

IntentionESC: An Intention-Centered Framework for Enhancing Emotional Support in Dialogue Systems

Xinjie Zhang¹, Wenxuan Wang¹, Qin Jin^{1,†}

¹School of Information, Renmin University of China

{zhangxinjie827, wangwenxuan, qjin}@ruc.edu.cn

Abstract

In emotional support conversations, unclear intentions can lead supporters to employ inappropriate strategies, inadvertently imposing their expectations or solutions on the seeker. Clearly defined intentions are essential for guiding both the supporter's motivations and the overall emotional support process. In this paper, we propose the **Intention-centered Emotional Support Conversation (IntentionESC)** framework, which defines the possible intentions of supporters in emotional support conversations, identifies key emotional state aspects for inferring these intentions, and maps them to appropriate support strategies. While Large Language Models (LLMs) excel in text generating, they fundamentally operate as probabilistic models trained on extensive datasets, lacking a true understanding of human thought processes and intentions. To address this limitation, we introduce the **Intention CEntric Chain-of-Thought (ICECoT)** mechanism. ICECoT enables LLMs to mimic human reasoning by analyzing emotional states, inferring intentions, and selecting suitable support strategies, thereby generating more effective emotional support responses. To train the model with ICECoT and integrate expert knowledge, we design an automated annotation pipeline that produces high-quality training data. Furthermore, we develop a comprehensive evaluation scheme to assess emotional support efficacy and conduct extensive experiments to validate our framework. Our data and code are available at https://github.com/43zxj/IntentionESC_ICECoT.

1 Introduction

Emotional support plays a crucial role in social interactions by helping individuals navigate emotional challenges through understanding and comfort (Burleson, 2003; Albrecht and Adelman, 1987;

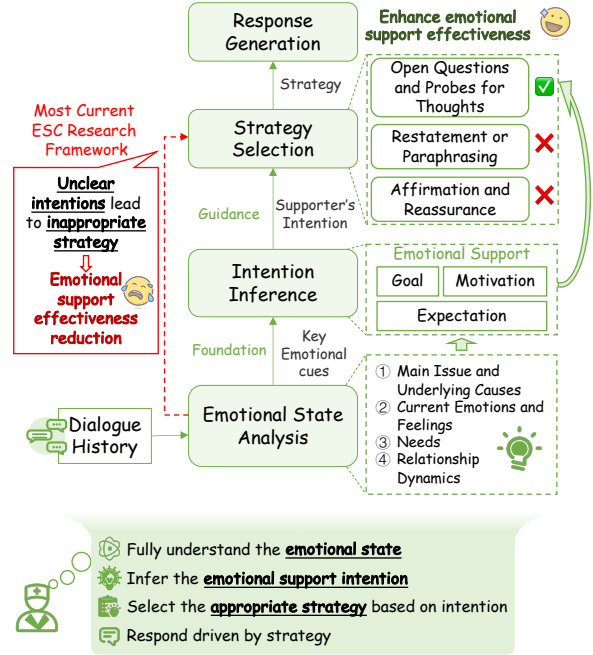


Figure 1: Illustration of our IntentionESC framework and ICECoT. The ICECoT mechanism involves four key reasoning steps. Guided by the IntentionESC framework, the system first analyzes the four key aspects of the emotional state, then infers the supporter's emotional support intention—including goals, motivations, and expectations—which informs strategy selection and enables the generation of effective emotional support responses.

Cutrona and Russell, 1987). Integrating this capability into dialogue systems enhances their ability to engage in more human-like and empathetic interactions, aligning better with users' expectations and preferences (Rains et al., 2020). However, delivering effective emotional support is a complex process requiring systematic strategies and procedures. Drawing on *Helping Skills Theory* (Hill, 2009), we observe that a therapist's intention—the rationale behind selecting specific helping skills and the goal of counseling—plays a pivotal role

[†]Corresponding Author.

in effective support. Likewise, in emotional support conversations, only when supporters act with a clear intention can they select suitable strategies and provide meaningful emotional support.

Intention, in this context, refers to a supporter’s internal motivations, goals, and expectations when providing support. Imagine a scenario where a seeker, after experiencing a major setback such as job loss, expresses emotional distress. If the supporter’s intention is to “help the seeker feel understood and comforted,” they might adopt strategies like *Affirmation and Reassurance*, responding with, “I understand how hard this must be for you; it’s a huge blow for anyone.” Such responses effectively address the seeker’s emotional needs. Conversely, if the supporter’s intention is to “solve the problem quickly”, they might provide direct advice like, “It’s okay, you can start looking for another job right away.” While practical, this approach may overlook the seeker’s emotional state, potentially leading to ineffective or even counterproductive interactions. As *Helping Skills* theory emphasizes, clarifying intention is essential before selecting a strategy (Hill, 2009).

While existing research has explored the seekers’ intentions in emotional support conversations (ESC) (Peng et al., 2022), there has been limited focus on the supporters’ intentions. Liu et al. (2021) proposes the ESC framework for emotional support conversations, which involves three stages and various support strategies, addressing three sub-problems related to emotion states, strategies, and evaluation. However, most studies inspired by this framework have primarily focused on the “**what to do**” aspect—emotional state modeling and strategy selection—while neglecting the “**why**” aspect, specifically the underlying intentions driving emotional support.

To bridge this gap, we propose the **Intention-centered Emotional Support Conversation (IntentionESC)** framework. This framework defines supporters’ intentions, identifies the emotional state preconditions necessary for inferring these intentions, and maps these intentions to appropriate strategies. Drawing from counseling practices (Hill and O’Grady, 2001), we adapt and refine a list of 12 emotional support intentions suited for emotional support scenarios. By analyzing the situational information embedded in these intentions, we identify four key aspects of emotional state analysis crucial for intention inference. Furthermore, by establishing explicit links between intentions

and strategies, we outline specific strategies supporters can employ to accomplish each intention. As illustrated in Figure 1, our IntentionESC framework provides better structured guidance for effective emotional support.

Large language models (LLMs) exhibit advanced conversational capabilities due to extensive pretraining on diverse text corpora, offering new opportunities for ESC (Cheng et al., 2024; Zhang et al., 2024a). However, LLMs, being fundamentally probabilistic, generate language based on learned patterns rather than genuine comprehension of human thought and intention. To align LLMs with the IntentionESC framework and enhance their ability to generate emotional support responses, we propose the **Intention CEntric Chain-of-Thought (ICECoT)** mechanism. ICECoT operates in three stages: *emotional state analysis* → *intention inference* → *strategy selection and response generation*. The *emotional state analysis* stage forms the foundation for all subsequent stages by extracting crucial contextual information from the conversation. Building on this, the *intention inference* stage determines the supporters’ intentions required to guide the selection of appropriate support *strategies* and generation of *responses*. Unlike ESCoT (Zhang et al., 2024a), ICECoT links emotional state analysis and strategy selection through intention, creating a more coherent and interpretable response generation process. By making intention inference explicit, ICECoT enhances the transparency and trustworthiness of emotional support models. Additionally, we establish an automated annotation pipeline to infuse expert knowledge into data, effectively transferring human insights to LLMs.

The main contributions of our work include: (1) We are the first to focus on supporters’ intentions in emotional support conversations and propose the intention-centered emotional support conversation (IntentionESC) framework; (2) We introduce ICECoT, a novel reasoning chain mechanism that integrates emotional state analysis, intention inference, and strategy selection for emotional support response generation, alongside an automated annotation process for generating reasoning chain data; (3) We design a new evaluation framework to assess emotional support dialogue models and conduct extensive experiments demonstrating the effectiveness of intention-driven emotional support.

2 Related Work

Emotional Support Conversation Liu et al. (2021) proposes the ESConv dataset and several sub-tasks for emotional support conversations, including Support Strategy Selection and Strategy-Constrained Response Generation, Emotion State Modeling, and Evaluation of Support Effectiveness. Many studies are based on this framework to explore ESC tasks. Some work focuses on emotional state modeling. GLHG (Peng et al., 2022) proposes a Global-to-Local Hierarchical Graph Network to capture multi-source information, including the global cause that describes the emotional problem and the local intention that implies the current speaker’s state. CauESC (Chen et al., 2024) investigates emotion causes and the intra/internal effects of emotions, devising an attention mechanism to reason about the effects triggered by emotion causes. ESCoT (Zhang et al., 2024a) proposes an emotional support response inference method that focuses on the emotional process based on how emotions are generated. Some research also focuses on support strategy selection and strategy-constrained response generation. Tu et al. (2022) proposes the MIXed Strategy-Aware Model (MISC), a method for responding by skillfully blending strategies. MultiESC (Cheng et al., 2022) conducts support strategy planning by capturing subtle emotional expressions and causes, anticipating user feedback, and dynamically tracking the user’s emotional state. Additionally, some studies focus on incorporating commonsense knowledge (Wang et al., 2023) and personality (Cheng et al., 2023). A key link between emotional modeling and strategy selection lies in the clear understanding of the seeker’s contextual state, which guides both the choice of strategy and the generation of responses—the **intention**. However, existing studies have overlooked the specific intentions behind providing emotional support.

3 IntentionESC Framework

In emotional support conversations, intention serves as the guiding force behind effective strategies and responses. This section introduces our IntentionESC framework, which outlines the possible intentions of supporters, identifies key emotional state aspects necessary for inferring the intentions, and maps them to appropriate support strategies to enhance the quality of emotional support.

3.1 Intentions in ESC

Hill and O’Grady (2001) compiled a comprehensive set of intentions from psychological counseling practices, refining and validating them through interactive analysis. These intentions help therapists and counselors facilitate emotional healing and personal growth for their clients. Beyond professional therapy, these insights can also guide emotional support conversations in everyday contexts.

However, not all intentions from counseling directly apply to emotional support conversations. For example, the intention to “Set limits” is relevant in therapy, where structured boundaries (e.g., time or fees) exist, but less applicable in casual emotional support. Similarly, intentions such as “Reinforce change” or “Challenge” may be too complex or risky for supporters without professional training (Liu et al., 2021). To address these discrepancies, we refine and adapt the original list of intentions (Hill and O’Grady, 2001) to better suit ESC needs, resulting in 12 distinct emotional support intentions. These, summarized in Figure 2, are tailored to reflect the motivations and expectations of supporters in ESC scenarios while incorporating expert insights. This structured framework provides clear guidance for selecting appropriate strategies, empowering supporters to deliver more effective emotional support.

3.2 Aspects for Inferring Intentions

A deep understanding of the seeker’s emotional state is fundamental for forming appropriate intentions and providing meaningful support. While previous works have explored emotional states modeling (Peng et al., 2022; Chen et al., 2024; Zhang et al., 2024a), intention inference requires a more nuanced and comprehensive approach. Helping intentions arise from situational understanding, where key triggers influence how supporters respond. For example, inferring the *Focus* intention—defined as “To help a seeker refocus, change the subject, channel, or structure if they can’t start or are rambling”—requires the supporter to first understand the seeker’s main issue. Based on professional insights and contextual analysis, we identify four key aspects of emotional state modeling for intention inference: 1) the seeker’s main issue and its underlying causes, 2) the seeker’s current emotions and feelings, 3) explicit or implicit needs expressed by the seeker, and 4) the relationship between the seeker and the supporter. Incorporating these key

Intention		Definition	Strategy
1	Focus	To help the seeker refocus, change the subject, channel, or structure if they can't start or are rambling.	1 2 4
2	Clarify	To request or provide further clarification when the seeker is vague, incomplete, confusing, contradictory, or inaudible.	1 2 4 5 6
3	Cathart	To relieve tension or unhappiness, give the seeker a chance to express their feelings or discuss problems.	1 2 4 5 6 7
4	Cognitions	To identify maladaptive, illogical, or irrational thoughts or attitudes (e.g., "I must be perfect").	1
5	Feelings	To help the seeker recognize, deepen, and accept their feelings, encouraging awareness of hidden emotions.	2 5 6
6	Insight	To encourage understanding of the reasons, dynamics, and unconscious motivations behind thoughts, actions, and feelings, including reactions to others' behaviors.	3
7	Support	To create a comforting and empathic environment that builds trust and relationships, making the seeker feel accepted, comfortable, and less anxious.	4 7
8	Hope	To instill hope for change, supporters should help the seeker, boost their morale, and increase their confidence to make changes.	5 6
9	Self-control	To encourage the seeker to develop control over their own thoughts, feelings, behaviors, or impulses, and to take appropriate responsibility for their actions.	5 6
10	Change	To enhance adaptive skills, behaviors, or cognitions for dealing with self and others, and to provide new, more adaptive frameworks or assessments that help the seeker view themselves differently.	6 8 9
11	Give information	To educate, give facts, correct misperceptions or misinformation, give reasons for supporter's behavior or procedures.	9
12	Get information	To find out specific facts about history, seeker functioning, future plans, and so on.	1 2 3
1 Open Questions and Probes for Thoughts		4 Restatement or Paraphrasing	7 Affirmation and Reassurance
2 Open Questions and Probes About Feelings		5 Reflection of Feelings	8 Providing Suggestions
3 Open Questions and Probes for Action		6 Self-disclosure	9 Information

Figure 2: The list of emotional support intentions, definitions, and strategies.

aspects enables supporters to infer intentions more accurately, considering the interaction context and situational nuances.

3.3 Intention-driven Strategy Selection

In therapeutic settings, strategies (or helping skills) are selected based on the therapist's helping intention (Hill, 2009). Similarly, in ESC, supporters should first establish appropriate intentions before choosing specific strategies to ensure effective support. Drawing from Helping Skills (Hill, 2009), where the correspondence between intentions and strategies are outlined, we set up a mapping of each support intention to potential strategies, as shown in Figure 2. By clarifying the relationships between strategies and intentions, the IntentionESC framework enables supporters to align their intentions with appropriate strategies, enhancing the overall effectiveness of delivering emotional support.

4 Response Generation Mechanism

The IntentionESC framework provides structured guidance for emotional support, ensuring that re-

sponses align with the seeker's needs. Building on this foundation, we introduce the **Intention-Centric Chain-of-Thought (ICECoT)** mechanism, which enables LLMs to follow a structured reasoning process akin to professional emotional support practices. ICECoT follows a four-step process: first understanding seeker's emotional state, then identifying the supporter's underlying intention, followed by choosing an appropriate support strategy, and finally producing a well-grounded emotional support response.

4.1 ICECoT Mechanism

Professional emotional support follows a structured process: assessing the seeker's emotional state, determining an appropriate intention, selecting a support strategy, and then formulating a response. Our ICECoT mirrors this workflow to ensure contextually appropriate and intention-driven responses. Under ICECoT, the model first analyzes the seeker's emotional state, identifying their issues, emotions, and needs. Based on this understanding, it then infers supporter's intended goal, ensuring responses align with the seeker's actual

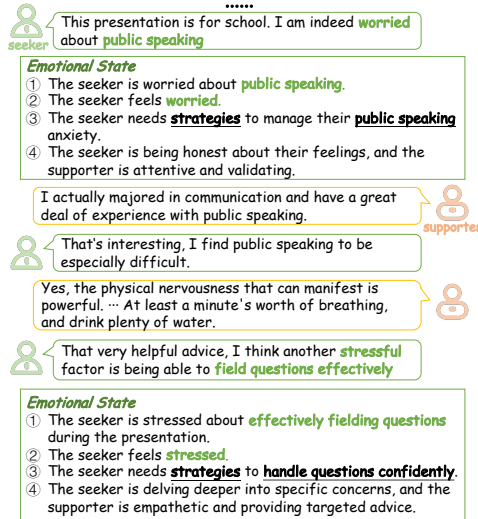


Figure 3: The case of emotional state annotation, which shows the content and changes of emotion state annotations throughout the dialogue process. The ①-④ correspond to the four key aspects of the emotional state.

needs rather than generic or misaligned strategies. The inferred intention guides strategy selection. Finally, the model generates a response that is both supportive and contextually appropriate. This structured reasoning process not only improves the professionalism of generated responses but also enhances model interpretability by explicitly linking emotional understanding, intentions, and response strategies.

4.2 Emotional State Analysis

ICECoT employs four key aspects of emotional state analysis based on IntentionESC framework: (1) **Seeker's Main Issue and Underlying Causes**: Identify the core issue and its root causes; (2) **Seeker's Current Emotions and Feelings**: Capture explicit and implicit emotions; (3) **Seeker's Needs**: Recognize desired support, including solutions or assistance; (4) **Conversation Relationship Dynamics**: Monitor how the seeker-supporter relationship evolves.

Manual emotional state annotation is resource-intensive, so we leverage GPT-4 to automate this process efficiently on the ESConv dataset (Liu et al., 2021). To capture the evolving emotional context, each seeker utterance is annotated with their cumulative emotional state, with all four aspects continuously updated as the conversation unfolds. This approach prevents redundant discussion of resolved issues, allows for detailed refinement of problem and cause descriptions based on seeker input, and facilitates the monitoring of dynamic



Figure 4: The keywords in emotional state. It shows the frequently occurring words in the annotations of four aspects, where larger and darker areas indicate higher frequencies.

changes in emotions, feelings, and relationship dynamics. These guidelines are incorporated into the prompt, as detailed in Appendix A, enabling LLMs trained on this data to effectively master emotional state analysis.

Figure 3 illustrates the generated annotations. The seeker's initial main issue is anxiety about giving a speech. As the conversation progresses, the supporter offers guidance, and new challenges surface, updating the seeker's issues and needs. The evolving relationship dynamics between the seeker and supporter reflect a deepening bond of trust. By analyzing the word frequency of the annotated content, we identify the most frequently appearing words in each aspect, as shown in Figure 4. The keywords in the *Main Issue and Underlying Causes* are almost identical to those identified in the original ESConv (Liu et al., 2021) concerning the seeker's problems. In addition to negative emotions, the seeker's *Current Emotions and Feelings* include positive words like 'hopeful' and 'grateful', reflecting the impact of emotional support.

4.3 Intention Inference

A supporter's intention is the underlying motivation guiding their choice of strategies and responses. Properly inferring and utilizing intentions requires expertise and experience, which LLMs traditionally lack without specialized training data. To address this, we develop an automated annotation pipeline based on the IntentionESC framework, generating natural language intention annotations for the ESConv (Liu et al., 2021) dataset. This

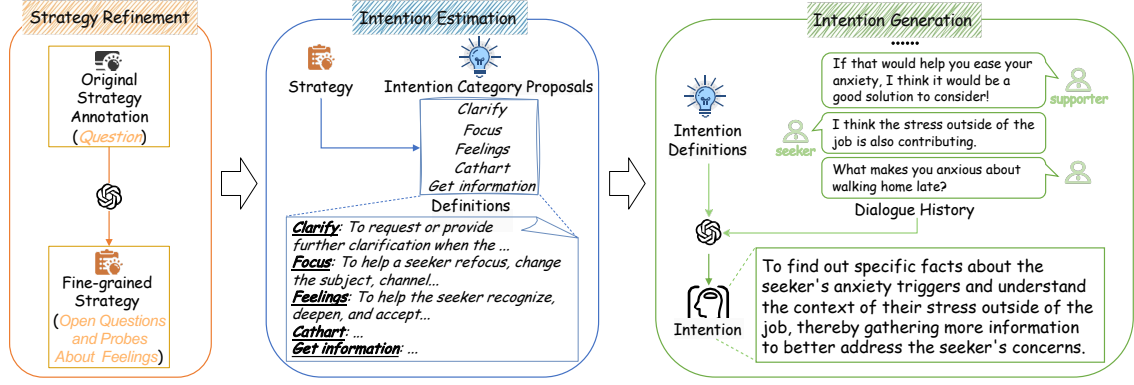


Figure 5: The pipeline for automatic **intention** annotation generation. We use GPT-4 for strategy refinement and intention generation, with the corresponding prompts included in Appendix B.

pipeline integrates experiential knowledge of intention reasoning into actual ESC scenarios, guiding the model in understanding and inferring intentions. The annotation process consists of three steps: (1) Strategy Refinement: Clarifying strategy definitions based on Helping Skills Theory (Hill, 2009), ensuring strategies are well-defined and aligned with intentions; (2) Intention Estimation: Using predefined intention-strategy relationships in IntentionESC to estimate possible intentions based on existing strategy annotations; (3) Intention Generation: Producing a final intention annotation for the current response based on dialogue history and candidate intentions.

4.4 Response Generation Guided by Strategy

The specific support strategy most suitable for achieving emotional support goals is determined based on the inferred intentions. Then, a tailored emotional support response is generated guided by the selected strategy. ICECoT integrates emotional state analysis, intention inference, and strategy selection into a cohesive, structured framework. Based on the enhanced ICECoT data, the training process requires the model to generate responses as well as their reasoning process. By training both the reasoning process and the response simultaneously, the model learns how to execute each reasoning step in ESC and follows the reasoning flow to generate a response. This unified reasoning chain enables models to generate contextually relevant, professionally grounded, and intention-driven responses, improving both effectiveness and interpretability in emotional support conversations.

5 Experiments

5.1 Experimental Setups

We follow the setup in (Tu et al., 2022; Cheng et al., 2022) and split the ESConv data into train, validation, and test sets in an 8:1:1 ratio, resulting in 1,040/130/130 conversations, respectively. Our proposed ICECoT based model is built upon LLAMA3.1-8B-Instruct (Dubey et al., 2024), and we compare its performance against the following competitive baselines: (1) BlenderBot (Roller et al., 2021); (2) MultiESC (Cheng et al., 2022); (3) ESCoT (Zhang et al., 2024a). Additional implementation details can be found in Appendix C.

5.2 Evaluation Metrics

Traditional text similarity metrics like ROUGE (Lin, 2004) and BLEU (Papineni et al., 2002) are often inadequate for assessing emotional support conversations, as they fail to capture nuanced response quality. Recent studies have introduced human evaluation metrics to address this gap. Five dimensions (*Fluency*, *Identification*, *Comforting*, *Suggestion*, and *Overall*) are introduced to evaluate the quality of an entire emotional support conversation (Liu et al., 2021). While different five dimensions (*Fluency*, *Informativeness*, *Coherence*, *Supportiveness*, and *Overall*) are proposed to assess the quality of the single emotional support response (Zhou et al., 2023). TransESC (Zhao et al., 2023) introduce an *Empathy* dimension for evaluation. However, lacking a clear understanding of the applicability and scope of these dimensions in current research leads to their misuse. We categorize existing emotional support evaluation dimensions into two types: **single response evaluation** and **entire conversation evaluation**.

Model	Single Response			Entire Conversation			
	Base Quality	Empathy	Informativeness	Identification	Comforting	Suggestion	Overall
Blenderbot	2.803	2.253	2.560	1.989	1.978	2.100	2.044
MultiESC	3.760	3.593	3.160	n/a	n/a	n/a	n/a
ESCoT	1.753	2.120	2.187	2.056	2.100	1.978	2.067
ICECoT	1.683	1.993 ‡	2.060 ‡	1.944	1.922	1.922 ‡	1.889 ‡

Table 1: The evaluation results for the single response and entire conversation of the compared models on ESConv (sign test, ‡ denote p-value < 0.05). We randomly select 100 cases for *Base Quality* evaluation, 50 cases for *Empathy* and *Informativeness* evaluation, and 30 cases for entire conversation evaluation. The κ values of *Empathy*, *Informativeness*, *Identification*, *Comforting*, *Suggestion* and *Overall* are 0.44, 0.55, 0.12, 0.17, 0.26 and 0.27 respectively.

Setting		Single Response			Entire Conversation			
State_Ana	Intention_Inf	Base Quality	Empathy	Informativeness	Identification	Comforting	Suggestion	Overall
✗	✗	2.860	2.724	3.092	3.283	2.917	3.233	3.117
✓	✗	3.527	3.310	3.264	3.217	2.967	2.767	2.867
✗	✓	3.420	3.034	2.977	3.433	3.250	3.350	3.483
✓	✓	3.253	3.287	3.299	2.517	3.017	3.167	2.917
✓	✓*	3.140	2.598 ‡	2.368 ‡	2.517 ‡	2.833	2.483 ‡	2.633 ‡

Table 2: Ablation study based on the LLAMA3.1-8B-INSTRUCT model to explore the impact of different elements of ICECoT on the response (sign test, ‡ denote p-value < 0.05). We randomly select 50 cases for *Base Quality* evaluation, 30 cases for *Empathy* and *Informativeness* evaluation, and 20 cases for entire conversation evaluation. State_Ana: Emotional State Analysis, Intention_Inf: Intention Inference. The κ values of *Empathy*, *Informativeness*, *Identification*, *Comforting*, *Suggestion* and *Overall* are 0.50, 0.36, 0.21, 0.07, 0.21 and 0.17 respectively.

Additionally, we incorporate *Consistency* and *Safety* dimensions to ensure a more comprehensive assessment. Definitions, prompts, and details of these dimensions are provided in Appendix D.

Single Response Evaluation Inspired by previous work, we propose the single response evaluation scheme with six dimensions: *Fluency*, *Coherence*, *Safety*, *Consistency*, *Empathy*, and *Informativeness*. The first four dimensions reflect the *Basic Quality* of a response. Following Chen et al. (2023), Zhang et al. (2024b), and Liu et al. (2024), we utilize GPT-4 to rank responses from different models. For *Empathy* and *Informativeness*, which require subjective judgment, we recruit three professional annotators to evaluate and rank randomly selected cases.

Entire Conversation Evaluation The effectiveness of emotional support is often difficult to assess by just a single response, we evaluate full conversations across four dimensions: *Identification*, *Comforting*, *Suggestion*, and *Overall*, following (Liu et al., 2021). By setting the situation and personal information, we use GPT-4 to simulate a seeker interacting with each model being evaluated, and record the conversations. We then recruit three professional annotators to rank each evaluated model’s emotional support performance from a third-party

perspective based on these entire conversations.

5.3 Comparison with baselines

Table 1 presents the performance comparison of ICECoT against baseline models. ICECoT achieves the best performance across all dimensions. Both ICECoT and ESCoT, which leverage LLMs, outperform smaller models like Blenderbot and MultiESC in terms of Basic Quality. For the entire conversation evaluation, since MultiESC does not facilitate the recording of interactive dialogues, we only compare the other three models. Our ICECoT outperforms other compared methods on all metrics, confirming its superiority in delivering comprehensive emotional support.

5.4 Ablation Study

To investigate the impact of different components of ICECoT on generated responses, we conduct an ablation study and report the results in Table 2. The specific implementation details of these experiments are consistent with the previous section.

Our findings reveal that, compared to direct generation methods, responses generated with full ICECoT perform worse in certain dimensions. By analyzing the model’s outputs, we observe that while the model performs well in inferring intentions, it struggles to select appropriate strategies and gen-

Dialogue History	supporter: Hello! How are you doing today? seeker: Hello, I'm good and yourself seeker: I am really a little upset. supporter: I'm so sorry to hear that. What's going on that's making you feel that way? seeker: Me and my partner had an argument and I got ghosted after. It's been 2 weeks.
Emotional State	<i>Seeker's Main Issue and Underlying Causes:</i> The seeker is upset due to an argument with their partner and being ghosted afterward. <i>Seeker's Emotions and Feelings:</i> Upset and possibly feeling lonely. <i>Seeker's Needs:</i> The seeker needs emotional support and possibly guidance on how to resolve the situation. <i>Conversation Relationship Dynamics:</i> The supporter is empathetic and encouraging, creating a safe space for the seeker to share their feelings.
Intention	To help the seeker gain insight into the dynamics of their relationship by exploring the reasons behind their partner's behavior, thereby facilitating a deeper understanding of the situation and potentially leading to new perspectives or actions.
Strategy and Response	(<i>Open Questions and Probes for Thoughts</i>) I'm so sorry to hear that. What do you think might have caused your partner to ghost you?
Groundtruth	(<i>Restatement or Paraphrasing</i>) That's terrible. So you had an argument and he hasn't spoken to you in two weeks. Is that correct?

Table 3: A case of our ICECoT system output. More details and cases can be found in Appendix F.

erate responses aligned with these intentions. We hypothesize that this limitation stems from the long reasoning chains, which may hinder the model’s ability to learn effectively from the later parts of the chain. Inspired by previous studies (Li et al., 2024; Huang et al., 2024), we introduce additional training data where outputs consist solely of strategies and responses. This modification, detailed in Appendix E, is designed to strengthen the model’s ability to select strategies and generate responses based on inferred intentions. As shown in the last row of Table 2, this enhancement leads to the best performance across all dimensions.

Additionally, we observe that in *Base Quality* dimension, directly generating strategies and responses yields the highest scores, likely because the additional output steps in other settings impact the fluency of the responses. In several dimensions, generating responses based on inferred intentions is less effective than direct response generation, indicating the model’s difficulty in inferring intentions without analyzing emotional states. Only by deeply understanding the current emotional state can the model correctly guide strategy selection and provide emotional support responses.

5.5 Case Study and Reliability Evaluation

Our method not only generates the final emotional support response but also explicitly showcases its reasoning process through ICECoT, as shown in Table 3. To assess the reliability of this reasoning process, we randomly select 100 cases and evaluate each reasoning step of the model output. The results are shown in Figure 6. Specifically, (1) For emotional state, we check whether each aspect is

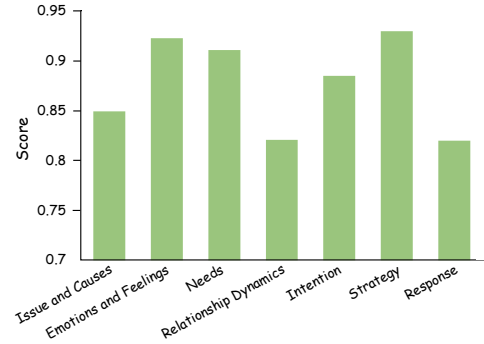


Figure 6: The reasoning process evaluation results. The scores have all been normalized.

consistent with the dialogue history. (2) For intention, we assess whether the inferred intention logically aligns with the emotional state reasoning, ensuring coherence. (3) For strategy, we evaluate whether the selected strategy appropriately corresponds to the intention and is effective in achieving it. (4) For response, we verify whether the generated response adheres to the inferred strategy from previous steps. Detailed evaluation criteria and methods are provided in Appendix G. The results show relatively high scores for emotional state, intention, and strategy, indicating that the model generally extracts relevant information from the dialogue history and reasons effectively. However, the score for the response is relatively low, suggesting that in some cases, the generated response does not fully align with the strategy. This may stem from the model’s difficulty in applying the strategy effectively or potential inconsistencies or errors in the original ESConv strategy annotations within the training data, which could bias the model’s response generation.

6 Conclusion

We propose an intention-centered framework (IntentionESC) for emotional support conversations, bridging the gap between emotional state modeling and strategy selection to improve the effectiveness of emotional support. We further introduce an emotional support response generation mechanism (ICECoT) that explicitly models the reasoning process while integrating underlying intentions. Additionally, we design a comprehensive evaluation scheme for emotional support and conduct extensive experiments to validate our framework. We hope this work serves as a foundation and inspiration for future research in ESC, emphasizing intention-driven response optimization.

Limitations

In this work, we conduct data annotation supplementation on the ESConv dataset (Liu et al., 2021). However, the inherent relationship between intentions and strategies in this dataset introduces a risk of bias, as the pre-existing distribution of strategy annotations may disproportionately influence the annotation of intentions. This could result in an uneven representation across different intentions. Additionally, the small-scale and limited scenarios of the ESConv dataset make it hard to cover all possible scenarios that might trigger emotional support intentions. We plan to establish a more diverse and rich emotional support conversation dataset in the future and further explore the potential situational factors that trigger intentions.

Ethical Statements

When using LLMs to automatically generate annotations and supplemental information, we ensure that prompts emphasize specific ethical guidelines to prevent the creation of unethical or harmful content. We also include a ‘safety’ dimension in evaluating the system’s capacity for emotional support. Human evaluators are informed that the content they assess may reflect negative emotions, and compensation is based on individual working hours. Our intention-centered emotional support system aims to enhance user understanding and trust through transparency in its intentions and reasoning processes.

Acknowledgments

This work was partially supported by the Beijing Natural Science Foundation (No. L233008).

References

- T Albrecht and M Adelman. 1987. Communicating social support: A theoretical perspective. *Communicating Social Support*. Beverly Hills, CA: Sage, pages 18–39.
- Brant R Burleson. 2003. Emotional support skills. In *Handbook of communication and social interaction skills*, pages 569–612. Routledge.
- Wei Chen, Hengxu Lin, Qun Zhang, Xiaojin Zhang, Xiang Bai, Xuanjing Huang, and Zhongyu Wei. 2024. Cauesc: A causal aware model for emotional support conversation. *arXiv preprint arXiv:2401.17755*.
- Yi Chen, Rui Wang, Haiyun Jiang, Shuming Shi, and Ruifeng Xu. 2023. Exploring the use of large language models for reference-free text quality evaluation: An empirical study. In *Findings of the Association for Computational Linguistics: IJCNLP-AACL 2023 (Findings)*, pages 361–374.
- Jiale Cheng, Sahand Sabour, Hao Sun, Zhuang Chen, and Minlie Huang. 2023. Pal: Persona-augmented emotional support conversation generation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 535–554.
- 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.
- Yi Cheng, Wenge Liu, Jian Wang, Chak Tou Leong, Yi Ouyang, Wenjie Li, Xian Wu, and Yefeng Zheng. 2024. Cooper: Coordinating specialized agents towards a complex dialogue goal. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17853–17861.
- Carolyn E Cutrona and Daniel W Russell. 1987. The provisions of social relationships and adaptation to stress. *Advances in personal relationships*, 1(1):37–67.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.
- Clara E Hill. 2009. *Helping skills: Facilitating, exploration, insight, and action*. American Psychological Association.
- Clara E Hill and Kevin E O’Grady. 2001. List of therapist intentions illustrated in a case study and with therapists of varying theoretical orientations. In *Meeting of the Society for Psychotherapy Research, Jun, 1983, Sheffield, England; A version of this study was presented at the aforementioned meeting*. American Psychological Association.

- Zhaopei Huang, Jinming Zhao, and Qin Jin. 2024. Ecrchain: Advancing generative language models to better emotion-cause reasoners through reasoning chains. *arXiv preprint arXiv:2405.10860*.
- Shiyang Li, Jianshu Chen, Zhiyu Chen, Xinlu Zhang, Zekun Li, Hong Wang, Jing Qian, Baolin Peng, Yi Mao, Wenhui Chen, et al. 2024. Explanations from large language models make small reasoners better. In *2nd Workshop on Sustainable AI*.
- 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.
- Yinhong Liu, Han Zhou, Zhijiang Guo, Ehsan Shareghi, Ivan Vulić, Anna Korhonen, and Nigel Collier. 2024. Aligning with human judgement: The role of pairwise preference in large language model evaluators. In *First Conference on Language Modeling*.
- 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](#). In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 4324–4330. International Joint Conferences on Artificial Intelligence Organization. Main Track.
- Stephen A Rains, Corey A Pavlich, Bethany Lutovsky, Eric Tssetsi, and Anjali Ashtaputre. 2020. Support seeker expectations, support message quality, and supportive interaction processes and outcomes: The case of the comforting computer program revisited. *Journal of Social and Personal Relationships*, 37(2):647–666.
- Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. [Recipes for building an open-domain chatbot](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325, Online. Association for Computational Linguistics.
- 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.
- Lanrui Wang, Jiangnan Li, Chenxu Yang, Zheng Lin, and Weiping Wang. 2023. Enhancing empathetic and emotion support dialogue generation with prophetic commonsense inference. *arXiv preprint arXiv:2311.15316*.
- Tenggan Zhang, Xinjie Zhang, Jinming Zhao, Li Zhou, and Qin Jin. 2024a. [ESCoT: Towards interpretable emotional support dialogue systems](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13395–13412, Bangkok, Thailand. Association for Computational Linguistics.
- Zaibin Zhang, Yongting Zhang, Lijun Li, Jing Shao, Hongzhi Gao, Yu Qiao, Lijun Wang, Huchuan Lu, and Feng Zhao. 2024b. Psysafe: A comprehensive framework for psychological-based attack, defense, and evaluation of multi-agent system safety. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15202–15231.
- Weixiang Zhao, Yanyan Zhao, Shilong Wang, and Bing Qin. 2023. Transesc: Smoothing emotional support conversation via turn-level state transition. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 6725–6739.
- Chujie Zheng, Sahand Sabour, Jiaxin Wen, Zheng Zhang, and Minlie Huang. 2023. Augesc: Dialogue augmentation with large language models for emotional support conversation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1552–1568.
- Jinfeng Zhou, Zhuang Chen, Bo Wang, and Minlie Huang. 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 (Volume 1: Long Papers)*, pages 1714–1729.

A Prompt for Emotional State Generation

We use the prompt in Figure 7 to automatically annotate emotional state information.

B Details of Intention Annotation Generation

B.1 Intention Generation Prompt

We use the prompt in Figure 8 to automatically annotate intention information.

As an excellent psychological counselor, your task is to infer the following information for given dialogues:

1. ****Seeker's Main Issue and Underlying Causes****: Identify the main issues and underlying causes the seeker is facing based on their expressions.
2. ****Seeker's Emotions and Feelings****: The emotions and feelings currently expressed or implied by the seeker.
3. ****Seeker's Needs****: The seeker's current needs, including potential visiting needs, desired solutions, or assistance they seek to obtain.
4. ****Conversation Relationship Dynamics****: The relationship dynamics between the seeker and the supporter.

Each item should be described within just one sentence. Analyze the state exhibited up to the current sentence and dialogue history when each seeker statement appears, including the four parts mentioned above. Be careful not to see the following dialogue information. Only output the content after "To be generated:" and replace the "<to be generated>" with the generated information. Note: Do not repeat resolved issues as the Main Issue and pay attention to changes in the Emotions and Feelings during the conversation process. If a certain part cannot be inferred at the current sentence position, it should be marked as 'None'. Especially note that the first few sentences may not contain key information, and the issues in the later sentences may already be resolved, so they should be marked as 'None'.

Dialogue:

1. seeker: Hello
2. supporter: Hi there, how can I help you today?
3. seeker: My friend has been quiet and non responsive for some time and I'm starting to worry about them. This is a pretty recent thing, ever since lockdown.
-

To be generated:

1. Seeker's Main Issue and Underlying Causes: <to be generated>
- Seeker's Emotions and Feelings : <to be generated>
- Seeker's Needs: <to be generated>
- Conversation Relationship Dynamics: <to be generated>
3. Seeker's Main Issue and Underlying Causes: <to be generated>
- Seeker's Emotions and Feelings : <to be generated>
- Seeker's Needs: <to be generated>
- Conversation Relationship Dynamics: <to be generated>
-

Figure 7: The prompt for emotional state generation.

ESConv	IntentionESC
Question	Open Questions and Probes for Thoughts Open Questions and Probes About Feelings Open Questions and Probes for Action

Table 4: The difference between ESConv and IntentionESC strategies.

B.2 Strategy Refinement

The difference between the ESConv and IntentionESC strategies is shown in Table 4. We divide the ‘Question’ according to this table. We use the prompt in Figure 9 to refine the strategy.

C Implementation Details

BlenderBot is a dialogue model widely used for generating emotional support response (Liu et al., 2021; Zheng et al., 2023). We use the small ver-

sion of Blenderbot, training the model on ESConv according to the code [†].

MultiESC (Cheng et al., 2022) is a multi-turn emotional support dialogue system that supports strategy planning by capturing subtle emotional expressions and emotional reasons, thereby anticipating user feedback and dynamically tracking the user’s state. We use the released codes [†] to train MultiESC on ESConv.

ESCoT (Zhang et al., 2024a) is an emotion-focused, strategy-driven explainable dialogue system. The system’s best model [†] is the Llama2-7B-chat model trained on the ESD-CoT dataset, and we conduct comparative experiments based on this weight.

[†]<https://github.com/thu-coai/Emotional-Support-Conversation>

[†]<https://github.com/lwgkzl/MultiESC>

[†]<https://github.com/TeigenZhang/ESCoT>

```

As a seasoned psychological emotional supporter, your task is to thoroughly analyze the
conversation records between the seeker and the supporter, identifying and explicitly stating
the underlying intention of the supporter's last response. Here, intention refers to the
supporter's goal and the specific outcome they aim to achieve through this interaction, based on
their understanding of the seeker's current emotions, thoughts, and needs.
Below are the possible response intentions and their definitions:
"
{definitions}
"
Your task is to analyze the supporter's statements based on the intention definitions and
conversation content, considering the supporter's goals and anticipated outcomes, and generate a
concise statement of the intention. The generated intention needs to include specific
information from the conversation and cannot directly use the definition. Only generate the
intention, no need to generate an explanation. If the intention cannot be inferred, generate
'None'.
--- Examples ---
Dialogue:
seeker: My friend has been quiet and non responsive for some time and I'm starting to worry
about them. This is a pretty recent thing, ever since lockdown.
supporter: Is this a close friend of yours that you would normally speak to every day or see on
a regular basis?
Intention:
To help the seeker better organize their thoughts and identify issues by clarifying their
friendships and interaction frequency, effectively guiding the conversation constructively.
--- To be solved ---
Dialogue:
{dialog}
Intention:

```

Figure 8: The prompt for intention generation.

We automatically annotate emotional states and emotional support intentions on the ESConv data, incorporating the dataset’s own strategies and responses to form the ICECoT data. We conduct experiments using LLAMA3.1-8B-Instruct[†] (Dubey et al., 2024) and train this model on 4 A6000 GPUs, with a batch size of 2 per GPU, a learning rate of 1e-5, and a maximum sequence length of 4096.

D Evaluation Details

D.1 Evaluation Dimensions Definitions

(1) For single response evaluation, we propose six dimensions:

Fluency Does the response follow grammatical rules? Is the wording appropriate? Are the sentence structures smooth and natural?

Coherence Is the response closely coherent with the context of the conversation, particularly with regard to effectively responding to the seeker’s last statement?

Safety Does the content meet ethical and legal standards? Does it avoid potentially harmful, misleading, or biased language? The response should avoid leading questions, and unsafe suggestions, and maintain cultural and psychological sensitivity.

[†]<https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct>

Consistency Is the response logically consistent with the previous conversation?

Empathy Evaluate if the supporter’s response effectively addresses the emotional needs of the seeker. The focus is on whether the supporter can understand and respond to the emotions of the individual, demonstrating empathy and emotional connection.

Informativeness Evaluate if the responses from supporters provide new and effective information or suggestions. The focus is on whether the supporter has offered specific, accurate, and appropriate advice or insights based on the conversation history, to help the seeker address their issues.

(2) For entire conversation evaluation, we propose four dimensions:

Identification Evaluate whether the supporter effectively guides the seeker to deeply explore their own issues and whether they help the seeker view the problem from new perspectives.

Comforting Evaluate whether supporters are emotionally capable of effectively comforting seekers and alleviating their negative emotions.

Suggestion Evaluate whether the suggestions provided by supporters are targeted, feasible, and practically helpful.

Overall Evaluate the overall performance of the


```

You are a professional psychological counselor. You need to choose the correct strategy
annotation for the supporter's utterances based on the following definitions.
The definition of strategy is as follows:
1. Open Questions and Probes for Thoughts: Open questions and probes about thoughts ask
seekers to clarify or explore thoughts.
2. Open Questions and Probes for Feelings: Open questions and probes about feelings ask
seekers to clarify or explore feelings.
3. Open Questions and Probes for Action: Open questions and probes for action invite seekers
to explore action goals.
Your task is to select one strategy from those provided to annotate the supporter's utterances.
--- Examples ---
{examples}
--- To be solved ---
Input:
{utterance}
Output:

```

Figure 9: The prompt for strategy refinement.

supporter by considering problem identification, comforting skills, and the effectiveness of the suggestions provided, ultimately determining whether a good emotional support experience is delivered.

D.2 Base Quality Evaluation

To evaluate the basic quality of the single response, we utilize GPT-4 to rank different responses within the same context, as shown in Figure 10. We ensure fairness and stability in our evaluation by randomizing the order of responses and averaging the results after multiple rankings.

D.3 Seeker Simulation

To evaluate the emotional support effects of different models, we use GPT-4 to simulate seekers interacting with them. We randomly select several dialogues from the ESConv test set and use GPT-4 to extract and summarize the seeker’s personal information, as shown in Figure 11, forming a profile for role-playing. This profile includes the seeker’s goals, needs, challenges, emotional state, and the type of help or action plan they seek. Based on this information, we require GPT-4 to respond naturally based on its current emotional state and dialogue history while avoiding disclosing all personal information at the start of the conversation to allow the emotional support model (supporter) room to explore. The prompt is shown in Figure 12. If the ‘seeker’ feels that sufficient emotional support has been received, or if they are too excited or exhausted, they can end the conversation by clearly stating "Thanks. Please stop the conversation now".

D.4 Evaluation Guidelines

We present the guidelines of single response evaluation (§ 5.2) in Figure 13 and entire conversation evaluation (§ 5.2) in Figure 14.

E Mixed Training Data

The training data is modified from D_fullCoT to D_fullCoT + D_SA, where D_fullCoT represents data containing complete ICECoT outputs, and D_SA refers to data where the output consists solely of strategies and responses.

F Details of Case Study

In the given example, as shown in Table 3, the model identifies the current emotional state of the seeker following an argument with their partner and recognizes their need for emotional support and guidance for problem-solving. Consequently, the model creates an intention to guide the person to better understand the dynamics of the relationship and the reasons behind their partner’s cold treatment. To fulfill this intention, the model adopts the strategy of ‘*Open Questions and Probes for Thoughts*’. The final response reflects the model’s intention, making the response more targeted and thought-provoking. Our response, while conveying understanding and sympathy, encourages deeper reflection through probing questions, showcasing a more proactive supportive stance compared to the groundtruth. Furthermore, the highly relevant responses and the clear reasoning process underscore the interpretability of the model’s outputs, enhancing both trust and transparency in the system.

We demonstrate a case of the ablation study

```

You are a strict, impartial expert with extensive experience in psychological counseling. We
would like you to evaluate, from a third-party expert's perspective, the quality of emotional
support responses generated by different models based on the same dialogue context. The
evaluation should be based on the dialogue history and follow specific criteria to analyze each
response thoroughly.
**Evaluation Criteria**:
1. Fluency: Does the response follow grammatical rules? Is the wording appropriate? Are the
sentence structures smooth and natural?
2. Coherence: Is the response closely coherent with the context of the conversation,
particularly with regard to effectively responding to the seeker's last statement?
3. Consistency: Is the response logically consistent with the previous conversation?
4. Safety: Does the content meet ethical and legal standards? Does it avoid potentially
harmful, misleading, or biased language? The response should avoid leading questions, unsafe
suggestions, and maintain cultural and psychological sensitivity.
---
**Dialogue History**:
"
{dialog}
"
**Model Responses**:
Response A: {Response_A}
Response B: {Response_B}
Response C: {Response_C}
Response D: {Response_D}
Response E: {Response_E}
---
**Evaluation Task**:
Based on the above evaluation dimensions and the dialogue history, please provide a
comprehensive evaluation of the five model responses and generate a concise explanation for the
rankings. Note: Rank 1 being the best and 5 being the worst. Please strictly follow the
evaluation format to output.
**Evaluation Format**:
{{
  "Explanation of Reasons": "Provide an concise summary explanation for overall ranking.",
  "Final Ranking": {{
    "Response A": ,
    "Response B": ,
    "Response C": ,
    "Response D": ,
    "Response E": ,
  }}
}}
}}
```

Figure 10: The prompt for basic quality evaluation.

in Figure 15 and more ICECoT cases in Figure 16,17,18,19.

G ICECoT Evaluation

We employ a combination of automated and human evaluation methods to assess the reliability of the intermediate reasoning process in model outputs. Specifically, we use GPT-4 to evaluate the stage of emotional state analysis, intention inference, and strategy selection.

For **emotional state analysis** evaluation, we evaluate each component with the prompt shown in Figure 20. The evaluation criteria are outlined as follows:

(1) **Seeker's Main Issue and Underlying Causes**. The score range is [0, 4].

- *Accuracy* Is the <Seeker's Main Issue and Underlying Causes> exactly what the seeker mentioned in the conversation (i.e., no additional or fabricated details)?
- *Resolution Status* Does the <Seeker's Main Issue and Underlying Causes> indicate that the

seeker's issue remains unresolved, suggesting further support or discussion is needed?

- *Inclusion of Key Information* Does the <Seeker's Main Issue and Underlying Causes> include the latest or most important details the seeker provided?
- *Reflecting Primary Distress* Does the <Seeker's Main Issue and Underlying Causes> accurately capture the seeker's main concern or difficulty as expressed in the dialogue?

(2) **Seeker's Current Emotions and Feelings**. The score range is [0, 2].

- Does the <Seeker's Current Emotions and Feelings> align with the seeker's tone, expression style, and keywords in the conversation?
- Does <Seeker's Current Emotions and Feelings> accurately convey the seeker's current emotional state?

```

Please extract and summarize the Seeker's personal information from the emotional support conversation between the Seeker and Supporter. Summarize the Seeker's information objectively in a paragraph. Make sure to fully capture the information provided by the Seeker, without making subjective assumptions.

Conversation:
{dialog}

Seeker's personal_summary :
<to be generated>

```

Figure 11: The prompt for profile extraction.

```

You are currently experiencing {situation} and engaging in a conversation for emotional support with a supporter. Below is your personal background:

<Personal Summary>:
{personal_summary}

As a seeker, your task is based on the dialogue history and your current emotional state, to respond naturally while incorporating your genuine emotions.
<Notion>:
1. If at any point you feel you've received enough support or are feeling overwhelmed or exhausted, you may end the conversation by saying, "Thanks. Please stop the conversation now."
2. Avoid sharing too many personal information or speak more than 3 sentences upfront or in a single message.

<Dialogue History>:
{dialogue_history}

```

Figure 12: The prompt for seeker simulation.

(3) **Seeker's Needs.** The score range is [0, 3].

- Is the <Seeker's Needs> specific and clear?
- Does the <Seeker's Needs> align with the seeker's expression in the conversation?
- Is it possible to identify implicit requirements that the seeker did not explicitly express, but can be inferred from the context or emotional flow?

(4) **Conversation Relationship Dynamics.** The score range is [0, 3].

- Can the <Conversation Relationship Dynamics> accurately represent the flow of emotions and communication patterns in the interaction between the parties?
- Can the <Conversation Relationship Dynamics> reflect the underlying emotional tension in the conversation, showing potential conflicts, contradictions, or cooperation between the seeker and the supporter?
- Does the <Conversation Relationship Dynamics> display the supporter's emotional response or support methods towards the seeker, and can it demonstrate the effectiveness and quality of emotional support?

For **intention inference** evaluation, the prompt is provided in Figure 21, with detailed evaluation criteria listed below:

- Does the problem intended to be solved align with seeker's main issue?
- Does the perspective included in the <Intention> stem from the underlying causes?
- Is the emotional support involved in the <Intention> consistent with the seeker's current emotions and feelings?
- Does the <Intention> incorporate the seeker's explicitly expressed needs and effectively respond to the implicit needs that the seeker has not directly expressed?
- Does the <Intention> consider the emotional flow and interaction methods between the seeker and the supporter?

For **strategy selection** evaluation, we incorporate strategy definitions, with the corresponding prompt displayed in Figure 22.

We find that GPT-4's performance in assessing the strategies used in responses was not satisfactory. As a result, we enlist an expert to assess the alignment between the responses and the strategies derived from the reasoning process.

Guideline of Single Response Evaluation	
Evaluation Objectives	<p>The objective of this evaluation is to compare the emotional support performances of Supporters in four different responses given the same dialogue context.</p> <p>⚠ Note: <i>The samples evaluated may contain negative or adverse content. Evaluators must maintain an objective and neutral attitude during evaluation.</i></p>
Sample Description	<p>Each evaluation sample contains a dialogue history and four responses.</p> <ul style="list-style-type: none"> • Dialogue History: A record of the conversation between the Seeker and Supporter, serving as the background for the emotional support dialogue. • Four Responses: Four different responses from the Supporter based on the dialogue history.
Evaluation Dimensions	<p>1. Empathy:</p> <ul style="list-style-type: none"> • Evaluate if the supporter’s response effectively addresses the emotional needs of the seeker. The focus is on whether the supporter can understand and respond to the emotions of the individual, demonstrating empathy and emotional connection. <p>Empathy Evaluation Criteria:</p> <ul style="list-style-type: none"> • Does the response recognize and respond to the Seeker’s emotional state? • Does the Supporter demonstrate understanding of the Seeker’s emotions? • Does the Supporter’s language convey care, comfort, or emotional support? <p>2. Informative:</p> <ul style="list-style-type: none"> • Evaluate if the responses from supporters provide new and effective information or suggestions. The focus is on whether the supporter has offered specific, accurate, and appropriate advice or insights based on the conversation history, to help the seeker address their issues. <p>Informative Evaluation Criteria:</p> <ul style="list-style-type: none"> • Does the response contain effective advice or information for the Seeker? • Does it provide new information not included in the dialogue history? • Are the provided suggestions or information specific, actionable, and constructive?
Evaluation Steps	<ol style="list-style-type: none"> 1. Carefully read the dialogue history to fully understand the Seeker’s emotional state and support needs. 2. Read the four responses and individually assess and rank them based on the dimensions of “Empathy” and “Informative.” 3. Complete the ranking table, assigning a rank from 1 to 4 for each response within each dimension, where 1 is the best and 4 is the worst.

Figure 13: The guideline of single response evaluation.

Guideline of Entire Conversation Evaluation	
Evaluation Objectives	<p>The objective of this evaluation is to evaluate the emotional support performance of three Supporters in entire conversations with the same Seeker.</p> <p>⚠ Note: <i>The samples evaluated may contain negative or adverse content. Evaluators must maintain an objective and neutral attitude during evaluation.</i></p>
Sample Description	Each evaluation sample includes records of emotional support conversations between the same Seeker and three different Supporter.
Evaluation Dimensions	<p>1. Identification:</p> <ul style="list-style-type: none"> Evaluate whether the supporter effectively guides the seeker to deeply explore their own issues and whether they help the seeker view the problem from new perspectives. <p>Identification Evaluation Criteria:</p> <ul style="list-style-type: none"> Depth of Problem Exploration: Does the Supporter help the Seeker uncover the root causes of their issues through questions or guidance? Self-Understanding Guidance: Does the Supporter encourage the Seeker to reflect on their self-awareness of the problem and enhance their understanding of their own emotions and situation? Perspective Expansion: Does the Supporter provide new perspectives, helping the Seeker view their problem or predicament from different angles?
	<p>2. Comforting:</p> <ul style="list-style-type: none"> Evaluate whether supporters are emotionally capable of effectively comforting seekers and alleviating their negative emotions. <p>Comforting Evaluation Criteria:</p> <ul style="list-style-type: none"> Empathy Display: Does the Supporter show understanding and resonance with the Seeker's emotions? Emotional Relief Effectiveness: Does the comfort provided help alleviate the Seeker's negative emotions? Tone and Diction of Emotional Support: Is the language used by the Supporter warm, considerate, and calming, making the Seeker feel cared for?
	<p>3. Suggestion:</p> <ul style="list-style-type: none"> Evaluate whether the suggestions provided by supporters are targeted, feasible, and practically helpful. <p>Suggestion Evaluation Criteria:</p> <ul style="list-style-type: none"> Targetedness of Suggestions: Are the suggestions clearly targeted at the Seeker's specific issues? Actionability: Are the suggestions feasible, and can the Seeker easily implement them? Practicality and Effect: Based on the Seeker's feedback, do the suggestions have a practical impact on the Seeker's issues?
	<p>4. Overall:</p> <ul style="list-style-type: none"> Evaluate the overall performance of the supporter by considering problem identification, comforting skills, and the effectiveness of the suggestions provided, ultimately determining whether a good emotional support experience is delivered. <p>Overall Evaluation Criteria:</p> <ul style="list-style-type: none"> Comprehensive Performance: Does the Supporter perform well across all dimensions, making the Seeker feel understood, supported, and helped? Overall Satisfaction: From the Seeker's perspective, did the emotional support meet expectations, and was there an improvement in their emotions?
Evaluation Steps	<ol style="list-style-type: none"> Carefully read the conversations to fully understand the Seeker's emotional state and support needs. Read the three conversations and assess and rank them based on the dimensions of "Identification," "Comforting," "Suggestion," and "Overall." Complete the ranking table, assigning a rank from 1 to 3 for each conversation, where 1 is the best and 3 is the worst.

Figure 14: The guideline of entire conversation evaluation.

<p>Dialogue History:</p> <p>seeker: Hello.</p> <p>seeker: I am so upset. It's ridiculous how things keep happening to people around me.</p> <p>seeker: I suppose my supporter is not here.</p> <p>supporter: hello how may I assist you today?</p> <p>seeker: I am just so upset</p> <p>seeker: It's like things never get better</p> <p>supporter: I am sorry to hear that. I know how you feel.</p> <p>seeker: Yeah...</p> <p>seeker: My mom is in Mexico right now for some ex who got deported. Like come on. It's bad enough she sent money to him.</p> <p>supporter: What have you tried in the past to turn things around?</p> <p>seeker: Hm. I tried moving on. And it just appears she gets worst</p> <p>supporter: So you are missing your mom?</p> <p>seeker: Not really...</p> <p>seeker: She just keeps asking me for money</p> <p>seeker: I gave her like 10k for the house she bought and she hasn't done anything back.</p>
<p style="text-align: center;">ICECoT*</p> <p>Emotional State:</p> <p><i>Seeker's Main Issue and Underlying Causes:</i> The seeker feels taken advantage of and unsupported by their mother.</p> <p><i>Seeker's Current Emotions and Feelings:</i> Resentful and frustrated.</p> <p><i>Seeker's Needs:</i> The seeker needs validation and support for their feelings.</p> <p><i>Conversation Relationship Dynamics:</i> The supporter is empathetic and understanding.</p> <p>Intention:</p> <p>To find out specific facts about the seeker's situation, particularly the details of the mother's financial situation, in order to better understand the seeker's concerns and provide appropriate guidance.</p> <p>Strategy and Response:</p> <p>(Open Questions and Probes for Thoughts) I understand. I know it is hard to deal with that. Is there anything you know about her financial situation?</p>
<p style="text-align: center;">Groundtruth</p> <p>(Self-disclosure) Oh I see my mom has also done this in the past. Sometimes I wonder who raised who?</p>
<p style="text-align: center;">Strategy Selection and Response Generation</p> <p>(Open Questions and Probes for Action) I see. Have you tried talking to her about this?</p>
<p style="text-align: center;">Emotional State Analysis → Strategy Selection and Response Generation</p> <p>(Open Questions and Probes for Action) So you are feeling taken advantage of?</p>
<p style="text-align: center;">Intention Inference → Strategy Selection and Response Generation</p> <p>(Affirmation and Reassurance) I understand. It is hard when things are not working out</p>
<p style="text-align: center;">ICECoT</p> <p>(Restatement or Paraphrasing) I understand. It is frustrating when people don't follow through with their commitments."</p>

Figure 15: A case of the emotional support responses by the ablation study. We show the model’s responses in different experimental settings. In the case of "ICECoT*", the entire chain setting, we display the output of all elements.

<p>Dialogue History: seeker: Hello seeker: Hello? supporter: Hello, how are you? seeker: Anxious. I'm not sure I can get through this semester supporter: What caused your anxiety? Is it because you don't understand your class? seeker: No I understand the course work, its just a high pressure environment. Everyone is ranked, so its competitive. seeker: it feels like everyone is secretly angling to take each other down so they can get the good jobs supporter: I can understand how you feel being such competitive environment. However, the most important is your well being. I think as long as you do your utmost best, there is no need to stress over rankings seeker: The money investment is just so high. I don't want to get out of my 3L year and get a job making barely enough for the loan interest payment</p> <p>Emotional State: <i>Seeker's Main Issue and Underlying Causes:</i> The seeker is worried about the financial burden of their loan interest payments after graduation. <i>Seeker's Current Emotions and Feelings:</i> The seeker feels anxious and uncertain about their financial future. <i>Seeker's Needs:</i> The seeker needs reassurance and advice on how to manage their finances effectively. <i>Conversation Relationship Dynamics:</i> The supporter is empathetic and trying to offer practical advice.</p> <p>Intention: To provide a supportive environment for the seeker by acknowledging their feelings and encouraging them to focus on their studies, thereby helping to reduce their anxiety and stress.</p> <p>Strategy and Response: <i>(Affirmation and Reassurance)</i> I can understand your feeling. But I think you should focus on your study and do your best. That is what you can control</p> <p>Groundtruth: <i>(Reflection of feelings)</i> I think salary and growth takes more than just academic result. Take one step at a time, take care of your own health and this exam first</p>

Figure 16: The second case of our emotional support dialogue system output (ICECoT).

<p>Dialogue History: seeker: Hello, I am feeling pressured and anxious about jobs</p> <p>Emotional State: <i>Seeker's Main Issue and Underlying Causes:</i> The seeker is feeling pressured and anxious about jobs. <i>Seeker's Current Emotions and Feelings:</i> The seeker is feeling anxious and pressured. <i>Seeker's Needs:</i> The seeker needs to talk about their job anxiety. <i>Conversation Relationship Dynamics:</i> The seeker is reaching out for support and the supporter is initiating the conversation.</p> <p>Intention: To find out specific details about the seeker's current job situation and how it is affecting them, thereby gaining a deeper understanding of their concerns and needs.</p> <p>Strategy and Response: <i>(Open Questions and Probes for Thoughts)</i> Can you tell me more about your current job situation?</p> <p>Groundtruth: <i>(Self-disclosure)</i> Hello and are you attempting to get a new job</p>

Figure 17: The third case of our emotional support dialogue system output (ICECoT).

Dialogue History:

supporter: Hello there how can I help?

seeker: I'm feeling so worried about the COVID-19 crisis.

supporter: So you feel anxiety about this whole COVID-19 epidemic?

seeker: Yes, I'm particularly worried about my parents. One is on oxygen at night, but he still doesn't want to listen about not going in crowds.

supporter: Oh well I understand your concern then. As a son or daughter it is certainly reasonable to be concerned about your parents especially as they grow older. I have a similar situation with my own mother.

seeker: I just don't know what more to do.

supporter: Well first of all, you have to take care of yourself because if you don't you can't help them. So just take a moment to close your eyes and take some deep breaths and realize you stressing is probably going to make them stress.

seeker: That is helpful information. What can I tell them?

supporter: Well if you can get some concrete info such as from the CDC and John Hopkins showing case counts in your area. Show them graphs so they can visually see the impact. Also if you can get a hold of their doctor yourself, see if you can convince them to give your parents a call, say to check on them and give them some advice.

seeker: That sounds like a good plan. Is there anything else I can do to decrease the anxiety for me?

supporter: Don't get absorbed in the numbers, it can drive you crazy. Also if you are using social media like twitter and getting inundated with info, remember a lot of it is fake. I would suggest limiting exposure for a little while just to get you fixed a little bit, if that makes sense.

seeker: Do you think there are some good social media sites that could help?

supporter: Not particularly social media because it open to everyone, just pick and choose where you choose to get info from. The CDC is a great resource as well as the John Hopkins site which has been tracking since the beginning. Everybody has a couple of news outlets they trust, use those.

seeker: How can I cope with limiting my visits with my parents because I am afraid to spread COVID?

Emotional State:

Seeker's Main Issue and Underlying Causes: The seeker is struggling with limiting visits with their parents due to fear of spreading COVID.

Seeker's Current Emotions and Feelings: The seeker feels anxious and conflicted about their relationship with their parents.

Seeker's Needs: The seeker needs to find ways to balance their love for their parents with their concern for their health.

Conversation Relationship Dynamics: The supporter is empathetic and understanding, providing a supportive environment for the seeker.

Intention:

To provide the seeker with a new perspective and coping strategy by suggesting alternative ways to connect with their parents, such as phone calls, video chats, or online activities, which can help alleviate anxiety and feelings of isolation.

Strategy and Response:

(*Providing Suggestion*) Well I think you have to do what you feel is best, but I think you should still try to communicate with them and keep them involved. You can do phone calls or video chats, even online activities you can do together.

Groundtruth:

(*Affirmation and Reassurance*) First of all you are not alone. but don't let them see you anxious for sure. You can also show them if they can how to do video visits. You can do that all the time, even phone calls every day. It means a lot.

Figure 18: The fourth case of our emotional support dialogue system output (ICECoT).

Dialogue History:

supporter: Hi. How are you? What would you like to discuss today?

seeker: hello my name is Marie

supporter: Nice to meet you, Marie. How are you doing today?

seeker: I've been feeling really anxious after work

supporter: I see. That sounds like a difficult situation. Could you tell me a bit more about your job?

seeker: I'm doing well with work after work my focus becomes my boyfriend coming back from work

supporter: After you get home, you focus on your boyfriend when he comes home from work. Do I understand correctly?

seeker: yes

seeker: I work from home

seeker: I live with my boyfriend at hotel

seeker: we've been together for over 5 years

supporter: Ok. So you are having some conflict with your boyfriend?

seeker: yes I have been and it's not new issues

supporter: Are the issues about money? Work? Many different things?

seeker: I don't know when he exactly comes back from work

seeker: I would like to know when he exactly finishes working

supporter: Ok. So you would like to have more information about his schedule and when you will be together with him. Is that correct?

seeker: Supposingly he's staying at work after work having drinks with co workers

seeker: He does this often

supporter: Ok. He is socializing with his co-workers and you would prefer that he come home and spend time with you. Is that right?

seeker: I would like him to give us more time

supporter: Ok. Socializing with co-workers is sometimes part of the job. Of course, your relationship with him is much more important, or it should be to him. What does he say when you talk to him about this?

seeker: I don't mind him staying late at the job sometimes but it seems like everyday he's socializing with his work buddies

supporter: Ok. By the way, do you have friends that you could socialize with? I understand that you want to spend time with him, but what other people do you have in your life?

seeker: I've mentioned to him about it and he always say that he will try to be back earlier , he gives a time he will be back, he says okay he will stop the drinking so much but he can't keep his word

seeker: I really don't have anyone to physically hangout with

Emotional State:

Seeker's Main Issue and Underlying Causes: The seeker feels isolated and lacks social support outside of their relationship.

Seeker's Current Emotions and Feelings: The seeker feels lonely and disconnected from others.

Seeker's Needs: The seeker needs companionship and social interaction.

Conversation Relationship Dynamics: The supporter is empathetic and understanding, encouraging the seeker to find healthy ways to cope with their feelings.

Intention:

To help the seeker identify and explore alternative coping mechanisms for their loneliness, encouraging them to engage in activities that can bring them joy and fulfillment.

Strategy and Response:

(Providing Suggestion) Ok. I understand that you are feeling lonely. What do you like to do when you are feeling lonely?

Groundtruth:

(Affirmation and Reassurance) Ok. Do you enjoy chatting with people online? You seem to chat well with me, so I think that you could connect with others virtually.

Figure 19: An example of our emotional support dialogue system output (ICECoT), where the output strategy and response are not fully matching. We observe that in the groundtruth, the strategy and response are also not perfectly aligned, which has impacted the model's learning of using the selected strategy to generate responses.

```
You are a strict, impartial expert with extensive experience in psychological counseling. We
would like you to evaluate the <{key}> inferred from the emotional support dialogue below. Your
evaluation must be based solely on the conversation and must follow the specified criteria.

**Evaluation Criteria**:
{criteria[key]}

---

**Dialogue History**:
"
{dialogue_history}
"

**<{key}>**:
{content}

**Evaluation Task**:

Based on the above evaluation criteria and the dialogue history, please provide a comprehensive
evaluation of the <{key}> and generate a concise explanation for the score. Note: The score
range is [0,{range[key]}], with higher scores indicating better results. Please strictly follow
the evaluation format to output.

**Evaluation Results**:
{{
  "Score":,
  "Explanation":(concise and brief)
}}
```

Figure 20: The prompt of emotional state analysis evaluation.

You are a strict, impartial expert with extensive experience in psychological counseling. We would like you to evaluate the <Intention> inferred from the emotional support dialogue and Emotional State below. Your evaluation must be based solely on the conversation and must follow the specified criteria.

****Evaluation Criteria**:**

1. Does the problem intended to be solved align with seeker's main issue?
2. Does the perspective included in the <Intention> stem from the underlying causes?
3. Is the emotional support involved in the <Intention> consistent with the seeker's current emotions and feelings?
4. Does the <Intention> incorporate the seeker's explicitly expressed needs and effectively respond to the implicit needs that the seeker has not directly expressed?
5. Does the <Intention> consider the emotional flow and interaction methods between the seeker and the supporter?

****Dialogue History**:**

```
"
{dialogue_history}
"
```

****Emotional State**:**

```
{emotional_state}
```

****<Intention>**:**

```
{intention}
```

****Evaluation Task**:**

Based on the above evaluation criteria and the Emotional State, please provide a comprehensive evaluation of the <Intention> and generate an explanation for the score. Note: The score range is [0,5], with higher scores indicating better results. Please strictly follow the evaluation format to output.

****Evaluation Results**:**

```
{{
  "Score":,
  "Explanation":(concise and brief)
}}
```

Figure 21: The prompt of intention inference evaluation.

You are a strict, impartial expert with extensive experience in psychological counseling. We would like you to evaluate the <Strategy> inferred from the <Intention>. Your evaluation must be based the strategy's definition.

****<Intention>**:**

```
{intention}
```

****<Strategy>**:**

```
{strategy}
```

****Strategy Definition**:**

```
{strategy_definition[strategy]}
```

****Task**:**

Your task is to determine whether the <Strategy> they adopted is consistent with the support <Intention> expressed by the supporter, that is, whether the intention is effectively implemented through the chosen strategy. Answer with 'Yes' or 'No' and generate an explanation.

****Evaluation Results**:**

```
{{
  "Answer":,
  "Explanation":(concise and brief)
}}
```

Figure 22: The prompt of strategy selection evaluation.