ACC-ERG:COORDINATING AFFECTIVE AND COGNITIVE SUPPORT FOR EMPATHETIC RESPONSE GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Empathetic response generation methods model affective and cognitive information to better understand users' states. However, generated responses may mismatch the context by overlooking emotional needs or offering generic comfort when advice is required. To address this issue, we propose ACC-ERG, an Affective–Cognitive Coordination framework that adjusts the relative emphasis on affective and cognitive support according to the dialogue context. First, we build affective and cognitive experts to learn specialized support representations. Second, we design a Proportion Gate to predict a weight that characterizes the relative emphasis on affective versus cognitive support. Finally, we fuse these representations to generate an empathetic response according to the weight. Moreover, we construct ACC-ERG dataset with weight annotations and responses aligned with weights. Experiments on EmpatheticDialogues and ACC-ERG datasets show our framework outperforms state-of-the-art methods and demonstrate that varying weight correspondingly shifts the relative emphasis of affective and cognitive support in generated responses.

1 INTRODUCTION

Empathetic Response Generation plays an important role in open-domain dialogue systems by understanding users' states and generating appropriate responses. Such responses typically involve both affective and cognitive aspects (Davis, 1983; Decety & Jackson, 2004), expressed through reassurance and encouragement (Elliott et al., 2018; Arican-Dinc & Gable, 2023) or explanation, advice, and guidance. Their effectiveness also depends on how well they match subjective experiences and the current dialogue contexts of individuals (Cutrona & Russell, 1990; High & Dillard, 2012).

Recent studies have increasingly focused on interactions between affective and cognitive information for empathetic response generation. CASE (Zhou et al., 2023) performs multi-granularity alignment between cognitive and affective representations; IAMM (Yang et al., 2024) captures higher-order interactions between cognitive and emotional states; Wang et al. (2025) models dependencies among emotion, emotion cause, and commonsense knowledge; and ReflectDiffu (Yuan et al., 2025) models the dynamic relation between emotion and response intent. These methods enrich the modeling of user states by connecting multiple empathy-related signals.

Despite these advances, generated responses may still mismatch the dialogue context. For example, a response may overemphasize reassurance when practical guidance is needed (Ye et al., 2025), or shift to advice before sufficiently acknowledging the user's emotions (Son et al., 2026). These cases indicate that improved user-state understanding does not by itself specify how affective and cognitive support should be composed within the next response. Existing ERG methods mainly improve state representation and response generation, but rarely model this response-level support composition explicitly, limiting direct control over the relative emphasis of the two support functions.

To address this issue, we propose an **A**ffective-**C**ognitive **C**oordination framework for **E**mpathetic **R**esponse **G**eneration, named **ACC-ERG**, which coordinates affective and cognitive support by adapting their relative emphasis to the dialogue context. First, we build affective and cognitive experts to learn specialized representations for the two support functions. Second, we design a

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



Figure 1: Illustration of affective–cognitive support coordination. Given the user context, ACC-ERG decomposes the response into affective and cognitive support components, assigns a context-dependent support weight and fuses them to generate an ideal response.

Proportion Gate to predict a continuous weight from the dialogue context, representing the relative emphasis on affective versus cognitive support in the target response. Finally, we explicitly fuse the two expert representations according to the predicted weight and complement this routing with residual interaction paths to preserve cross-expert information. In this way, the predicted weight directly controls the support direction, while the residual interaction retains complementary information from both experts. As shown in Figure 1, although the user expresses concern about layoffs and compensation, the current utterance explicitly seeks advice. ACC-ERG therefore predicts a lower affective weight of 0.3, placing greater emphasis on cognitive support related to compensation and subsequent actions while retaining affective acknowledgment. Since existing datasets do not directly provide supervision for such response-level support proportions, we further construct the ACC-ERG dataset with weight annotations and weight-aligned empathetic responses.

Our main contributions are summarized as follows:

- We introduce a continuous formulation of context-adaptive support coordination that models response-level affective-cognitive support composition, bridging the gap between understanding and appropriate response generation.
- We propose the ACC-ERG framework, which uses continuous weights to guide affective-cognitive representation fusion and response generation, enabling explicit and controllable adjustment of support emphasis across dialogue contexts.
- We construct the ACC-ERG dataset with weight annotations and corresponding weight-aligned responses, providing direct supervision for learning and evaluating affective-cognitive support composition.
- Experiments show consistent improvements over state-of-the-art baselines and demonstrate that weight intervention regulates the composition of affective and cognitive support in generated responses.

2 RELATED WORK

Empathic response generation aims to understand users' emotions and situations, and generate appropriate and supportive responses (Rashkin et al., 2019; Lin et al., 2019; Sabour et al., 2022; Welivita et al., 2023).Early studies primarily generate responses based on users' emotion states, such as MoEL (Lin et al., 2019), MIME (Majumder et al., 2020) and EmpDG (Li et al., 2020). To better understand dialogue contexts and emotion causes, subsequent research has increasingly incorporated cognitive information. Some studies identify and infer emotion causes (Kim et al., 2021; Gao et al., 2021; He et al., 2026), while others utilize commonsense and external knowledge to supplement information regarding user intentions and potential reactions (Sabour et al., 2022; Chen & Liang, 2022; Fu et al., 2023; Cai et al., 2024; Cao et al., 2025). Building on this foundation, further research has focused on interactions between various factors. KEMP (Li et al., 2022) jointly organizes context, emotion and knowledge using a graph structure; CASE (Zhou et al., 2023) aligns cognitive and affective representations at both coarse-grained and fine-grained levels; IAMM (Yang et al., 2024) captures cognitive and emotional states through iterative associative memory and second-order interaction attention; ECC (Wang et al., 2025) model the relationships among emotion, emotion cause and commonsense; ReflectDiffu (Yuan et al., 2025) models the dynamic connection between emotion and intent.

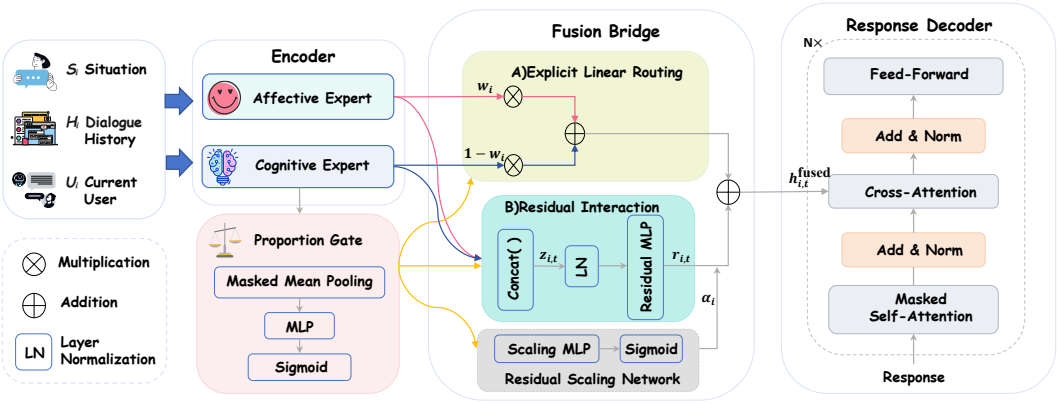


Figure 2: Architecture of ACC-ERG framework, which comprises four primary components: specialized Affective and Cognitive Experts, a Proportion Gate for weight prediction, a Fusion Bridge for representation coordination and a Response Decoder for empathetic response generation.

These works have improved understanding of affective and cognitive information. Despite it, generated responses may still not align with the current context, for example, by offering premature explanations or advice and relying on generic reassurances (Ye et al., 2025; Son et al., 2026). Therefore, further research needs transforming understood information into empathetic responses that align with users' needs and the dialogue context. Based on this observation, ACC-ERG investigates context-adaptive coordination of affective and cognitive support at the response level and employs a weighted fusion of two support representations to guide the generation of empathetic responses.

3 PRELIMINARIES

Problem Formulation. Given a dialogue instance $X_i = (S_i, H_i, U_i)$, where S_i denotes the situation description, H_i represents the dialogue history, and U_i is the current user utterance. Empathetic Response Generation aims to generate an appropriate response $Y_i = (y_{i,1}, \ldots, y_{i,T_i})$ to the current dialogue context. An ideal response should not only acknowledge and respond to the user's affective state, but also provide appropriate cognitive support, such as understanding, explanation, or guidance. In this work, we explicitly model the relative emphasis of these two support functions in the target response. For each dialogue instance, we associate the target response Y_i with a continuous support weight $w_i^* \in (0, 1)$, where a higher value indicates greater emphasis on affective support, while a lower value indicates greater emphasis on cognitive support. The corresponding proportion of cognitive support is $1 - w_i^*$. Accordingly, ACC-ERG learns to infer the context-adaptive support weight from X_i and generate a response Y_i whose affective–cognitive composition is consistent with the diaolgue context.

ACC-ERG Dataset. We construct an ACC-ERG dataset from EmpatheticDialogues (Rashkin et al., 2019) and KardiaBench (Yuan et al., 2026) to support weight prediction and weight-aligned response generation. We first use predefined rules to estimate support tendencies and sample contexts across cognitive-oriented, balanced, and affective-oriented groups. GPT-4o-mini is used to annotate the affective support weight $w_i^* \in \{0.1, 0.2, \ldots, 0.9\}$ and to generate a weight-aligned reference response Y_i^*. Both weight annotations and responses are manually verified. We also provide overall support intensity labels $\rho_i \in \{1, \ldots, 5\}$ and user speech-act labels a_i. The resulting dataset is $\mathcal{D}_{\text{ACC}} = \{(X_i, w_i^*, \rho_i, a_i, Y_i^*)\}_{i=1}^{N}$, containing 4,998 instances with an 8:1:1 split for train, validation, and test sets. Details of ACC-ERG dataset are shown in Appendix B.

4 METHODOLOGY

As shown in Figure 2, given the dialogue input $X_i = (S_i, H_i, U_i)$, ACC-ERG framework consists of three main components: (1) affective and cognitive experts that capture complementary support representations from the dialogue context, (2) a Proportion Gate that estimates the context-adaptive

relative emphasis of affective and cognitive support, and (3) a Fusion Bridge that coordinates the two representations under the predicted weight for response generation. In this work, the two experts are first trained separately and then keep freezing, while the Proportion Gate and Fusion Bridge are jointly optimized for empathetic response generation.

4.1 AFFECTIVE-COGNITIVE DUAL EXPERTS

We train two experts based on Flan-T5-XL (T5; Raffel et al., 2020) to obtain representations of affective and cognitive support that serve distinct functions. We inject independent LoRA (Hu et al., 2022) adapters into query and value projections within attention modules, updating only a small number of low-rank parameters to acquire specialized capabilities. This stage uses a standard conditional sequence generation objective. For input X, target sequence $Y = (y_1, \ldots, y_{|Y|})$, and trainable parameters Φ, we define $\mathcal{L}_{\text{gen}}(X, Y; \Phi) = -\sum_{t=1}^{|Y|} \log p_{\Theta_0,\Phi}(y_t \mid y_{<t}, X)$, where Θ_0 denotes the frozen backbone parameters and $|Y|$ is the number of tokens in the target sequence.

To enhance the ability to recognize and respond to emotions, we perform supervised training on the affective branch using EmpatheticDialogues dataset $\mathcal{D}_{\text{ED}}$ (Rashkin et al., 2019). Given the training input $X_i^{\text{aff}} = (E_i, S_i, H_i, U_i)$, where E_i denotes the emotion label, and the supervised target Y_i^{aff} is an affective-support response, the affective parameters are optimized as $\Phi_{\text{aff}}^* = \arg\min \mathbb{E}_{(X,Y)\sim\mathcal{D}_{\text{ED}}} [\mathcal{L}_{\text{gen}}(X, Y; \Phi_{\text{aff}})]$ and are then frozen.

To improve the understanding of users' situations and cognitive support capabilities, we perform two-stage training on the cognitive branch using KardiaBench $\mathcal{D}_{\text{KB}}$. In the first stage, the input is $X_i^{\text{ana}} = (P_i, E_i, S_i, H_i, U_i)$, where P_i denotes the user profile, and the supervision target is a structured cognitive analysis $A_i = (u_i, r_i)$, where u_i summarizes the user's intent and emotional cues, while r_i analyzes the user's background, potential causes, and possible support directions (Yuan et al., 2026). The first-stage parameters are optimized as $\Phi_{\text{cog}}^{(1)} = \arg\min \mathbb{E}_{(X,A)\sim\mathcal{D}_{\text{KB}}} [\mathcal{L}_{\text{gen}}(X, A; \Phi_{\text{cog}})]$.To transfer the structured cognitive analysis to response generation, the second stage is initialized from $\Phi_{\text{cog}}^{(1)}$. The input is extended to $X_i^{\text{rep}} = (P_i, E_i, S_i, H_i, U_i, A_i)$, and the supervised target Y_i^{cog} is a natural cognitive support response. The final cognitive parameters are optimized as $\Phi_{\text{cog}}^* = \arg\min \mathbb{E}_{(X,Y)\sim\mathcal{D}_{\text{KB}}} [\mathcal{L}_{\text{gen}}(X, Y; \Phi_{\text{cog}})]$.

After functional specialization, the input X_i is encoded by the dual experts to obtain token-level representations:

$$\mathbf{H}_i^{\text{aff}} = \text{Enc}\left(X_i; \mathbf{\Theta}_0, \mathbf{\Phi}_{\text{aff}}^*\right), \quad \mathbf{H}_i^{\text{cog}} = \text{Enc}\left(X_i; \mathbf{\Theta}_0, \mathbf{\Phi}_{\text{cog}}^*\right), \quad \mathbf{H}_i^{\text{aff}}, \mathbf{H}_i^{\text{cog}} \in \mathbb{R}^{n_i \times d}, \tag{1}$$

where $\text{Enc}(\cdot)$ denotes the encoder backbone with corresponding parameters; n_i is the input sequence length, and d is the hidden dimension.

4.2 PROPORTION GATE

For estimating the relative emphasis of affective support versus cognitive support in an ideal response, we design a Proportion Gate that takes the cognitive representation as input. Given the output $\mathbf{H}_i^{\text{cog}}$, we obtain $\mathbf{h}_i^{\text{cog}} \in \mathbb{R}^d$ through masked mean pooling and compute the weight as $w_i = \sigma\left(\text{MLP}_\phi\left(\mathbf{h}_i^{\text{cog}}\right)\right)$, where $w_i \in (0, 1)$ and ϕ denotes the learnable parameters of the gate.

We employ a confidence-weighted regression objective to accurately predict the weights. Given the annotated weight w_i^* and the confidence score c_i, we define

$$\mathcal{L}_{\text{reg}} = \mathbb{E}_{i\sim\mathcal{B}}\left[c_i \left(w_i - w_i^*\right)^2\right], \tag{2}$$

where $\mathcal{B}$ denotes the samples with weight annotations in the current training batch. This objective constrains the absolute error between the predicted and annotated weights while reducing the influence of low-confidence samples during training.

To further preserve the relative ordering of weights from different inputs, we introduce a pairwise ranking objective. For sample pairs satisfying $|w_i^* - w_j^*| \geq \delta_{\text{rank}}$, we adopt a ranking loss:

$$\mathcal{L}_{\text{rank}} = \mathbb{E}_{(i,j)\sim\mathcal{P}}\left[\text{softplus}\left(-\frac{sign(w_i - w_j)}{\tau}\right)\right], \quad \mathcal{P} = \left\{(i, j) \mid |w_i^* - w_j^*| \geq \delta_{\text{rank}}\right\}, \tag{3}$$

where $\delta_{\text{rank}} = 0.1$ corresponds to the minimum weight difference considered for ranking in the training data, and τ is a temperature coefficient controlling the sensitivity of the ranking loss.

4.3 FUSION BRIDGE

To coordinate affective and cognitive representations according to the predicted weight, inspired by conditional modulation (Perez et al., 2018) and residual adaptation (Rebuffi et al., 2017), we design a Fusion Bridge that combines explicit linear routing with residual interaction. We first perform weight-guided linear routing over the two expert representations:

$$\mathbf{h}_{i,t}^{\text{rte}} = w_i \mathbf{h}_{i,t}^{\text{aff}} + (1 - w_i)\mathbf{h}_{i,t}^{\text{cog}}, \tag{4}$$

and then we construct an interaction representation at each token position to capture cross-branch interactions beyond linear interpolation:

$$\mathbf{z}_{i,t} = \text{Concat}\left(\mathbf{h}_{i,t}^{\text{aff}}, \mathbf{h}_{i,t}^{\text{cog}}, \mathbf{h}_{i,t}^{\text{aff}} - \mathbf{h}_{i,t}^{\text{cog}}, \mathbf{h}_{i,t}^{\text{aff}} \odot \mathbf{h}_{i,t}^{\text{cog}}, w_i\right), \mathbf{z}_{i,t} \in \mathbb{R}^{4d+1}, \tag{5}$$

where $\text{Concat}(\cdot)$ denotes vector concatenation, and $\odot$ denotes element-wise multiplication. By including w_i as a conditioning variable, the residual network can adapt its cross-branch interaction to different support configurations. The residual interaction representation is then computed as:

$$\mathbf{r}_{i,t} = \text{MLP}_\psi\left(\text{LN}\left(\mathbf{z}_{i,t}\right)\right), \quad \mathbf{r}_{i,t} \in \mathbb{R}^d. \tag{6}$$

where $\text{LN}(\cdot)$ denotes layer normalization, and ψ denotes the learnable parameters of the residual Multi-Layer Perceptron.

To adapt the strength of residual correction to different support-weight configurations, we introduce a sample-level residual scaling network that predicts a scaling coefficient

$$\alpha_i = \sigma(\text{MLP}_\alpha\left([w_i;\, |w_i - 0.5|]\right)), \quad \alpha_i \in (0, 1) \tag{7}$$

where $|w_i - 0.5|$ measures the deviation of the current support weight from the affective–cognitive equilibrium point. The final fused representation is computed as: $\mathbf{h}_{i,t}^{\text{fus}} = \mathbf{h}_{i,t}^{\text{rte}} + \kappa\alpha_i \mathbf{r}_{i,t}$, where κ is a fixed scaling coefficient. Stacking the fused representations over all token positions yields $\mathbf{H}_i^{\text{fus}} = [\mathbf{h}_{i,1}^{\text{fus}}, \ldots, \mathbf{h}_{i,n_i}^{\text{fus}}]$. To make the fused representation directly available for response generation, we use $\mathbf{H}_i^{\text{fus}}$ as the encoder-side hidden states attended to by the decoder through cross-attention.

4.4 LEARNING STRATEGY

To ensure both weight prediction and dual experts coordination contribute to final response generation, we jointly train the Proportion Gate and Fusion Bridge. For the target response Y_i^*, the response-generation objective is:

$$\mathcal{L}_{\text{resp}} = \mathbb{E}_{i\sim\mathcal{B}}\left[-\log p\left(Y_i^* \mid \mathbf{H}_i^{\text{fus}}(\widetilde{w}_i)\right)\right], \tag{8}$$

where $\mathcal{B}$ denotes the current training batch and $\widetilde{w}_i$ is the support weight used by the Fusion Bridge during training. Since $\widetilde{w}_i$ participates in the computation of $\mathbf{H}i^{\text{fus}}$ through a differentiable routing path, gradients from $\mathcal{L}\text{resp}$ can be propagated through the Fusion Bridge to the Proportion Gate, enabling end-to-end joint optimization.

For mitigating the discrepancy between training with annotated weights and inference with predicted weights, we employ scheduled weight conditioning. At training stage k, for sample i, the weight used for fusion is defined as:

$$\widetilde{w}_i = b_i w_i^* + (1 - b_i)w_i, \quad b_i \sim \text{Bernoulli}(p_k), \tag{9}$$

where w_i^* is the annotated weight, w_i is the predicted weight, and p_k denotes the probability of using the annotated weight at stage k. The probability p_k decays linearly from 0.9 to 0.2 over training.

To ensure the fused representation changes consistently with the weight, we apply interventions to the predicted weight. For sample i and predicted weight w_i, we construct a lower weight w_i^- and a higher weight w_i^+ and obtain the corresponding fused representations through the Fusion Bridge. We then apply masked mean pooling to obtain the low and high weight fused vectors:$\mathbf{h}_i^-$, $\mathbf{h}_i^+$. We define the directional gains along the affective and cognitive dimensions as $g_i^{\text{aff}} = \cos(\mathbf{h}_i^+, \mathbf{h}_i^{\text{aff}}) -$

$\cos(\mathbf{h}_i^-, \mathbf{h}_i^{\text{aff}})$; $g_i^{\text{cog}} = \cos(\mathbf{h}_i^-, \mathbf{h}_i^{\text{cog}}) - \cos(\mathbf{h}_i^+, \mathbf{h}_i^{\text{cog}})$. A positive g_i^{aff} indicates that increasing the support weight moves the fused representation closer to the affective representation, whereas a positive g_i^{cog} indicates that decreasing the weight moves it closer to the cognitive representation. The direction constraint is defined as:

$$\mathcal{L}_{\text{dir}} = \frac{1}{2}\mathbb{E}_{i\sim\mathcal{B}}\left[\max\left(0, \gamma - g_i^{\text{aff}}\right) + \max(0, \gamma - g_i^{\text{cog}})\right], \tag{10}$$

where γ denotes the direction margin. This objective encourages local changes in the support weight to induce consistent shifts in the fused representation toward the corresponding support direction, even in the presence of nonlinear residual corrections.

The overall objective for joint optimization is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{resp}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}} + \lambda_{\text{rank}}\mathcal{L}_{\text{rank}} + \lambda_{\text{dir}}\mathcal{L}_{\text{dir}}, \tag{11}$$

where the coefficients λ_{reg}, λ_{rank}, and λ_{dir} control the contributions of the three auxiliary objectives relative to the response-generation objective. During joint training, only the parameters of the Proportion Gate and Fusion Bridge are optimized.

5 EXPERIMENTS

Datasets & Baselines. The ACC-ERG dataset evaluates models' ability to generate support weight-aligned empathetic responses. We select CEM, CASE, IAMM, ReflectDiffu and GPT-3.5-Turbo as representative baselines to cover relevant technical approaches, such as commonsense enhancement, multi-source interaction and emotion-intention modeling. All baselines except GPT-3.5-Turbo are trained and evaluated using the same sets. We further evaluate ACC-ERG on the standard EmpatheticDialogues split and compare it with representative baselines: MoEL (Lin et al., 2019), MIME (Majumder et al., 2020), CEM (Sabour et al., 2022), KEMP (Li et al., 2022), CASE (Zhou et al., 2023), IAMM (Yang et al., 2024), ReflectDiffu (Yuan et al., 2025), DialoGPT (Zhang et al., 2020), EmpGPT-3 (Lee et al., 2022) and PECER (Cai et al., 2024).The reported results for MoEL, MIME, CEM, KEMP, CASE, IAMM and ReflectDiffu are taken from ReflectDiffu (Yuan et al., 2025), while the results for DialoGPT, EmpGPT-3, and PECER are taken from PECER (Cai et al., 2024). More can be seen in Appendix C.1 and C.2.

Implementation Details. All experiments are implemented in PyTorch 2.5 and conducted on an NVIDIA RTX 3090 GPU. For dual-branch LoRA fine-tuning, we set the rank to 16, the scaling factor to 32, and the dropout rate to 0.05. During joint training, the model is trained for 4 epochs with a batch size of 2. The learning rates of the Proportion Gate and Fusion Bridge are set to 2e-4 and 5e-5. We use AdamW (Loshchilov & Hutter, 2019) with a weight decay of 0.01, bf16 precision, and gradient clipping with a maximum norm of 1.0. In the joint objective, λ_{reg}, λ_{rank}, and λ_{dir} are set to 0.5, 0.05 and 0.1. We train and evaluate the model using multiple random seeds.

Automatic Evaluations. We evaluate response generation with BLEU-1/2/3/4 (Papineni et al., 2002), PPL (Serban et al., 2015) and Distinct-1/2 (Li et al., 2016). For weight prediction, we use MAE, RMSE, Pearson and Spearman correlation to assess prediction error and ranking consistency. More can be seen in Appendix C.3

Human Evaluations. We randomly sample 120 dialogue contexts from the EmpatheticDialogues test set and conduct pairwise A/B comparisons between responses generated by ACC-ERG and baselines. Three independent evaluators with concealed model identities and candidate responses presented in random order select the better response according to four criteria: **Contextual Relevance**, **Empathy**, **Fluency** and **Appropriateness**: whether the response matches the user's current needs and context in emphasis of support provided. We report the win, tie, and loss rates of ACC-ERG against each baseline.

5.1 AUTOMATIC EVALUATION RESULTS & ABLATION STUDY

As shown in Table 1, ACC-ERG achieves the best BLEU-1/2/3/4 results on the ACC-ERG dataset, with relative improvements of 18.69%, 34.88%, 48.38%, and 62.65%, respectively, over the strongest baseline. The larger gains at higher n-gram orders suggest that weight-conditioned coordination improves higher-order agreement with reference responses, while the lower Distinct-1/2

scores indicate a trade-off between target-aligned generation and lexical diversity. As shown in Table 2, ACC-ERG achieves the best results on BLEU-1/2/3/4 and Distinct-1/2 on EmpatheticDialogues. Compared with the strongest baseline on each metric, it yields relative improvements of 0.51%, 10.84%, 46.01%, and 51.38% on BLEU-1 through BLEU-4, respectively, and improves Distinct-1/2 by 54.83% and 14.51%. The larger gains at higher n-gram orders, together with improved diversity, suggest that affective-cognitive coordination strengthens phrase-level reference agreement without restricting lexical variation. ACC-ERG framework does not achieve the lowest PPL, indicating that its gains are more pronounced in reference agreement and response diversity than in conditional likelihood. These results further indicate that the benefits of ACC-ERG framework extend beyond the ACC-ERG dataset to the standard EmpatheticDialogues benchmark.

To evaluate the contributions of the core components in ACC-ERG, we compare four ablated variants: **(1) w/o Affective Expert**: removes the specialized affective expert; **(2) w/o Cognitive Expert**: removes the specialized cognitive expert; **(3) w/o Proportion Gate**: removes sample-level support-weight prediction and uses unconditional fusion; and **(4) w/o Fusion Bridge**: replaces the Fusion Bridge with a trainable module of matched parameter count and retrains the model under the same setting, while retaining support-weight prediction and explicit linear routing.

Table 1: Model comparison and ablation results on the ACC-ERG dataset.

Models	B-1↑	B-2↑	B-3↑	B-4↑	D-1↑	D-2↑
CEM	30.05	18.34	13.02	9.12	2.88	9.59
CASE	31.78	19.51	14.00	10.69	2.48	8.03
IAMM	29.28	16.87	11.38	8.33	5.75	17.83
ReDiff	31.55	19.81	14.55	11.11	0.65	1.67
GPT3.5Turbo	<u>40.56</u>	<u>25.49</u>	<u>18.81</u>	<u>13.44</u>	**9.35**	**41.21**
ACC-ERG (ours)	**48.14**	**34.38**	**27.02**	**21.86**	9.03	34.71
w/o Cognitive Expert	47.17	33.69	26.36	21.21	**9.28**	**35.35**
w/o Affective Expert	47.86	33.95	26.46	21.22	9.23	35.20
w/o Proportion Gate	46.50	33.05	25.83	20.84	8.80	33.93
w/o Fusion Bridge	47.07	33.33	25.70	20.37	8.70	34.30

As shown in Table 1, ACC-ERG achieves the best BLEU scores among all variants. Removing either specialized expert reduces performance, confirming the complementary roles of affective and cognitive representations. Further drops are observed without the Proportion Gate or Fusion Bridge, with BLEU-4 decreasing to 20.84 and 20.37, respectively, indicating the benefits of context-adaptive support weighting and residual representation coordination. Some ablations yield slightly higher Distinct scores, suggesting a trade-off between lexical diversity and reference-response agreement. Overall, the results support the complementary contributions of the core components. Further ablations are reported in Appendix D.1.

Table 2: Automatic evaluation results on the EmpatheticDialogues dataset. Metrics include BLEU-1/2/3/4, PPL and Dist 1/2.

Models	MOEL	MIME	CEM	KEMP	CASE	IAMM	ReDiff	DialoGPT	EmpGPT3	PECER	**Ours**
B-1↑	18.02	19.82	16.12	17.92	19.66	19.51	<u>23.59</u>	15.74	16.38	21.01	**23.71**
B-2↑	8.67	8.86	7.29	8.54	8.95	8.74	<u>11.25</u>	6.86	7.67	10.01	**12.47**
B-3↑	4.35	4.43	4.06	4.38	4.92	4.86	5.35	3.94	3.91	<u>5.52</u>	**8.06**
B-4↑	2.73	2.77	2.03	2.71	2.90	3.32	<u>3.62</u>	2.36	2.15	3.40	**5.48**
PPL↓	36.81	36.93	36.33	36.59	35.97	25.94	24.56	18.84	–	**14.42**	<u>18.56</u>
D-1↑	0.43	0.51	0.62	0.66	0.70	0.88	0.98	2.71	3.18	<u>3.52</u>	**5.45**
D-2↑	1.76	1.92	2.39	2.43	2.66	3.05	4.35	12.01	14.91	<u>16.95</u>	**19.41**

5.2 WEIGHT PREDICTION & CONTROLLABILITY ANALYSIS

To evaluate whether the Proportion Gate can accurately predict weights from the dialogue context, we established four control conditions. **Global Mean** and **Source Mean** consistently output the average weights from the entire training set and the corresponding source dataset, respectively, serving as baselines to assess predictive performance based solely on global prior distributions. **Shuffled Input** randomizes the correspondence between test inputs and labels. **Current-only** retains only the current user utterance, while **History-only** retains only the historical context, allowing us to analyze

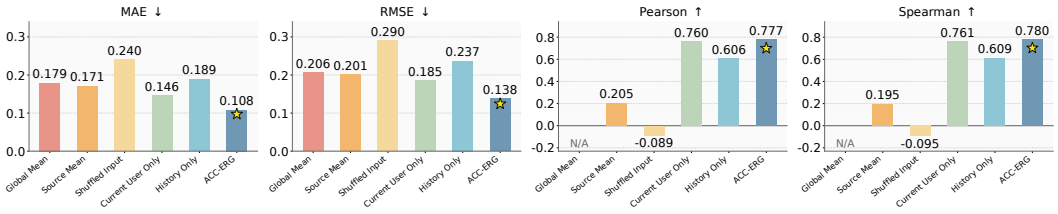


Figure 3: Weight prediction performance of the Proportion Gate under prior-based inputs, shuffled inputs and context ablations.

the contributions of different contextual information. Figure 3 demonstrates that ACC-ERG achieves the best results in terms of both weight prediction error and correlation metrics; this indicates that the Proportion Gate can recover support weights consistent with human annotations from dialogue semantics and that its advantages stem primarily from modeling the specific dialogue context rather than simply fitting the average distribution of the training set. Although the overall predicted distribution changes little after input shuffling, the correlation between predicted values and human labels drops significantly, further ruling out the possibility that the model relies solely on the global weight distribution. Contextual ablation results further reveal that the current user utterance provides the primary predictive signal, while the historical context plays a supplementary role.

To examine whether the weight controls affective–cognitive support emphasis, we use GPT-4o-mini to rate responses on four dimensions using a 1–5 scale. Affective Support evaluates emotional acknowledgment, validation, reassurance, and caring, while Cognitive Support evaluates situational understanding, explanation, reframing, advice, and guidance. Context Relevance and Fluency assess whether such functional changes preserve basic response quality. As shown in Figure 4, increasing the weight from 0.1 to 0.9 consistently raises affective support scores and lowers cognitive support scores, while relevance and fluency remain relatively stable. This indicates that the support weight effectively controls affective–cognitive emphasis without substantially degrading response quality.

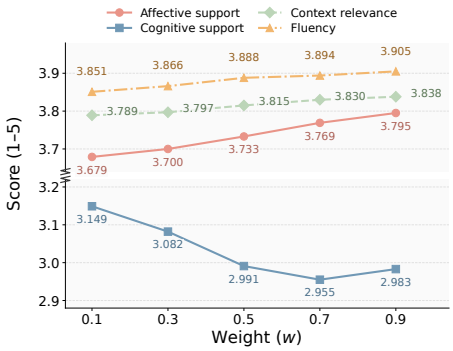


Figure 4: Weight intervention results across different support functions.

5.3 HUMAN EVALUATION & CASE STUDY

Table 3: Human A/B evaluations of ACC-ERG against baselines. All values are percentages.

	vs. CASE			vs. ReflectDiffu			vs. GPT-3.5-Turbo		
Aspect	Win	Tie	Lose	Win	Tie	Lose	Win	Tie	Lose
Emp.	**54.6**	24.6	20.8	**52.3**	27.4	20.3	**50.2**	22.7	27.1
Rel.	**52.7**	29.8	17.5	**48.6**	36.7	14.7	**38.3**	25.1	36.6
Flu.	**46.2**	40.4	13.4	**44.8**	42.7	12.5	**43.0**	14.2	42.8
App.	**56.4**	32.0	11.6	**55.9**	28.6	15.5	**46.8**	14.8	38.4

As shown in Table 3, ACC-ERG achieves higher win rates than CASE and ReflectDiffu across all four dimensions. Against GPT-3.5-Turbo, the advantage is most evident in Empathy and Appropriateness, while Relevance and Fluency remain broadly comparable. This pattern suggests that ACC-ERG primarily improves the context-appropriate support composition rather than quality of generic response alone.

Table 4 illustrates ACC-ERG's explicit, context-adaptive coordination of support through the predicted weight. A high weight ($w_i = 0.912$) accompanies emotional validation in the grief context, whereas a lower weight ($w_i = 0.378$) accompanies concrete suggestions in the advice-seeking con-

text. This adjustment helps align the response with the user's current needs and dialogue context. Beyond generating empathetic responses, ACC-ERG provides an interpretable signal of the intended support emphasis, illustrating its flexibility across different user needs.

Table 4: Case study comparison between ACC-ERG and baselines.

Emotion	**Devastated**	**Hopeful**
Context	Remember our beautiful dog at home? Well unfortunately he died recently. The family are so sad.	It's comforting to know that these feelings are normal. I guess it just means I'm really invested in this change, right? How do you think I can best manage that anxiety while still embracing the excitement?
CASE	I am so sorry to hear that.	I am sure you will be fine!
IAMM	I am sorry to hear that. I hope you guys do not get it.	I hope you have a lot of job.
ReDiff	That is so sad. I hope you have a good time.	You are a good job!
GPT3.5Turbo	I'm so sorry to hear that. Losing a pet can be incredibly difficult. How are you and your family holding up?	It's great that you're recognizing the importance of managing your anxiety. Have you thought about setting small, achievable goals to help you stay focused and motivated? Remember to take care of yourself along the way too. You're doing great!
ACC-ERG	*Predicted weight:* **0.912** I'm so sorry to hear about your beautiful dog; losing a beloved pet is incredibly painful, and it's completely understandable to feel sad right now.	*Predicted weight:* **0.378** It's great to hear that you're feeling invested in this change! To manage your anxiety, try journaling or reflecting on what you're grateful for during this time. It's okay to feel both excited and anxious, and it shows your commitment to this journey.
Reference	I am so sorry to hear about your dog. What happened to him?	Embrace that investment! Try grounding techniques to ease anxiety while celebrating your excitement. You're on a beautiful path!

6 CONCLUSION AND LIMITATIONS

We propose **ACC-ERG** (Affective-Cognitive Coordination for Empathetic Response Generation), which explicitly models the context-adaptive composition of affective and cognitive support in empathetic responses using continuous weights. ACC-ERG integrates dual-expert representations, a Proportion Gate, and a Fusion Bridge to jointly perform weight prediction and conditional representation coordination. We further introduce the ACC-ERG dataset, which provides supervision for weight prediction and weight-aligned empathetic response generation. Experimental results show that ACC-ERG consistently outperforms strong baselines in automatic evaluation, human evaluation, and weight intervention analysis, supporting the effectiveness and controllability of the affective-cognitive support coordination.

ACC-ERG focuses on support composition within individual responses and may be less proactive in eliciting feelings or clarifying needs. Incorporating dialogue history and user feedback could improve conversational initiative and support adaptation across turns. Future work could also incorporate overall support intensity as an additional control signal, allowing the model to jointly determine both the relative composition and the amount of support provided in a given context. Finally, extending the framework to finer-grained support functions while preserving interpretability and controllability remains an important direction.

AI USE STATEMENT

In this work, we use generative AI tools to assist with literature retrieval and discovery, dataset construction, automated response evaluation, and language editing. In particular, GPT-4o-mini is

used to assist with affective-cognitive support-weight annotation, generate weight-aligned response targets during dataset construction, and evaluate generated responses in the weight intervention analysis. Generative AI tools are also used to improve the clarity and readability of the manuscript. All AI-assisted annotations, generated data, evaluations, and text revisions are reviewed by the authors, and the authors take full responsibility for the final content of this work.

ETHICS STATEMENT

ACC-ERG is designed to study the coordination of affective cognitive support and controllable generation of empathetic responses. It is not intended for mental health diagnosis, risk assessment, or as a substitute for professional mental health services. The predicted support weights represent only the relative emphasis of affective and cognitive support in a generated response and should not be interpreted as direct measurements of the psychological state or clinical needs of a user.

The ACC-ERG dataset is constructed from existing dialogue datasets and involves LLM-assisted annotation followed by human review. Consequently, the resulting annotations and responses may inherit biases from the source data, annotation judgments, and underlying language models. The source datasets are used in accordance with their applicable licenses and terms of use, and the corresponding original works are properly cited. Deployment in sensitive or high-risk settings therefore requires additional evaluation across diverse populations, appropriate risk mitigation measures, and human oversight.

REFERENCES

Beyzanur Arican-Dinc and Shelly L Gable. Responsiveness in romantic partners' interactions. *Current Opinion in Psychology*, 53:101652, 2023.

Mingxiu Cai, Daling Wang, Shi Feng, and Yifei Zhang. Pecer: Empathetic response generation via dynamic personality extraction and contextual emotional reasoning. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 10631–10635. IEEE, 2024.

Huiying Cao, Yiqun Zhang, Shi Feng, Xiaocui Yang, Daling Wang, and Yifei Zhang. TOOL-ED: Enhancing empathetic response generation with the tool calling capability of LLM. In *Proceedings of the 31st International Conference on Computational Linguistics*, pp. 5305–5320, 2025.

Yangbin Chen and Chunfeng Liang. Wish i can feel what you feel: A neural approach for empathetic response generation. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pp. 922–933, Abu Dhabi, United Arab Emirates, 2022. Association for Computational Linguistics.

Yi Cheng, Wenge Liu, Wenjie Li, Jiashuo Wang, Ruihui Zhao, Bang Liu, Xiaodan Liang, and Yefeng Zheng. Improving multi-turn emotional support dialogue generation with lookahead strategy planning. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 3014–3026, 2022.

Carolyn E. Cutrona and Daniel W. Russell. Type of social support and specific stress: Toward a theory of optimal matching. In Barbara R. Sarason, Irwin G. Sarason, and Gregory R. Pierce (eds.), *Social Support: An Interactional View*, pp. 319–366. Wiley & Sons, 1990.

Mark H Davis. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, 44:113–126, 1983.

Jean Decety and Philip L Jackson. The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews*, 3:71–100, 2004.

Robert Elliott, Arthur C Bohart, Jeanne C Watson, and David Murphy. Therapist empathy and client outcome: An updated meta-analysis. *Psychotherapy*, 55:399–410, 2018.

Yahui Fu, Koji Inoue, Chenhui Chu, and Tatsuya Kawahara. Reasoning before responding: Integrating commonsense-based causality explanation for empathetic response generation. In *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 645–656, 2023.

Jun Gao, Yuhan Liu, Haolin Deng, Wei Wang, Yu Cao, Jiachen Du, and Ruifeng Xu. Improving empathetic response generation by recognizing emotion cause in conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pp. 807–819, November 2021.

Yuanyuan He, Guotai Huang, Wei Li, Jiali You, Jiawen Deng, and Fuji Ren. MvP-ECR: Multi-perspective emotion-cause reasoning for empathetic dialogue. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 1819–1827, 2026.

Andrew C High and James Price Dillard. A review and meta-analysis of person-centered messages and social support outcomes. *Communication Studies*, 63:99–118, 2012.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Dongjin Kang, Sunghwan Mac Kim, Taeyoon Kwon, Seungjun Moon, Hyunsouk Cho, Youngjae Yu, Dongha Lee, and Jinyoung Yeo. Can large language models be good emotional supporter? mitigating preference bias on emotional support conversation. In *Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers)*, pp. 15232–15261, 2024.

Hyunwoo Kim, Byeongchang Kim, and Gunhee Kim. Perspective-taking and pragmatics for generating empathetic responses focused on emotion causes. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 2227–2240, 2021.

Young-Jun Lee, Chae-Gyun Lim, and Ho-Jin Choi. Does GPT-3 generate empathetic dialogues? a novel in-context example selection method and automatic evaluation metric for empathetic dialogue generation. In *Proceedings of the 29th international conference on computational linguistics*, pp. 669–683, 2022.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and William B Dolan. A diversity-promoting objective function for neural conversation models. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pp. 110–119, 2016.

Qintong Li, Hongshen Chen, Zhaochun Ren, Pengjie Ren, Zhaopeng Tu, and Zhumin Chen. EmpDG: Multi-resolution interactive empathetic dialogue generation. In *Proceedings of the 28th international conference on computational linguistics*, pp. 4454–4466, 2020.

Qintong Li, Piji Li, Zhaochun Ren, Pengjie Ren, and Zhumin Chen. Knowledge bridging for empathetic dialogue generation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 10993–11001, 2022.

Zhaojiang Lin, Andrea Madotto, Jamin Shin, Peng Xu, and Pascale Fung. 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)*, pp. 121–132, 2019.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. Towards emotional support dialog systems. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)*, pp. 3469–3483, 2021.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Navonil Majumder, Pengfei Hong, Shanshan Peng, Jiankun Lu, Deepanway Ghosal, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. MIME: MIMicking emotions for empathetic response generation. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pp. 8968–8979, 2020.

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

Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Jonas Pfeiffer, Aishwarya Kamath, Andreas Rücklé, Kyunghyun Cho, and Iryna Gurevych. AdapterFusion: Non-destructive task composition for transfer learning. In *Proceedings of the 16th conference of the European chapter of the association for computational linguistics: main volume*, pp. 487–503, 2021.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21:1–67, 2020.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pp. 5370–5381, 2019.

Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. Learning multiple visual domains with residual adapters. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Sahand Sabour, Chujie Zheng, and Minlie Huang. CEM: Commonsense-aware empathetic response generation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 11229–11237, 2022.

Iulian Vlad Serban, Alessandro Sordoni, Yoshua Bengio, Aaron C. Courville, and Joelle Pineau. Hierarchical neural network generative models for movie dialogues. *CoRR*, abs/1507.04808, 2015.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017. URL `https://openreview.net/forum?id=B1ckMDqlg`.

Suhyune Son, Jungwoo Lim, Myunghoon Kang, Seongtae Hong, Yuna Hur, Evelyn H Zi, and Heuiseok Lim. I don't need solution. i need emotional support: Empathetic LLMs based on emotional validation. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 1–26, 2026.

Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. MISC: A mixed strategy-aware model integrating comet for emotional support conversation. In *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: Long papers)*, pp. 308–319, 2022.

Xu Wang, Bo Wang, Yihong Tang, Dongming Zhao, Jing Liu, Ruifang He, and Yuexian Hou. ECC: Synergizing emotion, cause and commonsense for empathetic dialogue generation. In *Proceedings of the 31st international conference on computational linguistics*, pp. 5475–5485, 2025.

Anuradha Welivita and Pearl Pu. A taxonomy of empathetic response intents in human social conversations. In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 4886–4899, 2020.

Anuradha Welivita, Chun-Hung Yeh, and Pearl Pu. Empathetic response generation for distress support. In *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 632–644, 2023.

Zhou Yang, Zhaochun Ren, Wang Yufeng, Haizhou Sun, Chao Chen, Xiaofei Zhu, and Xiangwen Liao. An iterative associative memory model for empathetic response generation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3081–3092, 2024.

Jing Ye, Lu Xiang, Yaping Zhang, and Chengqing Zong. From generic empathy to personalized emotional support: A self-evolution framework for user preference alignment. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pp. 18826–18853, 2025. doi: 10.18653/v1/2025.findings-emnlp.1024.

Jiahao Yuan, Zixiang Di, Zhiqing Cui, Guisong Yang, and Usman Naseem. ReflectDiffu: Reflect between emotion-intent contagion and mimicry for empathetic response generation via a rl-diffusion framework. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 25435–25449, 2025.

Jiahao Yuan, Zhiqing Cui, Hanqing Wang, Yuansheng Gao, Yucheng Zhou, and Usman Naseem. Kardia-r1: Unleashing llms to reason toward understanding and empathy for emotional support via rubric-as-judge reinforcement learning. In *Proceedings of the ACM Web Conference 2026*, pp. 9230–9240, 2026.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and William B Dolan. DIALOGPT: Large-scale generative pre-training for conversational response generation. In *Proceedings of the 58th annual meeting of the association for computational linguistics: system demonstrations*, pp. 270–278, 2020.

Weixiang Zhao, Yanyan Zhao, Shilong Wang, and Bing Qin. TransESC: Smoothing emotional support conversation via turn-level state transition. In *Findings of the association for computational linguistics: ACL 2023*, pp. 6725–6739, 2023.

Jinfeng Zhou, Chujie Zheng, Bo Wang, Zheng Zhang, and Minlie Huang. CASE: Aligning coarse-to-fine cognition and affection for empathetic response generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 8223–8237, 2023.

A EXTENDED RELATED WORK

Emotional Support Conversation. Research on emotional support conversation primarily focuses on modeling supportive behaviors and their organization over dialogue. Early work categorizes fine-grained empathetic intents (Welivita & Pu, 2020), while ESConv (Liu et al., 2021) formalizes emotional support through explicit strategies such as questioning, reflection of feelings, affirmation, suggestion, and information provision. Subsequent studies further investigate strategy combination, long-horizon planning, and strategy transition (Tu et al., 2022; Cheng et al., 2022; Zhao et al., 2023). Recent analyses also show that strategy selection can exhibit systematic preferences (Kang et al., 2024) or solution-oriented tendencies (Son et al., 2026), suggesting that selecting a plausible support behavior does not by itself determine how different support functions should be balanced within the final response.

ACC-ERG addresses a complementary level of modeling. Rather than selecting or planning discrete support behaviors, it explicitly represents the response-level composition of affective and cognitive support with a continuous support weight. This distinction allows ACC-ERG to model not only which kind of support behavior is appropriate, but also how strongly affective and cognitive support should be emphasized in the current context, providing a finer and directly controllable representation of support composition.

Modular Expert Fusion. Modular expert fusion focuses on combining complementary capabilities from specialized experts or adapters. Mixture-of-Experts dynamically routes inputs to different experts (Shazeer et al., 2017), while AdapterFusion aggregates knowledge from independently trained adapters (Pfeiffer et al., 2021). In empathetic response generation, MoEL combines emotion-specific experts according to predicted emotion distributions (Lin et al., 2019), whereas CEM (Sabour et al., 2022) and KEMP (Li et al., 2022) integrate emotional, commonsense, and contextual information to construct richer response representations. These methods primarily address how heterogeneous representations should be combined to improve generation quality.

ACC-ERG focuses instead on how the resulting response should balance affective and cognitive support. It introduces a continuous support weight with explicit response-level semantics to coordinate the two specialized experts, while the Fusion Bridge refines their interaction for response

generation. This design makes the fusion process directly interpretable and controllable in terms of support composition.

B ACC-ERG DATASET CONSTRUCTION

The ACC-ERG dataset is constructed as a derived annotation layer over EmpatheticDialogues (ED) and KardiaBench (KB). The dataset characterizes the desired composition of affective and cognitive support in the next empathetic response. Each instance is represented by the situation description, dialogue history, and current user utterance. The affective support weight is defined as

$$w_i^* = \frac{A_i}{A_i + C_i}, \tag{12}$$

where A_i and C_i denote the relative emphasis of affective and cognitive support, respectively. Overall support intensity is annotated separately and is not included in the support-weight definition.

B.1 SOURCE DATA AND UNIFIED REPRESENTATION

We construct the candidate pool from ED and KB. ED mainly provides open domain dialogues grounded in personal emotional experiences, whereas KB contains richer background information, contextual relations, and structured understanding and reasoning signals. Rather than treating the two sources as purely affective or cognitive, we retain instances spanning different context lengths, information complexities, and communicative purposes.

For each dialogue instance, we extract the *Situation*, the *Conversation History*, and the *Current User* utterance. The last user turn is treated as the current utterance, and up to four preceding turns are retained as dialogue history. Instances without a valid current user turn are discarded. The original reference response is preserved as metadata but is not used to determine support weights.

B.2 HEURISTIC STRATIFICATION FOR CANDIDATE SAMPLING

Because existing datasets do not provide explicit affective–cognitive support weights, we use a lightweight heuristic to improve coverage during candidate sampling. The heuristic prioritizes the Current User utterance with a coefficient of 1.00, while the Situation and Conversation History receive lower coefficients of 0.35 and 0.20, respectively, reflecting their supplementary contextual roles. Affective cues include explicit emotion expressions, acute distress, loss or isolation, self-worth threats, and requests for validation or connection. Cognitive cues include explicit advice seeking, procedural questions, decision comparison, action planning, causal explanation, perspective seeking, and problem solving. Cue matches are weighted by both cue type and input field and aggregated into independent affective and cognitive heuristic scores.

The resulting heuristic weight w_i^h is used only for stratification. Candidate instances are divided into five intervals $[0, 0.2)$, $[0.2, 0.4)$, $[0.4, 0.6)$, $[0.6, 0.8)$, $[0.8, 1.0]$. We then adopt a symmetric sampling distribution of $15\%/20\%/30\%/20\%/15\%$ across the five intervals. Within each interval, equal numbers of instances are sampled from ED and KB. This procedure yields 5,000 candidate instances, including 2,500 from each source. The sampling distribution is introduced only to improve coverage of different support configurations and does not represent the natural distribution of the original datasets.

B.3 ANNOTATION PROTOCOL

Pilot Annotation. Before large-scale annotation, we conduct two rounds of blinded pilot annotation to refine the annotation guidelines. Annotators independently distinguish affective support, cognitive support, overall support intensity, and the current speech act. The revised guideline asks annotators to consider two counterfactual questions: whether a purely analytical response would appear emotionally cold, and whether a purely comforting response would fail to address an explicit problem. Based on these judgments, annotators allocate relative emphasis between affective and cognitive support. The primary speech act is selected from eight categories: Emotional Disclosure, Validation Seeking, Social Sharing, Explanation Seeking, Advice Seeking, Decision Support, Action Planning, and Other. Speech acts describe what the user is doing in the current turn, whereas

Table 5: Statistics of the ACC-ERG dataset.

Split	Total	Eligible	ED	KB	Mean Weight
Train	4,055	4,054	2,026	2,029	0.530
Dev	475	475	237	238	0.529
Test	469	469	236	233	0.529

the support weight describes how the response should allocate affective and cognitive support; the two are therefore annotated separately.

LLM-Assisted Pre-annotation. After the annotation guideline is finalized, GPT-4o-mini is used to generate structured pre-annotations. Each request contains only the situation, conversation history, and current user utterance. Data-source identity, original emotion labels, heuristic weights, heuristic bins, and original reference responses are excluded from the prompt to prevent leakage from the candidate sampling procedure.

The model predicts an affective support emphasis from $\{10, 20, \ldots, 90\}$, overall support intensity on a five-point scale, primary and secondary speech acts, confidence, supporting evidence, and a brief rationale. We use temperature 0 and process one instance per request. Outputs are automatically validated for schema correctness, valid priority ranges, speech-act constraints, and consistency between affective and cognitive priorities. Among the 5,000 candidates, 4,999 receive valid pre-annotations. One additional instance is identified as having invalid context during subsequent review, yielding 4,998 eligible instances.

Human Review. The LLM outputs serve only as initial annotations. Human reviewers inspect the pre-annotations together with the full dialogue context and revise them when necessary. The final labels include the affective support weight w_i^*, overall support intensity ρ_i, and primary speech act a_i. Heuristic scores are retained only as audit fields and are never used as formal labels.

B.4 Data Partitioning

Dialogue groups are stratified using the data source, dominant support-weight label, and primary speech act, and are then divided into training, development, and test sets at an approximate $8{:}1{:}1$ ratio. The resulting split contains 4,055 training records, 475 development records, and 469 test records. One invalid training record is retained only for auditing and excluded from training, leaving 4,054 trainable training instances and 4,998 eligible instances in total. No original dialogue appears in more than one split.

B.5 Weight-Aligned Response Construction

After the weight labels are finalized, we construct weight-aligned responses for the training set. At this stage only, the teacher model is allowed to access the original reference response, which is treated as optional factual or topical evidence rather than as a generation target.

The teacher input contains the full dialogue context, the finalized affective and cognitive support weights, overall support intensity, and primary and secondary speech acts. For high affective-support weights, the generation instruction emphasizes validation and emotional acknowledgment before advice. For lower affective-support weights, the response remains supportive while placing greater emphasis on clarification, explanation, perspective taking, or actionable guidance. The teacher is instructed not to introduce unsupported facts, diagnoses, personal experiences, or guarantees. GPT-4o-mini generates weight-aligned responses for all 4,054 eligible training instances using temperature 0. The generated responses contain an average of 30.78 words. Development and test instances are not used to construct training targets.

C MORE DETAILS OF EXPERIMENTS

C.1 DATASETS DETAILS

EmpatheticDialogues (ED; Rashkin et al., 2019) is an open-domain empathetic dialogue dataset constructed around 32 categories of emotional situations. Following the standard split, the dataset contains 36,629 training instances, 5,712 development instances, and 5,242 test instances. ED dialogues primarily center on users' emotion experiences and expressions, while the reference responses frequently involve emotion recognition, validation, and supportive acknowledgment. We therefore use ED as the main training source for the affective expert, with the goal of strengthening the model's ability to represent users' affective states and generate affective-support responses.

KardiaBench (KB; Yuan et al., 2026) contains user background information, dialogue context, multi-turn interactions, and structured understanding and reasoning fields. Its richer personal background, event relations, and contextual cues provide useful supervision for situational understanding, causal reasoning, and support planning. We therefore use KB as the main training source for the cognitive expert, with the goal of strengthening the model's ability to represent user situations, infer potential causes, and generate cognitively grounded support.

C.2 BASELINES DETAILS

We compare ACC-ERG with the following representative empathetic response generation methods on EmpatheticDialogues benchmark:

- **MoEL:** Lin et al. (2019) combines multiple emotion-specific listeners according to the dialogue context to generate empathetic responses.
- **MIME:** Majumder et al. (2020) models response emotions through emotion grouping and emotional mimicry.
- **CEM:** Sabour et al. (2022) incorporates affective and cognitive commonsense knowledge to improve empathetic response generation.
- **KEMP:** Li et al. (2022) jointly models dialogue context, emotion, and external knowledge through a situational graph.
- **CASE:** Zhou et al. (2023) aligns cognitive commonsense with affective concepts using a coarse-to-fine mechanism.
- **IAMM:** Yang et al. (2024) captures multi-order contextual interactions through iterative associative memory.
- **ReflectDiffu:** Yuan et al. (2025) integrates emotion contagion, intent mimicry, and diffusion-based reflective generation.
- **DialoGPT:** Zhang et al. (2020) is a large-scale pretrained autoregressive dialogue model for open-domain response generation.
- **EmpGPT-3:** Lee et al. (2022) explores empathetic response generation with GPT-3 through prompt learning.
- **PECER:** Cai et al. (2024) jointly models dynamic personality and contextual emotion for empathetic response generation.

C.3 EVALUATION METRICS DETAILS

Response Generation. To evaluate the response generation performance of ACC-ERG, we use the following automatic metrics:

- **BLEU-1/2/3/4:** Papineni et al. (2002) measures n-gram overlap between generated responses and reference responses. Higher scores indicate stronger agreement with the reference at different n-gram orders.
- **Perplexity (PPL):** Following prior dialogue generation work (Serban et al., 2015), PPL measures the average uncertainty of the model when predicting the reference response conditioned on the dialogue context. Lower values indicate higher conditional likelihood assigned to the reference response.

- **Distinct-1/2:** Li et al. (2016) measures response diversity by computing the proportion of distinct unigrams and bigrams among all generated n-grams. Higher scores indicate greater lexical diversity in the generated responses.

Weight Prediction. To evaluate the weight prediction performance of the Proportion Gate, we use Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Pearson correlation, and Spearman rank correlation. Let the test set contain N instances, where w_i^* denotes the human-verified support weight of the i-th instance and w_i denotes the corresponding prediction from the Proportion Gate.

- **MAE:** MAE measures the average absolute deviation between predicted and human-verified weights:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |w_i - w_i^*| \,. \tag{13}$$

- **RMSE:** RMSE assigns greater penalties to larger prediction errors and is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (w_i - w_i^*)^2}. \tag{14}$$

- **Pearson Correlation:** Pearson correlation measures the linear agreement between predicted and human-verified weights:

$$r = \frac{\sum_{i=1}^{N}(w_i - \bar{w})(w_i^* - \bar{w}^*)}{\sqrt{\sum_{i=1}^{N}(w_i - \bar{w})^2}\sqrt{\sum_{i=1}^{N}(w_i^* - \bar{w}^*)^2}}, \tag{15}$$

  where $\bar{w}$ and $\bar{w}^*$ denote the mean predicted weight and mean human-verified weight, respectively.
- **Spearman Correlation:** Spearman correlation measures the monotonic agreement between the rankings induced by the predicted and human-verified weights. Let $R(w_i)$ and $R(w_i^*)$ denote their respective ranks. It is equivalently computed as the Pearson correlation between the rank variables:

$$\rho_s = \mathrm{corr}\left(R(w), R(w^*)\right). \tag{16}$$

Lower MAE and RMSE indicate more accurate numerical predictions, while higher Pearson and Spearman correlations indicate stronger agreement with the human-verified weights in terms of numerical variation and relative ordering, respectively.

D ADDITIONAL EXPERIMENTS RESULTS

D.1 ABLATION STUDY

Table 6 examines the main design choices of ACC-ERG from five perspectives: specialized experts, the Proportion Gate, the Fusion Bridge, decoder configuration, and training strategy. The full ACC-ERG model uses the cognitive representation as input to the Proportion Gate, applies the Fusion Bridge for weight-conditioned representation coordination, adopts the Cognition Decoder, and jointly optimizes the Gate and Fusion Bridge. Except for the single-expert variants, each controlled ablation modifies one target component while keeping the remaining settings unchanged.

Specialized Experts. Removing either specialized expert consistently reduces BLEU performance relative to the full model. In particular, BLEU-4 decreases from 21.86 to 21.21 without the cognitive expert and to 21.22 without the affective expert. The Cognition-only configuration retains slightly stronger BLEU-1/2 performance than the Affective-only configuration, whereas the latter yields marginally higher Distinct scores. This pattern is consistent with the functional specialization of the two branches: the cognitive expert emphasizes situational understanding and content organization, while the affective expert contributes complementary emotional-support information. Neither expert alone matches the full model, indicating that the main benefit comes from coordinating the two types of support rather than relying on either representation independently.

Proportion Gate. Removing the Proportion Gate causes a consistent decline across BLEU-1/2/3/4, with BLEU-4 decreasing from 21.86 to 20.84. This suggests that unconditional fusion cannot fully capture the variation in affective–cognitive support composition across dialogue contexts. By predicting a sample-level support weight, the Proportion Gate allows the relative contribution of the two experts to adapt to each input. The gain is smaller than that produced by the Fusion Bridge, indicating that the support weight provides useful guidance but depends on the subsequent fusion mechanism to affect generation effectively.

Gate Input Representation. We further compare the full model, which uses the cognitive representation as Gate input, with Dual-gate and Affective-gate variants. The full model achieves the highest BLEU-1/2/3/4 scores, while Affective-gate obtains the second-best BLEU-2/3/4 results and Dual-gate achieves the highest Distinct-2 score. These results suggest that the cognitive representation provides a more suitable basis for predicting support weights when the objective is reference-aligned generation. One possible explanation is that the cognitive branch explicitly encodes situational structure, user intent, and potential support direction, which are closely related to estimating how affective and cognitive support should be balanced. However, the higher diversity of Dual-gate shows that different Gate inputs may favor different generation properties rather than being uniformly superior across all metrics.

Fusion Bridge. Replacing the Fusion Bridge with a parameter-matched trainable alternative leads to a clear degradation in generation performance, with BLEU-1 decreasing from 48.14 to 47.07 and BLEU-4 from 21.86 to 20.37. The drop is particularly pronounced on higher-order BLEU metrics, suggesting that the Fusion Bridge contributes to coordinating affective and cognitive representations beyond what explicit linear routing alone can capture. Since support-weight prediction and linear routing are retained in this variant, the remaining gap can be attributed more directly to the nonlinear residual interaction introduced by the Fusion Bridge. This result supports its role as a complementary refinement mechanism.

Decoder Configuration. Replacing the Cognition Decoder used in the full model with either the Base Decoder or the Affective Decoder causes only modest changes in BLEU. The Affective Decoder remains close to the full model on BLEU-2/3/4, while the Base Decoder achieves higher Distinct-1/2 scores. In contrast, the full model retains the strongest BLEU performance overall. The relatively small differences among decoder variants suggest that the main improvement of ACC-ERG is not driven by decoder selection alone. Instead, once the affective and cognitive representations are effectively coordinated by the Fusion Bridge, different decoder initializations can consume the fused representation with similar effectiveness.

Training Strategy. Two-stage training performs competitively but remains below joint training on all BLEU metrics. For example, BLEU-1 decreases from 48.14 to 47.92 and BLEU-4 from 21.86 to 21.12, whereas Distinct-1/2 increase to 9.33 and 35.61. Joint optimization therefore favors reference-response agreement, while the two-stage setting produces slightly greater lexical diversity. This difference is consistent with the optimization structure of ACC-ERG: under joint training, gradients from the generation objective can simultaneously adjust support-weight prediction and representation coordination through the differentiable fusion path, reducing mismatch between independently optimized modules.

Overall, the ablation results show that the main components of ACC-ERG make complementary contributions to generation performance. The Proportion Gate provides consistent gains by adapting the relative contribution of the affective and cognitive experts to each dialogue context, while the two specialized experts supply complementary support representations. The Fusion Bridge yields the largest additional improvement among the controlled structural variants, suggesting that nonlinear residual interaction further refines the representation obtained from explicit weight-guided routing. Gate input, decoder configuration, and joint optimization provide smaller but consistent refinements. Together with the support-weight prediction and intervention analyses, these results suggest that the effectiveness of ACC-ERG comes from the coordinated interaction between explicit support weighting, specialized representations, and representation refinement rather than from any single component alone.

Table 6: Detailed ablation results of ACC-ERG. Best results are shown in bold and second-best results are underlined.

Component	Variant	B-1↑	B-2↑	B-3↑	B-4↑	D-1↑	D-2↑
Branch	Affective-only	47.17	33.69	26.36	21.21	9.28	35.35
	Cognition-only	47.86	33.95	26.46	21.22	9.23	35.20
Proportion Gate	No-Gate	46.50	33.05	25.83	20.84	8.80	33.93
	Dual-gate	47.87	33.94	26.33	21.08	9.18	**36.21**
	Affective-gate	47.78	34.02	26.63	21.46	9.03	34.62
Fusion Bridge	No Fusion Bridge	47.07	33.33	25.70	20.37	8.70	34.30
Decoder	Base Decoder	47.22	33.75	26.46	21.35	**9.49**	35.87
	Affective Decoder	47.75	34.11	26.64	21.47	9.22	35.21
Training	Two-stage	47.92	34.04	26.41	21.12	9.33	35.61
Full (ours)	ACC-ERG	**48.14**	**34.38**	**27.02**	**21.86**	9.03	34.71

D.2 Support Weight Learnability and Calibration

To examine whether the affective–cognitive support composition can be learned as a continuous variable from dialogue context, we evaluate the Proportion Gate on the 469-instance ACC-ERG test set. Each instance is associated with a human-verified affective support weight $w_i^* \in [0, 1]$, and the Gate predicts $\hat{w}_i = G_\theta(X_i), \hat{w}_i \in [0, 1]$. This analysis focuses only on the mapping from dialogue context to support weight and does not evaluate whether manipulating the weight changes the generated response. The latter is examined separately through controlled weight-intervention experiments.

We evaluate prediction accuracy using MAE and RMSE, and assess agreement with human annotations using Pearson and Spearman correlations. On the full test set, the Proportion Gate achieves an MAE of 0.108, an RMSE of 0.138, a Pearson correlation of 0.777, and a Spearman correlation of 0.780. These results indicate that the Gate captures both the numerical variation and the relative ordering of support weights across dialogue contexts. Figure 5 provides three complementary views of the prediction behavior.

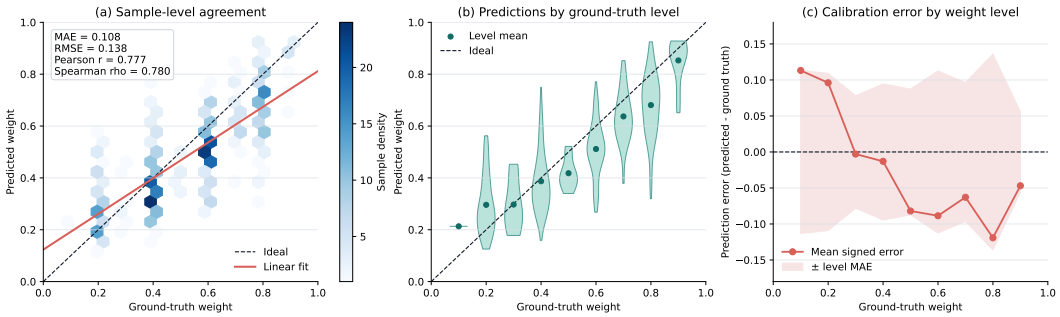


Figure 5: Weight prediction and calibration on the ACC-ERG test set. Figure (a) shows sample-level agreement between human-verified and predicted weights; figure (b) shows conditional prediction distributions for different annotated weight levels; and figure (c) shows the mean signed prediction error at each weight level.

Sample-level agreement. The left panel compares the predicted weight with the human-verified weight for each test instance. The dashed diagonal denotes perfect calibration, while the solid line shows the empirical linear fit. The clear positive trend, together with Pearson $r = 0.777$ and Spearman $\rho = 0.780$, indicates that the Gate recovers a substantial portion of the continuous variation in support composition rather than collapsing to a constant prediction. However, the shallower fitted slope also suggests that the predicted range is more conservative than the annotated range.

Conditional prediction distributions. The middle panel groups instances by their human-verified support weights and shows the corresponding distributions of Gate predictions. The predicted distributions generally shift upward as the annotated weight increases, and the level means follow a largely monotonic trend. This pattern indicates that the Gate learns graded support differences rather than a simple binary affective-versus-cognitive distinction. At the same time, the overlap between adjacent levels shows that the learned weight should be interpreted as a soft routing signal rather than an exact psychological measurement.

Calibration bias. The right panel shows the mean signed error $\mathbb{E}[\hat{w}_i - w_i^* \mid w_i^* = w]$ for each annotated weight level. Positive values indicate overestimation of affective support, whereas negative values indicate underestimation. The model tends to overestimate low support weights and underestimate medium-to-high weights, revealing a regression toward the center of the training distribution. This effect is most visible around the higher affective-support levels and suggests that the Gate is better calibrated for relative ordering than for exact absolute values.

Overall, the analysis shows that the Proportion Gate learns a meaningful continuous structure from dialogue context: predicted weights track both the magnitude and ordering of human-verified support composition. At the same time, the observed range compression indicates that the Gate is not perfectly calibrated, particularly at sparsely represented or extreme weight levels. Therefore, the support weight is best interpreted as an interpretable sample-level routing variable rather than as a precise estimate of a user's psychological state or needs.

D.3 DYNAMIC VS. FIXED ROUTING

To examine whether context-adaptive sample-level support weights are more effective than using a single global weight for all instances, we compare dynamic routing with four fixed-routing configurations on the 469 instances in the ACC-ERG test set. Dynamic routing uses the Proportion Gate to predict a support weight for each dialogue context, whereas Fixed-0.2, Fixed-0.5, Fixed-0.529, and Fixed-0.8 assign the same routing weight to every instance. Here, 0.529 corresponds to the mean human-verified support weight in the training set. All configurations share the same model parameters, inputs, prompt templates, and decoding settings, differing only in the routing weight.

As shown in Table 7, dynamic routing achieves the best results on PPL, BLEU-1/2/3/4, and ROUGE-L among the evaluated settings. Compared with the strongest fixed configuration on these metrics, dynamic routing reduces PPL from 3.127 to 3.075, increases BLEU-4 from 21.01 to 21.86, and improves ROUGE-L from 0.374 to 0.384. These results indicate that, among the fixed weights considered, no single global weight matches the overall likelihood and reference-response agreement achieved by context-dependent routing. Fixed-0.2 obtains the highest Distinct-1/2 scores, but performs worse on PPL, BLEU, and ROUGE-L. This result suggests that increased lexical diversity does not necessarily correspond to stronger agreement with the reference responses. Overall, the comparison supports the use of sample-level dynamic routing, indicating that adapting the relative emphasis of affective and cognitive support to individual dialogue contexts is more effective than applying the tested global weights uniformly across all instances.

Table 7: Automatic evaluation results under dynamic and fixed routing settings.

Routing	PPL↓	B-1↑	B-2↑	B-3↑	B-4↑	R-L↑	D-1↑	D-2↑
Fixed-0.2	3.222	44.91	31.14	23.73	18.61	0.364	**9.88**	**38.33**
Fixed-0.5	3.129	47.68	33.48	25.72	20.41	0.373	9.18	36.02
Fixed-0.529	3.127	47.79	33.80	26.05	20.74	0.374	9.17	35.83
Fixed-0.8	3.182	47.56	33.73	26.22	21.01	0.370	8.86	33.82
Dynamic(ours)	**3.075**	**48.14**	**34.38**	**27.02**	**21.86**	**0.384**	9.03	34.71

D.4 HYPERPARAMETER SENSITIVITY

We conduct a local sensitivity analysis to examine whether ACC-ERG depends strongly on a narrow hyperparameter configuration. We vary four key hyperparameters independently: the support-weight regression coefficient λ_{reg}, ranking coefficient λ_{rank}, directional regularization coefficient λ_{dir}, and residual scaling factor κ. The default setting is $\lambda_{\text{reg}} = 0.5, \lambda_{\text{rank}} = 0.05, \lambda_{\text{dir}} = 0.1, \kappa =$

0.5. For each experiment, only one hyperparameter is changed while all remaining training and decoding settings are kept fixed. All comparisons are conducted on the development set, and the test set is not used for hyperparameter selection.

Table 8: Local hyperparameter sensitivity of ACC-ERG on the development set.

Configuration	MAE↓	Spearman↑	B-4↑	PPL↓
Default	0.1095	0.7944	21.95	3.155
$\lambda_{\text{reg}} = 0.25$	0.1102	0.7900	22.05	3.154
$\lambda_{\text{reg}} = 1.0$	**0.1057**	0.7979	22.13	3.154
$\lambda_{\text{rank}} = 0$	0.1083	**0.8008**	22.33	**3.152**
$\lambda_{\text{rank}} = 0.1$	0.1106	0.7883	22.30	3.156
$\lambda_{\text{dir}} = 0$	0.1094	0.7940	22.04	3.155
$\lambda_{\text{dir}} = 0.2$	0.1093	0.7944	**22.34**	3.154
$\kappa = 0.25$	0.1096	0.7944	21.70	3.171
$\kappa = 0.75$	0.1092	0.7941	21.88	3.162

As shown in Table 8, the main evaluation metrics remain stable across all tested configurations. Gate Spearman varies only from 0.7883 to 0.8008, BLEU-4 from 21.70 to 22.34, and PPL from 3.152 to 3.171. No local perturbation causes a substantial performance collapse, indicating that ACC-ERG is not highly sensitive to a narrow hyperparameter choice.

Increasing λ_{reg} improves Gate prediction accuracy, with MAE decreasing from 0.1102 at 0.25 to 0.1057 at 1.0, while generation metrics remain nearly unchanged. This suggests that stronger support-weight supervision can improve weight estimation without introducing an evident trade-off in response generation.

The directional coefficient mainly affects the intended behavior of the fused representation rather than ordinary generation metrics. Removing the directional constraint reduces the proportion of samples satisfying the bidirectional margin from 94.84% to 71.25%, whereas increasing λ_{dir} to 0.2 raises it to 96.31%. This supports the role of directional regularization in encouraging weight changes to move the fused representation toward the corresponding affective or cognitive expert.

The residual scaling factor κ controls the strength of nonlinear correction in the Fusion Bridge. A smaller value ($\kappa = 0.25$) reduces BLEU-4 to 21.70 and increases PPL to 3.171, suggesting that overly weak residual interaction limits the benefit of nonlinear cross-branch refinement. The default value $\kappa = 0.5$ therefore provides a balanced setting between explicit linear routing and residual correction. The ranking coefficient shows only minor effects within the tested range, indicating that it mainly serves as an auxiliary regularizer rather than a primary source of performance gain.

Overall, these results show that ACC-ERG remains stable under moderate local hyperparameter changes. The observed trends are also consistent with the intended roles of different objectives: λ_{reg} mainly affects support-weight prediction, λ_{dir} constrains representation movement under weight intervention, and κ controls the balance between linear routing and nonlinear refinement.