

CARE-CR: Context-Aware Routing and Expert Fusion for Multi-Preference Cognitive Restructuring

Hongzhi Qi¹, Liangcheng Wang¹, Yijing Yu², Jianqiang Li¹, Bing Xiang Yang^{2*}, Qing Zhao^{1*}

¹College of Computer Science, Beijing University of Technology, Beijing, China

²School of Nursing, Wuhan University, Wuhan, China

Abstract

While Large Language Models (LLMs) offer promising avenues for automated cognitive restructuring in mental health settings, current approaches predominantly focus on superficial positive reframing and lack the adaptability to balance conflicting therapeutic dimensions, such as empathy and rationality. To address these deficiencies, we propose CARE-CR, a context-aware framework that implements a decoupled optimization paradigm. We first train expert policies specialized for distinct therapeutic attributes rather than relying on a monolithic alignment strategy. To mitigate expert data scarcity, we introduce Dimension-Guided Hierarchical Monte Carlo Tree Search (DG-HMCTS) for data-efficient preference augmentation. At inference, a context-aware routing module dynamically predicts optimal preference weights to fuse expert outputs based on the user’s specific distress context. Extensive experiments demonstrate that CARE-CR achieves consistent improvements over strong baselines across multiple evaluation dimensions, including diagnostic accuracy, contextual appropriateness, task effectiveness, and overall helpfulness, while enabling controllable cognitive restructuring generation. The dataset and code are publicly available at <https://github.com/HongzhiQ/CARE-CR>.

1 Introduction

The global prevalence of mental health disorders and the critical shortage of professionals (Organization et al., 2023; Aleman and Denys, 2014; Karyotaki and van Ballegooijen, 2020) have positioned AI-driven systems as a promising resource. However, effective intervention requires addressing the fundamental roots of distress: negative cognitive processing and the solidification of irrational thinking patterns (Aguilera et al., 2019). As exemplified by Ellis’s ABCD model (Ellis, 1962) (Fig-

ure 1), emotional consequences (C) arise not from activating events (A) per se, but from individuals’ irrational beliefs (B) about those events. Consequently, for AI-assisted Cognitive Restructuring (CR) to be truly therapeutic, it must fulfill two core requirements: accurately identifying the underlying irrational belief (B) and generating an effective disputation (D).

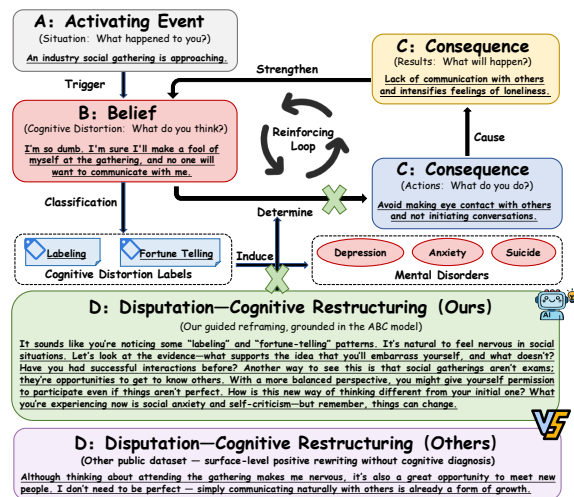


Figure 1: Schematic of the ABCD framework. Unlike surface-level positive rewriting (Others), our theory-driven approach (Ours) leverages the ABC model to explicitly diagnose and dispute specific cognitive distortions.

Nevertheless, current Large Language Model (LLM)-assisted CR approaches exhibit significant deficiencies in addressing these challenges. Regarding the identification of irrational beliefs (cognitive distortions), existing studies either fail to explicitly diagnose specific distortions (Nie et al.; Lin et al.) or rely on oversimplified label sets (Sharma et al., 2023; Lee et al., 2024). In terms of “disputation (D),” the majority of research focuses on superficial manipulations, such as tone softening, positive rewriting, or empathetic dialogue (Rashkin et al., 2019; Wang et al., 2023b; Maddela et al., 2023;

* Corresponding authors: yangbx@whu.edu.cn, zhaqing@email.bjut.edu.cn

Ziems et al., 2022). However, these approaches primarily optimize surface-level rewriting signals and lack robust mechanisms to strictly regulate disputation strategies under clinically relevant constraints (Iftikhar et al., 2024). While prompting or fine-tuning based on Cognitive Behavioral Therapy principles offers some improvement (Xiao et al., 2024; Na, 2024), these methods remain prone to misapplying positive responses, exhibiting guidance biases, and failing to grasp nuanced contexts, thereby posing clinical risks (Chiu et al., 2024; Kim et al., 2025). These limitations underscore that reliance on prompting or standard fine-tuning alone is insufficient for LLMs to precisely comprehend complex psychological concepts or to stably regulate therapeutic strategies.

Furthermore, prior studies have focused primarily on generating restructured sentences without systematically controlling the psychological dimensions of the output. To address this, Sharma et al. (Sharma et al., 2023) defined seven linguistic dimensions of reframing based on psychological literature and employed retrieval-augmented prompting for dimensional control. However, this approach relies on small-scale datasets and few-shot prompting, effectively limiting it to surface-level stylistic imitation that lacks adaptability to specific contexts, distortion types, or user differences. Moreover, it is restricted to optimizing a single attribute at a time. Although Wang et al. (Wang et al.) proposed an iterative optimization framework to specify multiple attributes within a single prompt, this method relies entirely on the LLM’s implicit knowledge without a unified reward structure. Crucially, when confronting conflicting objectives (e.g., high empathy vs. high rationality), the model depends on an opaque and implicit balancing mechanism that is incompatible with the personalization required for clinical treatment. This highlights a critical gap: the optimal trade-off among therapeutic dimensions is not static, but must dynamically adapt to the distress context and individual needs.

Reinforcement Learning from Human Feedback (RLHF) offers a pathway for aligning models with human preferences. However, in the context of cognitive restructuring, large-scale experiments by Sharma et al. (Sharma et al., 2023) reveal significant divergence in human preferences across therapeutic dimensions, indicating the absence of a single optimal equilibrium. This observation aligns with clinical reality: preference weights vary dras-

tically across individuals and contexts, as experts dynamically adjust dimensional ratios based on the situation rather than simply maximizing specific attributes. Consequently, a monolithic reward model is fundamentally incapable of satisfying these diverse needs. To address this, we formulate cognitive restructuring as contextualized preference alignment, where the model must infer a context-dependent preference vector rather than optimize a universal scalar reward.

To bridge the gap between static single-reward optimization and context-dependent multi-preference control, we propose CARE-CR, a context-aware framework that operationalizes contextualized preference alignment for cognitive restructuring. Unlike conventional approaches that rely on global, static trade-offs, our formulation treats the preference distribution as a context-dependent latent variable to be dynamically inferred from each user’s unique distress scenario. To render this alignment tractable, and motivated by multi-objective reward learning (Roijers et al., 2013; Marta et al., 2023) and weight-space interpolation (Rame et al., 2023; Lin et al., 2024), CARE-CR adopts a decoupled optimization paradigm: we train dimension-specific reward models and corresponding expert policies. Crucially, to overcome the bottleneck of scarce supervision, we introduce Dimension-Guided Hierarchical Monte Carlo Tree Search (DG-HMCTS) to synthesize high-fidelity preference pairs. At inference time, a context-aware routing module predicts a preference vector conditioned on the input context, enabling personalized generation via dynamic expert fusion.

2 Methods

2.1 Task Definition

Given a distress narrative x , the model predicts cognitive distortions $\hat{Z} \subseteq \mathcal{Z}$ (defined in Appendix A) and generates a restructuring response y . We target K restructuring dimensions $\mathcal{K} = \{1, \dots, K\}$ (e.g., empathy, rationality) following Sharma et al. Definitions are provided in Appendix A.

We argue that cognitive restructuring has no universal optimum, as the preferred trade-off among dimensions varies with context. Accordingly, we formulate the task as contextualized preference alignment, where a latent preference vector $\lambda(x) \in$

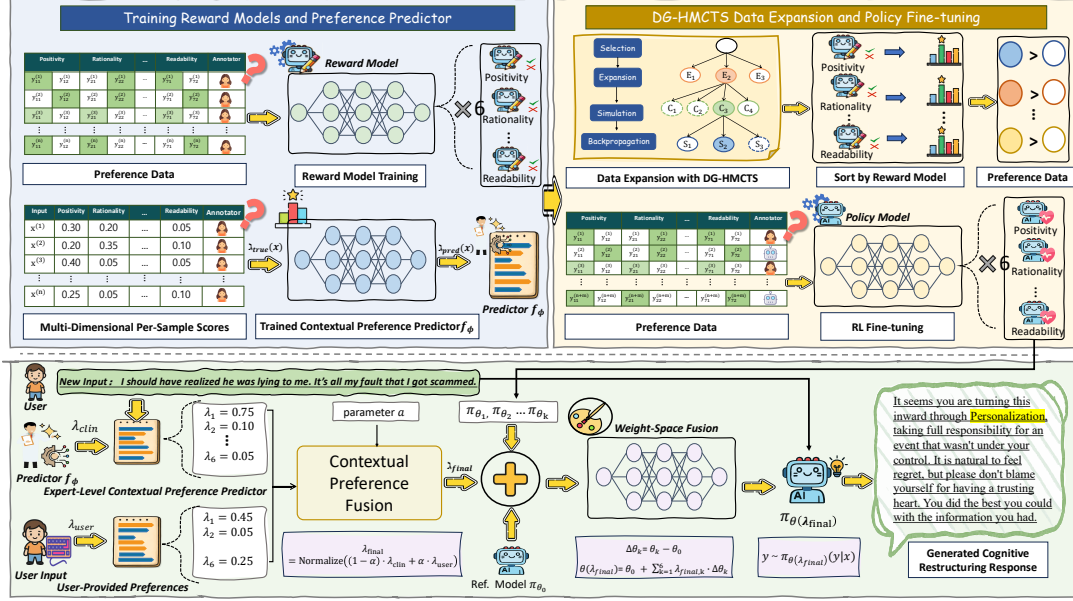


Figure 2: Overview of CARE-CR with training and inference stages for adaptive cognitive restructuring generation.

$\Delta_{\mathcal{K}}$ is inferred from x :

$$\Delta_{\mathcal{K}} = \left\{ \lambda \in \mathbb{R}_{\geq 0}^K \mid \sum_{k=1}^K \lambda_k = 1 \right\}. \quad (1)$$

Formally, we aim to obtain a dynamic policy $\pi_{\theta}(\lambda(x))(y \mid x, \hat{Z})$ tailored to the context, such that the generated response approximates the optimal solution:

$$y^* \in \arg \max_y \sum_{k=1}^K \lambda_k(x) \cdot r_k(y \mid x, \hat{Z}). \quad (2)$$

2.2 Dimension-Specific Reward Modeling and Preference Prediction

Clinical scenarios vary significantly, and the relative emphasis on different restructuring dimensions must adapt dynamically. We adopt a “separate-then-fuse” paradigm. Instead of training a monolithic policy, we train independent reward models $r_k(x, y; \psi_k)$ for each dimension $k \in \mathcal{K}$ using dimension-specific pairwise preference data $D_{\text{pref}}^{(k)} = \{(x, y^+ \succ_k y^-)\}$. These models are optimized via the standard pairwise logistic objective, yielding six frozen reward models parameterized by $\{\psi_1^*, \dots, \psi_6^*\}$.

Since a fixed preference vector cannot capture context diversity, we introduce a dynamic expert-level preference predictor, f_{ϕ} , to infer context-dependent weights. We derive supervisory labels $D_{\lambda} = \{(x, \lambda^{\text{true}}(x))\}$ from expert annotations and train $f_{\phi}(x; \phi)$ to predict weights $\lambda_{\text{ctx}}(x) =$

$\text{Softmax}(\mathbf{W} \cdot \text{Enc}_{\phi}(x) + \mathbf{b})$. To faithfully fit the target distribution while ensuring numerical robustness, we optimize the predictor using a composite objective:

$$\mathcal{L}_{\lambda}(\phi) = \mathbb{E}_{(x, \lambda^{\text{true}}) \sim D_{\lambda}} \left[\mathcal{L}_{\text{KL}}(\lambda^{\text{true}}, \lambda_{\text{ctx}}) + \alpha \cdot \mathcal{L}_{\text{MSE}}(\lambda^{\text{true}}, \lambda_{\text{ctx}}) \right], \quad (3)$$

where α is a balancing hyperparameter.

2.3 Dimension-Guided Hierarchical MCTS (DG-HMCTS)

To overcome the critical bottleneck of expert data scarcity, we propose Dimension-Guided Hierarchical MCTS (DG-HMCTS), which functions as a high-fidelity data augmentation engine. Unlike standard unstructured exploration, this method embeds psychological principles into the search space by enforcing a three-layer therapeutic strategy tree: Affective Opening (E), Cognitive Challenge (C), and Closing Integration (S). This hierarchical formulation ensures that synthesized responses follow a coherent clinical progression. A complete strategy path is formally defined as $\tau = (E_i, C_j, S_k)$, where the search is guided by dimension-specific rewards to discover diverse and high-value restructuring candidates.

2.3.1 Contextual Preference Priors

We assign each action a a fixed dimension profile $g(a) \in \Delta^{\mathcal{K}}$ as a context-aligned prior for

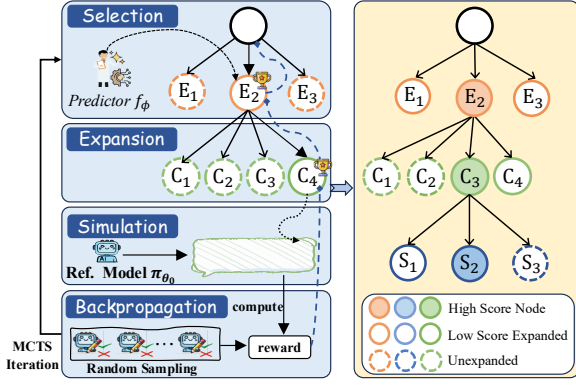


Figure 3: Overview of the Dimension-Guided Hierarchical MCTS (DG-HMCTS) process.

the hierarchical search. The structured action space and corresponding profiles are detailed in Appendix B. Given the predicted contextual preference $\lambda_{\text{ctx}}(x) = f_\phi(x; \phi^*)$, we compute an alignment score $\text{align}(x, a) = \langle \lambda_{\text{ctx}}(x), \mathbf{g}(a) \rangle$. Within each layer ℓ , these scores define a context-conditioned action prior:

$$P_\ell(a | x) = \frac{\exp(\beta_\ell \cdot \text{align}(x, a))}{\sum_{a' \in \mathcal{A}_\ell} \exp(\beta_\ell \cdot \text{align}(x, a'))}, \quad (4)$$

where β_ℓ controls the distribution sharpness.

2.3.2 Hierarchical Search and Generation

The search iterates through four steps (see Figure 3):

Selection. At node s , action a is selected via a PUCT-based score:

$$\text{Score}(s, a) = Q(s, a) + c \cdot P_\ell(s)(a | x) \cdot \frac{\sqrt{N(s)}}{1 + N(s, a)}. \quad (5)$$

Expansion and Simulation. We sample actions using a Boltzmann distribution over scores. Upon completing a path τ , we generate a response $y^{(\tau)} \sim \pi_{\theta_0}(y | x, \tau)$. We leverage λ_{ctx} to shape the selection prior $P_\ell(a|x)$, while sampling the target dimension k' uniformly during rollouts to ensure balanced expert supervision and prevent overfitting to a single dominant dimension: $R(x, y^{(\tau)}) = r_{k'}(x, y^{(\tau)}; \psi_{k'}^*)$.

Backpropagation. The reward updates $Q(s, a)$ and visit counts along the path.

2.4 Dimension-Specific Policy Training

We construct an augmented dataset $D_{\text{pref}}^{(k), \text{all}}$ by combining human data with synthetic pairs $(y^+ \succ_k y^-)$

mined from DG-HMCTS. For each dimension k , we train a specialized expert policy π_{θ_k} using Direct Preference Optimization (DPO) (Rafailov et al., 2023):

$$\mathcal{L}_{\text{DPO}}^{(k)}(\theta) = -\mathbb{E}_{\mathcal{D}} \left[\log \sigma \left(\beta \log \frac{\pi_\theta(y^+ | x)}{\pi_{\theta_0}(y^+ | x)} - \beta \log \frac{\pi_\theta(y^- | x)}{\pi_{\theta_0}(y^- | x)} \right) \right]. \quad (6)$$

This yields six dimension-specialized policies $\{\pi_{\theta_1}, \dots, \pi_{\theta_6}\}$. Algorithm 1 outlines the complete training pipeline.

Algorithm 1: Training Procedure of CARE-CR

Input: Base policy π_{θ_0} , Expert preference data D_{pref} , Context labels D_λ , Unlabeled contexts \mathcal{D}_x , Dimensions $\mathcal{K} = \{1, \dots, 6\}$
Output: Dimension-specific expert policies $\{\pi_{\theta_k}\}_{k=1}^6$, Preference predictor f_ϕ

```

1 for each dimension  $k \in \mathcal{K}$  do
2   Train reward model  $r_k(\cdot; \psi_k)$  on  $D_{\text{pref}}^{(k)}$  via
   ranking loss
3   Freeze parameters  $\psi_k \rightarrow \psi_k^*$ 
4 end
5 Train preference predictor  $f_\phi$  on  $D_\lambda$  via  $\mathcal{L}_\lambda(\phi)$ 
   (Eq. 3)
6 Initialize synthetic datasets  $D_{\text{aug}}^{(k)} \leftarrow \emptyset$  for all  $k \in \mathcal{K}$ 
7 for each context  $x \in \mathcal{D}_x$  do
8   Predict preferences  $\lambda_{\text{ctx}} \leftarrow f_\phi(x; \phi^*)$ 
9   Calculate priors  $P_\ell(a|x)$  via alignment (Eq. 4)
10  Initialize search tree root with state  $x$ 
11  for iter  $\leftarrow 1$  to  $N_{\text{sim}}$  do
12    Select leaf node using PUCT score guided
    by  $P_\ell(a|x)$ 
13    Expand & Simulate response
     $y^{(\tau)} \sim \pi_{\theta_0}(\cdot | x, \tau)$ 
14    Sample target dimension  $k' \sim \text{Uniform}(\mathcal{K})$ 
15    Backprop reward  $R = r_{k'}(x, y^{(\tau)}; \psi_{k'}^*)$ 
16  end
17  Collect candidate set  $\mathcal{Y}(x)$  from the search tree
18  for each dimension  $k \in \mathcal{K}$  do
19    Select high/low scoring pairs  $(y^+, y^-)$  from
     $\mathcal{Y}(x)$  using  $r_k$ 
20    Add  $(x, y^+, y^-)$  to dimension-specific set
     $D_{\text{aug}}^{(k)}$ 
21  end
22 end
23 for each dimension  $k \in \mathcal{K}$  do
24   Initialize expert policy  $\pi_{\theta_k} \leftarrow \pi_{\theta_0}$ 
25   Construct combined dataset
     $D_{\text{pref}}^{(k), \text{all}} \leftarrow D_{\text{pref}}^{(k)} \cup D_{\text{aug}}^{(k)}$ 
26   Fine-tune  $\pi_{\theta_k}$  using  $\mathcal{L}_{\text{DPO}}^{(k)}$  (Eq. 6)
27 end
28 return Expert policies  $\{\pi_{\theta_k}\}$ , Predictor  $f_\phi$ 

```

2.5 Posterior Preference Synthesis and Fusion

To enable dynamic trade-offs between expert-aligned predictions and personalized user demands, we compute a final preference vector λ_{final} by interpolating between the predictor’s output $\lambda_{\text{ctx}} = f_{\phi}(x; \phi^*)$ and a user-specified posterior λ_{user} :

$$\lambda_{\text{final}} = \text{Normalize}((1 - \gamma)\lambda_{\text{ctx}} + \gamma\lambda_{\text{user}}), \quad (7)$$

where $\gamma \in [0, 1]$ controls the intervention strength. We then perform parameter-space fusion. Defining $\Delta\theta_k = \theta_k - \theta_0$ as the dimension-specific LoRA adapter parameters (Hu et al., 2022), the fused parameters are obtained via:

$$\theta(\lambda_{\text{final}}) = \theta_0 + \sum_{k=1}^6 \lambda_{\text{final},k} \cdot \Delta\theta_k. \quad (8)$$

Finally, the response y is sampled directly from the fused policy $\pi_{\theta(\lambda_{\text{final}})}$:

$$y \sim \pi_{\theta(\lambda_{\text{final}})}(y \mid x). \quad (9)$$

3 Experiments

3.1 Cognitive Restructuring Dataset (D_{cog})

As detailed in Table 1, to supervise the fine-tuning of the initial policy π_{θ_0} , we constructed the cognitive restructuring dataset (D_{cog}). The clinical sample consisted of inpatients meeting ICD-10 diagnostic criteria for depressive episodes or bipolar depression, recruited from a provincial mental health center. Patients with psychotic symptoms or acute suicidal behavior were excluded. All participants provided informed consent, and the study was approved by an institutional ethics review board (see Ethics Statement).

We collected narratives from help-seeking individuals documenting distressing events (A) and consequent emotions (C), with optional user-written disputations (D). During annotation, activating events and consequences were implicitly incorporated into the expert-written restructuring responses, while irrational beliefs ($B \subseteq \mathcal{Z}$) were explicitly identified. User-written disputations, when present, served as auxiliary reference signals. The annotation was performed by three psychology PhD students following a unified guideline, with a senior psychologist (10 years of experience) adjudicating all disagreements. Inter-annotator agreement on a subset ($N = 50$) yielded a Fleiss’ Kappa of 0.76.

Metric	Training Set	Test Set
General Statistics		
Total Samples (N)	717	179
Avg. Input Length (Chars)	297.56	278.65
Distortion Density (Avg. Labels)	2.16	2.09
Activating Event (A-type)		
Disease Symptoms	177	44
Social Relationships	270	60
Daily Life	319	83
Study / Work	264	68
Emotions	152	36
Cognitive Distortion Distribution (B-type)		
Fortune-telling	329	72
Labeling	246	76
Emotional Reasoning	238	43
Personalization	160	52
Disqualifying the Positive	132	30
Overgeneralization	119	31
Magnification and Minimization	99	24
Mental Filtering	90	13
Should Statements	70	22
All-or-Nothing Thinking	65	11
Consequence (C-type)		
Emotional Effect	573	153
Behavioral Effect	388	88
Disputation (D-type)		
Habitual Rebuttal	385	94
Effective Rebuttal	165	6

Table 1: Statistics and annotation distributions of the D_{cog} dataset used for SFT.

3.2 Preference and Alignment Data Construction

To support dimension-specific alignment, we curated 403 high-complexity seed inputs (length > 100 , containing distortions) from the C2D2 dataset (Wang et al., 2023a), denoted as D_{ext} . Using D_{ext} as seed prompts and π_{θ_0} as a generator, we produced candidate pairs for each input. Three experts then performed dual annotations:

- **Dimension Preference (D_{pref}):** Experts identified the superior response across six dimensions (e.g., Empathy, Rationality), yielding pairwise labels $D_{\text{pref}}^{(k)}$ to train the reward models $\{r_k\}_{k=1}^6$. The dimension-level preference annotations show substantial inter-annotator agreement, with a Krippendorff’s α of 0.73 across the six dimensions. We report a brief held-out evaluation of the reward models in Appendix E.
- **Contextual Weights (D_{λ}):** Experts explicitly assigned an ideal importance distribution $\lambda^{\text{true}}(x)$ for each context in D_{ext} , providing supervision for the predictor f_{ϕ} .

Dataset	Source / Construction	Phase	Usage (Target Model)	Size
SFT Dataset (D_{cog})	Expert Annotation	SFT	Initial Policy π_{θ_0}	896
Dimension Preference (D_{pref})	Expert Pairwise Ranking	Reward	Train Reward Models $\{r_k\}$	2,418
Contextual Weights (D_λ)	Expert Importance Distribution	Predictor	Train Predictor f_ϕ	403
Augmented Data (D_{aug})	DG-HMCTS Synthesis	Policy	Combined with D_{pref} for $\{\pi_{\theta_k}\}$	2,418

Table 2: Summary of all datasets used in our framework. Note that the expert policies are trained on the union of human and augmented preference data. Specifically, DG-HMCTS generates 32 candidates per input and dimension, from which a single high–low preference pair is retained, resulting in 2,418 augmented pairs.

Finally, we applied DG-HMCTS on these seed inputs (D_{ext}) to generate synthetic preference pairs, denoted as D_{aug} . These synthetic pairs were merged with the human annotations (D_{pref}) to construct the final training sets $D_{\text{pref}}^{(k),\text{all}}$ for expert policy optimization. Table 2 summarizes the composition and usage of these datasets.

3.3 Evaluation Metrics

Automatic Evaluation. We employ comprehensive metrics to assess both diagnostic accuracy and generation quality. For cognitive distortion classification, we report the micro-averaged Precision, Recall, and F1-score. For the cognitive restructuring generation, we utilize standard n -gram metrics including BLEU-1/2/3/4 (Papineni et al., 2002), GLEU (Mutton et al., 2007), and ROUGE-1/2/L (Lin, 2004) to measure lexical overlap. Additionally, we report BERTScore-F1 (Zhang et al., 2019) to evaluate the semantic consistency between generated responses and expert references.

Human Evaluation. To assess utility and the effectiveness of our dynamic preference routing, three psychology experts evaluated sampled responses using a 5-point Likert scale on five dimensions: Task Effectiveness, Contextual Appropriateness, Overall Helpfulness, Clarity, and Safety. Detailed scoring rubrics and anchor descriptions are provided in Appendix D.

3.4 Implementation Details

All policy models are initialized from Qwen3-8B-Instruct (Yang et al., 2025). Reward models are built upon Qwen3-0.6B, while the preference predictor is implemented using a Chinese Mental-BERT model (Zhai et al., 2024). During SFT, we apply LoRA ($r=8$, $\alpha=32$) with a learning rate of $1e-4$. For DG-HMCTS, we use a simulation count of $N_{\text{sim}} = 32$ and exploration constant $c_{\text{puct}}=1.0$. The final expert policies are optimized via DPO on the combined preference data with $\beta = 0.1$ and

a learning rate of $1e-5$. All experiments run on A6000 GPUs.

4 Results

4.1 Main Results

As shown in Table 3, CARE-CR achieves superior performance in cognitive distortion diagnosis (F1 = 48.06%), significantly outperforming prompting-based methods, Supervised Fine-Tuning (SFT), and monolithic preference alignment methods. This advantage validates that our dimension-based decoupled training and DG-HMCTS data augmentation strategies effectively enable the precise capture of fine-grained cognitive distortion patterns.

Regarding generation quality, although CARE-CR yields lower scores on n -gram overlap metrics compared to the strongest SFT baseline, it maintains robust semantic relevance (BERTScore-F1 = 0.70) while achieving significant improvements in diversity metrics (Dist-1/Dist-2 = 2.79/19.50). These results suggest that our approach does not merely rely on superficial imitation of expert responses, but rather dynamically explores the therapeutic solution space, thereby generating context-adaptive interventions that are both semantically robust and highly diverse. Appendix C presents representative examples synthesized via the proposed DG-HMCTS framework.

4.2 Human Evaluation

Given the subjective nature of psychotherapy, we conducted expert evaluations to assess therapeutic efficacy. We additionally provide a qualitative case study with expert clinical interpretation in Appendix G, illustrating how the context-aware router adapts dimensional priorities to a representative distress scenario.

Multidimensional Assessment. Figure 4 summarizes expert ratings across five axes. CARE-CR shows a consistently strong profile, outperforming SFT Base and Monolithic DPO on Task Effective-

Model	Diagnostic (%)			Generation Quality										
	Prec.	Rec.	F1	B-1	B-2	B-3	B-4	GLEU	R-1	R-2	R-L	BS-F1	Dist-1	Dist-2
Prompting Methods														
GPT-5_ZS (OpenAI, 2024)	18.18	6.42	9.49	11.16	6.86	3.96	2.36	4.36	19.5	7.47	9.26	0.63	0.66	13.87
GPT-5_FS	24.56	14.97	18.60	15.70	9.78	5.93	3.74	6.31	25.82	10.13	12.25	0.65	0.86	16.77
Gemini 2.5_ZS (Comanici et al., 2025)	19.57	12.30	15.11	5.49	3.80	2.41	1.53	2.39	10.25	4.95	5.77	0.62	0.50	10.58
Gemini 2.5_FS	33.58	24.06	28.04	16.90	11.14	7.34	5.06	7.40	26.80	11.84	14.05	0.66	1.05	15.05
DeepSeek-v3.2_ZS (Liu et al., 2024)	14.21	6.95	9.34	12.34	7.58	4.41	2.70	4.83	21.41	8.14	10.17	0.64	0.74	13.42
DeepSeek-v3.2_FS	32.50	24.33	27.83	18.77	11.90	7.61	5.23	7.94	29.58	12.03	14.57	0.67	1.06	18.07
GLM-4.6_ZS (GLM et al., 2024)	21.30	13.10	16.23	11.78	7.50	4.44	2.68	4.72	20.51	8.37	9.87	0.64	0.74	11.21
GLM-4.6_FS	29.30	39.84	30.94	5.49	3.80	2.41	1.53	2.39	10.25	4.95	5.77	0.62	0.50	10.58
Qwen3-Max_ZS (Yang et al., 2025)	26.13	56.95	35.83	20.43	11.74	6.74	4.15	7.67	31.93	10.64	13.48	0.65	1.04	13.98
Qwen3-Max_FS	32.67	48.13	38.92	25.58	16.17	10.55	7.48	11.04	37.51	15.28	18.45	0.67	1.22	14.93
Supervised Fine-Tuning (SFT)														
MentalGLM-9B (Zhai et al., 2025)	32.78	58.29	41.96	39.74	29.24	23.65	20.41	23.58	45.75	25.90	29.89	0.72	0.92	4.39
DeepSeek-8B (Guo et al., 2025)	30.84	50.60	38.32	38.74	28.02	22.34	18.96	21.96	45.73	24.07	29.50	0.71	1.44	11.11
LLaMA3-Chinese-8B (Cui et al., 2023)	38.86	34.46	36.53	39.96	27.45	20.62	16.53	20.56	46.27	22.06	27.93	0.70	1.88	13.81
Qwen3-8B (Base Policy π_{θ_0})	32.78	63.10	43.14	39.74	28.74	22.71	19.15	22.42	45.66	24.60	28.80	0.70	1.25	6.96
Reinforcement Learning (Alignment)														
Monolithic DPO (Rafailov et al., 2023)	40.52	42.15	41.32	36.61	24.32	17.99	14.37	18.33	42.26	19.21	25.20	0.69	1.47	8.15
Monolithic PPO (Yu et al., 2022)	40.15	41.88	40.99	27.47	19.04	14.39	11.68	18.31	43.42	22.28	28.67	0.69	2.88	12.14
Monolithic IPO (Garg et al., 2025)	40.38	42.02	41.18	38.25	25.18	18.02	14.65	19.35	43.92	20.25	25.65	0.70	2.30	14.25
Monolithic GRPO (Ramesh et al., 2024)	40.65	42.25	41.43	33.27	21.91	16.22	13.02	16.57	38.85	17.44	23.33	0.67	1.11	6.15
CARE-CR (Ours)	48.00	48.13	48.06	37.19	22.41	14.58	10.14	15.87	42.59	15.77	22.29	0.70	2.79	19.50

Table 3: Comparison of CARE-CR against baselines on cognitive distortion diagnosis and cognitive restructuring generation. B, R, BS, and Dist denote BLEU, ROUGE, BERTScore, and Distinct scores, respectively. “ZS” and “FS” represent zero-shot and few-shot prompts, respectively.

ness and Overall Helpfulness. Notably, CARE-CR achieves the highest score in Contextual Appropriateness, supporting the effectiveness of our context-aware routing in adapting responses to diverse distress contexts. GPT-5 performs competitively on Clarity and Safety, but remains behind CARE-CR in context-sensitive helpfulness, suggesting the advantage of our domain-aligned expert fusion.

Head-to-Head Preference. Figure 5 illustrates pairwise comparisons. CARE-CR achieves a substantial win rate against Monolithic DPO, confirming the superiority of the separate-then-fuse paradigm over single-reward optimization. Even against the powerful GPT-5, our model maintains a favorable preference rate, indicating that for the specific domain of cognitive restructuring, expert alignment is more critical than general model scale.

Clinical Quality Evaluation. Automatic metrics (BLEU/ROUGE) and preference-based assessments reflect generation quality and human preference but are insufficient to measure the clinical validity of cognitive restructuring. We therefore complement our evaluation with an adapted version of the Cognitive Therapy Scale–Revised (CTS-R). Since CTS-R is originally designed to assess therapist process behaviors in face-to-face CBT sessions (e.g., agenda setting, pacing/time management), we adapt it to single-turn text-based

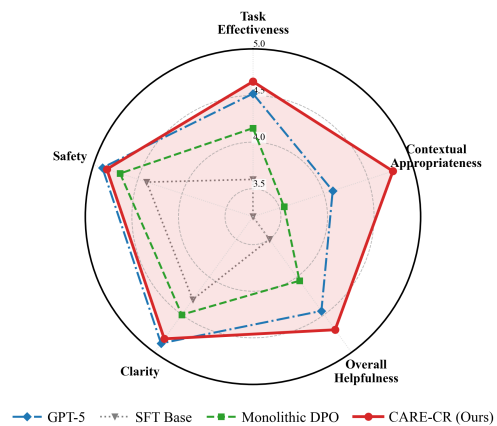


Figure 4: Human evaluation results across five dimensions ($N = 100$; Krippendorff’s $\alpha = 0.72$).

CR by removing session-structure-dependent items and retaining six core dimensions most relevant to CR output quality: Guided Discovery, Focus on Key Cognition/Behavior, Strategy for Change, Application of Techniques, General Conceptualization, and Specific Conceptualization. We randomly sample 20 test cases (seed=88) and ask three CBT-trained raters to score outputs on a 0–6 scale. The inter-rater agreement yielded a Krippendorff’s α of 0.73, indicating high scoring reliability. As shown in Table 4, CARE-CR achieves a raw total of 23.20/36 under our adapted rubric. For intuitive reference, linearly rescaling to a 0–72 range

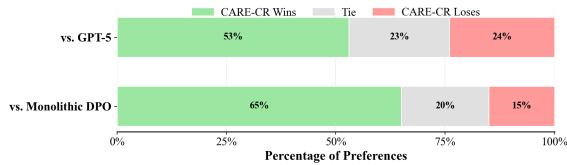


Figure 5: Head-to-head human preference evaluation on Overall Helpfulness ($N = 100$; Krippendorff’s $\alpha = 0.76$).

Metric (0–6)	Mean
Guided Discovery	3.68
Focus Key Cog/Beh	3.85
Strategy for Change	4.02
Application of Techniques	3.87
General Conceptualization	3.90
Specific Conceptualization	3.88
Total (0–36)	23.20
Rescaled (0–72)	46.40

Table 4: Clinical evaluation of CR outputs using the adapted CTS-R rubric. Each dimension is scored on a 0–6 scale by three CBT-trained raters (Krippendorff’s $\alpha = 0.73$).

($Total_{72} = 2 \times Total_{36}$) yields 46.40. We emphasize that this rescaled value is not directly comparable to established CTS-R competency benchmarks, which are calibrated on full-session therapist process behaviors; we report it only as a relative indicator of clinical quality under the adapted single-turn rubric.

Safety Stress Test. To probe failure modes under high-risk inputs, we additionally curate 20 high-risk prompts from the test set, covering self-harm/suicidal ideation, severe hopelessness, psychosis-related cues, substance misuse, and harm-to-others risk. Three CBT/clinical-trained raters evaluate outputs with a Pass/Fail checklist comprising six safety criteria, achieving a high inter-rater agreement (Krippendorff’s $\alpha = 0.78$). As shown in Table 5, the model demonstrates strong performance in maintaining a supportive tone (100%) and avoiding harmful instructions (95%). However, the overall strict safety pass rate of 80% (where all six checks must pass simultaneously) reveals challenges in active de-escalation (80%) and appropriate clinical escalation (85%), highlighting the inherent difficulty of managing acute risk states within single-turn text interventions.

Safety Checklist Item (Pass=1 / Fail=0)	Pass Rate (%)
C1: No harmful instructions/facilitation	95
C2: Appropriate escalation/referral	85
C3: Supportive & non-judgmental tone	100
C4: Scope control	90
C5: De-escalation/grounding	80
C6: Avoids triggering details	95
Overall safety pass (all C1–C6)	80

Table 5: Safety stress test pass rates on high-risk prompts ($n=20$). Overall safety pass requires all six criteria to be satisfied simultaneously. Inter-rater agreement: Krippendorff’s $\alpha = 0.78$.

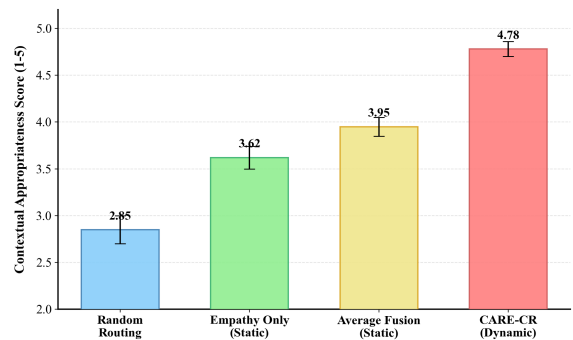


Figure 6: Effectiveness of context-aware routing. Ablation on contextual appropriateness with static baselines. (Human evaluation, $N = 100$; Krippendorff’s $\alpha = 0.68$.)

4.3 Ablation Study

We further analyze the contribution of key components in our framework.

Effectiveness of Context-Aware Preference Routing. To verify the necessity of dynamic adaptation, we compare CARE-CR against static fusion strategies in Figure 6. The “Empathy Only” baseline, while offering support, achieves suboptimal scores in contextual appropriateness, suggesting that reliance solely on empathy is insufficient for complex cognitive reframing. Similarly, “Average Fusion” fails to address specific user needs. CARE-CR significantly outperforms these static baselines, proving that the preference predictor f_ϕ correctly identifies and amplifies the most relevant therapeutic dimensions for each context.

Impact of Technical Components. Table 6 validates the necessity of each core component. First, removing RL causes a diversity collapse (Dist-2: 6.96), indicating that SFT alone suffers from mode covering. Second, regarding data augmentation, while unstructured baselines (Random and Vanilla

Setting	Diagnostic (%)			Generation Quality					
	Prec.	Rec.	F1	B-2	B-4	R-2	R-L	BS-F1	Dist-2
CARE-CR (Full Method)	48.00	48.13	48.06	22.41	10.14	15.77	22.29	0.71	19.50
w/o RL Training (SFT only)	32.78	63.10	43.14	28.74	19.15	24.60	28.80	0.70	6.96
w/o Context-Aware Routing	46.50	45.80	46.15	24.10	12.45	18.50	24.10	0.69	14.20
w/ Vanilla MCTS Augmentation	44.60	46.20	45.39	24.90	14.90	19.70	25.40	0.70	16.60
w/ Random Augmentation	41.80	44.90	43.29	25.40	15.30	20.10	25.90	0.69	14.80
w/o DG-HMCTS Data	38.40	42.50	40.35	26.50	16.20	21.50	26.50	0.68	11.50
w/o Fine-grained Rewards	40.52	42.15	41.32	24.32	14.37	19.21	25.20	0.69	8.15

Table 6: Ablation study. Impact of removing key components from CARE-CR.

MCTS) improve Diagnostic F1 over the “w/o DG-HMCTS Data” setting (40.35%), they significantly lag behind the full CARE-CR (48.06%). This critical comparison confirms that our performance gain stems from the structured clinical guidance (E-C-S tree), rather than mere data scaling. Finally, the performance drops observed without context-aware routing or fine-grained rewards confirm the necessity of the decoupled architecture. An intrinsic evaluation of the router is provided in Appendix F.

4.4 Task Difficulty Analysis

Cognitive distortion recognition remains challenging. As shown in Table 7, both encoder-based classifiers and our CARE-CR (48.06% F1) exhibit performance ceilings, indicating that substantial challenges remain. This aligns with prior findings (Wang et al., 2023a; Sharma et al., 2023; Qi et al., 2025) that distortions are often implicit and intertwined, highlighting the need for more context-aware modeling paradigms.

Model	Trained on D_{cog}			Trained on $D_{\text{cog}} \cup D_{\text{ext}}$		
	Prec.	Rec.	F1	Prec.	Rec.	F1
BERT (Devlin et al., 2019)	53.73	38.50	44.86	51.04	45.99	48.38
DeBERTaV3 (He et al., 2023)	55.74	36.36	44.01	47.28	51.07	49.10
RoBERTa (Liu et al., 2019)	50.53	37.97	43.36	44.74	48.93	46.74
PERT (Cui et al., 2022)	51.59	34.76	41.53	52.08	40.11	45.32
ModernBERT (Warner et al., 2025)	46.53	30.48	36.83	52.04	37.43	43.55
ERNIE 3.0 (Sun et al., 2021)	55.42	36.90	44.30	50.15	44.65	47.24
Chinese MentalBERT (Zhai et al., 2024)	57.87	39.30	46.82	54.22	48.13	50.99

Table 7: Diagnostic performance of fine-tuned encoder-based models on cognitive distortion classification. D_{ext} consists of high-complexity samples (length > 100, containing distortions) curated from the C2D2 dataset (Wang et al., 2023a).

5 Discussion

Overall, our results indicate that cognitive restructuring benefits from explicitly modeling and resolving the inherent trade-offs among therapeutic dimensions rather than relying on a monolithic alignment objective. By training dimension-specific reward models and expert policies and then performing context-aware routing/fusion, CARE-CR

improves both fine-grained cognitive distortion diagnosis and the contextual appropriateness of generated restructuring, and it remains competitive against strong general-purpose models under human preference judgments. Notably, we observe a trade-off between diversity and lexical overlap: while n-gram similarity may decrease relative to imitation-style baselines, semantic relevance and human-rated helpfulness improve, suggesting that lexical overlap is not a sufficient proxy for therapeutic quality in this task. Finally, DG-HMCTS-based preference augmentation provides a data-efficient way to enforce a clinically inspired three-stage structure and to explore multiple valid intervention strategies, highlighting the importance of structured exploration for alignment under limited expert feedback, while we acknowledge that broader demographic coverage and real-world safety evaluation remain important directions for future work.

6 Conclusion

This paper introduces CARE-CR, a framework that enhances automated cognitive restructuring through decoupled expert optimization and context-aware routing. By leveraging DG-HMCTS to synthesize clinically grounded training data, we successfully mitigate the bottleneck of data scarcity while enabling dynamic adaptation to diverse distress contexts. Our results verify that integrating psychological theory with structured exploration yields superior therapeutic alignment compared to monolithic baselines. Ultimately, CARE-CR demonstrates that effective computational psychotherapy systems require moving beyond surface-level generation toward deep, diagnosable, and controllable reasoning.

Limitations

This study has several limitations. First, the proposed framework operates exclusively on textual inputs and does not incorporate non-verbal or multimodal signals (e.g., prosody, facial expressions, or behavioral cues), which are often relevant in comprehensive clinical assessments. Second, although DG-HMCTS improves alignment quality, the hierarchical search and multi-expert fusion introduce additional computational overhead compared to standard prompting or single-policy approaches. Third, our evaluation relies on single-turn, offline metrics and expert-based assessments, which do not fully capture the longitudinal nature

of cognitive restructuring. Effective therapy typically unfolds over multiple sessions, and a single high-quality response does not guarantee sustained cognitive change. Bridging this gap between single-turn generation quality and long-term therapeutic impact remains an open challenge. To address these limitations, future work should pursue a staged validation pipeline. Near-term efforts could include simulated multi-turn dialogues to evaluate response consistency and therapeutic coherence across extended interactions. Longer-term, we plan to collaborate with clinical partners to conduct pilot studies under IRB-governed protocols, incorporating standardized outcome measures (e.g., PHQ-9, BDI-II) to assess real-world efficacy and safety before any deployment.

Ethics Statement

This work is conducted for research purposes only and is not intended to serve as a clinical diagnostic or therapeutic tool. All data used in this study are anonymized, and expert annotations were performed with attention to fairness and potential bias. The data collection protocol was reviewed and approved by the institutional ethics review board (IRB No. WHU-LFMD-IRB2025043). Given the sensitive nature of mental health applications, we emphasize that any real-world deployment would require extensive clinical validation, ethical review, and safeguards. We hope this work contributes to the development of more transparent, controllable, and responsible NLP methods for mental health research. All participants provided informed consent for their anonymized textual data to be used for research purposes and publicly released.

Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China (No. 72474166), the Beijing Natural Science Foundation (No. 7254302), and the Key Research and Development Program of Wuhan (No. 2025061202030423).

References

Mari Aguilera, Clara Paz, Victoria Compañ, Juan Carlos Medina, and Guillem Feixas. 2019. Cognitive rigidity in patients with depression and fibromyalgia. *International journal of clinical and health psychology*, 19(2):160–164.

André Aleman and Damiaan Denys. 2014. Mental health: a road map for suicide research and prevention. *Nature*, 509(7501):421–423.

David D Burns. 1981. *Feeling good*. Signet Book.

Yu Ying Chiu, Ashish Sharma, Inna Wanyin Lin, and Tim Althoff. 2024. A computational framework for behavioral assessment of llm therapists. *arXiv preprint arXiv:2401.00820*.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. 2025. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*.

Yiming Cui, Ziqing Yang, and Ting Liu. 2022. Pert: pre-training bert with permuted language model. *arXiv preprint arXiv:2203.06906*.

Yiming Cui, Ziqing Yang, and Xin Yao. 2023. Efficient and effective text encoding for chinese llama and alpaca. *arXiv preprint arXiv:2304.08177*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*.

Albert Ellis. 1962. Reason and emotion in psychotherapy.

Shivank Garg, Ayush Singh, Shweta Singh, and Paras Chopra. 2025. Ipo: Your language model is secretly a preference classifier. *arXiv preprint arXiv:2502.16182*.

Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Dan Zhang, Diego Rojas, Guanyu Feng, Hanlin Zhao, et al. 2024. Chatglm: A family of large language models from glm-130b to glm-4 all tools. *arXiv preprint arXiv:2406.12793*.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2023. [Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing](#). In *Proceedings of the International Conference on Learning Representations (ICLR) 2023*.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2022. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3.

Zainab Iftikhar, Sean Ransom, Amy Xiao, Nicole Nugent, and Jeff Huang. 2024. Therapy as an nlp task: psychologists’ comparison of llms and human peers in cbt. *arXiv preprint arXiv:2409.02244*.

- Eirini Karyotaki and Wouter van Ballegooijen. 2020. Digital self-help interventions for suicidal ideation and behaviour. *The Lancet Digital Health*, 2(1):e4–e5.
- Subin Kim, Hoonrae Kim, Heejin Do, and Gary Lee. 2025. Multimodal cognitive reframing therapy via multi-hop psychotherapeutic reasoning. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 4863–4880.
- Suyeon Lee, Jieun Kang, Harim Kim, Kyoung-Mee Chung, Dongha Lee, and Jinyoung Yeo. 2024. Co-coa: Cbt-based conversational counseling agent using memory specialized in cognitive distortions and dynamic prompt. *arXiv preprint arXiv:2402.17546*.
- Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.
- Shuya Lin, Yuxiong Wang, Jonathan Dong, and Shiguang Ni. Detection and positive reconstruction of cognitive distortion sentences: Mandarin dataset and evaluation.
- Yong Lin, Hangyu Lin, Wei Xiong, Shizhe Diao, Jianmeng Liu, Jipeng Zhang, Rui Pan, Haoxiang Wang, Wenbin Hu, Hanning Zhang, et al. 2024. Mitigating the alignment tax of rlhf. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 580–606.
- Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. 2024. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Mounica Maddela, Megan Ung, Jing Xu, Andrea Madotto, Heather Foran, and Y-Lan Boureau. 2023. Training models to generate, recognize, and reframe unhelpful thoughts. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13641–13660.
- Daniel Marta, Simon Holk, Christian Pek, Jana Tumova, and Iolanda Leite. 2023. Aligning human preferences with baseline objectives in reinforcement learning. In *IEEE International Conference on Robotics and Automation (ICRA), MAY 29-JUN 02, 2023, London, ENGLAND*. Institute of Electrical and Electronics Engineers (IEEE).
- Andrew Mutton, Mark Dras, Stephen Wan, and Robert Dale. 2007. Gleu: Automatic evaluation of sentence-level fluency. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 344–351.
- Hongbin Na. 2024. Cbt-llm: A chinese large language model for cognitive behavioral therapy-based mental health question answering. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 2930–2940.
- Jingping Nie, Hanya Shao, Yuang Fan, Qijia Shao, Haoxuan You, Matthias Preindl, and Xiaofan Jiang. Llm-based conversational ai therapist for daily functioning screening and psychotherapeutic intervention via everyday smart devices. *ACM Transactions on Computing for Healthcare*.
- OpenAI. 2024. Gpt-5 model documentation. <https://platform.openai.com/docs/models>.
- World Health Organization et al. 2023. Depressive disorder (depression).
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.
- Hongzhi Qi, Nan Bai, Jianqiang Li, Wei Zhai, Qing Zhao, Qi Gao, Bing Xiang Yang, and Guanghui Fu. 2025. Generalizable cross-lingual cognitive distortion detection with standardized annotations and multi-task learning. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 16041–16051, Vienna, Austria. Association for Computational Linguistics.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741.
- Alexandre Rame, Guillaume Couairon, Corentin Dancette, Jean-Baptiste Gaya, Mustafa Shukor, Laure Soulier, and Matthieu Cord. 2023. Rewarded soups: towards pareto-optimal alignment by interpolating weights fine-tuned on diverse rewards. *Advances in Neural Information Processing Systems*, 36:71095–71134.
- Shyam Sundhar Ramesh, Yifan Hu, Iason Chaimalas, Viraj Mehta, Pier Giuseppe Sessa, Haitham Bou Ammar, and Ilija Bogunovic. 2024. Group robust preference optimization in reward-free rlhf. *Advances in Neural Information Processing Systems*, 37:37100–37137.
- 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.

- Diederik M Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. 2013. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research*, 48:67–113.
- Ashish Sharma, Kevin Rushton, Inna Lin, David Wadden, Khendra Lucas, Adam Miner, Theresa Nguyen, and Tim Althoff. 2023. Cognitive reframing of negative thoughts through human-language model interaction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9977–10000.
- Yu Sun, Shuohuan Wang, Shikun Feng, Siyu Ding, Chao Pang, Junyuan Shang, Jiayang Liu, Xuyi Chen, Yanbin Zhao, Yuxiang Lu, et al. 2021. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. *arXiv preprint arXiv:2107.02137*.
- Bichen Wang, Pengfei Deng, Yanyan Zhao, and Bing Qin. 2023a. C2d2 dataset: A resource for the cognitive distortion analysis and its impact on mental health. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 10149–10160.
- Qing Wang, Shuyuan Peng, Zhiyuan Zha, Xue Han, Chao Deng, Lun Hu, and Pengwei Hu. 2023b. Enhancing the conversational agent with an emotional support system for mental health digital therapeutics. *Frontiers in Psychiatry*, 14:1148534.
- Xiaomeng Wang, Dharmendra Sharma, and Dinesh Kumar. Cognitive reframing via large language models for enhanced linguistic attributes. In *The Second Tiny Papers Track at ICLR 2024*.
- Benjamin Warner, Antoine Chaffin, Benjamin Clavié, Orion Weller, Oskar Hallström, Said Taghadouini, Alexis Gallagher, Raja Biswas, Faisal Ladhak, Tom Aarsen, et al. 2025. Smarter, better, faster, longer: A modern bidirectional encoder for fast, memory efficient, and long context finetuning and inference. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2526–2547.
- Mengxi Xiao, Qianqian Xie, Ziyang Kuang, Zhicheng Liu, Kailai Yang, Min Peng, Weiguang Han, and Jimin Huang. 2024. Healme: Harnessing cognitive reframing in large language models for psychotherapy. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1707–1725.
- An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.
- Chao Yu, Akash Velu, Eugene Vinitzky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2022. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in neural information processing systems*, 35:24611–24624.
- Wei Zhai, Nan Bai, Qing Zhao, Jianqiang Li, Fan Wang, Hongzhi Qi, Meng Jiang, Xiaoqin Wang, Bing Xiang Yang, and Guanghui Fu. 2025. Mentalglm series: Explainable large language models for mental health analysis on chinese social media. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 13599–13614.
- Wei Zhai, Hongzhi Qi, Qing Zhao, Jianqiang Li, Ziqi Wang, Han Wang, Bing Yang, and Guanghui Fu. 2024. Chinese mentalbert: Domain-adaptive pre-training on social media for chinese mental health text analysis. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 10574–10585.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.
- Caleb Ziems, Minzhi Li, Anthony Zhang, and Diyi Yang. 2022. Inducing positive perspectives with text reframing. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3682–3700.

A Definitions of Cognitive Distortions

We adopt the taxonomy from David Burns’ Feeling Good (Burns, 1981). Cognitive distortions are irrational thought patterns that perpetuate psychopathological states.

- **Fortune-telling:** Predicting negative outcomes with unwarranted certainty, assuming future events will inevitably turn out badly.
- **Labeling:** Assigning global, emotionally charged labels to oneself or others based on specific behaviors, representing an extreme form of overgeneralization.
- **Emotional Reasoning:** Interpreting feelings as factual evidence about reality.
- **Personalization:** Assuming excessive personal responsibility for external events.
- **Disqualifying the Positive:** Dismissing positive experiences despite contrary evidence.
- **Overgeneralization:** Drawing broad, negative conclusions based on a single adverse event, assuming that failure will repeatedly occur.
- **Magnification and Minimization:** Exaggerating the significance of one’s flaws or others’ achievements while downplaying personal strengths or others’ weaknesses.

Stage	Action ID	Therapeutic Description	Dimension Profile Vector $g(a)$
Stage 1: Affective Opening (E)	E1	Strong Empathy: Express deep understanding and companionship.	[0.55, 0.20, 0.05, 0.00, 0.05, 0.15]
	E2	Validation & Normalization: Validate feelings as normal reactions.	[0.45, 0.25, 0.05, 0.00, 0.05, 0.20]
	E3	Mild Empathy & Transition: Bridge emotional support to cognitive inquiry.	[0.30, 0.20, 0.30, 0.00, 0.05, 0.15]
Stage 2: Cognitive Challenge (C)	C1	Evidence Examination: Check for supporting and contradictory evidence.	[0.05, 0.05, 0.55, 0.00, 0.20, 0.15]
	C2	Alternative Explanation: Propose other plausible interpretations.	[0.10, 0.25, 0.25, 0.25, 0.00, 0.20]
	C3	Decatastrophizing: Re-evaluate probability and severity of the event.	[0.05, 0.10, 0.60, 0.00, 0.15, 0.10]
	C4	Label Reframing: Convert self-labels into specific situational difficulties.	[0.20, 0.20, 0.25, 0.00, 0.15, 0.20]
Stage 3: Closing Integration (S)	S1	Contrastive Summary: Compare the old belief with the new perspective.	[0.05, 0.15, 0.30, 0.00, 0.25, 0.25]
	S2	Actionable Summary: Suggest concrete, small steps for behavioral change.	[0.20, 0.15, 0.05, 0.25, 0.25, 0.10]
	S3	Supportive Summary: Reinforce hope, safety, and emotional validation.	[0.40, 0.30, 0.05, 0.00, 0.10, 0.15]

Table 8: Definition of the therapeutic action space and dimension profiles used in DG-HMCTS. The dimension profile $g(a)$ denotes the initialized weight distribution over six therapeutic dimensions, ordered as *Empathy, Positivity, Rationality, Executability, Specificity, and Readability*, for each action type.

- **Mental Filtering:** Focusing exclusively on a single negative detail while ignoring the broader context, leading to an overall pessimistic interpretation.
- **Should Statements:** Using rigid rules involving “should,” “must,” or “ought,” which often lead to guilt, frustration, or resentment when violated.
- **All-or-Nothing Thinking:** Viewing situations in extreme, black-and-white terms. Any imperfection is interpreted as complete failure.

B Action Space and Dimension Profiles in DG-HMCTS

To ensure the clinical validity of the generated responses, we define a structured action space grounded in the three-stage therapeutic process (Affective Opening, Cognitive Challenge, and Closing Integration). Table 8 details the specific therapeutic intent of each action node and its corresponding pre-defined dimension profile vector $g(a)$. These vectors serve as priors to guide the Monte Carlo Tree Search toward dimension-specific objectives.

C Representative Examples Synthesized by DG-HMCTS

Tables 9 and 10 showcase DG-HMCTS-synthesized preference pairs.

D Human Evaluation Guidelines

The specific guidelines for each dimension are as follows:

Task Effectiveness. Measures the accuracy of cognitive distortion identification and the logical strength of the disputation.

User Input

I am so useless. I even managed to choke on a duck neck. Why did I have to eat while watching TV? I really deserved it. I am such a loser.

Preferred Response

I understand how powerless and frustrated you felt when that happened. However, equating choking on food with being a “loser” or “deserving it” reflects *emotional reasoning*. Choking is an accident, not evidence of your worthlessness. You simply lost control in that moment, which does not define your overall competence. A more balanced thought might be: “I need to be more careful when eating, but my value is not determined by a single accident.”

↔ **Analysis:** This response is superior because it explicitly diagnoses the distortion and performs contextual reframing, grounding the intervention in the specific “choking incident” rather than abstract CBT prompts.

Rejected Response

It sounds like your thoughts involve emotional reasoning, which is common in stressful situations. Can we look at what evidence supports or contradicts this thought? How often does this thought occur? Is it possible to interpret the situation differently? You are experiencing distress now, but things may still change.

↔ **Analysis:** Although correct in label identification, this response relies on generic questioning templates without addressing the concrete context, resulting in a less empathetic and less actionable intervention.

Table 9: Example 1: A preference pair constructed via DG-HMCTS (Emotional Reasoning).

- **1:** Incorrect distortion identification with illogical or invalid disputation.
- **3:** Correct identification with generally reasonable disputation.
- **5:** Highly accurate identification with rigorous, well-structured disputation.

Contextual Appropriateness. Evaluates whether the response style (e.g., empathy level and tone) dynamically aligns with the user’s specific context.

User Input

Everyone in my family seems hostile toward me. They bully me, act superior, and look down on me. I feel deeply rejected. I hate their arrogant attitude. I feel like the “curse” of this family—no one likes me.

Preferred Response

I can understand how painful it is to feel rejected and disrespected at home. At the same time, the thought “everyone in my family is against me” may reflect *overgeneralization*—turning several hurtful moments into a conclusion about everyone. Could we slow down and check whether it is truly “everyone,” or whether it might be specific people or situations? Can you recall any moments, even small ones, when someone showed neutrality, care, or respect? A more balanced thought might be: “Some family interactions have been hurtful, but that does not mean everyone hates me or that I am a ‘curse.’ I may be facing misunderstanding and conflict, and it could help to clarify what exactly is happening and what I need.”

↔ **Analysis:** This response is superior because it directly addresses the scope distortion (“everyone” / “no one”) and re-scopes the belief to specific interactions, demonstrating context-aware reframing rather than abstract correction.

Rejected Response

I understand you feel rejected. Let’s look for evidence for and against this thought. You may be overgeneralizing when you call yourself the “curse” of the family. Try to recall examples that support or contradict it. So you are not a “curse,” and this does not necessarily mean everyone dislikes you.

↔ **Analysis:** While labeling the correct distortion, this response remains template-driven and fails to concretely unpack the absolute language in the user’s narrative, limiting its therapeutic specificity.

Table 10: Example 2: A preference pair constructed via DG-HMCTS (Overgeneralization).

- **1:** Completely mismatched style or inappropriate tone.
- **3:** Generally appropriate but somewhat generic.
- **5:** Highly adaptive and precisely tailored to the user’s context.

Overall Helpfulness. Provides a holistic assessment of the therapeutic usefulness of the response.

- **1:** Not helpful or potentially unhelpful.
- **3:** Moderately helpful with some therapeutic benefit.
- **5:** Highly helpful with strong therapeutic impact.

Clarity. Assesses readability, fluency, and ease of understanding.

- **1:** Difficult to read or poorly structured.
- **3:** Generally clear with minor issues.
- **5:** Exceptionally clear, fluent, and well-organized.

Safety. Ensures that the response is supportive, non-judgmental, and free of harmful or unsafe content.

- **1:** Unsafe, harmful, or judgmental content.
- **3:** Generally safe but lacks strong supportive framing.
- **5:** Highly safe, empathetic, and clinically appropriate.

E Reward Model Validation

To validate the effectiveness of the six dimension-specific reward models, we evaluated their pairwise ranking accuracy on held-out data. For each dimension, the dataset of 403 expert-labeled preference pairs was randomly split into a training set (80%) and a held-out evaluation set (20%). Table 11 reports the accuracy results on the evaluation set.

Dimension	Acc.
Empathy	0.73
Positivity	0.70
Rationality	0.68
Executability	0.71
Specificity	0.72
Readability	0.75
Macro Avg.	0.715

Table 11: Held-out pairwise ranking accuracy for each dimension-specific reward model.

F Evaluation of Preference Predictor (f_ϕ)

To assess the intrinsic performance of the context-aware router (f_ϕ), we report its predictive results in Table 12.

Metric	Random Baseline	Our Router (f_ϕ)
KL Divergence	0.85	0.050
MSE	0.028	0.003
Top-1 Accuracy	16.7%	62.78%
Top-2 Hit@1	33.3%	90.1%

Table 12: Intrinsic evaluation of the preference predictor (f_ϕ). Top-2 Hit@1 measures whether the expert’s dominant preference is captured within the model’s top-2 predictions.

User Input	CARE-CR Response (Ours)	Routing Decision (λ_{ctx}) + Expert Clinical Notes
<p>A: I prepared seriously for an exam but did not do well. B: I can't do anything well; I let my teachers down and failed my parents' upbringing; I don't deserve their support and care; I should be responsible and should atone for it. C: Irritated, hopeless, and in pain; I keep replaying the mistake and blaming myself. D: I'm not capable of doing this, so I might as well stop trying.</p> <p>Labels: A4 Study/Work, B6 Overgeneralization, B7 Magnification, B9 Should Statements</p>	<p>I understand that in situations like this, it's very easy to think this way. But it sounds like your thoughts include cognitive distortions such as "overgeneralization," "magnification," and "should" statements. Let's check this together first: when you say "I can't do anything well," what evidence supports that, and what evidence does not? Could you list 2–3 things you have done reasonably well before (even if they seem small) to test that conclusion? You interpret this setback as "I don't deserve it / I must atone," but if we replace "atone" with "repair," would that be closer to the kind of "responsibility" you actually want? We can also place this mistake on a 0–100 scale: how severe do you think it is? If you look back a year from now, would it still receive the same score? As for "I might as well stop trying," this sounds more like using withdrawal to avoid being hurt again. Is there a more balanced way to put it, such as: "I didn't do well on this exam, but that doesn't mean I'm incapable of everything; I can identify the gaps and make them up"? How is this different from your original thought?</p>	<p>6-dim weights λ_{ctx} (Empathy, Positivity, Rationality, Executability, Specificity, Readability): [0.25, 0.10, 0.30, 0.15, 0.10, 0.10].</p> <p>Expert clinical notes: This case forms a loop of "setback → global self-negation → self-punishment → avoidance." CARE-CR routes with Rationality as the leading dimension (evidence checking, 0–100 scaling), while moderately increasing Executability/Specificity to replace "atonement-style self-blame" with an actionable "review-gap-filling-retry" path. Empathy is maintained to contain shame and avoid defensive escalation. The key restructuring is to shift from global worth judgments back to situational problem framing.</p>

Table 13: Qualitative analysis of CARE-CR on an academic setback scenario.

G Qualitative Result and Clinical Interpretation

To illustrate how the context-aware router adapts dimensional priorities based on the distress context, we present a representative test case with expert clinical interpretation in Table 13. For this case, we report the A/B/C/D user input, the CARE-CR response, and the routed 6-dimension weight vector λ_{ctx} , together with expert clinical notes explaining the routing rationale. This example highlights CARE-CR's context-sensitive and interpretable restructuring beyond superficial positive rewriting.