knowledge $K_t$ and conduct the instruction $I_t$. The $PG$ models $I_t$, $K_t$ and $D_t$ to generate the soft prompt which enters into the base LLM to get the final response $u_t^{sp}$.

## Strategy Planner

Inspired by (Deng et al. 2024), instead of relying on LLMs to generate strategies, we adopt a lightweight model to explicitly learn the features of strategy planning in ESC, which offers more flexibility in optimization. However, unlike prior work that overlooks the complexity of strategy selection, we propose a model that considers multiple aspects of the seeker's state $s_t$ alongside the current dialogue history $D_t$ to predict an appropriate support strategy. Following the rules defined in data annotation, the base LLM of ESCA extracts four aspects represented by four discrete categorical variables: $s_t^{emo}, s_t^b, s_t^{tr}, s_t^{sc}$. They are encoded via the embedding layer and the multi-layer perception (MLP) layer to convert into four feature vectors: $H_t^{emo}, H_t^b, H_t^{tr}, H_t^{sc}$. We adopt RoBERTa (Liu et al. 2019) to encode the dialogue history $D_t$, extracting the hidden state of the [CLS] token as $H_t^D$ to present the current context semantics. Based on $H_t^D$, the attention layer calculates the attention weights of four aspects in the current turn, and then takes the weighted sum of them to get the final state feature $H_t^s$:

$$\alpha_i = \frac{\exp((W_d H_t^D)^T (W_s H_t^{asp}))}{\sum_{asp \in \{emo,tr,b,sc\}} \exp((W_d H_t^D)^T (W_s H_t^{asp}))} \quad (1)$$

$$H_t^s = \sum_{asp \in \{emo,tr,b,sc\}} \alpha_i (W_v H_t^{asp}) \quad (2)$$

$$P(\hat{o}_t) = softmax(\text{FFN}([H_t^S : H_t^D])) \quad (3)$$

FFN means the feed-forward network. $W_d, W_s, W_v$ are trainable weight matrices for attention and fusion. We concatenate $H_t^D$ with $H_t^S$ and feed the result into the FFN layer, followed by a softmax layer to obtain the probability distribution over all strategies. The strategy with the highest probability in $P(\hat{o}_t)$ is selected as the predicted strategy.

## Prompt Generator

Due to the lack of learning from ESC data, LLMs often struggle to accurately express support strategies and provide effective emotional support. Therefore, we first introduce multiple types of external knowledge to supplement strategy-relevant information. Then, the $PG$ models the strategy-aligned, context-aware knowledge feature and instruction feature for generating the soft prompt. It serves as the bridge to overcome the gap between ESC responses and general responses. Finally, the base LLM generates the final response conditioned on the above soft prompt.

**Knowledge Extraction** The support strategies in ESConv are categorized into three types based on their focus: context-related (question, restatement or paraphrasing, others), emotion-related (affirmation and reassurance, reflection of feeling), and information-related (self-disclosure, providing suggestions, information). Each strategy type requires distinct external knowledge to enhance its expression. Specifically, PsyQA (Sun et al. 2021), presented in a question–answer format, and ER (Sharma et al. 2020), composed of single-turn dialogues, are applied to information-related and emotion-related strategies, respectively. To facilitate effective knowledge retrieval, we construct a vector database using Sentence Transformers (Nikolaev and Padó 2023) on user queries in the knowledge base. The agent retrieves the top-$k$ most relevant knowledge samples as $K_t$ by computing the cosine similarity between the semantic embeddings of the seeker's current utterance and the user queries . For context-related strategies, $K_t$ is defined as the summary of the current turn's core topic, generated by the base LLM to provide contextual cues.

**Prompt Generation** Inspired by the architecture of Q-former (Li et al. 2023), we propose a prompt generator to extract two context-aware query features: knowledge prompt and instruction prompt, based on $D_t$, $K_t$, and $I_t$, where $I_t$ is mapped based on the predicted strategy $\hat{o}_t$. Firstly, $D_t$, $K_t$, and $I_t$ are embedded into $E_{D_t}$, $E_{K_t}$, and $E_{I_t}$ using the embedding layer of the base LLM. Here, two learnable query embeddings $IQ, KQ \in \mathbb{R}^{(L \times dim)}$ are introduced, where $dim$ is the hidden size of LLM's hidden state, $L$ is the query fixed length. We concatenate the queries and dialogue embeddings to form the input sequence $[IQ; KQ; E_{D_t}]$. This sequence is passed through a self-attention layer to allow interaction among the queries and the dialogue history. The outputs $KQ'$ and $IQ'$ are then entered into two cross-attention operations. The knowledge cross-attention uses the updated $KQ'$ as the query to attend over the knowledge embedding $E_{K_t}$. The resulting output is combined with the original query through a residual connection, followed by layer normalization and a feed-forward network to refine the representation. The instruction cross-attention follows the same process, using $IQ'$ to attend over the instruction embedding $E_{I_t}$. The process can be formulated as:

$$\hat{KQ} = \text{FFN}(\text{Norm}(KQ' + \text{Cross Attn}(KQ', E_{K_t}))) \quad (4)$$

$$\hat{IQ} = \text{FFN}(\text{Norm}(IQ' + \text{Cross Attn}(IQ', E_{I_t}))) \quad (5)$$

Finally, the soft prompt $pro_t = [\hat{IQ}; \hat{KQ}]$ contacting with $E_{D_t}$ is used as input to the base LLM for final generation.

## Training Strategy

The training process consists of two stages: supervised fine-tuning on labeled data and online RL. During the **supervised learning**, the strategy planner and the prompt generator are trained separately using the following objectives:

$$\mathcal{L}_p = -P(o_t) \log P(\hat{o}_t) \quad (6)$$

$$\mathcal{L}_{u_t} = -\sum_{n=1}^{N} \log \left( P\left( y_n \mid D_t, K_t, I_t, y_{<n} \right) \right) \quad (7)$$

Here, the strategy loss $L_p$ is the cross-entropy loss between the predicted strategy distribution $P(\hat{o}_t)$ and the true strategy distribution $P(o_t)$. The generation loss $L_{u_t}$ is the standard negative log-likelihood, where $y_n$ is the $n$-th response token. We only update the parameters of the prompt generator and keep the base LLM frozen.

After initial supervised training, we observe that the prompt generator still struggles to fully capture the strategy expression patterns. To address this, we further optimize $PG$

through **online RL**, where an advanced LLM agent acts as the seeker and interacts with our ESCA. This interactive setting enables $PG$ to learn from more diverse, flexible knowledge and conversation data.

The process of prompt generation can be formulated as a reinforcement learning problem, where the state $sta_t^{pg}$ comprises $K_t$, $I_t$ and $D_t$, the action is the soft prompt $pro_t$, the $PG$ serves as the stochastic policy network. Because $pro_t$ is the high-dimension vector, to obtain the distribution of the action, we employ the reparameterized Gaussian policy, where the output vector (soft prompt) serves as the mean, the variance $\phi$ is calculated by a learnable log-standard deviation. The action is sampled via the reparameterization trick to maintain differentiability:

$$pro_t = PG_\theta(sta_t^{pg}) + \sigma \odot \phi, \quad \phi \sim \mathcal{N}(0, I) \tag{8}$$

We also utilize LLM agent as the critic model to generate the rewards $r_t$ whether the responses perform strategies well. The scores 1, 0.5, and 0 are regarded as good performance, poor performance, and not following strategy respectively. After $T_d$ turns of interaction, the $PG$ is trained by the Proximal Policy Optimization (PPO) algorithm (Schulman et al. 2017), which employs a clipped surrogate objective to help stabilize training by preventing large policy updates. The policy loss $L_g$ is calculated as follows:

$$L_g = \mathbb{E}_t \left[ \min \left( r_g(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t \right) \right] \tag{9}$$

$$\hat{A}_t = \delta_V \left( sta_t^{pg} \right) + \gamma\lambda A_{\theta_l^{\text{old}}} \left( sta_{t+1}^{pg}, pro_{t+1} \right) \tag{10}$$

$$\delta_V(sta_t^{pg}) = r_t + \gamma V_\beta \left( sta_{t+1}^{pg} \right) - V_\beta \left( sta_t^{pg} \right) \tag{11}$$

where $r_g(\theta) = \frac{PG_\theta(pro_t|sta_t^{pg})}{PG_\theta^{\text{old}}(pro_t|sta_t^{pg})}$ is the ratio between current policy and old policy. $\delta$ represents the temporal difference of value network $V_\beta$, computed using the updated and current critic parameters. $\gamma$ is the discount factor. We utilize the FFN as $V_\beta$, taking the average-pooled $\hat{IQ}$ and $\hat{KQ}$ as input to estimate the expected total reward for the current state $sta_t^{pg}$. To train $V_\beta$, we minimize the mean squared error (MSE) between the predict value and the return $R_t$ calculated by the discounted cumulative reward, and the loss function is defined as:

$$L_{value} = \mathbb{E}_t \left[ \left( V_\beta \left( sta_t^{pg} \right) - R_t \right)^2 \right] \tag{12}$$

The total loss of the online RL combines the policy loss $L_g$ and the value loss $L_{value}$.

# Experimental Setting

## Comparison Models

To demonstrate the effectiveness of our model, we compare it with several state-of-the-art models in both response generation and strategy planning. To evaluate the performance of the response generation, we include PAL (Cheng et al. 2023), CauESC (Chen et al. 2024), D$^2$RCU (Xu et al. 2024), COOPER (Cheng et al. 2024) and DKPE (Hao and Kong 2025) as comparison models. In terms of strategy planning, we compare against PPDPP (Deng et al. 2024) and DPDP (He et al. 2024) two advanced methods which add a plug-in strategy planner to LLM.

## Evaluation Metrics

**Evaluation of Response Generation** For automatic evaluation, we report Distinct-1 (**D-1**), Distinct-2 (**D-2**) and Distinct-3 (**D-3**) (Li et al. 2016) to measure the response diversity. We utilize BLEU-3 (**B-3**), BLEU-4 (**B-4**) (Papineni et al. 2002) and ROUGE-L (**R-L**) (Lin 2004) to evaluate the similarity between generated responses and golden responses. In addition, we randomly choose 100 conversations from test dataset for human evaluation following (Zhao et al. 2023). Three annotators compare the generated responses between ESCA and comparison models from four perspectives: Suggestion (Sug.), Identification (Ide.), Comforting (Com.), and Overall (Ove.) to select the better responses. In case of disagreement among all three annotators, a fourth is introduced. We report the percentage of Win, Lose, and Tie for each aspect.

**Evaluation of Strategy Planning** Following the evaluation in (Deng et al. 2024; He et al. 2024), **Average Turn (AT)** and **Success Rate at turn $t$ (SR@$t$)** are used to evaluate the efficiency of strategy planning. AT reflects the average number of turns required to reach the goal. SR@$t$ measures the success rate within a predefined maximum number of turns, which is set to 10 in our experiments.

## Experimental Details

We conduct experiments on the extended ESConv dataset (Liu et al. 2021), annotated with four state aspects as described in the Data Preparation. The dataset is split into training, validation, and test sets with an 8:1:1 ratio. All experiments are run on a single Nvidia H100 GPU. The strategy planner is trained on RoBERTa-large with a learning rate of 1.5e-6 for 5 epochs, while the prompt generator is fine-tuned with a learning rate of 6e-6 for 10 epochs. LLaMA2-7b-chat serves as the base LLM for state extraction and response generation. During online RL, situations from the training set are sampled to let the LLM agent simulate seekers with emotional problems. The seeker's responses and critic values are provided by LLaMA3.3-70b-instruct via the Chat AI platform (Doosthosseini et al. 2024). The learning rate is set to 1e-6 and $\epsilon$ to 0.2. Comparison models are reproduced using their reported hyperparameters, with the temperature of 0.5. The generation performance results are averaged over three independent runs.

# Results and Analysis

## Overall Performance

Our research focuses on both the strategy planning and response generation for improving the emotional support capability of the ESCA. Therefore, we analyze the performance of response generation and strategy planning.

**Generation Performance** For evaluating the expression capability of ESCA, the automatic evaluation results are shown in Table 1. Our model achieves significantly higher Distinct scores compared to other models, indicating that the responses generated by ESCA are content-rich. After fine-tuning a prompt generator, ESCA achieves a relatively

| Models | D-1↑ | D-2↑ | D-3↑ | B-3↑ | B-4↑ | R-L↑ |
|---|---|---|---|---|---|---|
| PAL | 4.10 | 22.89 | 44.56 | 4.00 | 2.64 | 17.60 |
| CauESC | 3.16 | 20.02 | 37.90 | 3.48 | 2.24 | 15.82 |
| $D^2$RCU | 5.59 | 28.89 | 53.03 | **4.61** | **2.68** | 18.12 |
| DKPE | 5.40 | 27.46 | 50.26 | 4.40 | 2.53 | **18.51** |
| COOPER | 5.62 | 26.83 | 47.06 | 2.81 | 1.52 | 14.79 |
| **ESCA** | **6.15** | **30.51** | **55.82** | 4.24 | 2.50 | 17.87 |

Table 1: Results of the automatic evaluation on generation quality. The best and second-best results from all models are highlighted in bold and underlined respectively.

| **ESCA** vs | $D^2$RCU | | | COOPER | | |
|---|---|---|---|---|---|---|
| | Win | Lose | Tie | Win | Lose | Tie |
| Ide. | 52 | 40 | 8 | 70 | 18 | 12 |
| Com. | 53 | 37 | 10 | 69 | 26 | 5 |
| Sug. | 38 | 30 | 32 | 64 | 11 | 25 |
| Ove. | 50 | 44 | 6 | 70 | 18 | 12 |

Table 2: Results of the human evaluation on generation quality (all value in %). The Fleiss' kappa score (Fleiss and Cohen 2016) reaches 0.42, indicating a fair agreement.

high R-L and B-4 score, which means the responses are more similar to human's responses. However, COOPER utilized related topic candidates to help LLM generation, its responses exhibit relatively low consistency with real human responses. These results suggest that ESCA can provide rich content and also internalize the human response patterns in ESC. In addition, we select two models for further human evaluation, the results are shown in Table 2. COOPER, based on LLaMA-13b, its responses shows low win rates across all four human evaluation metrics. Compared with $D^2$RCU, which performs well in automatic evaluations, our model demonstrates a clear advantage in comforting ability and a slight advantage in the other metrics. The higher tie ratio in suggestion occurs mainly because, in some cases, neither model provides a suggestion.

**Strategy Planning Performance** For evaluating the strategy planning capability of the ESCA, we select current SOTA models on the strategy planning task as baselines and the results are shown in Table 3. All models use the LLaMA3.3-70b-instruct to determine whether the goal has been achieved after each round of interaction following the prompt as (He et al. 2024). ESCA achieves a significant improvement in SR@10 and maintains a lower AT relative to DPDP. Despite being based on the smaller LLaMA2-7b-chat model, ESCA consistently outperforms other methods.

## Ablation Analysis

To assess the contribution of each component, we conduct ablation experiments, with results summarized in Table 4. Firstly, we replace our strategy planner with a standard RoBERTa-large model without any seeker state information. This leads to a decrease in SR and an increase in AT, confirming that the seeker state information enhances strategy planning and helps goal achievement. Secondly, we train a prompt generator using only the instruction part without external knowledge. The results show a slight decline in response quality, and the SR and AT are also influenced. It indicates that external knowledge also works in our ESCA, and the generation capability is also essential for the final goal achievement. Thirdly, to evaluate the necessity of training the prompt generator, we remove $PG$ and directly insert retrieved knowledge into the prompt. This leads to a noticeable drop across all metrics as the responses become lengthy yet less focused, and the inference time increases due to the excessive input. All above experiments are conducted under supervised training, and we evaluate the effect of reinforcement learning as well. The results show that online RL further improves SR and enhances consistency with golden responses. Online RL helps $PG$ generate better soft prompts for more effective expression.

| Models | AT↓ | SR(%)↑ |
|---|---|---|
| PPDPP(LLaMA-7b) | 9.95 | 6.15 |
| PPDPP(LLaMA-13b) | 9.86 | 11.54 |
| DPDP(LLaMA-7b) | 9.71 | 20.00 |
| DPDP(LLaMA-13b) | 9.58 | 32.31 |
| **ESCA** | **8.96** | **47.44** |

Table 3: The performance of the goal achievement. The best and second-best results from all models are highlighted in bold and underlined respectively.

## Case Study

In order to observe the content generated by ESCA more intuitively, we show the interaction process with the LLM-based seeker in Table 5 and responses generated by different comparison models in Table 6. In the interactive case, ESCA first analyzes the seeker's current state and considers the dialogue history to determine the appropriate strategy. This state-aware planning process enhances the explainability of strategy selection. It subsequently retrieves relevant knowledge and generates an instruction to guide the response generation. This process continues iteratively until the dialogue goal is achieved or the maximum number of turns is reached. As shown in the example, ESCA selects appropriate strategies and produces informative and empathetic responses during the interaction. Meanwhile, the seeker's emotion intensity decreases and the trust increases. In the comparison case in Table 6, while all comparison models generate context-relevant responses, ESCA stands out by offering responses that are more empathetic and include rich content.

## Limitation and Future Work

Despite its effectiveness, ESCA still has some limitations now. Firstly, the accuracy of state extraction during interaction needs to be improved. During training, the dataset is annotated with relatively high-quality state aspects, but during real interaction, the four states are extracted by the base LLM itself, with the stable one-shot prompt limiting accuracy and robustness. To address this, we can retrieve relevant examples from the dataset for few-shot in-context

| state | kg | pg | rl | AT↓ | SR(%)↑ | D-1↑ | D-2↑ | D-3↑ | B-3↑ | B-4↑ | R-L↑ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| × | √ | √ | × | 9.33 | 37.31 | **6.37** | 31.35 | 57.24 | 3.90 | 2.31 | 17.53 |
| √ | × | √ | × | 9.26 | 40.33 | 6.44 | 31.50 | 56.77 | 3.64 | 2.16 | 17.29 |
| √ | √ | × | × | 9.57 | 32.31 | 4.37 | 19.90 | 34.75 | 2.16 | 1.24 | 13.54 |
| √ | √ | √ | × | 9.17 | 44.61 | 6.32 | **31.73** | **57.66** | 3.89 | 2.28 | 17.74 |
| √ | √ | √ | √ | **8.96** | **47.44** | 6.15 | 30.51 | 55.82 | **4.24** | **2.50** | **17.87** |

Table 4: Results of ablation study, where state, kg, pg and rl are short for state information, knowledge, prompt generator, and reinforcement learning. The best results are highlighted in bold.

| **Interaction Steps** |
|---|
| **Seeker**: My two friends are shaming me for not donating more per year to charities. One of them is the President of a local non-profit.<br>***States**: emotion: 5, behavior: statement-emotion, stage: precontemplation, trust: 3 → **Strategy**: Affirmation and Reassurance*<br>**Supporter**: Listen, I understand it can be stressful to deal with criticism from others. It's especially hard when it comes from someone you respect like a friend. |
| **Seeker**: Just hearing you acknowledge that makes me feel a bit better, but I'm still struggling to shake off the feeling that I'm somehow inadequate or selfish because of their comments.<br>***States**: emotion: 5, behavior: statement-emotion, stage: contemplation, trust: 4 → **Strategy**: Question*<br>**Supporter**: Makes sense, it's definitely understandable to feel that way. Have you ever tried talking to them about it? |
| **Seeker**: I've tried to explain my financial situation to them, but they just tell me that everyone has financial struggles and I should be prioritizing giving back to the community.<br>.....<br>***Strategy**: Affirmation and Reassurance*<br>**Supporter**: You're definitely not alone, and you're doing a great job! By volunteering at the animal shelter, you're giving back to the community in a way that's important to you, and that's something to be proud |
| **Seeker**:It's helpful to hear you say that, because it's making me realize that their opinions don't define my worth, and that I should be proud of the contributions I am making, even if they're not the kind they expect.<br>***States**: emotion: 2, behavior: statement-emotion, stage: preparation, trust: 4 → Goal completed !* |

Table 5: The example of interaction process between ESCA-based supporter and the LLM-based seeker

learning dynamically. Secondly, our current knowledge injection approach is simple. For instance, knowledge items from PsyQA are often long and may introduce noise. We plan to employ a base LLM to summarize and filter knowledge for better relevance and consistency in the future. Finally, the strategy planner is currently trained via supervised learning. Its performance could be further improved using reinforcement learning, where the reward can be designed based on seeker state transitions.

| | |
|---|---|
| Context | **Seeker**: Hello, I found out I' m losing my job in January. I' m so very worried.<br>...<br>**Supporter**: I know , it seems like everyone is in the same boat. You should know that you' re not alone in this, i am so sorry.<br>**Seeker**: I don't even know what the next step I need to take is - I've never been without a job like this. |
| Golden | Have you tried reaching out to a career counselor or recruiter? |
| $D^2$RCU | Have you considered looking for another job? There are so many jobs out there, you should be able to find something. |
| DKPE | I know that it's really hard to get new jobs, but i think you should start by looking for a job that you can apply for. |
| COOPER | Let ' s start by updating your resume and exploring job opportunities. |
| **ESCA** | I can understand that completely. I have been in the same situation before, and i know how it feels. But you have to try to stay positive and keep your head up. |

Table 6: The example of different models' ESC responses

## Conclusion

In this paper, we propose ESCA, a novel LLM-based Emotional Support Conversation Agent that jointly enhances strategy planning and response generation for emotional support. By introducing multiple aspects of the seeker's state, we improve the interpretability and effectiveness of strategy selection. Furthermore, we propose a strategy-aligned and context-aware prompt generator that integrates relevant knowledge and instructions to guide LLMs in producing more supportive and content-rich responses. Extensive experiments demonstrate that ESCA significantly improves both the quality of responses and the success rate of achieving emotional support goals.

## Acknowledgments

This work is supported by the National Key Research and Development Program of China (2019YFB1405302), the National Natural Science Foundation of China (No.61672144), and the China Scholarship Council program

(No.202406080007). This work has been partly funded by Horizon Europe COVER project (No. 101086228).

## References

Bagdon, C.; Karmalker, P.; Gurulingappa, H.; and Klinger, R. 2024. " You are an expert annotator": Automatic Best-Worst-Scaling Annotations for Emotion Intensity Modeling. *arXiv preprint arXiv:2403.17612*.

Bai, J.; Bai, S.; Chu, Y.; Cui, Z.; Dang, K.; Deng, X.; Fan, Y.; Ge, W.; Han, Y.; Huang, F.; et al. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.

Bosselut, A.; Rashkin, H.; Sap, M.; Malaviya, C.; Celikyilmaz, A.; and Choi, Y. 2019. COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, 4762–4779.

Chen, M.; Yu, X.; Shi, W.; Awasthi, U.; and Yu, Z. 2023a. Controllable Mixed-Initiative Dialogue Generation through Prompting. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 951–966.

Chen, W.; Lin, H.; Zhang, Q.; Zhang, X.; Bai, X.; Huang, X.; and Wei, Z. 2024. CauESC: A Causal Aware Model for Emotional Support Conversation. *arXiv preprint arXiv:2401.17755*.

Chen, W.; Zhao, G.; Zhang, X.; Bai, X.; Huang, X.; and Wei, Z. 2023b. K-ESConv: Knowledge Injection for Emotional Support Dialogue Systems via Prompt Learning. *arXiv preprint arXiv: 2312.10371*.

Cheng, J.; Sabour, S.; Sun, H.; Chen, Z.; and Huang, M. 2023. PAL: Persona-Augmented Emotional Support Conversation Generation. In *Findings of the Association for Computational Linguistics: ACL 2023*, 535–554.

Cheng, Y.; Liu, W.; Wang, J.; Leong, C. T.; Ouyang, Y.; Li, W.; Wu, X.; and Zheng, Y. 2024. Cooper: Coordinating Specialized Agents towards a Complex Dialogue Goal. In *Thirty-Eighth AAAI Conference on Artificial Intelligence*, 17853–17861.

Deng, Y.; Liao, L.; Chen, L.; Wang, H.; Lei, W.; and Chua, T. 2023a. 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*, 10602–10621.

Deng, Y.; Zhang, W.; Lam, W.; Ng, S.; and Chua, T. 2024. Plug-and-Play Policy Planner for Large Language Model Powered Dialogue Agents. In *The Twelfth International Conference on Learning Representations*.

Deng, Y.; Zhang, W.; Yuan, Y.; and Lam, W. 2023b. Knowledge-enhanced Mixed-initiative Dialogue System for Emotional Support Conversations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 4079–4095.

Doosthosseini, A.; Decker, J.; Nolte, H.; and Kunkel, J. M. 2024. Chat ai: A seamless slurm-native solution for hpc-based services. *arXiv preprint arXiv:2407.00110*.

Fleiss, J. L.; and Cohen, J. 2016. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient As Measures of Reliability. *Educational Psychological Measurement*, 33(3): 613–619.

Grattafiori, A.; Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Vaughan, A.; et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Guo, D.; Yang, D.; Zhang, H.; Song, J.; Zhang, R.; Xu, R.; Zhu, Q.; Ma, S.; Wang, P.; Bi, X.; et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Hao, J.; and Kong, F. 2025. Enhancing Emotional Support Conversations: A Framework for Dynamic Knowledge Filtering and Persona Extraction. In *Proceedings of the 31st International Conference on Computational Linguistics*, 3193–3202.

He, T.; Liao, L.; Cao, Y.; Liu, Y.; Liu, M.; Chen, Z.; and Qin, B. 2024. Planning like human: A dual-process framework for dialogue planning. *arXiv preprint arXiv:2406.05374*.

Hill, C. E. 1999. Helping skills: Facilitating exploration, insight, and action. *American Psychological Association*.

Li, J.; Galley, M.; Brockett, C.; Gao, J.; and Dolan, B. 2016. A Diversity-Promoting Objective Function for Neural Conversation Models. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 110–119.

Li, J.; Li, D.; Savarese, S.; and Hoi, S. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*, 19730–19742.

Lin, C. Y. 2004. ROUGE: A Package for Automatic Evaluation of summaries. In *Proceedings of the Workshop on Text Summarization Branches Out*.

Liu, S.; Zheng, C.; Demasi, O.; Sabour, S.; Li, Y.; Yu, Z.; Jiang, Y.; and Huang, M. 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*, 3469–3483.

Liu, Y.; Li, L.; Tu, Y.; Zhang, B.; Zha, Z.; and Huang, Q. 2025. Dynamic Strategy Prompt Reasoning for Emotional Support Conversation. *IEEE Trans. Multim.*, 27: 108–119.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Madsen, M.; and Gregor, S. 2000. Measuring human-computer trust. In *11th australasian conference on information systems*, volume 53, 6–8. Citeseer.

Nikolaev, D.; and Padó, S. 2023. Representation biases in sentence transformers. *arXiv preprint arXiv:2301.13039*.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318.

Peng, W.; Qin, Z.; Hu, Y.; Xie, Y.; and Li, Y. 2023. FADO: Feedback-Aware Double COntrolling Network for Emotional Support Conversation. *Knowl. Based Syst.*, 264: 110340.

Prochaska, J.; Norcross, J.; and DiClemente, C. 2005. Stages of change: Prescriptive guidelines. *Psychologists' desk reference*, 226–231.

Prochaska, J. O.; Norcross, J. C.; and DiClemente, C. C. 1994. *Changing for good: the revolutionary program that explains the six stages of change and teaches you how to free yourself from bad habits*. W. Morrow.

Saha, T.; Patra, A.; Saha, S.; and Bhattacharyya, P. 2020. Towards emotion-aided multi-modal dialogue act classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4361–4372.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Sharma, A.; Miner, A. S.; Atkins, D. C.; and Althoff, T. 2020. A computational approach to understanding empathy expressed in text-based mental health support. *arXiv preprint arXiv:2009.08441*.

Sun, H.; Lin, Z.; Zheng, C.; Liu, S.; and Huang, M. 2021. PsyQA: A Chinese Dataset for Generating Long Counseling Text for Mental Health Support. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 1489–1503.

Tu, Q.; Li, Y.; Cui, J.; Wang, B.; Wen, J.; and Yan, R. 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*, 308–319.

Väth, D.; Vanderlyn, L.; and Vu, N. T. 2023. Conversational Tree Search: A New Hybrid Dialog Task. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, 1256–1272.

Wang, L.; Li, J.; Yang, C.; Lin, Z.; and Wang, W. 2023. Enhancing Empathetic and Emotion Support Dialogue Generation with Prophetic Commonsense Inference. *arXiv preprint arXiv: 2311.15316*.

Welivita, A.; and Pu, P. 2022. HEAL: A Knowledge Graph for Distress Management Conversations. In *Proceedings of 36th AAAI Conference on Artificial Intelligence*, 11459–11467.

Xu, X.; Meng, X.; and Wang, Y. 2022. PoKE: Prior Knowledge Enhanced Emotional Support Conversation with Latent Variable. *arXiv preprint arXiv: 2210.12640*.

Xu, Z.; Chen, D.; Kuang, J.; Yi, Z.; Li, Y.; and Shen, Y. 2024. Dynamic demonstration retrieval and cognitive understanding for emotional support conversation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 774–784.

Yu, X.; Chen, M.; and Yu, Z. 2023. Prompt-Based Monte-Carlo Tree Search for Goal-oriented Dialogue Policy Planning. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 7101–7125.

Zhang, Q.; Naradowsky, J.; and Miyao, Y. 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*, 6665–6694.

Zhao, W.; Zhao, Y.; Wang, S.; and Qin, B. 2023. TransESC: Smoothing Emotional Support Conversation via Turn-Level State Transition. In *Findings of the 61th Association for Computational Linguistics: ACL 2023*, 6725–6739.

Zheng, C.; Sabour, S.; Wen, J.; Zhang, Z.; and Huang, M. 2022. Augesc: Dialogue augmentation with large language models for emotional support conversation. *arXiv preprint arXiv:2202.13047*.

Zhou, J.; Chen, Z.; Wang, B.; and Huang, M. 2023. Facilitating Multi-turn Emotional Support Conversation with Positive Emotion Elicitation: A Reinforcement Learning Approach. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 1714–1729.