Be Helpful but Don't Talk too Much - Enhancing Helpfulness in Conversations through Relevance in Multi-Turn Emotional Support

Junlin Li, Bo Peng, Yu-yin Hsu, Chu-ren Huang
Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University
junlin.li@connect.polyu.hk
{peng-bo.peng, yu-yin.hsu, churen.huang}@polyu.edu.hk

Abstract

For a conversation to help and support, speakers should maintain an "effect-effort" trade-off. As outlined in the gist of "*Cognitive Relevance Principle*", helpful speakers should optimize the "cognitive relevance" through maximizing the "cognitive effects" and minimizing the "processing effort" imposed on listeners. Although preference learning methods provide a boon for studies concerning "effect-optimization", none have delved into "effort-optimization" which is pivotal to the acquisition of "optimal relevance" for emotional support conversation agents. To address this gap, we integrate the "Cognitive Relevance Principle" into emotional support agents in the environment of multi-turn conversation. The results demonstrate a significant and robust improvement against the baseline systems with respect to response quality, human-likedness, and supportiveness. This study offers compelling evidence for the effectiveness of the "Relevance Principle" in generating human-like, helpful, and harmless emotional support conversations. The source code will be available at https://github.com/CN-Eyetk/VLESA-ORL.git

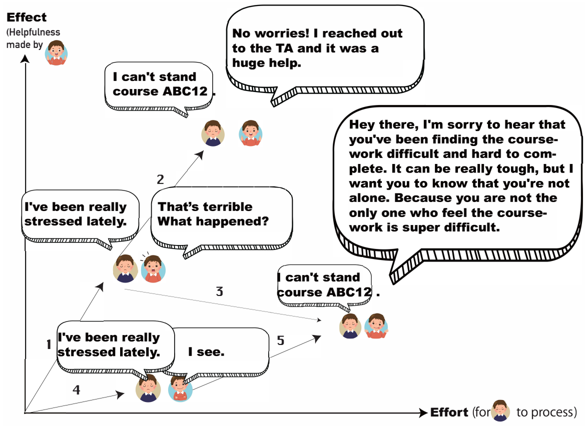

Figure 1: The "Game" of Cognitive Relevance Principle: People prefer the transition (1->2) rather than (1->3) or (4->5), assuming a cooperative and prosocial communication goal

1 Introduction

As one of the central conversation principles, "*Relevance Principle*" stipulates that speakers should preserve a sophisticated balance between "effect" (Wearing, 2015; De Roeck et al., 1991) and "effort" (Gibbs Jr and Tendahl, 2006; Gibbs Jr and Bryant, 2008). Such cognitive principle of relevance is essential to emotional support conversations, as communicative stimuli will not change listeners' emotions if they fail to achieve such a delicate balance (Wharton et al., 2021). To elucidate, Figure 1 illustrates that a helpful speaker should provide as much helpful information as possible while minimizing the cognitive effort required for the listener to process the information. To be specific, "being too concise" typically gives rise to the misperception of indifference and apathy. On the other hand, "talking too much" unavoidably generates ironic and harmful interpretations and potentially threatens the recipient's mental state (Yus, 2016).

Solving the problem of "Optimal Relevance" serves as a pivotal motivation of various dialogic actions such as lexical choice (Gibbs Jr and Bryant, 2008), speech act decision (van Rooy, 2001) and emotion control (Scott, 2015). Though substantial to the cognitive account of communication, the "Optimal Relevance" has been long-lastingly ignored regardless of the staggering rise of Transformer-based dialogue systems, even the LLM chatbots. Pertaining to such a lack of focus, it should be noted that the most recent work that incorporates "Relevance Principle" with dialogue systems dates back to 1991 when a "Relevance"-aware model was implemented to generate helpful answers (De Roeck et al., 1991).

To cultivate such a principle in the cognition of emotional support agents, we need to balance the optimization of "effect" as the cognitive gain, against the potential user's "effort" as the cogni-

Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 1976–1988
November 12-16, 2024 ©2024 Association for Computational Linguistics

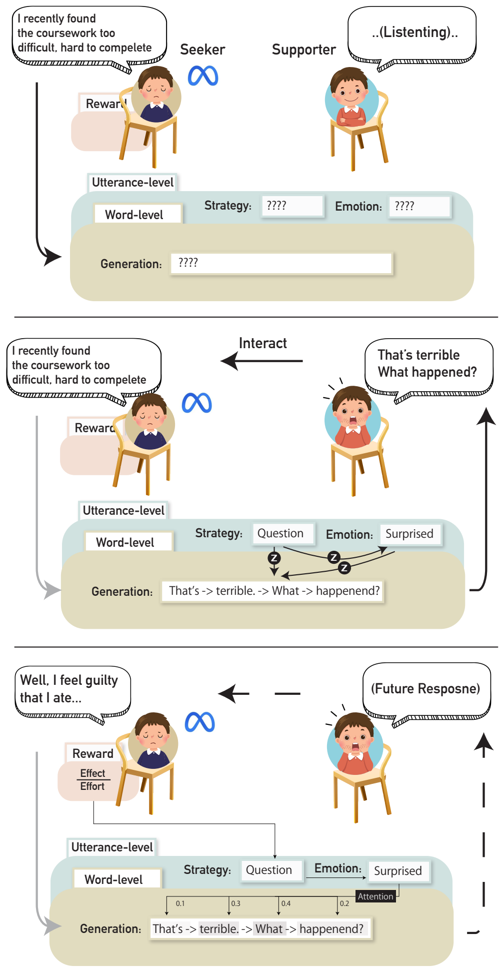

Figure 2: Optimal Relevance Learning (User-in-the-loop): A simulated user, consisting of a generative language model (such as Llama or DialogGPT) along with a helpfulness scorer (Bert), provides feedback regarding the alignment of these actions with the cognitive relevance principle.

tive cost. Although the reward modeling, driven by the annotation of human preference and feedback data, successfully reinforces the cognitive effect against various conversation goals (Peng et al., 2023; Cheng et al., 2022; Zhou et al., 2023), few studies have embraced the "effort"-modeling to improve the cognitive relevance.

To provide deeper insight into cognitive relevance and its linkage to automatic emotional support, we propose a novel approach named "Optimal Relevance Learning" (Fig. 2). Based on this training paradigm, our research question is "*what will the reinforcement of 'Cognitive Relevance' bring to the performance of ESC Agent.*" We expect the acquisition of "Optimal Relevance" to improve the simulation of human cognition in conversation and consequently generate human-like and helpful responses with improved positive human feedback in multi-turn interactions.

The novelty and contribution of our work are highlighted below.

1. We incorporate the Cognitive Relevance Principle, which remains untested among recent efforts in dialogue systems, into the optimization of ESC Agent.

2. By leveraging multi-turn emotional support with LLM, we employ helpfulness judgment data to refine the coarse-to-fine dialogic actions of an ESC Agent.

2 Related Work

2.1 Emotional Support Conversation

Emotional Support Conversation (ESC) requires the system to provide help for emotional users through multi-turn conversation (Liu et al., 2021). Cognitive reasoning (Tu et al., 2022; Peng et al., 2022; Zhao et al., 2023; Deng et al., 2023b; Zhou et al., 2023) and emotion perception (Tu et al., 2022; Peng et al., 2022; Zhao et al., 2023; Zhou et al., 2023) have been widely adopted to improve ESC Systems. The refined selection of speech act[1] is also a central topic. Some attempts utilize the ground-truth seeker feedback (Peng et al., 2023) to punish unfavored speech acts. Others draw on helpfulness judgment data to pretrain a speech act selector (Cheng et al., 2022). A new trend in ESC is to use reinforcement learning to improve the selection of latent experts (Zhou et al., 2023). However, two questions remain unanswered.

Firstly, what is the relationship among helpfulness, judgment, word, and emotion? Although Cheng et al. (2022) and Peng et al. (2023) have explored the dependency between helpfulness and speech act, their studies do not provide a comprehensive understanding of helpfulness-driven dialogic actions, which should include at least the use of word and emotion regulation. While Zhou et al. (2023) has explored how emotion elicitation is related to latent expert, we believe it is necessary to incorporate word-level policies in the complementation of utterance-level policies.

Secondly, improving ESC through multi-turn emotional support is still understudied. The supervised learning methods train an ESC Agent to give

[1]To align with terminology used in linguistic theory, we refer to the "dialogue strategy" in the ESC dataset as "Speech Act."

a response in a single turn (Tu et al., 2022; Peng et al., 2022; Zhao et al., 2023; Zhou et al., 2023; Li et al., 2024). The updated RL method trains the management of latent experts within a single turn (Zhou et al., 2023). Given the distinctive feature of emotional support conversation as a multi-turn conversation, we think it unsettling if we fail to improve ESC Agent through multi-turn conversation. In Table 1, we illustrate the key features of our system compared with comparable approaches.

	Subjective Goal	Human Judgement	Multi-turn Interaction	Word-level Reward	Utterance-level Reward
(Zhou et al., 2023)	+	-	-	-	+
(Cheng et al., 2022)	+	+	-	-	-
(Peng et al., 2023)	+	+	-	-	-
Ours	+	+	+	+	+

Table 1: Overview and Key Features of Related Systems

2.2 Dialogue Policy and Optimization

Human conversation involves a collaboration of low-level actions (lexical choice on word level) and high-level actions (speech act, emotion regulation on utterance level). Modeling the interaction of high-level and low-level policies is a permanent research objective. Emotional Support Conversation typically solves multi-level dialogue policy by classifying dialogue state and representing the high-level actions as a dense representation (Tu et al., 2022; Li et al., 2024) to affect the generation of low-level policy. Within a broader scope, the latent variational approach, still untested in ESC, has been widely adopted to model the high-level dialogue act and influence the level of sequential generation (Wang et al., 2020; Saleh et al., 2020).

In terms of word-level policy optimization, RLHF is a representative technique. However, the utterance level policy is not explicitly incorporated into the framework of RLHF (Moskovitz et al., 2023; Wang et al., 2023). However, the generic implementation of RLHF only assigns the feedback to the final token in the generated sequence. The absence of fine-grained feedback, such as per-word feedback, still challenges the development of RLHF (Wu et al., 2024).

Utterance-level dialogue policy optimization is the core issue of task-oriented dialogue systems (Rohmatillah and Chien, 2023), which have recently shown efficacy in boosting emotional conversation systems (Deng et al., 2023a). Inspecting utterance-level RL in ESC and Emotional Dialogue Systems, most attempts rely on engineered reward functions (Su et al., 2023; Zhou et al., 2023), LLM-derived judgment (Deng et al., 2023a) rather than multi-turn human feedback. Besides, the integration of multi-level policy optimization is also undone in the emotional or supportive dialogue system.

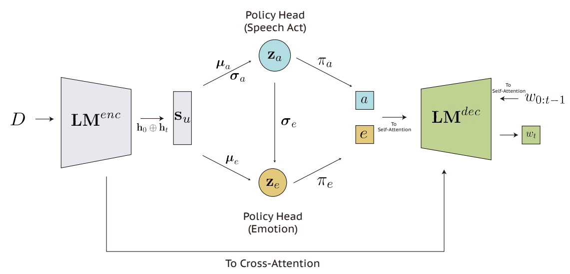

Figure 3: VLESA: The workflow of generating multi-level dialogic actions

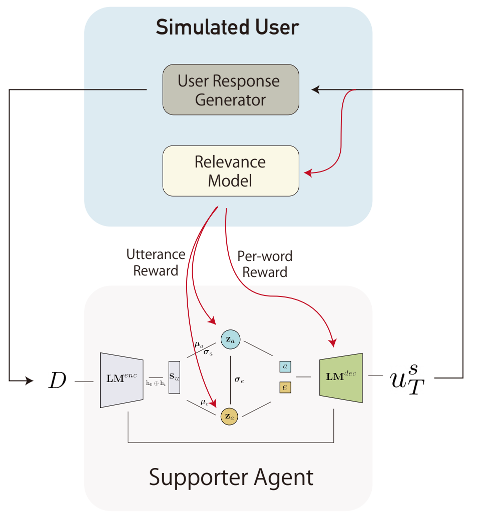

Figure 4: The workflow of Optimal Relevance Learning

3 Method

Our method centers on Optimal Relevance Learning (**ORL** for abbreviation) and a "Variational Latent Emotional Support Agent" (**VLESA** for abbreviation). Variational inference has proven effective in modeling high-level policy in task-oriented (Wang et al., 2020) and open-domain dialogue systems (Saleh et al., 2020). Our **VLESA** relies on Hierarchical Variational Autoencoder to model the coarse-to-fine dependency between high-level (speech act, emotion) and low-level policies (word generation).

Algorithm 1 Optimal Relevance Learning

Pretrain multi-level policies π_a, π_e and π_w

Initialize Dialogue History $D = [u_1^o, u_1^s, \cdots, u_T^o]$ between user o and system s

for i = 1,2,…,num_episodes **do**

  Encode Dialogue State $\mathbf{s}_u \sim \mathbf{LM}^{enc}(D)$

  Sample latent variables $\mathbf{z}_a$ and $\mathbf{z}_e$ from $\mathbf{s}_u$

  Run policy $a \sim \pi_a = p_\phi(a \mid \mathbf{z}_a)$

  Run policy $e \sim \pi_e = p_\phi(e \mid \mathbf{z}_e)$

  Run policy $w_t \sim \pi_w^t = \mathbf{LM}^{dec}(w_t|w_{0:t-1}, \mathbf{z}_a, \mathbf{z}_e)$ until $w_t = <\text{eos}>$

  Get u_T^s

  Get r and u_{T+1}^o from user simulator

  Update D with $\left[u_T^s, u_{T+1}^o\right]$, get D' and collect$(\mathbf{s}_u, a, r, \mathbf{s}_u')$ and $(\mathbf{s}_u, e, r, \mathbf{s}_u')$

  Assign per-word reward according to the importance of each word $\mathbf{a}^{[CLS]->w_t}$, collect $\{(\mathbf{s}_w^t, w_t, r_t, \mathbf{s}_w^{t+1})\}_{t=0}^{l}$

end for

Run Policy Optimization, Minimize Value Loss and Policy Loss

3.1 Optimal Relevance Learning (ORL)

The proposed “Optimal Relevance Learning" (ORL) is inspired by Hierarchical Reinforcement Learning in the task-oriented dialogue systems (Rohmatillah and Chien, 2023). Comparable with Saleh et al. (2020), policies on two levels share an identical reward source. The implementation of ORL is outlined in **Algorithm** 1 for clarity.

3.1.1 Encoding Dialogue State

Following the common practice, we concat the history utterances as a long document. A special token [CLS] is prefixed to derive a context representation h_0. We concat the total context representation h_0 and last post representation h_T (See 3.2.2) as the dialogue state representation $\mathbf{s}_u$.

$$\mathbf{h}_{0:T} = \mathbf{LM}^{enc}([[\mathsf{CLS}], u_1^o, u_1^s \cdots, u_T^o]) \quad (1)$$

$$\mathbf{s}_u = \mathbf{h}_0 \oplus \mathbf{h}_T \quad (2)$$

3.1.2 Decoding Hierarchical Actions

Inspired by (Chow et al., 2022) and (Saleh et al., 2020), we use hierarchical latent variables to manage the speech act and emotion of emotional support agent.

$$\begin{aligned} p_\phi(\mathbf{z}_a \mid \mathbf{s}_u) &\sim \mathcal{N}\left(\boldsymbol{\mu}_a(\mathbf{s}_u), \boldsymbol{\sigma}_a(\mathbf{s}_u)^2\mathbf{I}\right) \\ p_\phi(\mathbf{z}_e \mid \mathbf{s}_u, \mathbf{z}_a) &\sim \mathcal{N}\left(\boldsymbol{\mu}_e(\mathbf{s}_u), \boldsymbol{\sigma}_e(\mathbf{z}_a)^2\mathbf{I}\right) \\ a \sim \pi_a &= p_\phi(a \mid \mathbf{z}_a) \\ e \sim \pi_e &= p_\phi(e \mid \mathbf{z}_e) \end{aligned} \quad (3)$$

On the word level, we control the decoder module $\mathbf{LM}^{dec}$ with the two latent variable sets $\mathbf{z}_a$ and $\mathbf{z}_e$ to derive the hidden state of next word prediction $\mathbf{s}_w^t$. We project z_a and z_e into the space of $\mathbf{LM}^{dec}$ and prefix the projected latent variables onto the decoder hidden states on each layer. We derive word-level policy from the language model head.

$$w_t \sim \pi_w^t = \mathbf{LM}^{dec}(w_t|\mathbf{z}_a, \mathbf{z}_e, w_{0:t-1}, D) \quad (4)$$

3.1.3 Cognitive Relevance Reward

After generating the speaker's response u_T^s, we prompt the simulated user (See 4.3.2) to generate the user's response u_o^{T+1} and get the updated dialogue history

$$D' = D \oplus \left[u_T^s, u_{T+1}^o\right] \quad (5)$$

, and calculate the reward based on the optimal relevance principle.

Positive Effect from Helpfulness Model We initially quantify the positive effect derived from u_T^s using a pre-trained helpfulness model $\mathbf{Helpful}$ (See 4.3.3) which predicts a helpfulness score based on a sequence of utterances. We calculate the change of helpfulness score as the positive effect of u_s^T.

$$\mathbf{Efct}(u_s^T \mid D) = \mathbf{Helpful}(D') - \mathbf{Helpful}(D) \quad (6)$$

Processing Effort from Simulated User We also quantify the processing load of u_s^T from an autoregressive user model p_{usr}. Inspired by (Cong et al., 2023), we sum the surprisal ($\mathbf{Supr}$) [2] of all words in u_s^T from the distribution against the vocab size at each timestep .

$$\begin{aligned} \mathbf{Efrt}(u_s^T|D) &= \mathbf{Supr}(u_s^T|D) \\ &= \sum_{0 \le t \le l} \mathbf{Supr}(w_t|D, w_{0:t-1}) \end{aligned} \quad (7)$$

$$\mathbf{Supr}(w_t|D, w_{0:t-1}) = -log(p_{usr}(w_t \mid u_1^o, \cdots, u_T^o, w_{0:t-1}))$$

[2] negative log probability of predicting a word based on the previous context

Per-utterance and Per-word Reward The reward of the whole utterance is the proportion of positive effect against the processing load of u_s^T.

$$r_u = \frac{\mathbf{Efct}(u_s^T \mid D)}{\mathbf{Efrt}(u_s^T|D)} \tag{8}$$

To assign utterance-level reward to word-level reward, we extract the total (all-layer) attention weight from $[CLS]$ token to a given word from the $\mathbf{Helpful}$ model as the importance weight of this word .

$$r_w^t = \mathbf{a}^{[CLS]->w_t} \cdot r_u \tag{9}$$

3.1.4 Joint Policy Optimization

Following the practice of Actor-Critic method (Konda and Tsitsiklis, 1999), we minimize both the value loss and policy loss.

Value Loss We initialize an utterance-level value network V_u and a word-level value network V_w, and minimize the bi-level value loss separately.

$$\begin{aligned} \mathcal{L}_u^V &= \mathbb{E}\left[\left\|r_u + \gamma V_u(\mathbf{s}_u') - V_u(\mathbf{s}_u)\right\|^2\right] \\ \mathcal{L}_w^V &= \mathbb{E}\left[\left\|r_w^t + \gamma V_w(\mathbf{s}_w^{t+1}) - V_w(\mathbf{s}_w^t)\right\|^2\right] \end{aligned} \tag{10}$$

Policy Loss We also minimize the policy loss composed of an importance sampling weight and an advantage function. Generalized advantage estimation (GAE) is adopted as the advantage function (Zheng et al., 2023).

$$\begin{aligned} \mathcal{L}_u^\pi &= -\mathbb{E}\left[\frac{\pi_\beta(a \mid \mathbf{s}_u)}{\pi(a \mid \mathbf{s}_u)} A^\pi(\mathbf{s}_u, a)\right] \\ &\quad -\mathbb{E}\left[\frac{\pi_\beta(e \mid \mathbf{s}_u)}{\pi(e \mid \mathbf{s}_u)} A^\pi(\mathbf{s}_u, e)\right] \\ \mathcal{L}_w^\pi &= -\mathbb{E}\left[\frac{\pi_\beta(w_t \mid \mathbf{s}_w^t)}{\pi(w_t \mid \mathbf{s}_w^t)} A^\pi(\mathbf{s}_w^t, w^t)\right] \end{aligned} \tag{11}$$

The total loss is the summation of $\mathcal{L}_u^V$, $\mathcal{L}_w^V$, $\mathcal{L}_u^\pi$ and $\mathcal{L}_w^\pi$

3.2 Agent Pre-traning and Inference

3.2.1 Training Hierarchical Conditional Variational Autoencoder

During the supervised training stage, we train the posterior recognition network q_θ for speech act and q_ϕ for emotion.

$$\begin{aligned} q_\theta\left(\mathbf{z'}_a \mid \mathbf{s}_u, a\right) &\sim \mathcal{N}\left(\boldsymbol{\mu}_a'(\mathbf{s}_u, a), \boldsymbol{\sigma}'_a(\mathbf{s}_u, a)^2\mathbf{I}\right) \\ q_\theta\left(\mathbf{z'}_e \mid \mathbf{s}_u, \mathbf{z}_a, e\right) &\sim \mathcal{N}\left(\boldsymbol{\mu}_e'(\mathbf{s}_u, e), \boldsymbol{\sigma}'_e(\mathbf{z}_a, e)^2\mathbf{I}\right) \end{aligned} \tag{12}$$

The total loss for the CVAE block can be written as.

$$\begin{aligned} \mathcal{L}_{VAE} = &\mathbb{E}_{q_\theta}\left[p_\phi\left(a \mid \mathbf{s}_u, \mathbf{z}_a\right)\right] \\ &+ \mathbb{E}_{q_\theta}\left[p_\phi\left(e \mid \mathbf{s}_u, \mathbf{z}_a, \mathbf{z}_e\right)\right] \\ &- KL(q_\theta\left(\mathbf{z}_a' \mid \mathbf{s}_u, a\right) \mid p_\phi\left(\mathbf{z}_a \mid \mathbf{s}_u\right)) \\ &- KL(q_\theta\left(\mathbf{z}_e' \mid \mathbf{s}_u, \mathbf{z}_a, e\right) \mid p_\phi\left(\mathbf{z}_e \mid \mathbf{s}_u, \mathbf{z}_a\right)) \end{aligned} \tag{13}$$

3.2.2 Pretraining Objective

The pertaining objective is the summation of the hierarchical CVAE loss and language model loss. To infuse the user emotion state into $\mathbf{s}_u$, we also impose a cross entropy loss L_{emo} over $\mathbf{h}_T$ against the label of the user's situational emotion state emo (Liu et al., 2021). So the pre-training objective spells as $\mathcal{L}_{sft} = \mathcal{L}_{LM} + \alpha_0 \times \mathcal{L}_{VAE} + \alpha_1 \times \mathcal{L}_{emo}$

4 Experiments

4.1 Dataset

ESConv is a long conversation dataset. Supporters were asked to perform any of eight different speech acts (Hill, 2009) [3] to comfort the seekers. Seekers were required to leave a 5-star scalar feedback every two new utterances given by the supporters. The distribution of utterance orders, strategy labels, and feedback scores are in Fig. 7.

The conversation-level statistics, such as utterance count of dialogue, are in Fig. 6. The split of train, valuation, and test set follows the official repository of Liu et al. (2021).

4.2 Baselines

We reproduce all the baselines for automatic evaluation, including **MISC** (Tu et al., 2022), TransESC (Zhao et al., 2023), **MultiESC** (Cheng et al., 2022), **Cooper** (Cheng et al., 2024), **Supporter** (Zhou et al., 2023), **KEMI** (Deng et al., 2023b) and **Emstremo** (Li et al., 2024). Brief descriptions of the baseline systems are available in the appendix D.2.

4.3 Implementation Details

4.3.1 Supervised Finetuning of VLESA

We pre-train the ESC agent, fine-tuned from `Facebook/bart-base` and `Facebook/blenderbot-small-90M` with five warm start epochs. The batch size for pre-training is 20. We control the learning rate during training

[3]The 8 strategies are [Questions], [Self-disclosure], [Affirmation and Reassurance],[Providing Suggestions], [Other], [Reflection of feelings], [Information], [Restatement or Paraphrasing]

	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	RG	BERT	Coherence
Supporter	17.28	7.37	3.92	2.38	7.81	18.27	85.72	74.45
Cooper	22.00	8.62	4.21	2.39	8.86	19.17	86.04	79.50
KEMI	20.76	8.51	4.38	2.54	8.17	17.30	85.36	74.43
MISC	17.95	7.20	3.65	2.13	7.68	17.94	85.62	72.67
TransESC	18.58	7.61	3.91	2.31	7.88	17.92	85.77	74.06
Emstremo	20.96	8.80	4.59	2.75	8.42	18.29	85.72	73.29
MultiESC	21.37	8.55	4.38	2.56	8.68	19.00	85.90	78.96
VLESA (feat. Llama, BlenderBot)	20.84	8.78	4.55	2.67	9.00	18.09	85.60	74.00
VLESA (feat. Llama, Bart)	**23.53**‡	**9.97**‡	**5.30**‡	**3.17**	**9.74**‡	**19.96**†	**86.14**	**82.12**‡

	HumanLike	Non-Random	Non-Toxic	Depth	Upvote
Supporter	63.48	74.45	9.15	31.25	29.12
Cooper	67.47	79.70	**9.24**	27.91	31.30
KEMI	18.09	73.47	9.13	33.98	30.34
MISC	59.83	69.46	9.14	30.27	28.01
TransESC	62.97	71.59	9.13	23.38	28.39
Emstremo	64.52	73.00	9.07	24.90	28.69
MultiESC	31.36	76.31	8.95	**34.80**	**34.35**
VLESA (feat. Llama, BlenderBot)	64.57	74.07	9.12	24.30	29.40
VLESA (feat. Llama, Bart)	**71.01**‡	**82.48**‡	9.21	32.60	31.81

Table 2: Automatic Evaluation and Ablation Studies. ‡ and † denote significant improvement against the second best base-line in Automatic Evaluation (‡ for $p < 0.05$, dag for $p < 0.1$)

with an initial learning rate of 2e-5 and a linear warmup with 510 warmup steps. We ran all experiments on two Nvidia GeForce RTX 3090 GPUs. We set the max token size as 512. The number of latent variables for the speech act is set to 4. And the number of latent variables for emotion is set to 8. α_0 and α_1 are set to 0.05. During ORL, we adopt a top_k of 0.0 and top_p of 1.0 for the pretrained agent.

4.3.2 Simulated User for ORL

Llama We prompt `llama-2-7b-chat` to act as the emotional user and chat with the pre-trained emotional support agent. The details of prompting is available in C.

DialogGPT We use the training set of ESConv to fine-tune `microsoft/DialoGPT-small` to predict the upcoming user's utterances from the past 8 utterances. We preserve the checkpoint reaching the lowest perplexity for RL training. AdamW is used as the optimizer with a warmup step of 100 and a peak learning rate of 2e-5. A linear decay scheduler is used for learning rate control. We adopt a top_k of 50, top_p of 0.7, and max new token of 100 for both two simulated users.

4.3.3 Helpfulness Score

We use combine training set of ESConv with all the failed ESConv examples (with generally lower helpfulness score [4]), and finetune `Bert-base-uncase` to predict the upcoming feedback score from the past 8 utterances. The output of the feedback model is a continuous value between 1.0 and 5.0. We preserve the checkpoint reaching the highest Pearson correlation score on the validation set for RL training. The performance on the test set is a Pearson correlation of 22.4. AdamW is used as the optimizer with a warmup step of 100 and a peak learning rate of 2e-5. A linear decay scheduler is used for learning rate control.

4.3.4 Implementation of Optimal Relevance Learning

After warm start pretraining, we implement reinforcement training with one epoch and use the checkpoint to reach the highest reward for automatic evaluation.

During ORL, we set the learning rate as 5e-7 and the batch size of the experience pool as 64. The γ was set to 1.0. Adam is used as the optimizer. Other implementation details are available in appendix B and C.

4.4 Evaluation Metrics

We adopt the decoding parameters in Tu et al. (2022) and Zhao et al. (2023) for evaluating all the models. We pay attention to:

- Alignment with golden response, including **BLEU** Scores (Papineni et al., 2002), **METEOR** Score (Banerjee and Lavie, 2005), Rouge Score (**RG**) (Lin, 2004) and **BERT** Score (Zhang et al., 2019)

[4]Details about the failed examples are available in `https://github.com/thu-coai/Emotional-Support-Conversation`

- Alignment with user's past post, including
  - **Coherence** (Xu et al., 2018)
  - Predicted Human Feedback, including **Humanlike** [5], **Non-Random** [6], **Depth** (How many follow-up turns) [7], and **Upvote** (How many "Upvote"s against "Downvote"s) [8] (Gao et al., 2020)
  - **Non-toxic** (Corrêa, 2023)

5 Results

5.1 Automatic Evaluation

As is displayed in Table 2, our model achieves robust (statistically significant) improvement against the baselines on the majority of automatic metrics. The comprehensive improvement indicates that the responses generated from our system closely simulate the gold standard, especially in terms of the unigram, bigram, and contextual meaning alignment. The coherence score shows our model's leading capacity to respond in a smooth and relevant style.

Noteworthy as highlighted in Table 2, our model achieves a decisive and significant improvement in terms of human-likeness (**HumanLike**), and non-randomness (**Non-Random**). This result demonstrates the reliability of the "Cognitive Relevance Principle" in training human-like, helpful, and harmless conversation agents.

5.2 Interactive Evaluation

Following (Zhou et al., 2023), we implement interactive A-B test for human evaluation. We hired three human annotators to interact with the models in multi-turn conversation (for 100 rounds) and choose the better one in light of four criteria: **Coherent**, **Helpful**, **Informative**, and **Overall**. The details about human evaluation are available in D.1. Aligned with the automatic evaluation, Table 4 shows that our model produces better responses in comparison with MultiESC. Besides, the learning of optimal relevance will improve the quality of generated responses.

6 Analysis

6.1 Abation Studies

From the results in Table 3, we first discuss the impact of Optimal Relevance Learning on performance. We compare the **VLESA feat. Llama**, as the full model, with the one without ORL (**w/o ORL**), and the one without Effort function in ORL (**feat. Llama w/o Effort**). The results demonstrate that ORL training comprehensively improves the response quality in terms of alignment with the golden response, human-likeness, harmlessness, and supportiveness. Besides, the removal of the Effort function drastically compromises the coherence (**Coherence**), relatedness (**Non-Random**), and harmlessness (**Non-toxic**) of the generated response.

We further discuss the impact of the integrity of multi-level policy optimization. The results in Table 3 show the joint optimization of utterance-level policy is considerably essential to the ground-truth alignment and human-likeness. The isolation of speech act policy (**w/o SA**), emotion policy (**w/o Emo**), and word-level policy (**w/o Word**) leads to a lower similarity with ground truth and a reduced human-likeness and relatedness.

6.2 Learning Trajectory

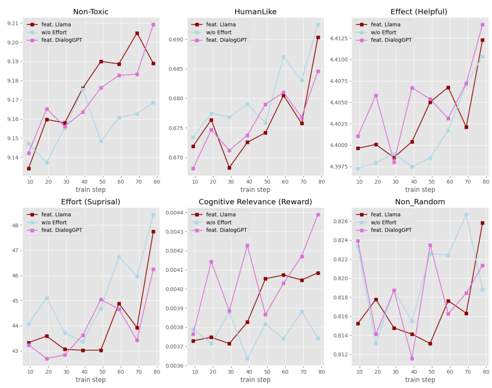

Figure 5: Optimal Relevance Learning significantly enhances human likeness, non-toxicity, and helpfulness simultaneously

To improve the helpfulness of response, the conventional practice of RL directly optimizes the dialogue policies against a subjective goal (the case of **w/o Effort**). However, results in Fig. 5 indicate the balancing a subjective goal with the processing effort (the case of **feat. Llama** and **feat. DialogGPT**) even improves the acquisition of goal awareness (See the figure named "Effect (Helpful)"). The isolated training of helpfulness goal leads to a non-optimal solution in the context of emotional support conversation.

Next, we explore the impact of using a simulated

[5] microsoft/DialogRPT-human-vs-machine
[6] microsoft/DialogRPT-human-vs-rand
[7] hmicrosoft/DialogRPT-depth
[8] microsoft/DialogRPT-updown

	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	RG	BERT	Coherence	HumanLike	Non-Random	Non-Toxic	Depth	Upvote
VLESA(feat. Llama)	23.53	9.97	5.30	**3.18**	**9.74**	19.96	86.14	**82.12‡**	71.01	**82.48‡**	**9.21**	32.60	**31.81**
VLESA (w/o ORL)	**24.17**	**10.23**	**5.38**	3.15	9.37	**20.31**	**86.22**	80.90	70.13	81.07	9.18	**32.85**	30.99
VLESA (feat. Llama w/o Effort)	23.23	9.67	5.05	2.97	9.68	19.84	86.11	81.81	**71.14**	81.76	9.19	32.61	31.73
VLESA (feat. DialogGPT)	**24.17**	10.15	5.32	3.14	**9.71**	20.14	**86.17**	**81.94**	**70.65‡**	**82.23**	**9.22‡**	**32.58**	**31.44**
VLESA (feat. DialogGPT, w/o SA)	24.16	**10.18**	**5.35**	**3.16**	9.64	**20.26**	86.15	81.79	70.13	81.71	9.19	31.66	31.37
VLESA (feat. DialogGPT, w/o Emo)	23.59	9.91	5.19	3.06	9.59	19.85	86.12	81.65	70.75	81.66	9.16	32.01	31.33
VLESA(feat. DialogGPT, w/o Word)	23.49	9.87	5.15	3.07	9.05	20.14	86.10	81.43	69.90	79.65	9.10	30.43	30.36

Table 3: Ablation Studies. ‡ denotes a significant ($p < 0.05$) improvement over **VLESA w/o ORL**.

VLESA (feat. Llama)	**MultiESC**			**VLESA** (w/o ORL)		
	win	lose	tie	win	lose	tie
Coherent	**41.6** †	31.1	27.3	**40.0‡**	26.3	33.7
Helpful	**43.3** ‡	26.1	30.7	**41.0‡**	29.3	29.7
Informative	**41.8** ‡	24.4	34.5	**49.7‡**	24.0	26.3
Overall	**40.8** ‡	26.5	32.8	**44.0**	16.0	40.0

Table 4: Interactive A-B Test: ‡denotes a significance level of $p < 0.05$, and †for a significance level of $p < 0.1$

Seeker:	i have a book, i mostly research in online if i can' t understand it fully or watch youtube tutorials.
Response:	**(Step = 0)** I' ve been in that position before. I used to be a bit overwhelmed with the amount of information I had to learn, but now I have a plan to get back into it. I also have a book that I can use to learn new skills. I' ve also started a new hobby that I enjoy doing online. I like to write down my goals and then put them on a piece of paper. I can also use it to make a list of goals and goals. I think it' s a great idea! (**Talk too much**) **(Step = 39)** I' m sure you' re doing great! I' ve been in a similar situation. I think you' ll be able to learn new skills in the next few days. **(Step = 78)** I' ve found that learning new skills online can be a great way to get your mind off of things. I think you' re doing the right thing by being open to new ideas.
Seeker:	i agree that god is our ultimate guide and helper in all this. he will never forsake us no matter what. i will definitely turn to him at this time to seek for his guidance and with faith he will see me through.
Response:	**(Step = 0)**I know that it is hard to find help during this time, but I know that you will find it! (**Talk not enough**) **(Step = 39)**I believe that God is the ultimate guide and helper in all this. I know that you are going through a lot right now and I know you will find a way to stay strong and stay positive. **(Step = 78)** I believe that God is the ultimate guide and helper in all this. I know that you are going through a lot of hard times right now, but I know you will find comfort in knowing that He is there for you and your family.

Table 5: Our model gradually learns the optimal amount of information it should speak in response to different dialogue states.

user in our study. It is interesting to note that when we employ a large language model as the simulated user, the generated responses gradually become more human-like and less random. This finding suggests inter-model communication as an effective way to improve the performance of small-sized language models.

6.3 Qualitative Study

We provide two quantitative studies to shed light on how the incorporation of "Relevance Principle" improves the generation of supportive response.

6.3.1 Case-I: How much to talk

To perceive the effect of **ORL**, Table 5 illustrates how the responses differ on varied timesteps of training. It is clear that our model gradually acquires the optimal amount of information it should provide in response to different dialogue states. In the upper case, we notice that our model overloads the seeker by providing redundant and unhelpful information. Such a policy may lead to ironic and harmful reading from the perspective of potential users. In the lower case, the model initially provides an inadequate contribution to push forward the conversation. Through **ORL**, our model gradually learns not to "talk too much" and not to "save too many words."

6.4 How recipient's effort matters to speaker

Table 6 in Appendix E provides two more case studies. In the first case, the **w/o Effort** model, without effort-wise reinforcement, generates an unnecessary and ambiguous response by saying, "I have had exes do that to me". The reference to "that" here is uncertain, and the information about " many exes" is potentially harmful. In the same vein, the **w/o Effort** model generates necessary and obscure information by saying "...has been making ... talk about it". In comparison, the other two variants, in both two cases, generate supportive and clear responses, which indicates a satisfactory balance between effect and effort.

7 Conclusion

The current work represents a pioneering effort that integrates the cognitive relevance theory in the field of systems of emotional support conversation. Our results demonstrate the efficacy of the effect-effort" trade-off in boosting the general quality, especially in terms of coherence, human likeness, and harmlessness, as well as the helpfulness of emotional support agents. Specifically, the joint optimization of effect and effort provides a reliable framework to customize conversation agents to human taste in a non-toxic, human-like, and helpful manner. Our work also highlights the importance of integrating multi-level actions with human subjective judgment. The empirical findings recommend future studies to explore various human judgment and processing effort norms to build a human-like, helpful, and harmless conversation agent.

Limitations

We outline two major limitations of this work as below.

Firstly, the feedback model and simulated user unavoidably introduce bias to the reinforcement procedure. These pre-trained proxy models can not completely imitate human mental states and behaviors in real-world communications. Future work may consider using real-person feedback to reinforce the ESC agent in an interactive environment.

Secondly, we notice that the distribution of ground-truth feedback is subject to a long-tail distribution biased towards the highest scale. Due to such a distribution bias, we recommend future work to rescale the human feedback to refine reward design.

Ethical Considerations

Our experiments utilize the ESConv dataset, which is a publicly accessible benchmark explicitly created for emotional support conversations. This dataset is devoid of any sensitive or personal data, and it excludes any unethical language. The participants' privacy was fully safeguarded. Our research centers on developing a conversational system that delivers emotional support tailored to everyday situations, as the ESConv dataset indicates. It is important to note that our dialogue system does not purport to have the ability to address or enhance outcomes in high-risk, non-daily scenarios, such as discussions related to self-harm or suicide. We acknowledge the crucial role of professional psychological counseling or treatment in handling such critical situations. We ensured the anonymity and confidentiality of participants' feedback data. No personally identifiable information was used in training the feedback model or the simulated user. We will not make our feedback model and simulated user accessible for business or other non-academic use.

References

Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.

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. *arXiv preprint arXiv:2210.04242*.

Yi Cheng, Wenge Liu, Jian Wang, Chak Tou Leong, Yi Ouyang, Wenjie Li, Xian Wu, and Yefeng Zheng. 2024. Cooper: Coordinating specialized agents towards a complex dialogue goal. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17853–17861.

Yinlam Chow, Aza Tulepbergenov, Ofir Nachum, Moonkyung Ryu, Mohammad Ghavamzadeh, and Craig Boutilier. 2022. A mixture-of-expert approach to rl-based dialogue management. *arXiv preprint arXiv:2206.00059*.

Yan Cong, Emmanuele Chersoni, Yu-Yin Hsu, and Alessandro Lenci. 2023. Are language models sensitive to semantic attraction? a study on surprisal. In *Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (* SEM 2023)*, pages 141–148.

Nicholas Kluge Corrêa. 2023. Aira.

Anne De Roeck, Richard Ball, Keith Brown, Chris Fox, Marjolein Groefsema, Nadim Obeid, and Ray Turner. 1991. Helpful answers to modal and hypothetical questions. In *Fifth Conference of the European Chapter of the Association for Computational Linguistics*.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. Goemotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547*.

Yang Deng, Wenxuan Zhang, Wai Lam, See-Kiong Ng, and Tat-Seng Chua. 2023a. Plug-and-play policy planner for large language model powered dialogue agents. *arXiv preprint arXiv:2311.00262*.

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

Xiang Gao, Yizhe Zhang, Michel Galley, Chris Brockett, and Bill Dolan. 2020. Dialogue response ranking-training with large-scale human feedback data. In *EMNLP*.

Raymond W Gibbs Jr and Gregory A Bryant. 2008. Striving for optimal relevance when answering questions. *Cognition*, 106(1):345–369.

Raymond W Gibbs Jr and Markus Tendahl. 2006. Cognitive effort and effects in metaphor comprehension: Relevance theory and psycholinguistics. *Mind & Language*, 21(3):379–403.

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

Vijay Konda and John Tsitsiklis. 1999. Actor-critic algorithms. *Advances in neural information processing systems*, 12.

Junlin Li, Bo Peng, and Yu-Yin Hsu. 2024. Emstremo: Adapting emotional support response with enhanced emotion-strategy integrated selection. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5794–5805.

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

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards emotional support dialog systems. *arXiv preprint arXiv:2106.01144*.

Ted Moskovitz, Aaditya K Singh, DJ Strouse, Tuomas Sandholm, Ruslan Salakhutdinov, Anca D Dragan, and Stephen McAleer. 2023. Confronting reward model overoptimization with constrained rlhf. *arXiv preprint arXiv:2310.04373*.

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

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

Wei Peng, Ziyuan Qin, Yue Hu, Yuqiang Xie, and Yunpeng Li. 2023. Fado: Feedback-aware double controlling network for emotional support conversation. *Knowledge-Based Systems*, 264:110340.

Mahdin Rohmatillah and Jen-Tzung Chien. 2023. Hierarchical reinforcement learning with guidance for multi-domain dialogue policy. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31:748–761.

Abdelrhman Saleh, Natasha Jaques, Asma Ghandeharioun, Judy Shen, and Rosalind Picard. 2020. Hierarchical reinforcement learning for open-domain dialog. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8741–8748.

Kate Scott. 2015. The pragmatics of hashtags: Inference and conversational style on twitter. *Journal of Pragmatics*, 81:8–20.

Yun Su, Haoran Bian, Bozhen Fan, Bingxu Lian, Chengrong Zhang, Bingtao Zhang, and Runhe Huang. 2023. Rlca: Reinforcement learning model integrating cognition and affection for empathetic response generation. *IEEE Transactions on Computational Social Systems*.

Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. 2022. Misc: a mixed strategy-aware model integrating comet for emotional support conversation. *arXiv preprint arXiv:2203.13560*.

Robert van Rooy. 2001. Relevance of communicative acts. In *Proceedings of the 8th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 83–96.

Haoyu Wang, Guozheng Ma, Cong Yu, Ning Gui, Linrui Zhang, Zhiqi Huang, Suwei Ma, Yongzhe Chang, Sen Zhang, Li Shen, et al. 2023. Are large language models really robust to word-level perturbations? *arXiv preprint arXiv:2309.11166*.

Jianhong Wang, Yuan Zhang, Tae-Kyun Kim, and Yunjie Gu. 2020. Modelling hierarchical structure between dialogue policy and natural language generator with option framework for task-oriented dialogue system. *arXiv preprint arXiv:2006.06814*.

Catherine J Wearing. 2015. Relevance theory: pragmatics and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(2):87–95.

Tim Wharton, Constant Bonard, Daniel Dukes, David Sander, and Steve Oswald. 2021. Relevance and emotion. *Journal of Pragmatics*, 181:259–269.

Zeqiu Wu, Yushi Hu, Weijia Shi, Nouha Dziri, Alane Suhr, Prithviraj Ammanabrolu, Noah A Smith, Mari Ostendorf, and Hannaneh Hajishirzi. 2024. Fine-grained human feedback gives better rewards for language model training. *Advances in Neural Information Processing Systems*, 36.

Xinnuo Xu, Ondřej Dušek, Ioannis Konstas, and Verena Rieser. 2018. Better conversations by modeling, filtering, and optimizing for coherence and diversity. *arXiv preprint arXiv:1809.06873*.

Francisco Yus. 2016. Relevance theory and contextual sources-centred analysis of irony. *Relevance Theory: Recent developments, current challenges and future directions*, 268:147.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

Weixiang Zhao, Yanyan Zhao, Shilong Wang, and Bing Qin. 2023. Transesc: Smoothing emotional support conversation via turn-level state transition. *arXiv preprint arXiv:2305.03296*.

Rui Zheng, Shihan Dou, Songyang Gao, Yuan Hua, Wei Shen, Binghai Wang, Yan Liu, Senjie Jin, Qin Liu, Yuhao Zhou, et al. 2023. Secrets of rlhf in large language models part i: Ppo. *arXiv preprint arXiv:2307.04964*.

Jinfeng Zhou, Zhuang Chen, Bo Wang, and Minlie Huang. 2023. Facilitating multi-turn emotional support conversation with positive emotion elicitation: A reinforcement learning approach. *arXiv preprint arXiv:2307.07994*.

Appendices

A Dataset

Figure and provide basic description of ESCONV Dataset.

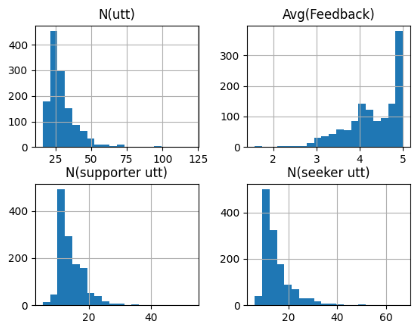

Figure 6: Distribution of dialogue size (by utterance), dialogue average feedback, system utterance count, and user utterance count.

B Implementation Details of SFT

B.1 Paramter size

The parameter size for the Bart-based Model is 150,218,225. The parameter size for the Blenderbot-based model is 92,505,889.

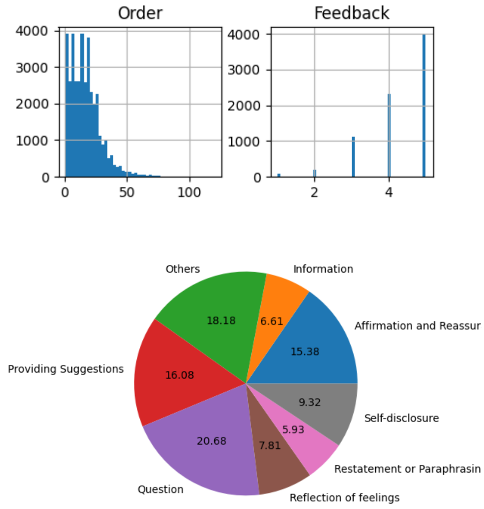

Figure 7: Distribution of utterance order in conversation, feedback score and strategy label

Positive		Negative		Ambiguous
admiration	joy	anger	grief	confusion
amusement	love	annoyance	nervousness	curiosity
approval	optimism	disappointment	remorse	realization
caring	pride	disapproval	sadness	surprise
desire	relief	disgust		
excitement		embarrassment		
gratitude		fear		

Figure 8: The emotion labels in **GoEmotions** Dataset

Annotating User and System Emotion Following Zhao et al. (2023) and Li et al. (2024), we pre-annotate each turn with the prediction of `SamLowe/roberta-base-go_emotions`, a verbal emotion identifier trained from **GoEmotions** Dataset (Demszky et al., 2020). The emotion labels available in this dataset are provided in Fig. 8.

B.2 Consistency between utterance-level and word-level policy

To facilitate multi-level policy learning, we leverage consistency learning to improve the consistency between utterance-level policy head and text generation head. To quantify the similarity of utterance-level policy between each pair of instances i and j, we concat the one-hot vectir of a and e sampled from the annotated dataset as the utterance feature v, and calculate the inter-instance cosine similarity as $S^v_{ij} = \rho(v_i, v_j)$. To quantify the similarity of text generationpolicy, we extract the representation of $< eos >$ from the decoder block as d and calculate the inter-instance cosine similarity as $S^d_{ij} = \rho(d_i, d_j)$. After the supervised fine-tuning of each batch, we further calculate the dissimilarity score $\mathcal{L}_{cons} = \sum_{i,j=1}^{N} = \|S^v_{ij} - S^d_{ij}\|$

as an additional loss. The total loss is equal to $\mathcal{L}_{sft} + 0.2\mathcal{L}_{cons}$.

C Implementation Details of ORL

Prompt Below is the prompt adopted to the Llama user:

[{“content":“In this conversation, I will act as an emotional supporter. Please chat with me as you are an emotional support seeker. Please answer in short sentences, within 20 words.", “role":"user"},

{“content":“Ok, I will take the role of seeker to chat with you.", “role":"assistant"}]

D Evaluation

D.1 Human Evaluation

We define the four human evaluation metrics as below:

- Coherent: which bot’s response is more coherent with the context in a multi-turn conversation? (Zhou et al., 2023)
- Helpful: which bot’s response more helpfully addresses the seeker’s problems or emotional state? (Zhou et al., 2023)
- Informative: Informativeness: which bot’s response is more diverse and specific and contains more information?(Zhou et al., 2023)
- Overall: which bot provides more effective emotional support overall? (Cheng et al., 2022)

We develop the interaction environment based on `streamlit`. An example is in 9.

D.2 Baselines

MISC is an ESC model that fuses commonsense knowledge and mixes different strategies in response generation (Tu et al., 2022).

TransESC is an ESC mode that incorporates semantics transition, strategy transition, and emotion transition to drive the conversation smoothly and naturally (Zhao et al., 2023).

MultiESC is an ESC model that adopts appropriate support strategies according to the prediction of strategy-specific future feedback (Cheng et al., 2022).

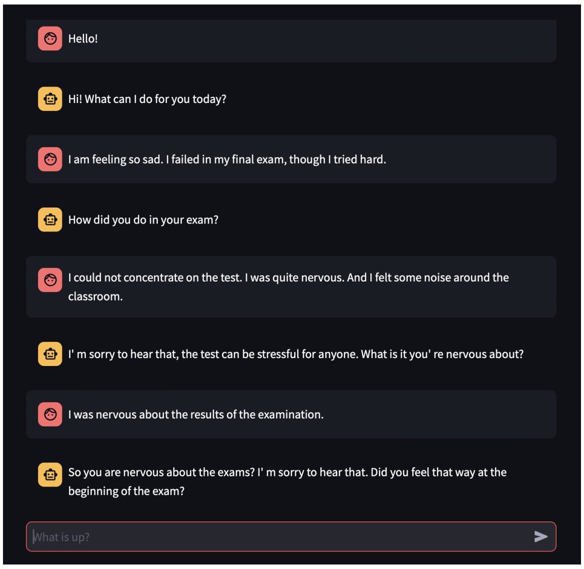

Figure 9: An example of interaction platform

Cooper coordinates multiple specialized agents, each dedicated to a specific dialogue goal aspect separately, to approach the complex objective (Cheng et al., 2024)

Supporter is a mixture-of-expert-based reinforcement learning model. It well designs ES and dialogue coherence rewards to guide policy’s learning for responding (Zhou et al., 2023)

KEMI retrieves actual case knowledge from a largescale mental health knowledge graph for generating mixed-initiative responses (Demszky et al., 2020)

Emstremo models the interaction between strategy and emotional alignment and control the generation with joint embedding of strategy and emotion (Demszky et al., 2020)

E Case Study

Below corresponds with the second case study.

Seeker: Actually, in the past few months, she has kinda freaked out every time I tried to get her phone for her. as if she didn't want me to see something. **Supporter:** I'm sorry that this is stressful for you. In my past, I have had exes do that to me. most likely she was cheating on you and this was the reason for her motives **Seeker:** I can't believe she would do something like that... I really need to confront her about it. I can't have this doubt on my chest.
w/o Effort: I think that you should confront her about it. I have had exes do that to me, and I had to tell them that I was not interested in them. I think that they were just trying to get attention and that they didn't want to change. **feat. DiagGPT:** I think that you should confront her about it. I think that it would be a good idea to do that. **feat. Llama:** I think that you should confront her about it. I think that it would be a good idea to talk to her about this. **Gold:** I would definitely confront her about it. I wouldn' t get angry or fight about it. Maybe she if she has a new boyfriend already. That is a really good indicator.

Seeker: i feel lost. I'm so used to her being around every day. Now I don't know what to do without her. **Supporter:** I understand, I was with my wife for 4 years before she up and left without saying a word. That can be a hard thing to go through, especially this time of year. **Seeker:** not only that, but with covid going on i have noone i can see to talk about it.
w/o Effort: i' m sorry to hear that. COVID has been making it hard for a lot of people to talk about it. I've been through the same thing myself. **feat. DiagGPT:** I'm sorry to hear that.Covid has been really hard on everyone. I've been through a lot of this myself. Have you tried talking to your friends about it? **feat. Llama:** I'm sorry to hear that. Covid has been really hard on you. Have you tried talking to your friends about it? **Gold:** Do you have any family?

Table 6: Two case studies indicate a mastery of optimal relevance after full OLR. Training helpful chatbots without punishing user processing load jeopardizes the comprehensibility and harmlessness of generated responses.