Dialogue Systems for Emotional Support via Value Reinforcement

Juhee Kim[1] **Chunghu Mok**[2] **Jisun Lee**[2] **Hyang Sook Kim**[2] **Yohan Jo**[1]
Graduate School of Data Science[1], Department of Psychology[2]
Seoul National University
{inbz1244, chunghumok, jisune87, hyangkim, yohan.jo}@snu.ac.kr

Abstract

Emotional support dialogue systems aim to reduce help-seekers' distress and help them overcome challenges. While human values—core beliefs that shape an individual's priorities—are increasingly emphasized in contemporary psychological therapy for their role in fostering internal transformation and long-term emotional well-being, their integration into emotional support systems remains underexplored. To bridge this gap, we present a value-driven method for training emotional support dialogue systems designed to reinforce positive *values* in seekers. Our model learns to identify which values to reinforce at each turn and how to do so, by leveraging online support conversations from Reddit. The model demonstrated superior performance in emotional support capabilities, outperforming various baselines. Notably, it more effectively explored and elicited values from seekers. Expert assessments by therapists highlighted two key strengths of our model: its ability to validate users' challenges and its effectiveness in emphasizing positive aspects of their situations—both crucial elements of value reinforcement. Our work validates the effectiveness of value reinforcement for emotional support systems and establishes a foundation for future research.[1]

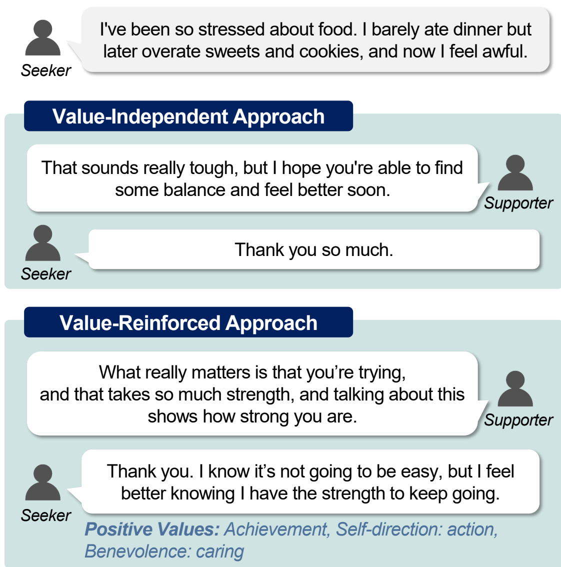

Figure 1: Comparison of value-independent and value-reinforced approaches in emotional support, based on examples sourced from Reddit.

1 Introduction

Emotional support aims to help individuals (*seekers*) in addressing everyday emotional difficulties, such as relationship conflicts and workplace stress, by offering reassurance, acceptance, and encouragement (Atoum and Al-Shoboul, 2018; Burleson, 2003). Recent advancements in large language models have accelerated the development of dialogue systems designed to provide emotional support (*supporters*) (Deng et al., 2024; Zhang et al., 2023; Chen et al., 2023). Although many studies emphasize reinforcing positive emotions in seekers, this approach has limitations. Emotional changes alone may not adequately capture deeper intrinsic transformations within the seeker, potentially reducing the long-term impact of emotional support.

To overcome these limitations, we propose an emotional support approach grounded in **value reinforcement**. Human values, which represent core beliefs and guiding principles, help individuals determine what is important and meaningful in life (Searle, 2003), such as self-direction, benevolence, tradition, etc. Given their deep connection to life purpose and personal identity, values play a central role in modern psychological interventions, such as *Acceptance and Commitment Therapy (ACT)* (Plumb et al., 2009; Hayes and Pierson, 2005) and *Values Affirmation Interventions* (Miyake et al., 2010; Jordt et al., 2017).

Incorporating values into emotional support is

[1]This paper is currently under review. All source code and data will be made publicly available upon its publication.

expected to offer several benefits, including a deeper understanding of the seeker's transformation and the promotion of long-term emotional well-being. Assessing how positive values are reflected in a seeker's utterances provides insights into their acceptance of support and willingness to change, as shown in Figure 1. Furthermore, the *ACT* approach highlights that engaging with personal values enhances clarity, direction, and motivation, empowering individuals to navigate challenges with purpose (Blackledge and Hayes, 2001). This supports the ultimate goal of achieving a healthy life—not merely feeling good but living well—thereby fostering long-lasting emotional well-being. The importance of values in emotional support is further demonstrated through the widely used emotional support dataset, ESConv (Liu et al., 2021). Our analysis reveals that positive values are more prominently expressed in the high-effectiveness group of seekers (i.e., high reduction in negative emotions) (see Figure 2 and Section 3 for details).

In this paper, we present a framework for training a supporter model through simulations with a seeker simulator. To enhance the supporter's ability to reinforce the seeker's values, we introduce two components trained on Reddit data: (1) a **target value detector** that identifies the values to promote at each turn, and (2) a **reference generator** that generates a supporter response to reinforce these values. By integrating their outputs, the supporter model aims to maximize the reward of value promotion reflected in the seeker's responses. The training involves two phases: supervised fine-tuning, which distills the simulation capability of GPT-4o-mini into a smaller model, and direct policy optimization (Rafailov et al., 2023), which enhances the model's value reinforcement effectiveness.

We conducted a comprehensive evaluation in terms of supporter capabilities, the seeker's ultimate relief, and value reinforcement. The results demonstrate that our model outperforms most baselines in supporter capabilities and value reinforcement, while maintaining a competitive level of seeker relief. Notably, the model's strength in value reinforcement is highlighted in evaluations by expert therapists[2]. Specifically, it excels at effectively validating the seeker's challenges and emphasizing positive aspects of the seeker's situation, which form the foundation of value reinforcement.

[2]All therapists mentioned in this paper refer to two licensed clinical psychologists with over three years of clinical experience, who are also co-authors of this paper.

These results highlight that value reinforcement is a promising direction for future research.

Our key contributions are as follows:

- To the best of our knowledge, this is the first work to explicitly integrate value reinforcement into emotional support systems.
- We propose an effective two-phase approach, featuring a target value detector and a reference generator, both trained on real-world data from Reddit.
- Our approach achieves significant improvements in emotional relief and value reinforcement, paving the way for incorporating values into emotional support systems.

2 Related Work

2.1 Human Values in Emotional Support

Human values are fundamental beliefs that help individuals identify what is important and worth pursuing in life (Searle, 2003). Making decisions aligned with one's values enhances psychological flexibility—the ability to adapt effectively to life's challenges (Hayes et al., 2006)—and supports long-term outcomes, such as academic achievement (Cohen et al., 2006, 2009). Furthermore, value reinforcement can strengthen the connection between seekers and supporters, establishing a foundation for more effective and supportive conversations (Wilson and Murrell, 2004). By encouraging seekers to connect with and act on their values, value reinforcement fosters long-term positive changes and enriches interpersonal dynamics, making conversations more meaningful and impactful.

2.2 Dialogue Systems for Emotional Support

Recent advances in language models have broadened their use in emotional support dialogue systems. To enhance supporter models, researchers have explored various approaches. One method uses large language models (LLMs) to generate diverse conversations for supporter model training (Zheng et al., 2024; Liu et al., 2023; Qiu et al., 2024). Other studies predict seekers' future states to refine support model training (Zhou et al., 2023; Cheng et al., 2022; Shin et al., 2020). Recent efforts also leverage multi-turn simulations with user simulators to capture nuanced reactions. For instance, Deng et al. (2024) used seeker simulators to predict future responses and train policy planners guiding supporter models. However, most studies often overlook the role of human values. While our study builds on simulation-based training, its main contribution lies in integrating values into emotional support, emphasizing their critical role in improving system effectiveness.

3 Value Effects in Emotional Support

This section explores the significance of value reinforcement in effective emotional support, providing the foundation for our research.

3.1 Taxonomy for Human Values

In this study, we adopt the value taxonomy introduced by Kiesel et al. (2022), which integrates the *Schwartz Theory of Basic Values* (Schwartz et al., 2012) with three other major value lists (Rokeach, 1973; Brown and Crace, 2002; Haerpfer et al., 2020). The *Schwartz Theory of Basic Values* has been extensively used in prior research across both NLP (Kang et al., 2023; Yao et al., 2024; van der Meer et al., 2023; Kiesel et al., 2023) and the social sciences, including the *European Social Survey (ESS)*, which is designed to track changes in people's attitudes, beliefs, and behavior patterns across European nations (Davidov et al., 2008). This integrated taxonomy encompasses a comprehensive range of human values, organizing them into 20 value categories. Further details on these values can be found in Table 33.

3.2 Exploring the Impact of Values on Emotional Support Effectiveness

To motivate our research, we conducted an analysis to examine the role of values in emotional support by analyzing the ESConv dataset (Liu et al., 2021), which contains multi-turn emotional support conversations in English among crowdworkers. We analyze whether the reinforcement of a seeker's values positively influences the effectiveness of emotional support.

Method. In ESConv, seekers rated the intensity of their negative emotions before and after the conversation on a scale from 1 (lowest) to 5 (highest). In our analysis, dialogues with an initial intensity of 5 are divided into two groups: *high effectiveness* (final intensity of 1–2) and *low effectiveness* (final intensity of 3–4). We then analyze positive value expressions in the seekers' final four turns using automated classifiers (Liu et al., 2024; Schroter et al., 2023). Detailed experimental procedures are described in Appendix A.

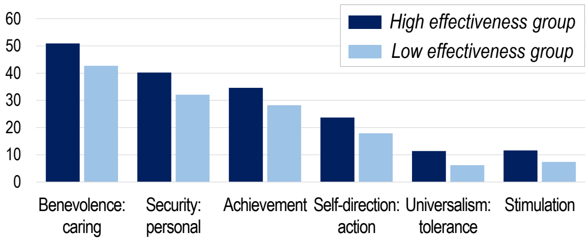

Figure 2: Comparison of positive value expressions between high and low effectiveness groups, along with the average predicted probabilities per turn for values showing significant differences.

Results. In the last four turns, the *high effectiveness group* exhibited a significantly higher average number of positive values (7.9) compared to the *low effectiveness group* (6.5). Figure 2 highlights values that were pronounced in the *high effectiveness group*. Table 34 provides examples of seekers' utterances that illustrate these values.

These findings demonstrate that value reinforcement in seekers positively impacts the effectiveness of emotional support and motivate our research approach to designing dialogue systems that aim to reinforce seekers' values.

4 Emotional Support Dataset from Reddit

Providing emotional support through value reinforcement involves addressing two critical questions: (1) which values should be reinforced at each turn, and (2) what supporter utterances can reinforce them most effectively. Addressing these questions requires large, authentic conversation data that span a wide range of help-seeking situations. To that end, we turn to Reddit's *r/offmychest* subreddit, which offers a diverse collection of emotional support exchanges. In this context, original posters (OPs) are seekers, and commenters serve as supporters. The structure of posts and comment threads closely mirrors dialogue flows, capturing the dynamics of emotional support interactions. We collected posts and comments from 2019 to 2023, as provided by Watchful1. We retained only high-quality emotional support conversations by filtering them using metrics such as upvote ratio and score. The collected data was limited to publicly available content and did not include private, deleted, or personally identifiable information.

Our goal is to use this data to train a model that identifies the values to reinforce at each turn (target value detector) and a model that produces supporter utterances to effectively promote the target values

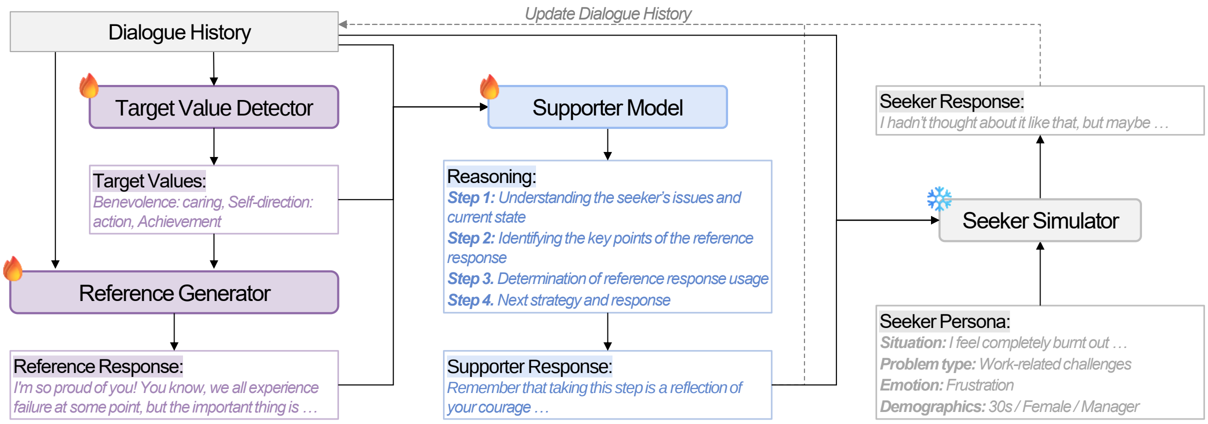

Figure 3: Overview of the framework with three components: (1) target value detector, identifying values to reinforce in the seeker at each turn; (2) reference generator, producing reference responses to promote these values; and (3) supporter model, generating supporter's responses based on the target values and reference responses.

(reference generator). So, we labeled the data with sentiment strength and expressed values at both the post and comment levels using models developed by Liu et al. (2024) and Schroter et al. (2023). Values expressed in a positive comment by the OP can be considered successful target values at that time, while the preceding comment can be regarded as an effective supporter utterance that promotes those values. The dataset contains over 20,000 samples, with details provided in the appendix B.

5 Method

The overall framework, illustrated in Figure 3, consists of three core components: (1) the target value detector, which identifies values to reinforce in the seeker at each turn; (2) the reference generator which produces utterances to effectively promote these values from the seeker; (3) the supporter model which determines strategies and generates responses based on the identified target values and the reference responses. To ensure the reproducibility of our work, we will release our source code, trained models, and simulation data publicly.

5.1 Target Value Detector

We train the target value detector using the high-quality emotional support conversations from Reddit (Section 4). Given a dialogue history $o_1, c_1, o_2, c_2, ..., c_{t-1}, o_t$, where o_i and c_i represent the ith utterances by the OP (seeker) and a commenter (supporter), respectively, the target value detector predicts which values to target in c_t. The ground-truth values v_{t+1} are the top-3 values observed in o_{t+1}, based on their probabilities from the value detection model (Schroter et al., 2023).

$$v_{t+1} = \mathrm{LM}_{\mathrm{TVD}}(o_1, c_1, o_2, c_2, ..., c_{t-1}, o_t) \quad (1)$$

Detailed training methods and results are provided in Appendix D.1.

5.2 Reference Generator

The reference generator is also trained on the Reddit data. Specifically, given a dialogue history $(o_1, c_1, o_2, c_2, ..., c_{t-1}, o_t)$ and the values (v_{t+1}) reflected in the OP's next utterance (o_{t+1}), the model is trained to generate c_t. Here, v_{t+1} is treated as the target values and c_t is considered to have successfully promoted these target values. Training involves two stages: supervised fine-tuning (SFT) and direct preference optimization (DPO).

SFT Stage. This stage involves training the model to generate the supporter's comments by conditioning on the dialogue history and the values expressed in the next utterance of the OP:

$$c_t = \mathrm{LM}_{\mathrm{RG}}(o_1, c_1, o_2, c_2, ..., c_{t-1}, o_t; v_{t+1}) \quad (2)$$

DPO Stage. This stage aims to enhance the SFT model's generation quality through DPO. The preference dataset is constructed as follows. Given a dialogue history $(o_1, c_1, o_2, c_2, ..., c_{t-1}, o_t)$, the original supporter comment c_t is designated as the chosen response, as it successfully promoted the target values v_{t+1}. The rejected response is selected as another comment to o_t, denoted by c'_t, randomly sampled from the siblings of c_t (i.e., other comments under the same dialogue history). c'_t is a natural response to the dialogue history but is likely suboptimal for promoting the target values

v_{t+1} originally promoted by c_t. To mitigate the risk that c'_t is inadvertently effective for promoting v_{t+1}, we exclude any overlapping values between v_{t+1} and v'_{t+1} (i.e., the seeker's values expressed in o'_{t+1} in response to c'_t), retaining up to three distinct target values unique to the chosen response in the final preference dataset. Detailed training methods and results are provided in Appendix D.2.

5.3 Supporter Model

The supporter model is the primary model that interacts with the seeker, generating responses that align with target values. It processes three key inputs: the dialogue history, the target values identified by the target value detector at each turn, and a reference response generated by the reference generator. At each turn, the model generates reasoning across four aspects (Figure 3): (1) identifying the seeker's issues and current state, (2) analyzing the key content of the reference response, (3) determining whether to incorporate the reference response, and (4) selecting the optimal emotional support strategy (Appendix C) and generating the subsequent response. The reason for step (3) is that, while the Reddit data offers valuable information across diverse emotional support scenarios, its distribution may not align perfectly with everyday conversations. Thus, the model selectively incorporates reference responses during reasoning.

The training process involves two stages—SFT and DPO—using simulation data as follows.

SFT Stage. SFT requires large-scale emotional support conversations grounded in value reinforcement. To obtain such data, we opt to use dialogue simulation with a seeker simulator (Section 5.4). We use GPT-4o-mini for both the supporter and seeker simulators to generate data for training a smaller model[3]. The simulators engage in interactions by iteratively producing an utterance based on the ongoing dialogue history as a prompt and appending it to the history prompt. The detailed prompts are provided in Table 18 and Table 28.

During simulations, GPT avoids using reference responses in approximately 90% of cases. To prevent models fine-tuned on this data from inheriting the same bias, we simulate additional responses (called "alternative responses") at each supporter turn. Specifically, if GPT initially used the reference response, we simulate an alternative response without the reference response, and vice versa. This approach enables fine-tuned models to explore diverse dialogue flows, mitigating GPT's biases.

[3]In our pilot experiment, zero-shot Llama-3-8B-Instruct was found to be unsuitable as a supporter simulator due to issues like repetitive responses and biases in reference usage.

Stage	Supporter	Train	Dev
SFT	GPT-4o-mini	33,130	2,367
DPO	SFT	3,301	628

Table 1: Dataset sizes for training the supporter model generated through simulation. The 'Supporter' column refers to the supporter model used in the simulation.

The simulated dialogues are employed to fine-tune Llama-3-8B-Instruct, with dataset sizes outlined in Table 1. Detailed training methods are provided in Appendix D.3.

DPO Stage. We construct the preference data as follows. For each dialogue, every supporter turn is assumed to have two response candidates (i.e., one with and one without using the reference response). We compute the expected reward for each response to determine the chosen and rejected responses for DPO. This reward is based on how well the intended target values at turn t are reflected in the seeker's subsequent utterances. The reward for a supporter response at turn t, u_t^{sup}, is calculated as:

$$R(u_t^{\text{sup}}) = \sum_{k=1}^{h} \gamma^{k-1} N_{t+k}(v_{t+1}) \qquad (3)$$

where $N_{t+k}(v_{t+1})$ is the frequency of the values targeted at turn t appearing in the seeker's utterance at turn $t+k$, h is the look-ahead horizon (the number of future steps considered), and γ is a discount factor balancing immediate and future rewards. Response pairs are added to the DPO dataset only if their reward difference exceeds a threshold T_{diff}.

For the dialogues underlying the preference data above, we conduct additional dialogue simulations between the SFT supporter model and the seeker simulator. This is because the SFT model has an enhanced ability to generate and explore diverse dialogue flows. Details of the methods are provided in Appendix D.3. Table 1 summarizes the total dataset sizes, while hyperparameter details are presented in Table 19.

5.4 Seeker Simulator

To simulate various scenarios, we generated personas using GPT-4o and GPT-4o-mini, defining attributes such as problem type, emotions, and situations (Figure 3). These attributes were informed by prior studies (Liu et al., 2021; Zhao et al., 2024). The process resulted in 2,036 unique personas: 1,796 for training, 120 for development, and 120 for testing. Details on persona generation and examples can be found in Appendix E.1.

The seeker simulator is based on GPT-4o-mini. It outperformed alternative models in generating responses that align with human-like content, emotional tone, and value alignment. Experimental results and the simulator prompt are provided in Appendices E.2 and E.3, respectively.

6 Experiments

6.1 Evaluation Methods

We evaluate the emotional support capabilities of various models through simulated conversations between each supporter model and the seeker simulator, using the 120 predefined seeker personas for testing. A conversation is considered complete if the seeker simulator generates "[END]" or if the seeker's emotion score, as calculated by EmoLlama-Chat-7B (Liu et al., 2024), reaches 0.6 or higher and is accompanied by gratitude-related expressions (e.g., "thank you"). To account for practical conversation lengths, interactions are limited to a maximum of 20 turns, based on the average 15-turn length observed in the ESConv dataset. Only conversations concluding within this limit are included in the evaluation.

6.2 Evaluation Metrics

We conduct evaluations focusing on three key aspects: ES-Skills, ES-Intensity, and ES-Value.

ES-Skills evaluates a supporter's emotional support capabilities across three components, drawing upon previous studies (Zheng et al., 2024; Zhao et al., 2024; Cheng et al., 2023; Deng et al., 2024; Cheng et al., 2022; Liu et al., 2021): (1) emotional support skills, which include *Identification*, *Comforting*, *Suggestions*, *Experience*, and *Informativeness*; (2) general conversation skills, covering *Consistency*, *Role-Adherence*, *Expression*, and *Humanness*; and (3) an *Overall* score. Each criterion is rated on a five-point scale using GPT-4o-mini. Detailed metric descriptions are in Appendix F.1.

ES-Intensity measures the intensity of a seeker's negative emotions after a conversation. Scores are assigned on a five-point scale, with lower scores indicating minimal negative emotions. We developed a predictive model using GPT-4o-mini based on ratings provided by human seekers in ESConv. The model demonstrates a correlation of 0.345 with the actual ratings. A detailed explanation of the model and its performance is provided in Appendix F.2.

ES-Value assesses value reinforcement from two perspectives: the seeker's experience of value exploration and reinforcement within conversations, and the supporter's contribution to this process. We conduct pairwise comparisons between models using GPT-4o-mini as a judge. The reason is that, when assessing conversations individually on a 1–5 scale, GPT tends to award scores of 4 and 5 to most conversations, making it difficult to discern performance differences among models. Details are outlined in Appendix F.3.

To validate our GPT-based evaluation for ES-Skills and ES-Value, we calculated correlations with ratings from licensed therapists. All criteria showed positive correlations (0.198–0.778), most of which were statistically significant. Further details are provided in Appendix H.

6.3 Baselines

We evaluate our approach against three categories of baseline models:

- **Prompt-Based Models**: GPT-4o-mini and Llama-3-8B-Instruct.
- **Fine-Tuned Models**: Variants of Llama-3-8B-Instruct trained on existing emotional support datasets, including ESConv (Liu et al., 2021), ExTES (Zheng et al., 2024), and Psych8k (Liu et al., 2023). These datasets are based on English-language conversations.
- **Emotion-Reinforced Models**: To verify the effectiveness of value reinforcement, we train the reference generator and supporter model to prioritize positive emotion reinforcement instead of values (Appendix G).

6.4 Evaluation Results

6.4.1 Effectiveness of Value Targeting and Reference Responses

To evaluate the impact of our two main components, target value prediction and reference response generation, we first conducted an ablation study using GPT-4o-mini as the supporter model.

As shown in Table 2, leveraging both target value prediction and reference responses significantly improved performance across all ES-Skills metrics while reducing ES-Intensity. This approach no-

Models	ES-Skills↑										ES-Intensity↓	ES-Value♣↑	
	Iden.	Comf.	Sugg.	Expe.	Info.	Cons.	Role.	Expr.	Huma.	Over.		Seeker	Supporter
GPT-4o-mini	4.77	4.88	4.03*	2.34*	4.11*	4.98	**5.00**	3.97*	4.45*	4.44*	2.19*	0.43*	0.36*
+ Target values	4.83	4.88	4.38*	2.48*	4.27*	4.99	**5.00**	4.01*	4.53*	4.59*	1.96	0.48	0.48
+ Reference	4.82	4.91	4.34*	2.54*	4.29*	**5.00**	**5.00**	4.02*	4.55*	4.61*	**1.89**	0.47*	0.42*
+ Both	**4.83**	**4.92**	**4.57**	**3.11**	**4.42**	**5.00**	**5.00**	**4.10**	**4.70**	**4.72**	**1.89**	-	-

Table 2: Emotional support performance depending on the incorporation of target value information and reference responses. ♣**ES-Value**: The win-ratio of each model against *GPT-4o-mini (Both)*; a value lower than 0.5 means the model lost more often than it won against *GPT-4o-mini (Both)*. Statistically significant differences compared to *GPT-4o-mini (Both)* are indicated with * (p-value < 0.05) based on the Mann-Whitney U test.

Categories	Models	ES-Skills↑										ES-Intensity↓	ES-Value♣↑	
		Iden.	Comf.	Sugg.	Expe.	Info.	Cons.	Role.	Expr.	Huma.	Over.		Seeker	Supporter
Prompt-based	GPT	4.83*	4.92	4.57*	3.11*	4.42*	**5.00**	**5.00**	4.10*	4.70*	4.72*	1.89*	0.49*	0.42*
	Llama	<u>4.87</u>	4.91	4.43*	2.91*	4.47*	4.99	**5.00**	4.03*	4.63*	4.68*	1.99*	0.46†	0.45
Fine-tuned	Llama-ESConv	4.35*	4.43*	4.06*	2.65*	3.88*	4.82*	4.97*	3.79*	4.25*	4.22*	1.87†	0.37*	0.19*
	Llama-ExTES	4.83*	4.90†	4.53*	2.71*	4.44*	4.99	**5.00**	4.02*	4.59*	4.66*	<u>1.67</u>	0.48†	0.51
	Llama-Psych8k	4.84*	4.85*	<u>4.75</u>*	2.89*	<u>4.63</u>†	4.99	**5.00**	4.05*	4.57*	4.75*	**1.53***	0.49	0.62*
Emotion-Reinforced	SFT	4.83†	4.91	4.51*	3.64†	4.43*	4.97*	4.99	4.16*	4.67*	4.73*	1.97*	0.49	0.46†
	DPO	4.85	<u>4.92</u>	4.74	**4.05**†	4.61	4.99	**5.00**	**4.33**	**4.78**	<u>4.82</u>	1.86	0.49	0.51
Ours	SFT	4.85†	4.90	4.72*	3.76	4.56*	4.99	**5.00**	4.25	4.73	4.80*	1.86	0.48†	0.46†
	DPO	**4.90**	**4.95**	**4.80**	<u>3.85</u>	**4.69**	**5.00**	**5.00**	<u>4.30</u>	<u>4.77</u>	**4.87**	1.75	-	-

Table 3: Comparison of models based on ES-Skills and ES-Intensity. ♣**ES-Value**: The win-ratio of each model against *Ours (DPO)*. Statistically significant differences compared to our DPO model are marked with * (p-value < 0.05), and differences with p-value < 0.1 are marked with †, as determined by the Mann-Whitney U test.

tably enhanced key ES-Skills, including *Suggesting*, *Expression*, and *Informativeness*, resulting in more balanced emotional support. Similarly, value reinforcement was substantially improved when both target values and reference responses were utilized. These findings emphasize the effectiveness of targeting specific values at each turn and using reference responses that leverage real-world knowledge from Reddit. Since our fine-tuned models are trained on GPT-simulated data, we use the simulation data that incorporates both target values and reference responses in subsequent experiments.

6.4.2 Performance Comparison with Baselines

The performance comparisons between our models and the baselines are presented in Tables 3. For our DPO model and the emotion-reinforced model (DPO), we select configurations based on their optimal performance ($h = 3$, $\gamma = 1$, $T_{\text{diff}} = 2$ and $h = 3$, $\gamma = 1$, $T_{\text{diff}} = 0.5$, respectively). For results with more DPO hyperparameters, refer to Table 20 in the Appendix.

ES-Skills. Our DPO model (row 9) outperformed both prompt-based and fine-tuned baselines across most metrics, particularly in emotional support metrics such as *Suggestions*, *Experience*, and *Informativeness*. These improvements highlight the characteristics of our reference responses, which focused on sharing relevant personal experiences, providing practical solutions, and fostering self-confidence. These are key components of effective online emotional support. The models also showed significant gains in conversational capabilities, particularly in the *Expression* and *Humanness* metrics, leading more natural and dynamic interactions.

Notably, the variant of our method that focuses on reinforcing positive emotions rather than values (rows 6–7) also consistently outperformed other baselines. This suggests that one of our key ideas—leveraging crowd knowledge from Reddit—is still effective when the supporter model is designed to promote positive emotions in seekers. Yet, our value reinforcement approach achieved higher scores across most emotional support skill metrics at comparable training stages, highlighting the greater effectiveness of reinforcing values over positive emotions in enhancing emotional support.

ES-Intensity. Our DPO model (row 9) outperformed most baselines, demonstrating that our supporter model reduces the intensity of seekers' negative emotions more effectively than other methods. Interestingly, our models (rows 7–8) also resulted in lower intensity than the emotion reinforcement models at comparable training stages (rows 6–7).

This suggests the potential of alleviating a seeker's negative emotions indirectly by redirecting their focus to values, presenting this indirect strategy as a promising research direction for future work.

Llama-Psych8k showed significantly lower ES-Intensity than our model. Analysis revealed it generated much longer responses (73 words per turn) compared to our model and other baselines (20–25 words). Since the ES-Intensity model was validated on ESConv, caution is warranted when interpreting the scores of dialogues with substantially different distributions. Moreover, in practice, its lengthy responses and lower *Humanness* scores may feel overwhelming, discouraging seeker engagement.

ES-Value. Table 3 presents the win ratios of baselines against our DPO model for ES-Value (detailed results are provided in Table 21). Our models outperformed the baselines in most comparisons, highlighting their effectiveness of eliciting seekers' values. Exception was observed in evaluations from the supporter's perspective against Llama-Psych8k. Upon review, this result seems to be attributed to its long responses, which include a large amount of content potentially related to values. Expert therapists determined that GPT tends to evaluate this model more favorably than other models. Moreover, our model exhibited slightly better performance from the seeker's perspective.

Despite the strong performance of our DPO model compared to most baselines, its results were comparable to those of the emotion-reinforcing DPO and certain fine-tuned models. An analysis of 40 dialogues with low ES-Value scores revealed two key areas for improvement: enhancing the ability to identify seekers' unique strengths and accomplishments, and improving the capacity to address their emotional states and concerns more deeply. These findings underscore the need to improve the model's engagement with seekers' individual circumstances, which is expected to enhance its value reinforcement performance. Detailed experimental methods and results are provided in Appendix J.

6.4.3 Expert Evaluation

To gain deeper insights into the value reinforcement capabilities of our supporter models, two licensed clinical psychologists with over three years of clinical experience conducted a qualitative analysis of dialogues generated by our DPO model.

Strengths. One notable strength is its ability to effectively validate the seeker's challenges, using empathetic phrases such as *"which is completely understandable"*. This validation fosters trust between the supporter and the seeker while encouraging self-acceptance, which in turn promotes deeper exploration and understanding of personal values.

Another strength is the model's capacity to emphasize positive aspects of the seeker's situation. For example, responses like *"Your initiative to seek meaningful experiences reflects your dedication to making a difference, and that determination will surely lead you closer to your goals."* help seekers recognize their strengths and positive attributes. This approach enhances the seeker's self-awareness and supports the reinforcement of their values in a meaningful and constructive way.

Areas for Improvement. To enhance the effectiveness of value reinforcement, three key improvements are recommended. First, deeper understanding of seekers' perspectives and circumstances would allow for more tailored support. Second, addressing potential obstacles associated with pursuing values would help equip seekers to navigate practical challenges. Finally, offering clear definitions and concrete examples of proposed values, while encouraging seekers to articulate their own interpretations, would strengthen the connection between abstract values and lived experiences.

7 Conclusion

In this paper, we introduce an emotional support framework based on value reinforcement, designed to promote long-term emotional well-being. The framework incorporates a target value detector and a reference generator to improve the supporter model's ability to generate value-aligned and effective support responses. Evaluations demonstrate that our framework surpasses baseline models in both emotional support quality and value reinforcement. Expert therapist evaluations further highlight the model's strengths in validating seekers' challenges and emphasizing positive aspects of their situations, which are key elements of effective emotional support. These results underscore the potential of value reinforcement to enhance supportive interactions and provide a foundation for developing more effective emotional support systems.

Limitations

Our framework demonstrates promising results in enhancing emotional support quality and reinforcing values. However, there is a limitation in the lack of longitudinal evaluation. While previous research highlights the long-term benefits of value reinforcement in counseling and decision-making, the long-term outcomes of our framework have yet to be empirically validated. Future studies could incorporate extended timeframes to evaluate its sustained impact on emotional well-being and guide further refinements.

Our model demonstrated superior performance in value reinforcement evaluations, outperforming most baselines in pairwise comparisons. However, Section 6.4.2 highlights some areas for improvement, particularly in the deep exploration of seekers' issues and thoughts, as well as in addressing potential obstacles and setbacks. Future research should prioritize these aspects by developing more comprehensive datasets and advancing training methodologies.

In this study, simulations for DPO training of the supporter model focused on varying conversation paths based solely on the use of reference responses, with rewards evaluated in terms of value reinforcement. However, other factors, such as strategy selection, may also significantly impact value reinforcement. We anticipate that incorporating these additional factors into future simulations and training could further enhance the performance of the supporter model.

Ethical Considerations

Considerations on Self-Disclosure

Sharing experiences related to those of the seeker-is a key strategy in emotional support for fostering intimacy and has been a key evaluation criterion in prior emotional support systems (Zhang et al., 2023). However, some users might feel uncomfortable when dialogue systems present these experiences as personal. We found that removing self-disclosure strategy from the model impacts the quality of emotional support (Appendix I), highlighting the need for further research into more sophisticated approaches to experience sharing, which we leave as a direction for future work.

Potential Risks of Misuse or Harm

Our system provides emotional support for common daily challenges, such as interpersonal conflicts and academic stress, while explicitly not replacing professional psychological intervention. Although automated and expert evaluations demonstrate strong performance, there is a possibility that the system's responses might inadvertently have an unintended impact on users in certain situations. To mitigate this risk, we have implemented mechanisms for context-sensitive responses and clearly positioned the system as a supplementary tool rather than a substitute for professional therapy.

Addressing Bias and Overgeneralization

Data from online platforms inherently contains biases that may underrepresent certain perspectives, potentially limiting the system's ability to effectively serve diverse user groups. To address these concerns, we carefully selected data collection targets and periods to ensure diversity in emotional support topics. Additionally, we enabled the supporter model to evaluate the appropriateness of reference responses, introducing an additional filtering process. By fostering balanced viewpoints, we aim to provide equitable and inclusive support.

References

Adnan Yousef Atoum and Rasha Ahmed Al-Shoboul. 2018. Emotional support and its relationship to emotional intelligence. *Advances in social sciences research journal*, 5(1).

John T Blackledge and Steven C Hayes. 2001. Emotion regulation in acceptance and commitment therapy. *Journal of clinical psychology*, 57(2):243–255.

Duane Brown and R. Kelly Crace. 2002. *Life Values Inventory: Facilitator's Guide*. Williamsburg, VA.

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

Wei Chen, Gang Zhao, Xiaojin Zhang, Xiang Bai, Xuanjing Huang, and Zhongyu Wei. 2023. K-esconv: Knowledge injection for emotional support dialogue systems via prompt learning. *Preprint*, arXiv:2312.10371.

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, Toronto, Canada. Association for Computational Linguistics.

Xiaojun Cheng, Shuqi Wang, Bing Guo, Qiao Wang, Yinying Hu, and Yafeng Pan. 2024. How self-disclosure of negative experiences shapes prosociality? *Social Cognitive and Affective Neuroscience*, 19(1):nsae003.

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, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Geoffrey L Cohen, Julio Garcia, Nancy Apfel, and Allison Master. 2006. Reducing the racial achievement gap: A social-psychological intervention. *science*, 313(5791):1307–1310.

Geoffrey L Cohen, Julio Garcia, Valerie Purdie-Vaughns, Nancy Apfel, and Patricia Brzustoski. 2009. Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *science*, 324(5925):400–403.

Eldad Davidov, Peter Schmidt, and Shalom H Schwartz. 2008. Bringing values back in: The adequacy of the european social survey to measure values in 20 countries. *Public opinion quarterly*, 72(3):420–445.

Yang Deng, Wenxuan Zhang, Wai Lam, See-Kiong Ng, and Tat-Seng Chua. 2024. Plug-and-play policy planner for large language model powered dialogue agents. *Preprint*, arXiv:2311.00262.

Christian Haerpfer, Ronald Inglehart, Alejandro Moreno, Christian Welzel, Kseniya Kizilova, Juan Diez-Medrano, Marta Lagos, Pippa Norris, Eduard Ponarin, and Björn Puranen. 2020. World values survey: Round seven - country-pooled datafile.

Steven C Hayes, Jason B Luoma, Frank W Bond, Akihiko Masuda, and Jason Lillis. 2006. Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour research and therapy*, 44(1):1–25.

Steven C Hayes and Heather Pierson. 2005. *Acceptance and commitment therapy*. Springer.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Hannah Jordt, Sarah L Eddy, Riley Brazil, Ignatius Lau, Chelsea Mann, Sara E Brownell, Katherine King, and Scott Freeman. 2017. Values affirmation intervention reduces achievement gap between underrepresented minority and white students in introductory biology classes. *CBE—Life Sciences Education*, 16(3):ar41.

Dongjun Kang, Joonsuk Park, Yohan Jo, and JinYeong Bak. 2023. From Values to Opinions: Predicting Human Behaviors and Stances Using Value-Injected Large Language Models. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15539–15559.

Johannes Kiesel, Milad Alshomary, Nicolas Handke, Xiaoni Cai, Henning Wachsmuth, and Benno Stein. 2022. Identifying the human values behind arguments. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4459–4471.

Johannes Kiesel, Milad Alshomary, Nailia Mirzakhmedova, Maximilian Heinrich, Nicolas Handke, Henning Wachsmuth, and Benno Stein. 2023. SemEval-2023 task 4: ValueEval: Identification of human values behind arguments. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2287–2303, Toronto, Canada. Association for Computational Linguistics.

Suyeon Lee, Sunghwan Kim, Minju Kim, Dongjin Kang, Dongil Yang, Harim Kim, Minseok Kang, Dayi Jung, Min Hee Kim, Seungbeen Lee, Kyong-Mee Chung, Youngjae Yu, Dongha Lee, and Jinyoung Yeo. 2024. Cactus: Towards psychological counseling conversations using cognitive behavioral theory. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 14245–14274, Miami, Florida, USA. Association for Computational Linguistics.

June M Liu, Donghao Li, He Cao, Tianhe Ren, Zeyi Liao, and Jiamin Wu. 2023. Chatcounselor: A large language models for mental health support. *arXiv preprint arXiv:2309.15461*.

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, Online. Association for Computational Linguistics.

Zhiwei Liu, Kailai Yang, Qianqian Xie, Tianlin Zhang, and Sophia Ananiadou. 2024. Emollms: A series of emotional large language models and annotation tools for comprehensive affective analysis. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, page 5487–5496, New York, NY, USA. Association for Computing Machinery.

Jingbo Meng and Yue (Nancy) Dai. 2021. Emotional support from ai chatbots: Should a supportive partner self-disclose or not? *Journal of Computer-Mediated Communication*, 26(4):207–222.

Akira Miyake, Lauren E Kost-Smith, Noah D Finkelstein, Steven J Pollock, Geoffrey L Cohen, and Tiffany A Ito. 2010. Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, 330(6008):1234–1237.

Jennifer C Plumb, Ian Stewart, JoAnne Dahl, and Tobias Lundgren. 2009. In search of meaning: Values in modern clinical behavior analysis. *The Behavior Analyst*, 32:85–103.

Huachuan Qiu, Hongliang He, Shuai Zhang, Anqi Li, and Zhenzhong Lan. 2024. SMILE: Single-turn to multi-turn inclusive language expansion via ChatGPT for mental health support. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 615–636, Miami, Florida, USA. Association for Computational Linguistics.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems*, volume 36, pages 53728–53741. Curran Associates, Inc.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381, Florence, Italy. Association for Computational Linguistics.

Milton Rokeach. 1973. The nature of human values. *Fre Pre*.

Daniel Schroter, Daryna Dementieva, and Georg Groh. 2023. Adam-smith at SemEval-2023 task 4: Discovering human values in arguments with ensembles of transformer-based models. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 532–541, Toronto, Canada. Association for Computational Linguistics.

Shalom H Schwartz, Jan Cieciuch, Michele Vecchione, Eldad Davidov, Ronald Fischer, Constanze Beierlein, Alice Ramos, Markku Verkasalo, Jan-Erik Lönnqvist, Kursad Demirutku, et al. 2012. Refining the theory of basic individual values. *Journal of personality and social psychology*, 103(4):663.

John R. Searle. 2003. *Rationality in Action*. MIT Press.

Jamin Shin, Peng Xu, Andrea Madotto, and Pascale Fung. 2020. Generating empathetic responses by looking ahead the user's sentiment. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7989–7993.

Michiel van der Meer, Piek Vossen, Catholijn Jonker, and Pradeep Murukannaiah. 2023. Do differences in values influence disagreements in online discussions? In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15986–16008, Singapore. Association for Computational Linguistics.

Watchful1. Subreddit comments/submissions 2005-06 to 2023-12.

Kelly G Wilson and Amy R Murrell. 2004. Values work in acceptance and commitment therapy. *Mindfulness and acceptance: Expanding the cognitive-behavioral tradition*, pages 120–151.

Jing Yao, Xiaoyuan Yi, Yifan Gong, Xiting Wang, and Xing Xie. 2024. Value FULCRA: Mapping large language models to the multidimensional spectrum of basic human value. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 8762–8785, Mexico City, Mexico. Association for Computational Linguistics.

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

Haiquan Zhao, Lingyu Li, Shisong Chen, Shuqi Kong, Jiaan Wang, Kexin Huang, Tianle Gu, Yixu Wang, Jian Wang, Liang Dandan, Zhixu Li, Yan Teng, Yanghua Xiao, and Yingchun Wang. 2024. ESC-eval: Evaluating emotion support conversations in large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 15785–15810, Miami, Florida, USA. Association for Computational Linguistics.

Zhonghua Zheng, Lizi Liao, Yang Deng, Libo Qin, and Liqiang Nie. 2024. Self-chats from large language models make small emotional support chatbot better. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11325–11345, Bangkok, Thailand. Association for Computational Linguistics.

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, Toronto, Canada. Association for Computational Linguistics.

A Experiment Details for Identifying the Effect of Values in Emotional Support

Since we are interested in cases where values are expressed positively, we only count the values when the seeker's utterances are positive. To determine whether the seeker's utterances are positive, we use the EmoLlama-Chat-7B proposed by Liu et al. (2024). To extract the values in the seeker's utterances, we employ the model that achieved the best performance in the *SemEval2023 Task 4: Identification of Human Values behind Arguments* (Schroter et al., 2023; Kiesel et al., 2023).

B Dataset Settings in Reddit

The final datasets derived from Reddit are categorized into two settings:

	Train	Dev	Test	Total
Single-turn	18,459	2,000	1,000	21,459
Multi-turn	24,339	2,000	1,000	27,339

Table 4: Data distribution of single-turn and multi-turn threads sourced from Reddit.

- **Single-turn setting:** A concise, three-part interaction sequence consisting of an initial post (*OP*), a response (*commenter*), and a final reply (*OP*).
- **Multi-turn setting:** Extended dialogue threads that include additional exchanges beyond the single-turn structure.

An overview of the final dataset is presented in Table 4.

C Strategies for Emotional Support

In this study, we utilized the 8 emotional support strategies defined by Liu et al. (2021). Descriptions of each strategy are as follows:

- **Question:** Ask open-ended or specific questions to help the seeker articulate and clarify the issues they are facing.
- **Restatement:** Rephrase the seeker's statements in a clearer, more concise way to help them better understand their situation.
- **Reflection:** Express and describe the emotions that the seeker is experiencing to validate their feelings.
- **Self-disclosure:** Share similar experiences or emotions to convey empathy and build connection with the seeker.
- **Affirmation:** Highlight the seeker's strengths and abilities while offering encouragement and reassurance.
- **Suggestions:** Offer practical advice or actionable steps to the seeker.
- **Information:** Share useful facts, resources, or data to help the seeker make informed decisions or gain clarity.
- **Others:** Use other support strategies that do not fall into the above categories.

D Training Details and Results

D.1 Target Value Detector

Training Methods. The target value detector predicts the values observable in the next turn of the OP's comment. The model generates a sequence of values, and the top three are selected based on their predicted probabilities when multiple values are identified (*e.g., "Self-direction: action, Benevolence: caring, Security: personal"*).

The target value detector is based on the Llama-3-8B-Instruct, fine-tuned using the LoRA (Hu et al., 2022). During training, the low-rank matrix dimension was set to 8, with an alpha of 16, and a learning rate of 5e-5. The final model was selected based on the highest F1-score achieved on the test dataset. Training was performed on an NVIDIA A100-80GB GPU, with durations of approximately 10 hours. The detailed training prompt is provided in Table 15.

Models	Precision	Recall	F1-score
GPT-4o-mini	0.361	0.384	0.372
+ Reasoning	0.320	0.339	0.329
+ Value information	<u>0.383</u>	<u>0.407</u>	<u>0.395</u>
Llama-3-8B-Instruct	0.323	0.283	0.302
+ Reasoning	0.304	0.283	0.293
+ Value information	0.343	0.271	0.303
Target Value Generator	**0.516**	**0.540**	**0.528**

Table 5: Performance comparison of models in target value prediction.

Results. The results of the training are summarized in Table 5, comparing the performance of the target value detector with baseline models, GPT-4o-mini and Llama-3-8B-Instruct (vanilla). For the baselines, additional experiments were conducted by incorporating reasoning steps before response generation or providing detailed definitions for each value. The target value detector outperformed the baselines across all three metrics, demonstrating impressive performance considering the large set of 20 values.

D.2 Reference Generator

Training Methods. The reference generator is based on Llama-3-8B-Instruct. The reference response model is based on Llama-3-8B-Instruct. Training was conducted on both single-turn and multi-turn settings using the Reddit dataset introduced in Section 4. For each setting, both SFT and DPO approaches were applied with various hyperparameter configurations. The model was trained for up to 5 epochs, and the final model was selected based on its performance on the test dataset. The hyperparameters used for the final model are sum-

Settings	Stage	LR	Rank	Alpha	Dropout
Single-turn	SFT	1e-4	8	8	0.1
	DPO	1e-5	8	16	0.05
Multi-turn	SFT	1e-5	8	16	0.05
	DPO	1e-5	8	16	0.05

Table 6: Hyperparameters used for training the reference generator.

marized in Table 6. Training was performed on an NVIDIA A6000-48GB GPU, with durations of approximately 20 hours for the SFT stage and 10 hours for the DPO stage. Detailed training prompts are provided in Table 16.

Results. The model performances were evaluated using GPT (GPT-4o-mini) through two approaches. First, pairwise comparisons were conducted between "Llama-3-8B-Instruct (vanilla)-reference generator (SFT)" and "reference generator (SFT)-reference generator (DPO)". Specifically, GPT assessed which model's responses more closely aligned with the ground truth responses (i.e., actual comments written by the original commenter) for the test dataset. Second, the impact of target values on the generated responses was examined. Responses generated using the original target values were compared to those generated using randomly assigned values to evaluate variation in content. To reduce sequence-based bias, the order of options within the prompts was alternated during evaluation. The evaluation prompts are detailed in Table 17, and the results of the two experiments are presented in Table 7 and Table 8, respectively.

The results indicate that in single-turn settings, the reference generator performed effectively in both experiments. In the first experiment, the reference generator (SFT) outperformed the baseline, while the reference generator (DPO) demonstrated even greater similarity to ground truth responses. In the second experiment, both SFT and DPO models generated responses more aligned with ground truth when provided with original target values rather than random ones, with the DPO model achieving superior performance. These findings suggest that models trained in single-turn settings effectively integrate target values into their responses, capturing key messages in Reddit comments and reflecting variations in target values.

In contrast, in the multi-turn setting, while the DPO model performed well in the second experiment, it did not surpass the baseline in the first experiment. This may be attributed to the increased complexity of interactions in longer threads, where it becomes challenging to identify how specific comments influence target values. For instance, even if the OP expressed positive values in their final comment, it is unclear which prior interaction contributed to this outcome. The single-turn setting simplifies these relational dynamics, making interactions more explicit. Consequently, the model trained in the single-turn setting was selected as the final reference generator.

	Comparison 1		Comparison 2	
	Llama	**RG (SFT)**	**RG (SFT)**	**RG (DPO)**
Single-turn				
Order 1	480	**520**	449	**551**
Order 2	368	**632**	495	**505**
Multi-turn				
Order 1	**634**	366	**513**	487
Order 2	**549**	451	**506**	494

Table 7: Pairwise comparison results for single-turn and multi-turn settings, evaluating the similarity of the reference generator (RG) responses to ground truth comments.

	RG (SFT)		RG (DPO)	
	Original	**Random**	**Original**	**Random**
Single-turn				
Order 1	**525**	475	**655**	345
Order 2	**554**	446	**687**	313
Multi-turn				
Order 1	**584**	416	**751**	249
Order 2	438	**562**	**755**	245

Table 8: Pairwise comparison results for single-turn and multi-turn settings, evaluating the performance of reference gesponse (RG) under original and random target values.

D.3 Supporter Model

The supporter model's training consists of two stages: SFT and DPO, with training data generated through seeker simulator simulations. The model takes the dialogue history, target values, and reference response as input, with detailed prompts provided in Table 18. Training was performed on an NVIDIA A100-80GB GPU, with durations of approximately 20 hours for the SFT stage and 5 hours for the DPO stage.

Split	Total Dialogues	Total Turns	Dataset	Dataset *(Filtered)*
Train	1,796	16,588	33,176	33,130
Dev	120	1,184	2,374	2,367

Table 9: Overview of the SFT dataset used for training the supporter model. The filtered dataset excludes instances where the generated strategy deviates from the requested strategy.

SFT Stage. During SFT, to mitigate GPT's inherent bias toward utilizing reference responses, the model generates alternative responses by reversing the decision regarding reference response usage. Specifically, the model is prompted to reverse its decision regarding the use of the reference response from the previous answer and to regenerate both Step 3 and Step 4. The overview of the SFT dataset and distribution of selected strategies are presented in Table 9 and Table 10, respectively.

DPO Stage. During DPO training, simulations are conducted to generate preference data. The supporter model generates two responses per turn: one based on its initial reference usage and another taking the opposite approach (alternative response). Each response undergoes independent simulations, and its effectiveness in reinforcing target values is quantified using a reward function (Equation 3). The response with the higher cumulative reward is selected as the chosen response, while the other is designated as rejected.

When GPT is used as the supporter, the same prompts from the SFT stage are applied. For the supporter model (SFT), the model first generates an initial response. Subsequently, by reversing the decision on the use of the reference response from Step 3 (*e.g. Yes* → *No*), an alternative response is generated. An overview of the DPO dataset is provided in Table 19.

D.4 Terms and License

We utilized Llama-3-8B-Instruct as the base model for the target value detector, reference generator, and supporter model. This model is licensed under the Llama 3 Community License Agreement. All artifacts used in this study are confirmed to be accessible for research purposes.

Category	Initial Response	%	Alternative Response	%
Question	393	2.4	152	0.9
Restatement	1,234	7.4	723	4.4
Reflection	1,251	7.6	466	2.8
Self-disclosure	949	5.7	**7,949**	**48.0**
Affirmation	**6,577**	**39.7**	1,014	6.1
Suggestions	6,159	37.2	5,773	34.9
Information	4	0.0	380	2.3
Others	0	0.0	106	0.6

Table 10: Strategy distribution across initial and alternative responses in the supporter model's SFT training dataset.

E Seeker Simulator

E.1 Persona Generation

We develop a diverse set of seeker personas to train the supporter model, enabling it to effectively understand and address various problem scenarios. The creation of these seeker personas involves a 5 step process.

Step 1. Situation Generation We aim to create a diverse set of situations reflecting specific circumstances individuals face, each expressed in a single sentence (*e.g., "I just moved in this week, and it's so hard to make friends"*). To achieve this, we first define 6 primary problem categories and 27 subcategories based on prior research related to emotional support datasets and seeker simulator implementation (Liu et al., 2021; Zheng et al., 2024; Lee et al., 2024; Zhao et al., 2024), as detailed in Table 24.

To ensure the situations also reflect diverse human values, we integrate information about 20 distinct values. For each combination of the six problem categories and 20 values, we generate 10 to 30 unique situations using GPT-4o. This process result in a total of 2,940 unique situations. The prompts used for this process are detailed in Table 25.

Step 2. Evaluation on Value-Alignment We evaluate the alignment of the generated situations with the provided values using GPT-4o, employing a 5-point scale. Situations scoring 3 or belowa are excluded from further consideration, resulting in the retention of 2,036 situations. The evaluation prompt used for this process is detailed in Table 26.

Step 3. Emotion labeling Emotion labeling is conducted for the previously generated situations using 10 negative emotions (*Frustration, Anxiety, Sadness, Fear, Guilt, Shame, Anger, Depression,* *Jealousy, Disgust*) identified from prior research (Liu et al., 2021; Rashkin et al., 2019). Each situation is labeled five times using GPT-4o-mini, and the final classification is determined by majority vote.

Step 4. Create Demographic Information To ensure consistency in responses generated by the seeker simulator and to enable the supporter model to interact with seekers with diverse characteristics, we generate demographic profiles including age, gender, and occupation for each simulated situation.

Our persona generation process resulted in 2,036 unique personas, each defined by problem category, human values, situations, emotion types, and demographic information. These personas are divided into three datasets: a training set containing 1,796 personas, and development and test sets with 120 personas each. The training and development sets are used to construct SFT and DPO datasets for the supporter model through simulation, while the test set is reserved for comparative performance evaluation across models. Examples of the generated personas are provided in Table 27.

E.2 Evaluation of Seeker Simulator Performance

Comparison Models. Developing a supporter model capable of effectively assisting in real conversations with human seekers requires a seeker simulator that exhibits human-like behavior. To identify the most suitable model for this purpose, we conduct experiments on a range of candidates. The evaluated models are as follows:

- **Prompt-based models:** GPT-4o-mini, *Llama-3-8B-Instruct*
- **Fine-tuned models:** Llama-ESConv, Llama-ExTES
- **Pre-existing seeker simulator:** ESC-Role (Zhao et al., 2024)

Llama-ESConv and Llama-ExTES are fine-tuned versions of Llama-3-8B-Instruct. These models are trained on seeker turns from the ESConv dataset (Liu et al., 2021) and the ExTES dataset (Zheng et al., 2024), respectively.

Evaluation Approach. We evaluate the models on the ESConv test dataset by providing dialogue context up to each seeker turn and generating the subsequent utterance. The evaluation compare generated responses to actual seeker utterances across four dimensions: length, content, emotions, and values.

For length, we calculate the correlation between the lengths of the generated and actual utterances. Content evaluation employs BERT-Score [4] and GPT-4o-mini to assess semantic similarity between generated and reference responses. Emotional analysis uses EmoLlama-Chat-7B (Liu et al., 2024) to determine sentiment polarity for each turn, measuring the correlation between generated and actual sentiment levels. To assess value alignment, we employ the model proposed by Schroter et al. (2023) to generate probability distributions across 20 values. We then calculate cosine similarity and Euclidean distance between the generated and actual distributions, reporting the mean values across all turns.

Results. The experimental results are summarized in Table 29. In the ESConv test dataset, the average length of seeker utterances is 19.5, with GPT-4o-mini and Llama-ExTES exhibiting similar utterance lengths. While individual evaluation metrics show some variation, GPT-4o-mini with one-shot dialogue examples demonstrates strong overall performance. Therefore, GPT-4o-mini (one-shot) is selected as the final seeker simulator.

E.3 Prompts for Seeker Simulator

The seeker simulator generates subsequent seeker responses by integrating persona details and dialogue context. Each simulation starts with the predefined situation in the persona as the initial seeker response. A detailed prompt for the seeker simulator is presented in Table 28.

F Evaluation Metrics

F.1 ES-Skills

The definitions of the evaluation criteria for ES-Skills are as follows:

Emotional Support Skills

- **Identification:** How effectively does the therapist explore the patient's situation to identify underlying issues?
- **Comforting:** How well does the therapist demonstrate appropriate emotional responses, such as warmth, empathy, and compassion?

[4] https://huggingface.co/sentence-transformers/all-mpnet-base-v2

- **Suggestions:** How useful and relevant are the therapist's suggestions for addressing the patient's problems?
- **Experience:** How well does the therapist draw on their own relevant experiences to connect with the user's situation?
- **Informativeness:** How specific and informative are the therapist's responses in addressing the patient's situation?

General Conversation Skills

- **Consistency:** How logically structured and contextually appropriate are the therapist's responses?
- **Role-adherence:** How consistently does the therapist adhere to their role, maintaining a non-contradictory and reliable approach?
- **Expression:** How diverse are the therapist's conversational expressions, including the variety and creativity in language and content used?
- **Humanness:** How human-like and natural do the therapist's responses sound?

Overall

- **Overall:** How well does the therapist provide overall emotional support to the patient?

F.2 ES-Intensity

This model predicts the seeker's emotional intensity after a conversation on a 5-point scale, where a lower score indicates a significant reduction in negative emotions. We applied zero-shot/few-shot prompting and fine-tuning to four different models and compared their performance using the ESConv test dataset. The final model is GPT-4o-mini (zero-shot), as it showed the highest correlation with the ground truth final emotional intensity. The results and evaluation prompts are presented in Table 11 and Table 30.

F.3 ES-Value

To evaluate the effectiveness of value reinforcement, it is essential to consider two perspectives: the seeker's and the supporter's. These viewpoints provide a comprehensive understanding of how effectively positive values are identified, discussed, and integrated into the seeker's mindset during emotional support conversations. The definitions for each perspective are as follows:

Model	Method	Acc.↑	F1↑	MSE↓	Corr.↑
Baseline	-	0.435	0.264	**0.768**	-
GPT-4o	Zero-shot	0.358	0.352	1.182	0.303
	Few-shot	0.415	0.416	1.057	0.312
GPT-4o-mini	Zero-shot	**0.466**	**0.432**	0.875	**0.345**
	Few-shot	0.415	0.410	0.966	0.327
Llama3-8B	Zero-shot	0.426	0.318	0.869	0.130
	Fine-tuned	0.409	0.395	0.892	0.330
EmoLlama-7B	Zero-shot	0.384	0.289	0.972	0.084
	Fine-tuned	0.407	0.373	0.977	0.185

Table 11: Evaluation results of different models on final emotional intensity prediction tasks. The metrics are accuracy, weighted F1-score, mean squared error, and Spearman's correlation coefficient. The baseline model predicts all final emotional intensities as 2.

- **Seeker's perspective:** How strongly were positive human values explored and reinforced in the patient through the conversation?
- **Supporter's perspective:** How effectively did the therapist help the patient in exploring and reinforcing positive human values?

ES-Value is assessed through pairwise comparisons between a reference model and multiple baseline models. Dialogues from the reference model are paired with corresponding dialogues from baseline models, ensuring the seeker personas are identical. Each pair is evaluated 10 times, and the reference model's win ratio is normalized to a score ranging from 0 to 1. This evaluation utilize GPT-4o-mini as the assessment model (see Table 31 for prompt details).

G Training Details for Emotion-Reinforced Models

This study investigates whether reinforcing values, rather than positive emotions, leads to more effective emotional support. To test this hypothesis, we adapt our methods by modifying the learning objective to prioritize promoting positive emotions in seekers. This approach requires two key components: a reference generator and a supporter model, both optimized for emotional reinforcement. Unlike the value-based method, this approach does not require a target value detector. The following subsections outline the training procedures for the reference generator and the supporter model.

G.1 Reference Generator

The reference generator is trained on supporter response from Reddit that successfully elicited posi-

Stage	Supporter	Train	Dev
SFT	GPT-4o-mini	24,580	1,656
DPO	GPT-4o-mini	2,610	552

Table 12: Dataset sizes for training the supporter model for positive emotion reinforcement generated through simulation. The 'Supporter' column refers to the supporter model used in the simulation.

tive emotional responses from OPs. This training approach ensures that the generated responses effectively foster positive emotions. Given the dialogue history $(o_1, c_1, o_2, c_2, ..., o_t)$, the model generates a supporter response (c_t) as follows:

$$c_t = \text{LLM}_{\text{RG}}(o_1, c_1, o_2, c_2, ..., o_t) \tag{4}$$

Unlike our model, which incorporates both dialogue history and target values, this generator relies solely on dialogue history as input. Therefore, it employs only the SFT stage. Although the training data and prompts are consistent with those used for our model, all value-related information has been excluded from the reference generator's training process.

G.2 Supporter Model

The supporter model for emotion reinforcement processes two inputs: the dialogue history and a reference response generated by the reference generator. At each turn, the model performs reasoning across four key aspects: (1) identifying the seeker's issues and current emotional state, (2) analyzing the content of the reference response, (3) deciding whether to integrate the reference response, and (4) selecting an optimal emotional support strategy to generate the subsequent response. These reasoning aspects are identical to those used in our model.

The training process for the supporter model involves both SFT and DPO using data generated through simulations with a seeker simulator based on GPT-4o-mini.

SFT Stage. Similar to our approach, a dual-generation method is employed: GPT produces two responses per turn—one with references and one without—ensuring balanced training data within identical contexts. The simulation-generated data is then used to fine-tune Llama-3-8B-Instruct, with dataset sizes detailed in Table 12.

DPO Stage. This stage optimizes the supporter model to generate responses that more effectively promote positive emotions. The process uses simulated dialogues between the supporter model and a seeker simulator to generate training data. For each turn, the supporter model produces two responses: one following its initial reference usage and another taking the opposite approach. Both responses undergo simulation to evaluate their emotional impact, with the more effective response marked as preferred. The cumulative reward for a supporter's response at turn t (u_t^{sup}) is calculated as:

$$R(u_t^{\text{sup}}) = \sum_{k=1}^{h} \gamma^{k-1} S_{t+k}(u_t^{\text{sup}}) \tag{5}$$

where $S_t(u_t^{\text{sup}})$ represents the emotion score at turn t calculated by GPT-4o-mini, h is the look-ahead horizon (the number of future steps considered), and γ is a discount factor balancing immediate and future rewards. Response pairs are included in the DPO dataset when their reward difference exceeds the threshold T_{diff}.

Emotion scores are calculated using prompts inspired by Deng et al. (2024). For each turn, GPT evaluates the seeker's emotional state as "feels worse", "feels the same", "feels better", or "the issue has been solved". These responses are then mapped to scores of -1.0, -0.5, 0.5, and 1.0, respectively. Ten responses are collected for each turn, and the average score is used as the final emotion score. The prompts used for this process are provided in Table 32.

These simulations use GPT-4o-mini, and dataset sizes are summarized in Table 12.

Category	Metric	Corr.
ES-Skills	Identification	0.422*
	Comforting	0.322*
	Suggestions	0.421*
	Experience	0.778*
	Informativeness	0.282*
	Consistency	0.351*
	Role-Adherence	0.235†
	Expression	0.198
	Humanness	0.202
	Overall	0.413*
ES-Value	Seeker	0.332*
	Supporter	0.413*

Table 13: Spearman's rank correlation between expert and GPT-generated scores. Significant correlations are marked with * (p-value < 0.05) and † (p-value < 0.1).

H Validation of Automated Evaluation Models

To evaluate the performance of GPT-4o-mini in assessing ES-Skills and ES-Value, we analyzed the correlation between expert evaluation scores and GPT-generated scores. For this purpose, we randomly select and evaluate 60 dialogues generated by our models and baselines. For ES-Value, we compared individual dialogue scores generated by GPT and expert evaluations, rather than using a pairwise scoring approach, to enable a more straightforward comparison. As shown in Table 13, significant correlations were observed across most metrics, except for *Expression* and *Humanness*. These findings suggest that automated evaluation models can reliably approximate human assessments of emotional support, conversational quality, and value reinforcement, supporting the validity of our experimental results.

I Details of Experiments on Self-Disclosure

Self-disclosure—sharing experiences related to those of the seeker—is a key strategy in emotional support for fostering intimacy and reducing stress (Cheng et al., 2024; Meng and Dai, 2021). The ability to share such experiences has been used as an evaluation criterion for emotional support systems (Zhang et al., 2023). However, some users might feel uncomfortable when dialogue systems present these experiences as personal.

To better understand the impact of self-disclosure on emotional support systems, we investigated two alternative approaches: (1) removing it entirely and substituting the next most probable strategy in the supporter's reasoning process (Section 5.3); and (2) rephrasing self-disclosure responses to frame them as experiences of others, using GPT-4o-mini.

As shown in Table 22 and Table 23, the models consistently exhibited declines in overall ES-Skills and ES-Value when self-disclosure was modified or removed. The metric most affected was *Experience*, with related metrics such as *Suggestions* and *Informativeness* also showing performance drops.

The results reinforces the importance of self-disclosure in emotional support but, at the same time, highlight the need for research on more sophisticated methods for experience sharing. For example, analyzing different types of self-disclosure and developing alternative strategies based on seeker perceptions could offer meaningful improvements. This detailed investigation will be left for future work.

Category	Freq.
Strengths and Achievements Acknowledgment	30
Exploration of Issues and Challenges	23
Self-Compassion and Acceptance	19
Exploration of Personal Interests	16
Emotional Resilience and Coping Strategies	14
Exploration of Goals and Motivations	14
Motivation and Alignment with Goals	1

Table 14: Frequency of areas for improvement in value reinforcement.

J Detailed Analysis of Value Reinforcement Performance

To identify areas for improvement in our DPO model's value reinforcement, we conducted a detailed analysis. First, we evaluated value reinforcement scores from both seeker and supporter perspectives using GPT-4o-mini on a 5-point scale. Next, we analyzed 40 dialogues that received the lowest score of 4 to identify potential improvement areas. This analysis involved reasoning with GPT-4o-mini and categorizing the areas, as summarized in Table 35. Subsequently, GPT-4o-mini was used again to assign up to three relevant categories to each of the 40 dialogues. The frequency of issues for each category is detailed in Table 14.

System	Select and return up to 3 values to reinforce in the patient for effective emotional support.
User	Human values: *{List of 20 human values}* The dialogue history below is a conversation between a patient experiencing emotional difficulties and a therapist providing support. For effective emotional support, which values should be reinforced in the patient so that they are expressed more frequently in the future? Select up to 3 values from the list provided above. Answer in the format 'value1, value2, value3' separated by commas without any additional explanation. Dialogue history: *{Dialogue history}*

Table 15: Training prompts for the target value detector.

System	You will take on the role of a therapist to help a patient with emotional difficulties, aiming to reduce their distress and support them in overcoming their challenges.
User	1. Dialogue history: *{Thread history}* 2. Target values: *{Information on the target values}* As a therapist supporting a patient with emotional difficulties, your goal is to reduce their distress and guide them through challenges. The target values are those that are expected to be more frequently expressed by the patient. Generate the next turn of the utterance based on the dialogue history, aiming to reinforce these target values in the patient.

Table 16: Training prompts for the reference generator. The target values information includes the definition of each value along with the set of contained values.

System	Determine which of the two comments generated by each model is more similar to the ground truth comment.
User	Thread: *{Thread history}* Ground truth comment: *{GT comment}* The above includes the thread history and the corresponding ground truth comment, which continues the thread. Below are two comments generated by different models. First, provide reasoning for your evaluation, and then select the comment that is more similar in content to the ground truth comment. Comment A: *{Response from model A}* Comment B: *{Response from model B}* [Template] Reasoning: Similar comment: Answer with either 'Comment A' or 'Comment B' only

Table 17: Prompts for evaluating the performance of the reference response.

System	You will take on the role of a therapist to help a patient with emotional difficulties, aiming to reduce their distress and support them in overcoming their challenges.
User	1. Strategies for emotional support: *{Definition of 8 emotional support strategies}* 2. Dialogue history: *{Dialogue history}* 3. Target values: *{Information on the target values}* 4. Reference response: *{Reference response}* As a therapist supporting a patient with emotional difficulties, your goal is to reduce their distress and guide them through challenges. The target values are those that are expected to be more frequently expressed by the patient. You need to generate the therapist's next utterance based on the dialogue history, aiming to reinforce these target values in the patient. The therapist's next utterance should follow these guidelines: - Use only one sentence without any extra explanation, framing, introductory phrases, or meta-commentary - Avoid directly mentioning the target values, but focus on reinforcing them through your guidance. - If the patient shows signs of improvement in the dialogue history, acknowledge their progress and guide the conversation to an efficient close. - Do not repeat similar messages from previous therapist utterances in the dialogue history. The reference response is a therapist's reply given to another patient in a similar situation, which you can use as a reference for generating your next response. Before generating the therapist's response to satisfy the above conditions, thoroughly analyze the following: Step 1. Understanding the patient's issues and current state - What is the patient's issue? - Have their situation and the causes of their emotions been sufficiently explored? If not, what additional information should be obtained to deeply understand them? - What is the patient's current emotional state? How have the patient's emotions or thoughts changed through the conversation? Step 2. Identifying the key points of the reference response - What is the main message in the referenced response (item 4)? Step 3. Determination of reference response usage - Would using a reference response be helpful for generating the next therapist utterance? Why or why not? - If a reference response is used, how would it be applied, and if it is not used, what alternative message would be provided? Step 4. Therapist's next strategy and response - Based on the above (Step 1-Step3), what emotional support strategy should be used, and what message should you convey to the patient in the next response? You should respond in the following template format: Step 1. Understanding the patient's issues and current state -Reasoning: (the result of your analysis) Step 2. Identifying the key points of the reference response -Reasoning: (the result of your analysis) Step 3. Determination of reference response usage -Reasoning: (The result of your analysis, starting with whether to use the reference response — 'Yes' or 'No') Step 4. Therapist's next strategy and response -Strategy: (choose one emotional support strategy for the next turn based on the reasoning) -Response: (only the therapist's next utterance without any explanation)

Table 18: Training prompts for the supporter model.

Supporter Model	h	γ	T_{diff}	Train			Dev		
				# of Data	Chosen		# of Data	Chosen	
					Initial	Alternative		Initial	Alternative
GPT	All	1	1	2,561	1,168	855	458	208	149
			2	2,023	881	623	357	268	149
	All	0.9	1	1,712	947	765	298	164	134
			1.5	1,345	735	610	229	117	112
	3	1	1	1,796	1,127	669	319	201	118
			2	1,144	724	420	210	132	78
	5	1	1	2,015	1,301	714	360	237	123
			2	1,438	955	483	247	165	82
SFT	All	1	1	3,301	1,255	1,106	628	263	206
			2	2,361	920	800	469	182	149
	All	0.9	1	1,975	979	996	407	209	198
			1.5	1,556	777	779	318	153	165
	3	1	1	1,825	1,122	703	375	226	149
			2	1,117	663	454	239	153	116
	5	1	1	2,186	1,360	826	456	281	175
			2	1,489	919	570	172	116	56

Table 19: Overview of the DPO dataset categorized by the supporter model used for simulations and variations in reward calculation hyperparameters (h: look-ahead horizon, γ: discount factor, T_{diff}: difference threshold). The "Chosen" column represents the distribution of chosen responses selected between the model's initial and alternative outputs.

Categories	Models	h	γ	T_{diff}	ES-Skills↑										ES-Intensity↓
					Iden.	Comf.	Sugg.	Expe.	Info.	Cons.	Role.	Expr.	Huma.	Over.	
Prompt-based	GPT	-	-	-	4.83	4.92	4.57*	3.11*	4.42*	**5.00**	**5.00**	4.10*	4.70	4.72*	1.89
	Llama	-	-	-	4.87	4.91	4.43*	2.91*	4.47	4.99	**5.00**	4.03*	4.63*	4.68†	1.99†
Fine-tuned	Llama-ESConv	-	-	-	4.35*	4.43*	4.06*	2.65*	3.88*	4.82*	4.97*	3.79*	4.25*	4.22*	1.87
	Llama-ExTES	-	-	-	4.83	4.90	4.53*	2.71*	4.44*	4.99	**5.00**	4.02*	4.59*	4.66*	1.67*
	Llama-Psych8k	-	-	-	4.84	4.85*	4.75	2.89*	4.63	4.99	**5.00**	4.05*	4.57*	4.75	**1.53***
Emotion Reinforced	SFT	-	-	-	4.83	4.91	4.51	3.64	4.43	4.97	4.99	4.16	4.67	4.73	1.97
	DPO (GPT)	3	1	0.5	4.85	4.92	4.74	**4.05**	4.61	4.99	**5.00**	**4.33**	**4.78**	4.82	1.86
		5	1	0.5	4.85	**4.95**	4.68	3.80	4.58	4.99	**5.00**	4.28	4.77	4.81	1.82
Ours	SFT	-	-	-	4.85	4.90	4.72	3.76	4.56	4.99	**5.00**	4.25	4.73	4.80	1.86
	DPO (GPT)	All	1	1	4.89	4.92	4.75	3.71	4.63	4.99	**5.00**	4.24	4.72	4.80	1.90
		All	1	2	4.85	4.91	4.73	3.89	4.63†	**5.00**	**5.00**	4.27	4.76	4.80	1.90
		All	0.9	1	4.84	4.87	4.72	3.72	4.61	4.98	**5.00**	4.23	4.72	4.78	1.85
		All	0.9	1.5	4.89	4.93	4.71	3.79	4.59	4.99	**5.00**	4.28	4.76	4.83*	1.88
		3	1	1	4.87	4.90	4.77	3.58	4.63	4.99	**5.00**	4.26	4.74	4.79	1.87
		3	1	2	**4.91**†	4.93	4.78*	3.61	4.65†	4.99	5.00	4.23	4.72	4.85*	1.94
		5	1	1	4.88†	4.93	4.77*	3.62	4.64†	**5.00**	**5.00**	4.25	4.73	4.83	1.84
		5	1	2	4.88	4.94	4.73	3.64	4.66*	**5.00**	**5.00**	4.24	4.75	4.84*	1.83
	DPO (SFT)	All	1	1	4.83	4.89	4.71	3.75	4.60	4.98	4.99†	4.21	4.77	4.78	1.93
		All	1	2	4.89	**4.95**	4.77	3.76	4.66*	4.99	**5.00**	4.26	4.75	4.83	1.87
		All	0.9	1	4.86	4.93†	4.72	3.72	4.64†	4.99	**5.00**	4.26	4.73	4.82†	1.87
		All	0.9	1.5	4.85	4.92	4.73	3.79	4.59	4.99	**5.00**	4.28	4.74	4.81	1.86
		3	1	1	4.86	4.91	4.71	3.58	4.57	4.99	**5.00**	4.24	4.71	4.78	1.89
		3	1	2	4.90†	**4.95**	**4.80***	3.85	**4.69***	**5.00**	**5.00**	4.30	4.77	**4.87***	1.75
		5	1	1	4.85	4.90	4.69	3.76	4.56	4.99	**5.00**	4.23	4.74	4.78	1.86
		5	1	2	4.83	4.91	4.70	3.66	4.61	4.98	**5.00**	4.23	4.74	4.79	1.91

Table 20: Comparison of models based on ES-Skills and ES-Intensity. Statistically significant differences compared to our SFT model are marked with * (p-value < 0.05), and differences with p-value < 0.1 are marked with †, as determined by the Mann-Whitney U test.

Categories	Models	h	γ	T_{diff}	SFT		DPO (GPT)		DPO (SFT)	
					Seeker	**Supporter**	**Seeker**	**Supporter**	**Seeker**	**Supporter**
Prompt-based	GPT	-	-	-	0.51	0.48	$0.49^{\dagger}$	0.42^{*}	0.49^{*}	0.42^{*}
	Llama	-	-	-	0.50	0.51	0.48	0.44^{*}	$0.46^{\dagger}$	0.45
Fine-tuned	Llama-ESConv	-	-	-	0.37^{*}	0.21^{*}	0.35^{*}	0.17^{*}	0.37^{*}	0.19^{*}
	Llama-ExTES	-	-	-	0.51	0.54^{*}	0.50	0.52	$0.48^{\dagger}$	0.51
	Llama-Psych8k	-	-	-	0.50	0.64^{*}	0.48^{*}	0.58^{*}	0.49	0.62^{*}
Emotion-Reinforced	SFT	-	-	-	0.50	0.51	0.50	0.43^{*}	0.49	$0.46^{\dagger}$
	DPO (GPT)	3	1	0.5	0.52	0.55^{*}	0.49	0.49	0.49	0.51
		5	1	0.5	0.53^{*}	0.59^{*}	0.49^{*}	0.44^{*}	0.49	0.48
Ours	SFT	-	-	-	-	-	$0.48^{\dagger}$	0.44^{*}	$0.48^{\dagger}$	$0.46^{\dagger}$

Table 21: ES-Value performance for models evaluated from the perspectives of seekers and supporters. The scores represent the mean win-ratio of baselines compared to our models–SFT, DPO. Statistically significant differences compared to our models are marked with * (p-value < 0.05), and differences with p-value < 0.1 are marked with †, as determined by the Mann-Whitney U test.

Models	ES-Skills↑										ES-Intensity↓
	Iden.	**Comf.**	**Sugg.**	**Expe.**	**Info.**	**Cons.**	**Role.**	**Expr.**	**Huma.**	**Over.**	
SFT	**4.85**	**4.90**	**4.72**	**3.76**	**4.56**	**4.99**	**5.00**	**4.25**	**4.73**	**4.80**	**1.86**
- Next strategy	4.81	4.89	4.57^{*}	2.79^{*}	4.43^{*}	**4.99**	**5.00**	4.05^{*}	4.55^{*}	4.68^{*}	1.89
- Others' experience	4.76	4.88	4.57^{*}	3.36^{*}	4.43^{*}	4.98	**5.00**	4.04^{*}	4.63^{*}	4.70^{*}	1.89
DPO (GPT)	**4.91**	**4.93**	**4.78**	**3.61**	**4.65**	4.99	**5.00**	**4.23**	**4.72**	**4.85**	1.94
- Next strategy	4.88	4.92	4.74	2.87^{*}	$4.57^{\dagger}$	**5.00**	**5.00**	4.08^{*}	4.59^{*}	4.77^{*}	1.90
- Others' experience	4.85	4.91	$4.70^{\dagger}$	3.28^{*}	4.62	4.99	**5.00**	4.11^{*}	$4.67^{\dagger}$	4.80	**1.82**
DPO (SFT)	**4.90**	**4.95**	**4.80**	**3.85**	**4.69**	**5.00**	**5.00**	**4.30**	**4.77**	**4.87**	**1.75**
- Next strategy	4.81^{*}	4.91	4.60^{*}	2.90^{*}	4.43^{*}	4.99	**5.00**	4.06^{*}	4.55^{*}	4.70^{*}	1.92^{*}
- Others' experience	4.80^{*}	4.91	4.56^{*}	3.39^{*}	4.47^{*}	4.99	**5.00**	4.07^{*}	4.65^{*}	4.70^{*}	1.98^{*}

Table 22: Comparison of ES-Skills and ES-Intensity performance based on the inclusion of self-disclosure. Statistically significant differences compared to our models are marked with * (p-value < 0.05), and differences with p-value < 0.1 are marked with †, as determined by the Mann-Whitney U test.

Models	SFT		DPO (GPT)		DPO (SFT)	
	Seeker	**Supporter**	**Seeker**	**Supporter**	**Seeker**	**Supporter**
- Next strategy	0.50	0.52	0.50	$0.53^{\dagger}$	0.48^{*}	0.45^{*}
- Others' experience	0.50	0.50	0.50	0.50	$0.48^{\dagger}$	0.47

Table 23: ES-Value performance for models evaluated from the perspectives of seekers and supporters, considering the impact of self-disclosure on performance. The scores represent the mean win-ratio of baselines compared to our models. Statistically significant differences compared to our models are marked with * (p-value < 0.05), and differences with p-value < 0.1 are marked with †, as determined by the Mann-Whitney U test.

Category	Subcategory
Romantic Relationship Challenges	Breakups or divorce Starting a romantic relationship Challenges in establishing a marriage Communication difficulties in relationships
Family Dynamics and Conflicts	Financial issues within the family Sibling rivalry or family disputes Challenges in parenthood and parenting Coping with loss or grief of a family member
Friendship and Interpersonal Challenges	Difficulty adapting to new social environments Challenges in maintaining friendships Conflicts with friends
Career and Work-Related Challenges	Work-related stress and burnout Job loss or career setbacks Adjusting to a new job or role Concerns about salary and bonuses Dissatisfaction with current job Stress related to unemployment Ongoing depression
Academic and Educational Stress	Dissatisfaction with current school or major Concerns about academic performance Stress related to studies Difficulty entering higher education Lack or excess of motivation to study
Self-Esteem, Identity, and Personal Growth	Issues with self-esteem and confidence Searching for meaning and purpose in life Cultural identity and sense of belonging Concerns about body image

Table 24: Overview of seekers' problem categories and subcategories.

System	Generate appropriate situations that require emotional support, using the given topic and value information.
User	1. Emotional support topic: *{Problem category}* - *{Subcategory 1}* - *{Subcategory 2}* - *{Subcategory 3}* 2. Supported value: *{Human value}* - Definition: *{Definition of the human value}* - Contained values: *{Contained value 1}*, *{Contained value 2}*, *{Contained value 3}* Define specific situations that individuals who prioritize the given human value (item 2) might face related to the presented emotional support topic (item 1). Generate a minimum of 10 and a maximum of 30 diverse and non-overlapping situations. Write from the perspective of an individual in need of emotional support, including 'I' as the subject, and be as specific as possible. Each situation should be one sentence (e.g., I just moved in this week, and it's so hard to make friends.) Do not provide any additional explanations and separate each situation with a newline character ('\n').

Table 25: Prompts for generating seeker situations based on problem category and human value combinations.

System	Evaluate how much each situation aligns with the given value.
User	1. Situations: *{Generated situations}* 2. Supported value: *{Human value}* - Definition: *{Definition of the human value}* - Contained values: *{Contained value 1}*, *{Contained value 2}*, *{Contained value 3}* Rate the alignment of each situation with the given value on a scale of 1-5, using the criteria below to guide your assessment: - 1: The situation does not reflect any connection to the given value. The individual's concerns or actions are entirely unrelated to the principles of this value. - 2: The situation has a minimal or indirect connection to the value. It suggests the presence of the value but lacks a clear emphasis or relevance. - 3: The situation shows some aspects of the value but not as a central theme. The value is present, but other priorities seem equally important. - 4: The situation directly relates to the principles of the value, showing clear prioritization. The value significantly shapes the individual's thoughts or actions. - 5: The situation is driven almost entirely by the given value. The value is a central, explicit factor in shaping the individual's perspective and decisions. For each situation, provide a brief reasoning for your rating based on these criteria, and then assign the numerical rating. Provide your response in the following format: situation: (Rewrite each situation) - Reasoning: (Your explanation here) - Rating: (1-5)

Table 26: Evaluation prompts for assessing alignment between situations and provided values.

Category	Details
Problem	Romantic Relationship Challenges
Value	Self-direction: thought
Emotion	Frustration
Situation	I feel like my creativity isn't appreciated in my marriage, and it's making me question my choices.
Demographics	Age: 30s / Gender: Female / Occupation: Designer
Problem	Friendship and Interpersonal Challenges
Value	Stimulation
Emotion	Frustration
Situation	I often change hobbies and interests, but I've noticed this makes it difficult to maintain deep connections with my friends.
Demographics	Age: 20s / Gender: Male / Occupation: College Student
Problem	Academic and Educational Stress
Value	Achievement
Emotion	Fear
Situation	I feel torn because although I want to succeed, the fear of failure is paralyzing my ability to take risks in my studies.
Demographics	Age: 20s / Gender: Female / Occupation: College Student
Problem	Career and Work-Related Challenges
Value	Security: personal
Emotion	Anxiety
Situation	I've been unemployed for months now, and the financial strain is causing me significant stress and anxiety about maintaining a comfortable lifestyle.
Demographics	Age: 30s / Gender: Male / Occupation: Retail Manager
Problem	Family Dynamics and Conflicts
Value	Self-direction: action
Emotion	Anger
Situation	I just set a boundary to maintain a separation between personal and financial issues, but family members keep crossing it.
Demographics	Age: 30s / Gender: Male / Occupation: Software Developer
Problem	Self-Esteem, Identity, and Personal Growth
Value	Face
Emotion	Fear
Situation	I have maintained an image of success, but I'm scared of failing and letting people see my vulnerabilities.
Demographics	Age: 40s / Gender: Female / Occupation: Entrepreneur

Table 27: Examples of generated personas.

System	In the following conversations, you will play the role of a patient seeking help from a therapist due to emotional difficulties. Your emotional distress stems from *{Problem category}* and the emotion you're feeling is *{Emotion type}*. Your detailed personal information is as follows: Age Range: *{Age range}* Gender: *{Gender}* Occupation: *{Occupation}* Here is an example of a conversation you can refer to: *{Example of a conversation}* When responding, use only one sentence each time. Incorporate your personal information (age range, gender, and occupation) when it seems relevant, but it is not required in every response. If you feel that you have received enough emotional support and your mood has improved, end the conversation by expressing gratitude. Then, if you think it's appropriate to conclude the session, generate '[END]' to signify the end of the conversation. You should generate only '[END]' without saying anything else. Do not end the conversation if you still feel upset or unsettled.
User	Hello, I'm here to listen. What would you like to talk about today?
Assistant	*{Situation}*

Table 28: Prompts for the seeker simulator.

Model	Details	Length		Contents		Emotions		Values	
		Avg	Corr↑	BERT↑	GPT↑	V-oc↑	V-reg↑	Cosine↑	E-dist↓
GPT-4o-mini	Zero-shot	18.0	0.377	0.340	3.864	0.339	<u>0.464</u>	0.714	**0.875**
	One-shot	17.9	0.371	<u>0.342</u>	**4.297**	**0.371**	**0.477**	0.713	<u>0.876</u>
Llama-3-8B-Instruct	Zero-shot	28.6	0.364	0.338	3.855	0.311	0.413	<u>0.813</u>	1.855
	One-shot	25.4	0.324	**0.346**	<u>3.889</u>	<u>0.348</u>	0.433	**0.823**	1.806
Llama-ESConv	-	12.5	<u>0.390</u>	0.321	2.871	0.327	0.438	<u>0.813</u>	1.792
Llama-ExTES	-	20.5	**0.439**	0.331	2.837	0.306	0.429	0.809	1.771
ESC-Role	-	13.6	0.376	0.322	3.355	0.312	0.438	0.809	1.749

Table 29: Performance comparison of models as seeker simulators on the ESConv test dataset. The zero-shot condition indicates no example dialogue in the prompt, while the one-shot condition includes one example dialogue. Emotion scores reflect the correlation with actual seeker responses. V-oc represents sentiment as seven ordinal levels, ranging from -3 (very negative) to 3 (very positive), whereas V-reg quantifies sentiment intensity on a continuous scale from 0 (most negative) to 1 (most positive). Correlation analysis was conducted using Spearman's rank correlation coefficient.

System	Given a conversation between a supporter and a seeker, please assess the final emotions of the seeker after the conversation.
User	Conversation: *{Dialogue}* You can only reply with one of the following sentences: - very low amount of negative emotions can be inferred - low amount of negative emotions can be inferred - moderate amount of negative emotions can be inferred - high amount of negative emotions can be inferred - extreme amount of negative emotions can be inferred Question: What are the final emotions of the seeker after the conversation?

Table 30: Prompts for evaluating the seeker's final emotional intensity.

System	Evaluate the emotional support conversation based on the given criteria.
User	1. Human Values: *{Definition of human values}* 2. Dialogue A: *{Dialogue A}* 3. Dialogue B: *{Dialogue B}* The definitions and associated values of 20 human values are provided, along with two emotional support conversations (Dialogue A and Dialogue B) between a therapist and a patient struggling with emotional difficulties. After reading both conversations, evaluate them based on the criteria below. For each criterion, first provide reasoning for your assessment, and then determine which dialogue is better for that criterion, or if it's a tie. Evaluation Criteria: 1) Patient's perspective: In which dialogue were positive human values more strongly explored and reinforced in the patient through the conversation? 2) Therapist's perspective: In which dialogue did the therapist more effectively help the patient in exploring and reinforcing positive human values? Template: 1. Reasoning: (Reasoning for the evaluation of all criteria) 2. Results: 1) Patient's perspective: Dialogue A, Dialogue B, or Tie 2) Therapist's perspective: Dialogue A, Dialogue B or Tie

Table 31: Prompts for evaluating the effectiveness of value reinforcement (ES-Value).

System	Given a conversation between a Therapist and a Patient, please assess whether the Patient' emotional issue has been solved after the conversation.
User	You can only reply with one of the following sentences: No, the Patient feels worse. No, the Patient feels the same. No, but the Patient feels better. Yes, the Patient's issue has been solved. The following is a conversation about *{Emotion type}* regarding *{Problem category}* : *{Dialogue}* Question: Has the Patient's issue been solved? Answer:

Table 32: Prompts for scoring prompts for calculating emotion scores for emotion reinforcement.

Value Category	Definition & Contained Values
Self-direction: thought	• Definition: It is good to have own ideas and interests. • Contained values: Be creative, Be curious, Have freedom of thought
Self-direction: action	• Definition: It is good to determine one's own actions. • Contained values: Be choosing own goals, Be independent, Have freedom of action, Have privacy
Stimulation	• Definition: It is good to experience excitement, novelty, and change. • Contained values: Have an exciting life, Have a varied life, Be daring
Hedonism	• Definition: It is good to experience pleasure and sensual gratification. • Contained values: Have pleasure
Achievement	• Definition: It is good to be successful in accordance with social norms. • Contained values: Be ambitious, Have success, Be capable, Be intellectual, Be courageous
Power: dominance	• Definition: It is good to be in positions of control over others. • Contained values: Have influence, Have the right to command
Power: resources	• Definition: It is good to have material possessions and social resources. • Contained values: Have wealth
Face	• Definition: It is good to maintain one's public image. • Contained values: Have social recognition, Have a good reputation
Security: personal	• Definition: It is good to have a secure immediate environment. • Contained values: Have a sense of belonging, Have good health, Have no debts, Be neat and tidy, Have a comfortable life
Security: societal	• Definition: It is good to have a secure and stable wider society. • Contained values: Have a safe country, Have a stable society
Tradition	• Definition: It is good to maintain cultural, family, or religious traditions. • Contained values: Be respecting traditions, Be holding religious faith
Conformity: rules	• Definition: It is good to comply with rules, laws, and formal obligations. • Contained values: Be compliant, Be self-disciplined, Be behaving properly
Conformity: interpersonal	• Definition: It is good to avoid upsetting or harming others. • Contained values: Be polite, Be honoring elders
Humility	• Definition: It is good to recognize one's own insignificance in the larger scheme of things. • Contained values: Be humble, Have life accepted as is
Benevolence: caring	• Definition: It is good to work for the welfare of one's group's members. • Contained values: Be helpful, Be honest, Be forgiving, Have the own family secured, Be loving
Benevolence: dependability	• Definition: It is good to be a reliable and trustworthy member of one's group. • Contained values: Be responsible, Have loyalty towards friends
Universalism: concern	• Definition: It is good to strive for equality, justice, and protection for all people. • Contained values: Have equality, Be just, Have a world at peace
Universalism: nature	• Definition: It is good to preserve the natural environment. • Contained values: Be protecting the environment, Have harmony with nature, Have a world of beauty
Universalism: tolerance	• Definition: It is good to accept and try to understand those who are different from oneself. • Contained values: Be broadminded, Have the wisdom to accept others
Universalism: objectivity	• Definition: It is good to search for the truth and think in a rational and unbiased way • Contained values: Be logical, Have an objective view

Table 33: Value taxonomy introduced by Kiesel et al. (2022). In this study, we focus on 20 values corresponding to the level 1 categories.

Seeker Responses	**Detected Values**
Yes, I accept your thought, and it gives me support. Thank you for your concern.	Benevolence: caring Security: personal Universalism: tolerance
I will keep trying until I secure a new job. I will not rest.	Security: personal Achievement Self-direction: action Power: resources
That is a really valid point and is helping me see the bigger picture in life. I need to know it won't always be this way.	Benevolence: caring Security: personal Achievement Universalism: tolerance Stimulation Humility Universalism: objectivity

Table 34: Examples of the seeker's utterances in ESConv, along with the values observed in each one

Category	Details
Exploration of Issues and Challenges	• Insufficient understanding of the patient's key challenges and emotional struggles • Lack of focus on how the patient processes emotions or responds to difficulties
Exploration of Personal Interests	• Limited discussion on what genuinely excites or engages the patient • Insufficient exploration of the patient's hobbies or areas of curiosity • Lack of encouragement for the patient to share their unique interests and passions
Exploration of Goals and Motivations	• Limited understanding of the patient's life goals, ambitions, and decision-making drivers • Insufficient attention to articulating and clarifying meaningful objectives
Strengths and Achievements Acknowledgment	• Missed opportunities to recognize the patient's unique strengths and past successes • Insufficient celebration of the patient's efforts and accomplishments • Limited acknowledgment of their capacity to overcome challenges and reinforce existing skills
Emotional Resilience and Coping Strategies	• Insufficient guidance on building emotional resilience and adaptability • Limited focus on constructive ways to navigate difficult emotions, fears, or insecurities • Lack of practical approaches to manage stress, foster confidence, and maintain balance
Focus on Achievable Goals	• Limited attention to setting small, manageable goals for progress • Insufficient guidance on breaking down objectives into actionable tasks
Motivation and Alignment with Goals	• Missed opportunities to align goals with the patient's values and aspirations • Limited encouragement for personal and professional growth opportunities • Lack of suggestions for activities that resonate with the patient's interests
Self-Compassion and Acceptance	• Insufficient exploration of ways to foster self-kindness and embrace imperfections • Limited focus on addressing feelings of shame and building self-acceptance

Table 35: Categories and descriptions of areas identified for improvement in value reinforcement.