

Rational Sensibility: LLM Enhanced Empathetic Response Generation Guided by Self-presentation Theory

Anonymous ACL submission

Abstract

The development of Large Language Models (LLMs) provides human-centered Artificial General Intelligence (AGI) with a glimmer of hope. Empathy serves as a key emotional attribute of humanity, playing an irreplaceable role in human-centered AGI. Despite numerous researches aim to improve the cognitive empathy of models by incorporating external knowledge, there has been limited attention on the sensibility and rationality of the conversation itself, which are vital components of the empathy. However, the rationality information within the conversation is restricted, and previous methods of extending knowledge are subject to semantic conflict and single-role view. In this paper, we design an innovative encoder module inspired by self-presentation theory in sociology, which specifically processes sensibility and rationality sentences in dialogues. And we employ a LLM as a rational brain to decipher profound logical information preserved within the conversation, which assists our model in assessing the balance between sensibility and rationality to produce high-quality empathetic response. Experimental results demonstrate that our model outperforms other methods in both automatic and human evaluations.

1 Introduction

Empathetic response generation, the capacity to perceive emotion of individuals and response accordingly, is integral in the pursuit of intelligent agents (Hofmann et al., 2010). In psychology, sensibility and rationality are crucial components of an individual’s empathy. Lack of sensibility makes it ineffectual to emotionally relate to users. And conversely, the absence of rationality may result in emotional empathy and the symptom of unmitigated communion (Fritz and Helgeson, 1998). However, rational sensibility, known as cognitive empathy, allows for a better comprehension of users while decreasing one’s own negative emotional experience (Smith, 2006).

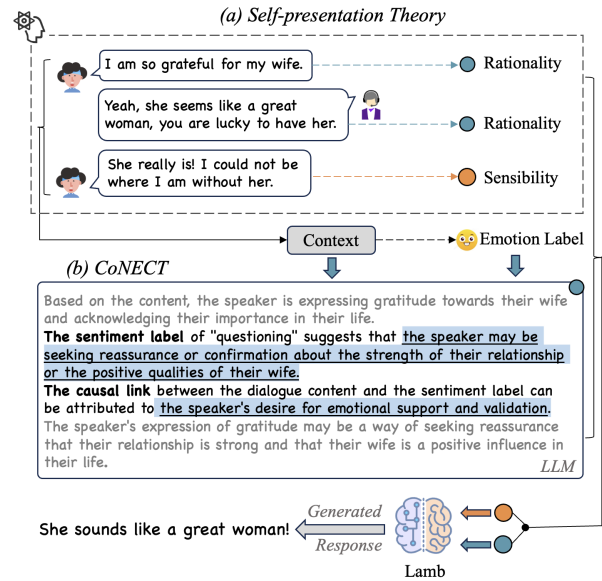


Figure 1: An example of empathetic response from EMPATHETICDIALOGUES dataset. (a) According to the sociological theory of self-presentation, categorizing conversation utterances into sensibility and rationality groups. (b) the chain of emotion-aware empathetic prompting based on LLM.

Although many researchers endeavor to enhance the cognitive ability of models through external knowledge (Ghosal et al., 2020; Zhou et al., 2021; Sabour et al., 2022), limited attention has been directed towards the inherent expression within the conversation, such as sensibility and rationality. Besides, the rational thinking of the dialog itself provides limited information and cannot furnish in-depth perception, such as intention and purpose. Previous methods usually adopt COMET (Bosselut et al., 2019) to enhance the commonsense reasoning ability of models, which is constructed on two large knowledge graphs, ATOMIC (Sap et al., 2019) and ConceptNet (Speer et al., 2017), including a huge amount of social commonsense knowledge. However, this method exists two limitations: semantic conflict and single-role view. Common-

sense knowledge can be divided into five categories: intent, need, effect, want, and react. These categories are relatively independent and difficult to ensure semantic consistency, which is the main reason for semantic conflict(Cai et al., 2023). Furthermore, previous methods are limited to focusing on a single-turn utterance to expand external knowledge with a single-role view, e.g., speaker or listener. To tackle the above issues, we seek solutions based on the intersection of sociology and the large language model (LLM).

In sociology, self-presentation theory(Baumeister and Hutton, 1987; Jensen Schau and Gilly, 2003) divides sentences into categories: sensibility and rationality. In sensibility, the emotionally charged utterances are used to accentuate the individual’s personality traits, depict personal state and project desired character. In rationality, sentences primarily serve the purpose of transmitting information, expressing opinions, and engaging in logical thought processes in an objective manner. As illustrated in Figure 1(a), the dialogue is divided into sensibility and rationality sentences with the aim to achieve more accurate context encoding.

Furthermore, we employ the LLM as a rational brain to provide logical reasoning for empathetic response without limitation of semantic conflict and single-role view. Particularly, we design an innovative module, rational representation enhancement, which are mainly composed by Chain of Emotion-aware Empathetic prompting(CoNECT). As shown in Figure 1(b), CoNECT employs an affective indicator to facilitate the evaluation and correlation of contextual connections, based on the extensive knowledge within LLM. This approach enables empathetic reasoning and boasts notable advantages: 1) **Rational reasoning.** Examine the portrayal of emotional labels within historical context and analyze the intrinsic psychological expression. 2) **Multi-role view.** Offer a comprehensive perspective when perceiving the emotions and intentions from the view of speaker and listener on a global scale.

In this paper, we present LLM enhanced empathetic response generation model guided by self-presentation theory, **Lamb**. The model consists of three modules: 1) Rational-sensible Encoder. Emotionally charged sentences are identified by the pretrained emotion-cause model. Then, we utilize a designed attention mechanism to combine the sensible and rational sentences. 2) Ra-

tional Representation Enhancement. In the context of a conversation and the corresponding emotion label, we employ the CoNECT and COMET methodologies to generate rational knowledge for Lamb. 3) Rational-sensible Decoder. We employ a cross-attention mechanism that allows the decoder to perceive context, commonsense knowledge and CoNECT data.

Our main contributions are listed as follows:

- Guided by self-presentation theory, we focus on the sensible and rational expression of the conversation itself to enhance cognitive empathy.
- We introduce the chain of emotion-aware empathetic data into the empathetic response generation task, which provides multi-role aware rational reasoning.
- Our model, Lamb, uses weighted context and external knowledge to understand speaker’s emotional state and generate suitable empathetic response.
- Experiments demonstrate that Lamb generates more empathetic response compared with the state-of-the-art methods.

2 Related Work

The objective of the empathetic response generation is to equip the model with the capability to deliver suitable emotional value(Hofmann et al., 2010). Numerous work has addressed this task in terms of perception enhancement and cognition enhancement.

2.1 Perception Enhancement Methods

Methods for perception enhancement typically analyze the sentiment words in conversations and distill fine-grained emotion control information(Lin et al., 2019; Ghosal et al., 2020; Li et al., 2020; Zhao et al., 2023). For example, multi-granularity sentiment labels of conversation and utterance in data can help the model capture user sentiment efficiently(Wang et al., 2022a). E-CORE(Fu et al., 2023) focus on exploring intrinsic sentiment by emotion correlation learning, utilization, and supervising. Besides, CAB(Gao et al., 2023) split the empathy response generation into three parts: cognition, affection and behavior. Additionally, ESCM(Yang et al., 2023) uses dynamic emotion-semantic vectors and dependency trees to guide

the model generate empathetic responses. Furthermore, Yufeng et al. (2024) categorizing emotions into fine-grained trait and state emotion to control sentiment representation more accurately. Nevertheless, despite the potential for fine-grained emotional control data to reveal perspective-specific emotion states, the sensible representation of dialogue at the utterance level remains underutilised.

Based on the self-presentation theory in psychology, we divide the conversation into sentences with two types of attributes: sensible and rational, and explore their importance for empathetic response(Fritz and Helgeson, 1998; Smith, 2006).

2.2 Cognition Enhancement Methods

Cognitive enhancement can improve the model’s comprehension and learning capacity.(Tahir et al., 2023; Zhou et al., 2023). For example, CEM (Sabour et al., 2022) introduces commonsense knowledge inference to this task for the first time through the pre-trained COMET model. The scope of cognition enhancement methods is further expanded with the help of ambiguous information filtering(Cai et al., 2023) and graph neural network structure(Li et al., 2022). However, these approaches are constrained by limitations of semantic contraction and single-role view(Cai et al., 2023).

Chain-of-Thought Prompting(CoT) are typically used to facilitate the logical resolution of complex problems, elucidating the rationale behind the decision-making process(Wei et al., 2022; Wang et al., 2022b, 2023). The chain-of-empathy method employs CoT technology to enhance the cognitive capacity of the empathetic response model, though it does not examine the logical interrelationship between emotion and conversation history(Lee et al., 2023). In our work, we utilize the chain of emotion-aware empathetic prompting method to enhance the cognitive ability of the model.

3 Method

As shown in Figure 2, Lamb is consists of three models: 1) rational-sensible encoder, 2) rational representation enhancement, 3) rational-sensible decoder.

3.1 Task Formulation

Given a dialogue history $U = [S_1, L_1, S_2, L_2, \dots, L_{N-1}, S_N]$ of $2N-1$ utterances, the goal of the empathetic response

generation is to predict the overall sentiment label e_{tar} and generate empathetic response $Y = [y_1, y_2, \dots, y_{m_y}]$. In dialogue history U , $S_i = [s_1^i, s_2^i, \dots, s_{m_{s_i}}^i]$ and $L_i = [l_1^i, l_2^i, \dots, l_{m_{l_i}}^i]$ represent the i -th utterance of speaker and listener, composed by m_{s_i} and m_{l_i} tokens respectively.

3.2 Rational-sensible Encoder

The first group encompasses sentences that used to underscore individual characteristics by sensible expression. On the other hand, the second group is composed of sentences that primarily emphasize the rational reasoning. When selecting the first group sentences, our purpose is to align the global sensible experiences of both the speaker and listener, which enables the model to better understand the speaker’s emotion and respond more empathy. Therefore, we adopt the pretrained model SEEK(Wang et al., 2022a) to prelabel comprehensive sentiment for each conversation, inspired by Welivita and Pu (2020). Then, we use a pretrained emotion-cause model RECCON(Poria et al., 2021) to detect relevant sensible expression in context.

As described in Figure 2, we set U as input of the SEEK model, and obtain predicted global sentiment label e_{ano} accordingly. Next, we use e_{ano} as the target sentiment of the model RECCON, and look for cues that evoke the sentiment in U . The final output is the sensible sentences D that implies the expression of the sentiment, which is a subset of U .

$$e_{ano} = SEEK(U) \quad (1)$$

$$D = RECCON(e_{ano}, U) \quad (2)$$

We feed U and D into the Bart encoder to obtain the context representation $R_U \in \mathfrak{R}^{l_U \times d}$ and the cause representation $R_D \in \mathfrak{R}^{l_D \times d}$.

$$R_U = Bart_{enc}(S_1 \oplus \langle \backslash s \rangle \oplus L_1 \oplus \langle \backslash s \rangle \dots S_N) \quad (3)$$

$$R_D = Bart_{enc}(D) \quad (4)$$

Additionally, in order to assist the model in discerning between these two categories of sentences, we incorporate the method of joint modeling through the designed attention mechanism to acquire distinct weights for them, and output finally representation $R_G \in \mathfrak{R}^{l_G \times d}$.

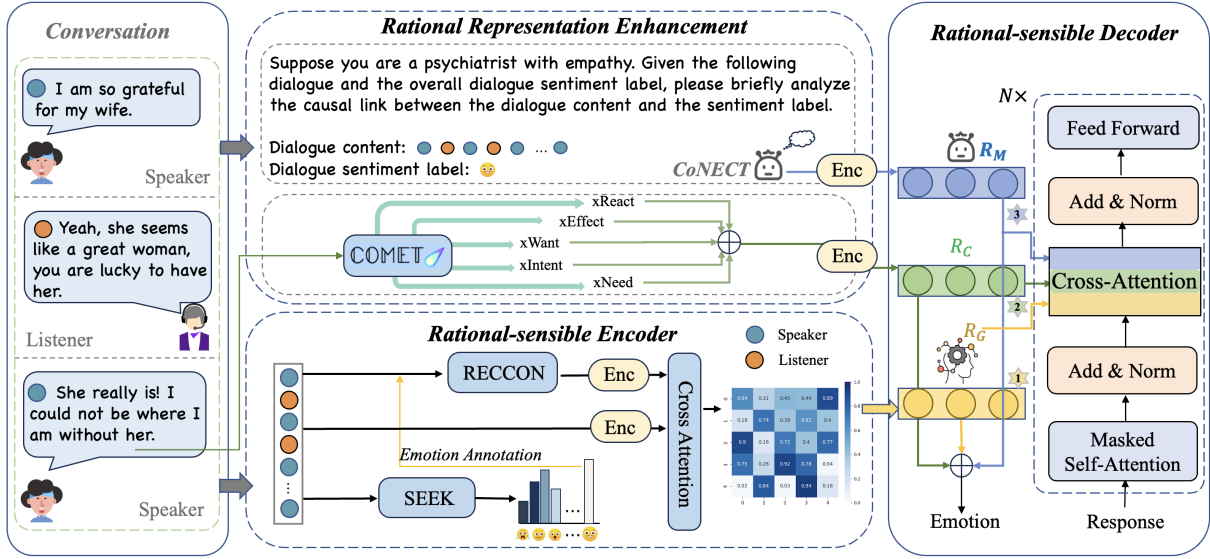


Figure 2: The model architecture of Lamb, which consists of three parts: (a) Rational-sensible Encoder based on self-presentation theory, (b) Rational Representation Enhancement by CoNECT and COMET, (c) Rational-sensible Decoder.

$$R_G = \text{softmax}\left(\frac{(W_q R_U) \cdot (W_k R_D)^T}{\sqrt{2d}}\right) \cdot (W_v R_D) \quad (5)$$

, where $W_q \in \mathbb{R}^{d \times d}$, $W_k \in \mathbb{R}^{d \times d}$ and $W_v \in \mathbb{R}^{d \times d}$ are randomly initialised tensor matrix. l_U , l_D and l_G is the number of corresponding tokens. d is the dimension of feature vector.

3.3 Rational Representation Enhancement

In this module, we dedicate to enhancing rational information by CoNECT and COMET's commonsense knowledge.

3.3.1 COMET

Following previous works, we adopt the pre-trained model COMET to generate commonsense knowledge. Given the last utterance S_N of the historical context U , we alternately select x_{Intent} , x_{Effect} , x_{Want} , x_{React} and x_{Need} cognitive attributes for inferencing, obtaining C_r commonsense knowledge where $r \in \{x_{Intent}, x_{Effect}, x_{Want}, x_{React}, x_{Need}\}$:

$$C_r = \text{COMET}(S_N) \quad (6)$$

After that, we append the $[CLS]$ token to the first position for each COMET relation. Then we fed them into the encoder getting feature matrix $R_C \in \mathbb{R}^{l_C \times d}$, where l_C is the number of tokens in set r and special tokens.

$$R_C = \text{Bart}_{enc-r}([CLS] \oplus C_r) \quad (7)$$

As a result, we embed the commonsense knowledge in the decoder, so as to provide more effective complementary information to the model.

3.3.2 CoNECT

As shown in Figure 2, the CoT prompt we used can be divided into three parts:

Character. We assign LLM to role-play a compassionate psychologist who is expected to possess theoretical knowledge and behavioral paradigms in professional domain.

Causal Chain. In order to improve contextual coherence and achieve more precise comprehension of dialogues, we extend the scope of external knowledge sources by incorporating the complete context into the prompt.

Global Sentiment Label. In conjunction with the causal chain, the method offers guidance for the logical behavior of the analytical model.

In practice, we use the prompt C_P as the input of LLM, and the output of the LLM is the CoNECT data C_M .

$$C_P = \text{Prompt}_{template}(U, e_{ano}) \quad (8)$$

$$C_M = \text{LLM}(C_P) \quad (9)$$

Afterwards, we pass C_M through the encoder to acquire unified feature representation $R_M \in \mathbb{R}^{c \times d}$,

where c is the length of CoNECT data.

$$R_M = \text{Bart}_{enc}(C_M) \quad (10)$$

3.4 Rational-sensible Decoder

The rational-sensible decoder facilitates the model’s ability to balance sensible and rational cognition, thereby generating empathetic responses that are more consistent with the listener’s role.

To begin with, the model perceives the sensible and rational cognitive aspects of the context with R_G , which offers essential background information. Subsequently, the model extends the basic commonsense knowledge with the assistance of the R_C . Lastly, the model incorporates information from the R_M to enrich and intensify the rational cognitive expression.

The target response $Y = [y_1, y_2, \dots, y_{m_y}]$ with length m_y , which is generated token by token by Bart decoder based above representation R_G , R_C and R_M . During training, we adopt the standard negative log-likelihood(NLL) loss on the target response Y :

$$\mathcal{L}_{nll} = - \sum_{t=1}^{m_y} \log(y_t | (R_G, R_C, R_M), y_{<t}) \quad (11)$$

3.5 Training Objectives

The objective of our training program is composed of two parts: emotion classification and empathetic response generation.

3.5.1 Emotion Classification

For emotion classification task, we utilize average pooling to get the knowledge vector $p_k \in \mathbb{R}^d$:

$$p_k = \text{Average - pooling}(R_C) \quad (12)$$

Then, in order to optimize knowledge acquisition, we strive for the fusion of historical conversations, empathetic thought chains, and common sense knowledge.

$$R_F = R_G[0] \oplus R_M[0] \oplus p_k, \quad (13)$$

$$P_{e_{pre}} = \text{softmax}(W_\theta(R_F)) \quad (14)$$

After that, we subsequently pass R_F through a linear operation $W_\theta \in \mathbb{R}^{d \times q}$, followed by a softmax calculation to produce the distribution $P_{e_{pre}} \in \mathbb{R}^q$, where q is the number of emotion categories:

In training phase, we optimize the object by minimizing the Cross-Entropy loss between the emotion category and the ground truth label e_{tar} :

$$\mathcal{L}_{emo} = -\log(P_{e_{pre}}(e_{tar})) \quad (15)$$

3.5.2 Final Training Objective

In the training phase, all the parameters of our model are optimized with \mathcal{L}_{emo} and \mathcal{L}_{nll} :

$$\mathcal{L} = \mathcal{L}_{nll} + \mathcal{L}_{emo} \quad (16)$$

4 Experiments

4.1 Datasets

Our experiments are conducted on the widely used EMPATHETICDIALOGUES(Rashkin et al., 2019), comprising 25k multi-turn empathetic conversations between a speaker and a listener, with an average of 4.31 turn per dialog. This dataset provides 32 evenly distributed emotion labels, and each conversation is assigned a related label.

To ensure the fairness of the comparison experiments, we use the same dataset division as the previous research method, dividing the dataset into training, validation and test sets in an 8:1:1 ratio(Cai et al., 2023).

4.2 Evaluation Metrics

In order to confirm the effectiveness of Lamb, we implement a dual evaluation strategy, a comprehensive evaluation from automation and human perspectives. The detailed experiment setting is described in Appendix E. And the compared baseline models are listed in Appendix B.

4.2.1 Automatic Evaluation

Following previous work, we choose Perplexity(PPL), corpus-level BLEU(B-N), sentence-level ROUGE(R-N), Distinct-n(Dist-N) and Accuracy(Acc) as our main automatic metrics. 1) Perplexity is used to assess the fluency and comprehensibility of the text generated by the model, and a lower PPL value means that the text is more natural. 2) The BLEU and ROUGE score indicate the degree of similarity between the generated text and the ground-truth text. A higher score indicates a greater degree of similarity. 3) Dist-N is used to assess the diversity of the content, with higher metric values indicating that the model can produce more diverse and richer representation. 4) Acc is used to measure the accuracy of the model for emotion classification.

Table 1: Results of automatic evaluation. Optimal outcomes are highlighted in bold, and suboptimal outcomes are indicated by underlining. The evaluation of the LLM utilizes a meticulously crafted prompt for inference.

Models	PPL	B-1	B-2	B-3	B-4	R-1	R-2	Dist-1	Dist-2	Acc
MoEL(Lin et al., 2019)	36.60	18.07	8.30	4.37	2.65	18.24	4.81	0.59	2.64	31.74
MIME(Ghosal et al., 2020)	37.24	18.60	8.39	4.54	2.81	17.08	4.05	0.47	1.66	30.96
EmpDG(Li et al., 2020)	37.43	19.96	9.11	4.74	2.80	18.02	4.43	0.46	1.99	31.65
CEM(Sabour et al., 2022)	36.33	16.12	7.29	4.06	2.03	15.77	4.50	0.62	2.39	36.84
SEEK(Wang et al., 2022a)	36.78	10.77	4.40	2.02	1.08	12.74	2.94	0.68	2.81	42.74
CASE(Zhou et al., 2023)	35.20	15.59	7.22	3.80	2.24	17.33	4.67	0.65	3.37	38.99
E-CORE(Fu et al., 2023)	33.03	-	-	-	-	-	-	0.72	3.49	42.59
KEMP(Li et al., 2022)	36.39	16.72	7.17	3.77	2.33	16.11	3.31	0.66	3.07	36.57
CAB(Gao et al., 2023)	34.36	19.23	8.55	4.36	2.57	17.50	4.13	1.13	4.23	40.52
ESCM(Yang et al., 2023)	34.82	-	-	-	-	-	-	1.19	4.11	41.19
DCKS(Cai et al., 2023)	18.58	18.75	9.12	5.38	3.57	19.14	5.45	1.57	6.02	48.69
CTSM(Yufeng et al., 2024)	34.56	-	-	-	-	-	-	2.00	7.34	43.41
Qwen1.5-MoE-A2.7B-Chat(Team, 2024)	-	7.73	2.39	1.03	0.52	10.45	1.12	2.76	22.41	-
LLaMA2-13B-Instruct(Touvron et al., 2023)	-	11.69	4.03	1.79	0.93	13.27	1.83	2.91	18.92	-
LLaMA3-8B-Instruct(Touvron et al., 2023)	-	13.17	4.42	1.92	1.02	14.12	1.68	2.69	18.70	-
Qwen1.5-72B-Chat(Team, 2024)	-	14.19	4.85	2.27	1.23	13.83	1.97	3.29	22.68	-
Mixtral-8x7B-Chat(Jiang et al., 2024)	-	14.66	4.72	2.11	1.10	14.54	1.69	3.30	21.13	-
Lamb(CoNECT based on LLaMA2-13B-Instruct)	<u>20.14</u>	22.16	10.55	6.01	3.78	19.39	5.47	2.38	10.32	51.36
+ Compared with LM	-	▲ 2.93	▲ 1.43	▲ 0.63	▲ 0.21	▲ 0.25	▲ 0.02	▲ 0.38	▲ 2.98	▲ 2.67
+ Compared with LLM	-	▲ 7.50	▲ 5.83	▲ 3.74	▲ 2.55	▲ 4.85	▲ 3.50	-	-	-
& CoNECT based on LLaMA3-8B-Instruct	<u>20.22</u>	22.56	11.03	6.33	3.82	19.73	5.68	2.41	10.41	52.16
+ Compared with LM	-	▲ 3.32	▲ 1.91	▲ 0.95	▲ 0.25	▲ 0.59	▲ 0.23	▲ 0.41	▲ 3.07	▲ 3.47
+ Compared with LLM	-	▲ 7.87	▲ 6.31	▲ 4.06	▲ 2.59	▲ 5.19	▲ 3.71	-	-	-
& CoNECT based on Qwen1.5-72B-Chat	<u>20.15</u>	22.27	10.88	6.32	3.60	19.54	5.71	2.29	10.33	51.69
+ Compared with LM	-	▲ 3.04	▲ 1.76	▲ 0.94	▲ 0.03	▲ 0.40	▲ 0.26	▲ 0.29	▲ 2.99	▲ 3.0
+ Compared with LLM	-	▲ 7.61	▲ 6.16	▲ 4.05	▲ 2.37	▲ 5.0	▲ 3.74	-	-	-
& CoNECT based on Mixtral-8x7B-Chat	<u>20.18</u>	22.24	10.87	6.35	3.79	19.51	5.73	2.37	10.35	51.75
+ Compared with LM	-	▲ 3.01	▲ 1.75	▲ 0.97	▲ 0.22	▲ 0.37	▲ 0.28	▲ 0.37	▲ 3.01	▲ 3.06
+ Compared with LLM	-	▲ 7.58	▲ 6.15	▲ 4.08	▲ 2.56	▲ 4.97	▲ 3.76	-	-	-

4.2.2 Human Evaluation

To further verify the effectiveness and reliability of our method, we conduct manual evaluation work. Specially, we set four evaluation dimensions, and the score of each dimension is quantitatively evaluated from 1 to 5. 1) Coherence(**Coh.**): Measures the level of relevance between the text generated by the model and the gold response. 2) Empathy(**Emp.**): Demonstrates a greater comprehension of the speaker’s circumstances and conveys a more fitting sentiment. 3) Informativeness(**Inf.**): Measures the richness of information contained in generated responses. 4) Continuity(**Cont.**): Indicates the intensity of the speaker’s desire to continue the conversation. Particularly, we randomly select 1000 sets of test data and randomly shuffled their order to ensure the fairness and objectivity of the evaluation. Then we distribute the data to three qualified and experienced human evaluators, whom score the model’s responses and get a average score. In this way, we can ensure the accuracy and scientificity of the assessment results, which in turn strongly supports the effectiveness of our methods. The detailed basis of scoring is shown in the Appendix D.

4.3 Automatic Evaluation Results

Our CoNECT module conduct several sets of experiments based on different LLMs: LLaMA2-13B-Instruct, LLaMA3-8B-Instruct, Qwen1.5-72B-Chat and Mixtral-8x7B-Chat. Compared with LM: MoEL, MIME, EmpDG, CEM, SEEK, CASE, E-CORE, KEMP, CAB, ESCM, DCKS, and CTSM, Lamb exceed the baseline models on most metrics, proving the robustness and effectiveness of our method. We observe that all four variants of Lamb achieve significant performance, which indicates that the improvement comes from the architecture of Lamb rather than relying on a specific LLM.

The enhancement in both Dist-1 and Dist-2 indicates a greater diversity in our model responses, thereby mitigating the appearance of hollow response. Furthermore, achieving the highest scores on both BLEU-N and ROUGE-N metrics reflects better consistency between model responses and historical conversations while also maintaining diversity. Additionally, our model exhibits a notable increase of 2.67 points in emotion classification accuracy, showcasing its capability to accurately capture speaker sentiment. Despite a slight decrease

Table 2: Ablation study

Models	PPL ↓	B-1 ↑	B-2 ↑	B-3 ↑	B-4 ↑	R-1 ↑	R-2 ↑	Dist-1 ↑	Dist-2 ↑	Acc ↑
Vanilla	18.58	18.75	9.12	5.38	3.57	19.14	5.45	1.57	6.02	48.69
Vanilla + Self-pres	17.74	21.02	10.17	5.84	3.74	19.88	5.56	2.01	8.52	50.65
Vanilla + CoNECT	20.30	19.95	9.79	5.72	3.69	19.62	5.69	2.25	9.96	51.30
Lamb	20.14	22.16	10.55	6.01	3.78	19.39	5.22	2.38	10.32	51.36

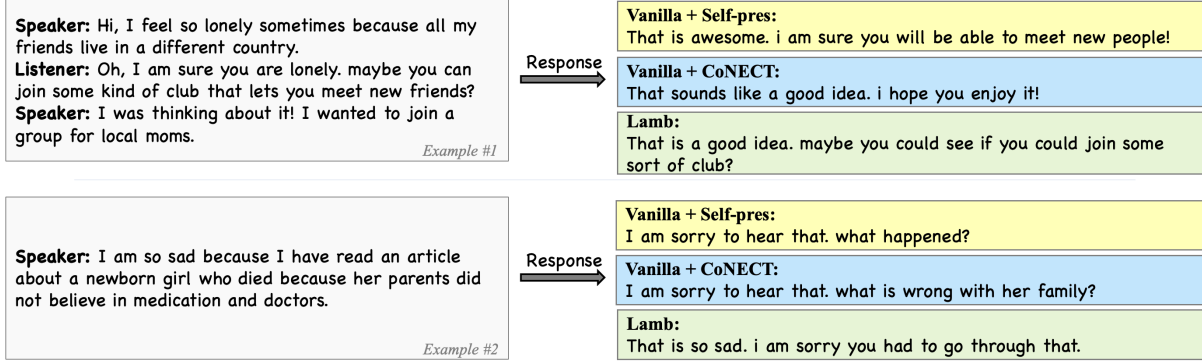


Figure 3: Effect of Self-presentation and CoNECT module.

435 in our PPL due to the complex structure of the
 436 Lamb, the overall improvement in other evaluation
 437 metrics reaffirms the effectiveness of our approach.
 438 We will discuss the results of the comparison with
 439 LLM in Section 4.6.

440 4.4 Ablation Studies

441 In order to prove the validity of the CoNECT and
 442 self-presentation module, we conduct two ablation
 443 experiments. The following is an introduction to
 444 the comparison models used in the experiment:

445 1) **Vanilla:** Removing both Self-presentation
 446 module and CoNECT module.

447 2) **Vanilla + Self-pres:** Based on vanilla, the
 448 self-presentation module is added.

449 3) **Vanilla + CoNECT:** Based on vanilla, the
 450 CoNECT module is added.

451 The Table 2 demonstrates that removing either
 452 the Self-presentation module or the CoNECT mod-
 453 ule result in a significant decrease in most eval-
 454 uation metrics, indicating the effectiveness of both
 455 modules. When comparing the CoNECT and Self-
 456 presentation module, the approach with CoNECT
 457 excels in diversity metrics but underperforms in
 458 BLUE and ROUGE metrics. This observation sug-
 459 gests that the CoNECT module enhances the model
 460 by providing more contextual information with ex-
 461 ternal knowledge. The Self-pres module, on the
 462 other hand, directs the model to focus more on the
 463 dialog content to improve semantic understanding.

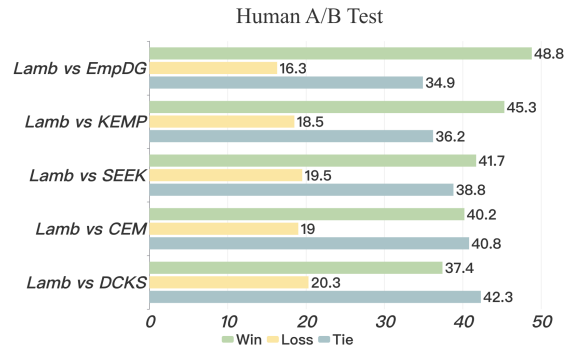


Figure 4: Results of Human A/B test.

464 The synergistic effect of these two modules boosts
 465 the model’s performance in empathetic response
 466 generation.

467 To further validate the effectiveness of our pro-
 468 posed self-pres and CoNECT modules, we provide
 469 a detailed analysis with two sets of case ablation
 470 experiments, as shown in Figure 3.

471 **Self-presentation Module.** In example-1, the
 472 model integrates both perceptual and rational cog-
 473 nition by conducting a detailed analysis of the
 474 conversation context to better understand and re-
 475 spond to the user’s expressed emotions of "long-
 476 ing for friends". This approach not only strength-
 477 ens the emotional bond between the model and
 478 the user but also proves the effectiveness of the
 479 self-presentation theory in real-world scenarios.
 480 In example-2, the model effectively utilizes self-

presentation module to express empathy, and leads speaker to express their deeper feeling of sadness by asking questions. This strategy not only demonstrates the model’s role as a superb listener but also emphasizes its human-like qualities.

CoNECT. In example-1, CoNECT recognizes that users were joining social groups out of loneliness. The model responds positively to the decision, expressing optimism about its potential to enhance user satisfaction. These responses not only address the emotional needs of users but also demonstrate CoNECT’s expertise in handling complex emotional issues. In example-2, CoNECT analyzes parents’ distrustful attitude towards medication and doctor’s advice, clearly pointing out that this is a key factor leading to the unfortunate death of a newborn. As a result, the model responds by questioning family members to provide emotional support to the user.

Table 3: Results of human evaluation.

Models	Coh.	Emp.	Inf.	Cont.
EmpDG	3.22	3.10	2.99	3.07
CEM	3.41	3.49	3.12	3.18
SEEK	3.40	3.62	3.19	3.33
KEMP	3.56	3.66	3.35	3.52
DCKS	4.10	3.94	3.76	3.87
Lamb	4.31	4.21	4.78	4.32
+ <i>Improvement</i>	+ 0.21	+ 0.27	+ 1.02	+ 0.45

4.5 Human Evaluation Results

As illustrated in Table 3, Our model improves on all human assessment indicators. The incorporation of CoNECT enhances the model’s capacity for empathizing, along with an increase in the volume of information conveyed in the response. Furthermore, reinforced by self-presentation theory, the model can focus on the sensibility clues in historical content, improving the contextual coherence of the generated response and thus motivating the speaker’s willingness to continue the conversation. Additionally, the pairwise response comparison results are shown in Figure 4. The results further confirm that the responses generated by Lamb are more preferred by human judges. Please refer to the Appendix A for a more detailed case study.

4.6 Comparing with LLM

Although we employ LLaMA2-13B-Chat inference data as an aid during our study, our gener-

You are a professional psychologist. Please give a reply from the perspective of a friend based on the patient's situation. Be careful to avoid a large number of professional vocabulary in your reply, and the expression should be natural.

The following is a set of historical conversations between you and your best friend. Please respond according to the context and pay attention to fully consider the emotional information in the historical conversations. Remember, response should not exceed 30 words.


[conversation] 

Figure 5: Prompt used for evaluating LLM empathetic ability.

Table 4: Compared with closed source LLMs.

Models	BLEU-2	BLEU-4
ChatGPT(+ 0-short)	6.19	1.86
ChatGPT(+ 1-short)	6.79	2.12
ChatGPT(+ 5-short)	7.85	2.65
GPT-3(+ 0-short)	6.88	2.22
GPT-3(+ 1-short)	6.71	2.16
GPT-3(+ 5-short)	8.51	2.83
GPT-3.5(+ 0-short)	8.51	2.80
GPT-3.5(+ 1-short)	5.62	1.99
GPT-3.5(+ 5-short)	9.37	3.26
GPT-4(+ 0-short)	9.83	3.25
Lamb	10.55	3.78

ative model is still based on BART, a much less parameters model than LLM. Considering the large amount of data processed during LLM training and the sampling strategy used, there is a significant advantage over small models, thus, we do not compare perplexity and diversity metrics with them.

The compared open-source LLMs utilize a meticulously crafted system prompt in order to optimising its capacity of empathetic response, as shown in Figure 5. In the bottom of Table 1, our model is best in all BLUE and ROUGE metrics. Furthermore, our model outperforms compared closed-source LLM in BLUE-2 and BLUE-4 in Table 4. For fairness, we maintain the setting of GPT-series LLM are consistent of (Qian et al., 2023).

5 Conclusions

In this research, we enhance the generation of empathetic responses through the incorporation of self-presentation module and CoNECT. The results of both automated and human evaluations indicate the superior quality of responses generated by Lamb.

In the future, we will strive to strengthen the integration of multidisciplinary knowledge and further delve the significance of sensibility and rationality in cognitive processes.

544 **6 Limitations**

545 There are two main limitations in our work. First
546 of all, we employed COMET commonsense knowl-
547 edge in the rational representation enhancement
548 module, however, we did not take a strong approach
549 to mitigate the problem of semantic conflicts be-
550 tween the information it generates. Secondly, auto-
551 mated evaluation is difficult to score the empathy
552 of generated responses, existing a certain degree of
553 limitation.

554 **7 Ethics Considerations**

555 We obtain our data from EMPATHETICDIA-
556 LOGUES, a publicly available dataset that does
557 not include any personal information. Our human
558 assessments are carried out by three experienced
559 annotators and they are fairly compensated for their
560 work.

561
562
563
564
565

566
567
568
569

570
571
572
573
574
575

576
577
578
579
580
581
582

583
584
585
586

587
588
589
590
591

592
593
594
595
596
597

598
599
600
601
602
603

604
605
606
607

608
609
610

611
612
613
614
615
616

References

Roy F. Baumeister and Debra G. Hutton. 1987. [Self-Presentation Theory: Self-Construction and Audience Pleasing](#), pages 71–87. Springer New York.

Steven Bird, Ewan Klein, and Edward Loper. 2009. Analyzing text with the natural language toolkit. [Natural Language Processing with Python](#), page 504.

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. [COMET: Commonsense transformers for automatic knowledge graph construction](#). In [Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics](#), pages 4762–4779.

Hua Cai, Xuli Shen, Qing Xu, Weilin Shen, Xi-aomei Wang, Weifeng Ge, Xiaoqing Zheng, and Xiangyang Xue. 2023. [Improving empathetic dialogue generation by dynamically infusing commonsense knowledge](#). In [Findings of the Association for Computational Linguistics: ACL 2023](#), pages 7858–7873.

Heidi L Fritz and Vicki S Helgeson. 1998. Distinctions of unmitigated communion from communion: self-neglect and overinvolvement with others. [Journal of Personality and Social Psychology](#), 75:121.

Fengyi Fu, Lei Zhang, Quan Wang, and Zhendong Mao. 2023. E-core: Emotion correlation enhanced empathetic dialogue generation. In [Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing](#), pages 10568–10586.

Pan Gao, Donghong Han, Rui Zhou, Xuejiao Zhang, and Zikun Wang. 2023. [Cab: Empathetic dialogue generation with cognition, affection and behavior](#). In [Database Systems for Advanced Applications: 28th International Conference, DASFAA 2023](#), page 597–606.

Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. [COSMIC: Commonsense knowledge for eMotion identification in conversations](#). In [Findings of the Association for Computational Linguistics: EMNLP 2020](#), pages 2470–2481.

Stefan G Hofmann, Alice T Sawyer, and Angela Fang. 2010. The empirical status of the “new wave” of cognitive behavioral therapy. [Psychiatric Clinics](#), 33(3):701–710.

Hope Jensen Schau and Mary C Gilly. 2003. We are what we post? self-presentation in personal web space. [Journal of consumer research](#), 30:385–404.

Albert Q Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, et al. 2024. [Mixtral of experts](#). [arXiv preprint arXiv:2401.04088](#).

Yoon Kyung Lee, Inju Lee, Minjung Shin, Seoyeon Bae, and Sowon Hahn. 2023. Chain of empathy: Enhancing empathetic response of large language models based on psychotherapy models. [arXiv preprint arXiv:2311.04915](#).

Qintong Li, Hongshen Chen, Zhaochun Ren, Pengjie Ren, Zhaopeng Tu, and Zhumin Chen. 2020. [EmpDG: Multi-resolution interactive empathetic dialogue generation](#). In [Proceedings of the 28th International Conference on Computational Linguistics](#), pages 4454–4466.

Qintong Li, Piji Li, Zhaochun Ren, Pengjie Ren, and Zhumin Chen. 2022. Knowledge bridging for empathetic dialogue generation. In [Proceedings of the AAAI Conference on Artificial Intelligence](#), volume 36, pages 10993–11001.

Zhaojiang Lin, Andrea Madotto, Jamin Shin, Peng Xu, and Pascale Fung. 2019. [MoEL: Mixture of empathetic listeners](#). In [Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing \(EMNLP-IJCNLP\)](#), pages 121–132.

Soujanya Poria, Navonil Majumder, Devamanyu Hazarika, Deepanway Ghosal, Rishabh Bhardwaj, Samson Yu Bai Jian, Pengfei Hong, Romila Ghosh, Abhinaba Roy, Niyati Chhaya, et al. 2021. Recognizing emotion cause in conversations. [Cognitive Computation](#), 13:1317–1332.

Yushan Qian, Weinan Zhang, and Ting Liu. 2023. Harnessing the power of large language models for empathetic response generation: Empirical investigations and improvements. In [Findings of the Association for Computational Linguistics: EMNLP 2023](#), pages 6516–6528.

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.

Sahand Sabour, Chujie Zheng, and Minlie Huang. 2022. Cem: Commonsense-aware empathetic response generation. In [Proceedings of the AAAI Conference on Artificial Intelligence](#), volume 36, pages 11229–11237.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019. Atomic: An atlas of machine commonsense for if-then reasoning. In [Proceedings of the AAAI conference on artificial intelligence](#), volume 33, pages 3027–3035.

Adam Smith. 2006. Cognitive empathy and emotional empathy in human behavior and evolution. [The Psychological Record](#), 56:3–21.

673	Robyn Speer, Joshua Chin, and Catherine Havasi. 2017.	Weixiang Zhao, Yanyan Zhao, Xin Lu, and Bing Qin.	729
674	Conceptnet 5.5: An open multilingual graph of	2023. Don't lose yourself! empathetic response gen-	730
675	general knowledge. In Proceedings of the AAAI	eration via explicit self-other awareness . In Findings	731
676	conference on artificial intelligence, volume 31, page	of the Association for Computational Linguistics:	732
677	4444–4451.	ACL 2023 , pages 13331–13344.	733
678	Sharjeel Tahir, Syed Afaq Shah, and Jumana Abu-	Jinfeng Zhou, Chujie Zheng, Bo Wang, Zheng	734
679	Khalaf. 2023. Artificial empathy classification: A	Zhang, and Minlie Huang. 2023. CASE: Align-	735
680	survey of deep learning techniques, datasets, and	ing coarse-to-fine cognition and affection for em-	736
681	evaluation scales. arXiv preprint arXiv:2310.00010 .	pathetic response generation . In Proceedings of	737
682	Qwen Team. 2024. Introducing qwen1.5 .	the 61st Annual Meeting of the Association for	738
683	Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier	Computational Linguistics , pages 8223–8237.	739
684	Martinet, Marie-Anne Lachaux, Timothée Lacroix,	Pei Zhou, Pegah Jandaghi, Bill Yuchen Lin, Justin Cho,	740
685	Baptiste Rozière, Naman Goyal, Eric Hambro,	Jay Pujara, and Xiang Ren. 2021. Probing common-	741
686	Faisal Azhar, et al. 2023. Llama: Open and effi-	sense explanation in dialogue response generation.	742
687	cient foundation language models. arXiv preprint	In Findings of the Association for Computational	743
688	arXiv:2302.13971 .	Linguistics: EMNLP 2021 , pages 4132–4146.	744
689	Boshi Wang, Sewon Min, Xiang Deng, Jiaming Shen,		
690	You Wu, Luke Zettlemoyer, and Huan Sun. 2023.		
691	Towards understanding chain-of-thought prompting:		
692	An empirical study of what matters . In Proceedings		
693	of the 61st Annual Meeting of the Association		
694	for Computational Linguistics (Volume 1: Long		
695	Papers), pages 2717–2739.		
696	Lanrui Wang, Jiangnan Li, Zheng Lin, Fandong Meng,		
697	Chenxu Yang, Weiping Wang, and Jie Zhou. 2022a.		
698	Empathetic dialogue generation via sensitive emo-		
699	tion recognition and sensible knowledge selection .		
700	In Findings of the Association for Computational		
701	Linguistics: EMNLP 2022 , pages 4634–4645.		
702	Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc		
703	Le, Ed Chi, Sharan Narang, Aakanksha Chowdh-		
704	ery, and Denny Zhou. 2022b. Self-consistency im-		
705	proves chain of thought reasoning in language mod-		
706	els. arXiv preprint arXiv:2203.11171 .		
707	Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten		
708	Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou,		
709	et al. 2022. Chain-of-thought prompting elicits		
710	reasoning in large language models. Advances in		
711	Neural Information Processing Systems , 35:24824–		
712	24837.		
713	Anuradha Welivita and Pearl Pu. 2020. A taxonomy		
714	of empathetic response intents in human social con-		
715	versations . In Proceedings of the 28th International		
716	Conference on Computational Linguistics , pages		
717	4886–4899.		
718	Zhou Yang, Zhaochun Ren, Wang Yufeng, Xiaofei Zhu,		
719	Zhihao Chen, Tiecheng Cai, Wu Yunbing, Yisong		
720	Su, Siboj Ju, and Xiangwen Liao. 2023. Exploit-		
721	ing emotion-semantic correlations for empathetic re-		
722	sponse generation . In Findings of the Association		
723	for Computational Linguistics: EMNLP 2023 , pages		
724	4826–4837.		
725	Wang Yufeng, Chen Chao, Yang Zhou, Wang Shuhui,		
726	and Liao Xiangwen. 2024. Ctsm: Combining trait		
727	and state emotions for empathetic response model.		
728	arXiv preprint arXiv:2403.15516 .		

A Case Study

The responses generated by Lamb and other models are shown in Table 5. The response generated by KEMP is not relevant to the history of the dialog. Although SEEK and DCKS responses expressed congratulations to the speaker, they did not establish a connection to concrete things in the conversation and therefore contained relatively few semantic information.

In our method, CoNECT analyzes the speaker’s state of mind from a global perspective, and speculates that "the happiest thing in life" and "something you will never forget" are the key words in the conversation, which add color to the joy brought by the birth of the first child. The keywords emphasize the joy of the birth of the first child. By further integrating the self-presentation theory, Lamb recognized the speaker’s joyfulness while closely focusing on the key messages in the conversation, which effectively stimulated the speaker’s willingness to continue the communication. Table 6 and Table 7 are two more empathetic response case.

In comparison to responses from LLaMA2-13B-Chat that demonstrate strong emotion, our response has the following advantages:

(1) Lamb directly expresses sympathy for Speaker’s feelings. This simplicity avoids excessive explanation or unnecessary emotional burden that might be caused.

(2) In Table 6, Lamb echoes Speaker’s feelings about “finally meeting someone after waiting for a long time” without shifting to Listener’s own experience. This allows the Speaker to feel that their feelings have been responded to and valued.

(3) The information presented in Table 5 shows that LLaMA2-13B-Chat shares personal experiences. This can make the speaker feel like the other person is talking about themselves instead of paying attention to the speaker’s feelings. As a result, this may divert the conversation away from the original emotional expression of the speaker.

B Baseline Models Details

The following are the models we compared in the experiments. We use the official codes and follow the implementations.

1) MoEL(Lin et al., 2019): The model designs a corresponding decoder for each emotion to generate a response, and all the results are synthesised to produce the final result. 2) MIME(Ghosal et al., 2020): Depending on the positive or negative po-

larity of the emotion and the contextual content, the model is able to simulate the user’s emotions and enable the generation of empathetic responses by introducing randomness. 3) EmpDG(Li et al., 2020): A model that includes an empathetic information generator and a sentiment discriminator. The function of the generator is to ensure that the model has the ability to generate diverse content, while the function of the discriminator is to ensure that the information produced by the generator matches the empathetic sentiment in the context. 4) CEM(Sabour et al., 2022): The first model introduces the COMET pre-trained model to acquire common sense knowledge in empathetic response generation task, and uses both intent and cognitive types to categorise the knowledge. 5) SEEK(Wang et al., 2022a): A model focuses on sentence-level sentiment information and utilizes attention mechanisms to provide the model with multi-granularity sentiment label information. 6) CASE(Zhou et al., 2023): A model employ external message, COEMT and ConceptNet, to enhance the ability of cognitive and emotion. 7) E-CORE(Fu et al., 2023): A method focuses on exploring intrinsic sentiment by emotion correlation learning, utilization, and supervising. 8) KEMP(Li et al., 2022): A model that uses ConceptNet and VRC-NED as external knowledge sources and performs contextual modelling through a graph neural network structure. 9) CAB(Gao et al., 2023): A model split the empathy response generation into three parts: cognition, affection and behavior. 10) ESCM(Yang et al., 2023): Using dynamic emotion-semantic vectors and dependency trees to guide the model generate empathetic responses. 11) DCKS(Cai et al., 2023): A model uses adaptive module for commonsense knowledge selection to ensure consistency between the model responses and the history context. 12) CTSM(Yufeng et al., 2024): A model that categorizes emotions into fine-grained trait and state emotion to enhance the ability of perceiving sentiment.

C Empathetic Response Derived from Different Prompts

To maximize the empathetic response capabilities of the model, we developed several iterations of prompts. In practice, given the same contextual information, we assessing the impact of various prompt based on GPT-4-0613. The detailed comparative results is shown in Table 8.

D Details of Human Evaluation

We apply human evaluation to assess the Coherence(**Coh.**), Empathy(**Emp.**), Informativeness(**Inf.**) and Continuity(**Cont.**) of responses from empathetic models.

For each evaluation indicator, we specify a scoring dimension from 1 to 5, as follows:

Rating 1: The generated response completely fails to meet the specific requirement.

Rating 2: The response generally meets the specific requirements, but has some shortcomings.

Rating 3: The response does meet the required requirements.

Rating 4: The response meets specific requirements while also making some additional contribution or merit.

Rating 5: The response exceeds expectations and provides excellent content or solutions while meeting specific requirements.

E Codes of Compared Method

We use PyTorch to implement our model. The encoder and decoder are from base version of BART as same as (Cai et al., 2023). The CoNECT data are derived from LLaMA2-13b-Chat model on NVIDIA-A100 GPU. Besides, we train the empathetic response model Lamb using Adam optimizer with initial learning rate 0.00005 in 5 epochs and the batch size is set to 16. All empathetic response experiments is conducted on NVIDIA-4090 GPU and Intel(R) Xeon(R) Gold 6338 CPU @ 2.00GHz. Qwen1.5-MoE-A2.7B-Chat, LLaMA2-13B-Instruct, LLaMA3-8B-Instruct and Mixtral-8x7B-Chat are constructed based on vLLM-v0.3.2 and sampling temperature is 0.8, top-p is 0.95. We use the official DCKS code and reproduced it in our experimental environment. See the Appendix E for detailed code links.

The existing official codes that are involved in this work include: DCKS¹, CAB², KEMP³, CASE⁴, SEEK⁵, CEM⁶. The evaluation metrics from Natural Language Toolkit(Bird et al., 2009), same as Cai et al. (2023).

¹<https://github.com/Hanscal/DCKS>

²<https://github.com/geri-emp/CAB>

³<https://github.com/qtli/KEMP>

⁴<https://github.com/jfzhuyoo/case>

⁵<https://github.com/wlr737/EMNLP2022-SEEK>

⁶<https://github.com/Sahandfer/CEM>

Table 5: An example of EMPATHETICDIALOG and the responses generated from Lamb and other SOTA models. Blue text and orange text denotes the rational and emotional expression separately.

Context	<p>Speaker: One of the times i remember feeling the most blissed out in life was right after the birth of my first child.</p> <p>Listener: That is a very blessed day. It is something you will never forget.</p> <p>Speaker: Obviously there are the demands of a new child – but that feeling of finally meeting someone you waited so long for, and the love surrounding the whole situation. truly something to remember.</p>
Emotion CoNECT	<p>Grateful</p> <p>Based on the content of the dialogue, it is clear that the speaker is expressing feelings of gratitude and contentment. The use of phrases such as "most blissed out in life" and "something you will never forget" suggest a strong positive emotion. Additionally, the speaker mentions the birth of their first child, which is a significant life event that can evoke powerful emotions. The sentiment label "grateful" accurately captures the speaker’s emotional state. Overall, the causal link between the dialogue content and the sentiment label is strong and straightforward.</p>
KEMP SEEK DCKS	<p>I am sure you will do great !</p> <p>That is a great accomplishment !</p> <p>That is great to hear. i am happy for you.</p>
LLaMA2-13B-Chat	<p>Oh my gosh, I can totally relate! Having my first child was one of the most surreal and magical experiences of my life. The pure joy and love I felt in that moment was indescribable. It’s like nothing else mattered except for that little bundle of happiness in my arms.</p>
Lamb	<p>Yes, it is a great feeling to finally meet someone you waited so long for.</p>
Gold	<p>Yes, I could not agree more. it is remarkable how your feeling suddenly change.</p>

Table 6: Empathetic response case-2.

Context	<p>Speaker: Hi, I feel so lonely sometimes because all my friends live in a different country</p> <p>Listener: Oh, I am sure you are lonely. Maybe you can join some kind of club that lets you meet new friends?</p> <p>Speaker: I was thinking about it! I wanted to join a group for local moms</p>
Emotion CoNECT	<p>Lonely</p> <p>Based on the dialogue content, the speaker is expressing a feeling of loneliness due to their friends living in a different country. The sentiment label is questioning, as the speaker is seeking suggestions for meeting new friends in their local area. The causal link between the dialogue content and the sentiment label is that the speaker's feeling of loneliness is driving their desire to seek out new social connections.</p>
KEMP SEEK DCKS	<p>I am sorry to hear that . I hope you can find a new friends !</p> <p>I am sure they are going to be a good friend .</p> <p>I am sure you will have a great time.</p>
LLaMA2-13B-Chat	<p>Oh, that sounds like a great idea! There are probably a lot of other moms in the same situation as you who would love to meet new friends too. Have you tried reaching out to any local mom groups or clubs in your area? They might have meetups or events where you can connect with other moms and make some new friends.</p>
Lamb	<p>That is a good idea. Maybe you could see if you could join some sort of club?</p>
Gold	<p>That is a good idea ! This way you can also meet friends for yourself , but also maybe meet new friend 's for your children to hang out with while you do with their moms !</p>

Table 7: Empathetic response case-3.

Context	<p>Speaker: I am so sad because I have read an article about a newborn girl who died because her parents did not believe in medication and doctors</p>
Emotion	<p>Sad</p>
KEMP SEEK DCKS	<p>I am so sorry to hear that . I hope you can find a way to get a new one ?</p> <p>Oh wow , that is a lot of people .</p> <p>I am sorry to hear that.</p>
LLaMA2-13B-Chat	<p>Oh no, that's heartbreaking! I can't imagine how devastating it must be for the parents to lose their child like that. It's understandable that you're feeling sad after reading that. Can I give you a hug?</p>
Lamb	<p>That is so sad. I am sorry you had to go through that.</p>
Gold	<p>Ugh , those articles always get me too ... : (what was wrong with her ?</p>

Table 8: The effect of different versions of prompt on the empathetic response of LLM generation. Red text represent keywords of prompt.

Context	
Speaker: i cheated on my girlfriend recently . i feel horrible about myself right now . Listener:	
Prompt	Response
You are an expert listener with empathy. Please give an appropriate response based on the provided conversation history.\n\nspeaker: i cheated on my girlfriend recently . i feel horrible about myself right now .\nlistener:"	It's important to take responsibility for your actions and acknowledge the pain and hurt that you've caused. Consider apologizing to your girlfriend and taking steps to make amends, such as being transparent about your actions, being honest about what led to the infidelity, and making a commitment to be faithful in the future.
Prompt	Response
You are an expert listener with empathy. Please give an appropriate response based on the provided conversation history. Remember, the length of the response should match the scenario of daily conversation.\n\nspeaker: i cheated on my girlfriend recently . i feel horrible about myself right now .\nlistener:	It's also crucial to reflect on why you cheated and address any underlying issues that may have contributed to your behavior. This could involve seeking therapy or counseling to work through any emotional or psychological issues, as well as communicating with your partner about what you need to change in order to rebuild trust and move forward.
Prompt	Response
You are an expert listener with empathy. Please give an appropriate response based on the provided conversation history. Remember, the length of the response should match the scenario of daily conversation that usually is composed by 2 or 3 sentences.\n\nspeaker: i cheated on my girlfriend recently . i feel horrible about myself right now .\nlistener:	Remember that forgiveness is a process and it may take time for your girlfriend to heal and trust you again. Be patient and supportive, and consider taking concrete steps to show her that you're committed to making things right. Ultimately, it's up to your girlfriend to decide whether she can forgive you, but by taking responsibility and working to make things right, you can at least demonstrate your sincerity and commitment to the relationship.
Prompt	Response
You are a professional psychologist. Please give a reply from the perspective of a friend based on the patient's situation. Be careful to avoid a large number of professional vocabulary in your reply, and the expression should be natural.\n\nThe following is a set of historical conversations between you and your best friend. Please respond according to the context and pay attention to fully consider the emotional information in the historical conversations when replying.\n\nspeaker: i cheated on my girlfriend recently . i feel horrible about myself right now .\nlistener:	Hey buddy, it sounds like you're going through a tough time. Nothing I say can change what happened, but I'm here for you. You did something you regret, it's plain to see that you're feeling guilty, which shows you've realised your mistake. It's important now to think about what you want to do moving forward and how you can avoid repeating this in future. Consider coming clean with your girlfriend if it feels right. Sometimes, owning up to our mistakes and facing the consequences can actually help us grow as individuals. But remember, whatever you decide, I'm here for you, man.
Prompt	Response
You are a professional psychologist. Please give a reply from the perspective of a friend based on the patient's situation. Be careful to avoid a large number of professional vocabulary in your reply, and the expression should be natural.\n\nThe following is a set of historical conversations between you and your best friend. Please respond according to the context and pay attention to fully consider the emotional information in the historical conversations. Remember, response should not exceed 30 words.\n\nSpeaker: i cheated on my girlfriend recently . i feel horrible about myself right now .\nListener:	I'm really sorry to hear that you're going through this. It's important to confront your feelings and be honest with her about it.