北京理工大学
信息系统及安全对抗实验中心
信息安全与对抗技术研究所
Beijing Forest Studio

认知扭曲干预策略研究

www.isclab.org.cn

博士研究生 陈星星

2026年9月27日

问题回溯

- 总结反思
  - 算法讲解不够深入
  - 无关动作太多影响报告效果

- 相关内容
  - 陈星星《认知扭曲识别研究》——2026.04.06

内容提要

- 预期收获
- 题目内涵解析
- 研究背景与意义
- 研究历史与现状
- 知识基础
- 算法原理
  - CoPoLLM
  - CARE-CR
- 特点总结与工作展望
- 参考文献

预期收获

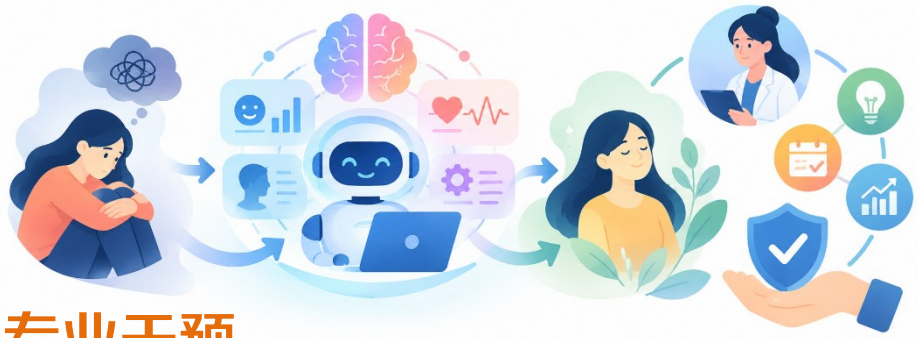
- 预期收获
  - 1.认识大语言模型心理支持从情绪回应到认知干预的发展与挑战
  - 2.理解认知干预中的认知诊断、策略决策与个性化回复生成机制
  - 3.掌握强化学习、多偏好建模与安全约束在心理支持智能体中的典型应用

内涵解析与研究目标

- 内涵解析
  - 认知行为疗法：通过识别和修正不合理认知模式，改善个体情绪与行为反应
  - 认知扭曲：用户对自身、他人或事件形成的系统性思维偏差，是负面情绪和不良行为的重要认知来源
  - 认知重构：采用证据检验、替代解释、去灾难化等方式，对错误认知进行针对性修正与重新解释
- 研究目标
  - 提高认知诊断与策略干预能力
    - 显式建模认知扭曲类型、强度与风险状态
    - 学习不同认知状态下的专业干预策略
  - 提高认知重构的个性化适配能力
    - 建模共情、理性、积极性、可执行性等多维治疗偏好
    - 融合不同治疗专家，实现上下文适配的认知重构

研究背景与意义

- 研究背景
  - 现有心理支持大语言模型主要关注情绪识别与共情回复，对用户潜在认知扭曲及其严重程度缺乏细粒度建模，难以开展针对性的认知行为干预
  - 认知重构同时包含共情、理性、可执行性等多维且可能冲突的治疗目标，固定偏好或单一奖励难以适应不同用户与心理情境
  - 心理专家标注成本高、专业要求强



- 研究意义
  - 推动心理支持大语言模型从情绪陪伴向认知诊断与专业干预
  - 提高模型在复杂心理状态下的策略决策、安全控制与可解释性
  - 为构建具备认知理解、策略规划、个性化生成和风险控制能力的心理支持智能体提供方法基础

研究历史与现状

Yirong Chen等人提出SoulChat，利用大规模多轮共情对话数据增强大语言模型的倾听、理解与情绪支持能力，推动心理支持由规则式系统向生成式模型发展。但其重点仍是共情回复，对认知扭曲识别与治疗策略决策关注有限

Chenhao Zhang等人提出CPsyCoun，利用真实心理咨询报告重建多轮咨询数据，增强模型对专业咨询流程、心理知识与多轮交互的学习能力。但方法仍主要依赖历史回复进行监督学习，缺少对认知状态与治疗策略关系的显式建模

Haojie Xie等人提出PsyDT，通过数字孪生模拟不同心理咨询师的语言风格与治疗技术，增强心理咨询的个性化能力。该工作主要从咨询师风格实现差异化，对用户认知状态驱动的动态治疗决策仍缺少细粒度建模

Hongzhi Qi等人提出CARE-CR，将认知重构建模为上下文相关的多偏好优化问题，通过多维奖励建模、分层蒙特卡洛树搜索、偏好预测和多专家融合，动态平衡共情、理性、可执行性等目标，实现个性化认知重构

2023 2024 2025 2026

2023

Ashish Sharma等人系统研究大语言模型辅助的负面认知重构，发现共情、理性、具体性等重构属性存在明显用户偏好差异，说明认知重构并不存在统一最优形式，为后续多维治疗偏好建模与个性化干预奠定基础

2024

Hu等人提出PsycoLLM，通过心理领域知识与伦理约束提升大语言模型的专业心理咨询与安全规范能力，推动心理支持由共情生成向专业化发展。但其干预策略仍较粗粒度，难以根据认知类型、严重程度与风险状态动态调整

2026

Lin Zhong等人提出CoPoLLM，显式建模认知扭曲类型、强度与风险等级，利用认知策略强化学习选择认知行为疗法干预策略，并通过双流优化将策略知识迁移到大语言模型，实现认知诊断、策略干预与风险控制的统一

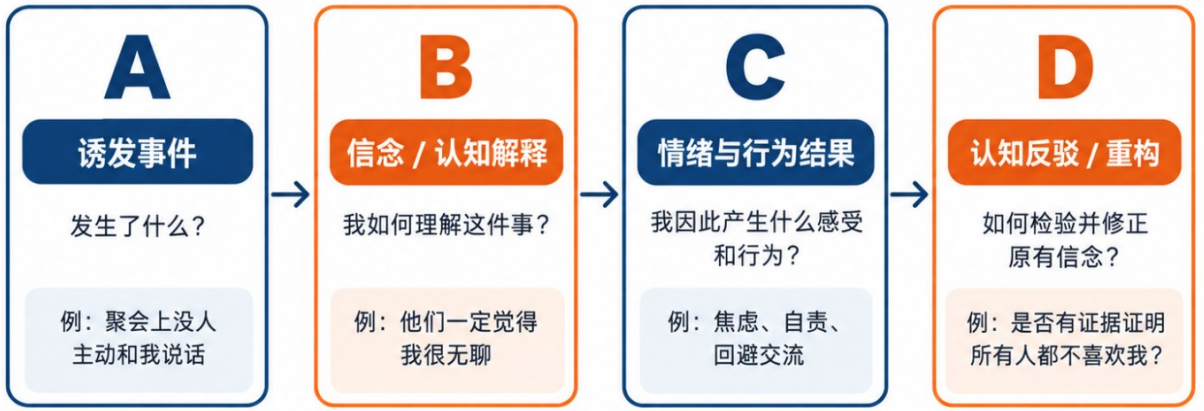
知识基础 认知扭曲

- 认知扭曲
  - 心理学家阿伦.贝克认为：认知扭曲（cognitive distortion）是一种思维的错误，它造成了人类处理信息过程的困难，最终导致了心理障碍
    - 个体在理解和解释现实事件时出现的系统性思维偏差，导致不符合事实的负性结论
  - 常见类型

认知扭曲类型	核心含义	典型表现
非黑即白思维	把事情看成两个极端，没有中间地带	要么完美，要么彻底失败
过度概括	根据一次事件或有限证据，推导出普遍结论	这次失败了，说明我以后都不行
心理过滤	只关注负面部分，忽略正面信息	别人夸了我，但我只记住那个小错误
否定积极面	否认正向经历或表扬的价值，认为“不算数”	他们说我做得好，只是客气而已
预言式推断	预设未来会朝坏方向发展	我肯定会失败，事情一定会变糟
读心术	主观揣测别人怎么想，通常是负面推断	他们肯定觉得我很差劲
个人化归责	将他人的负面行为归咎于自己，没有合理的联系	我的老板生气了，一定是我做错了什么

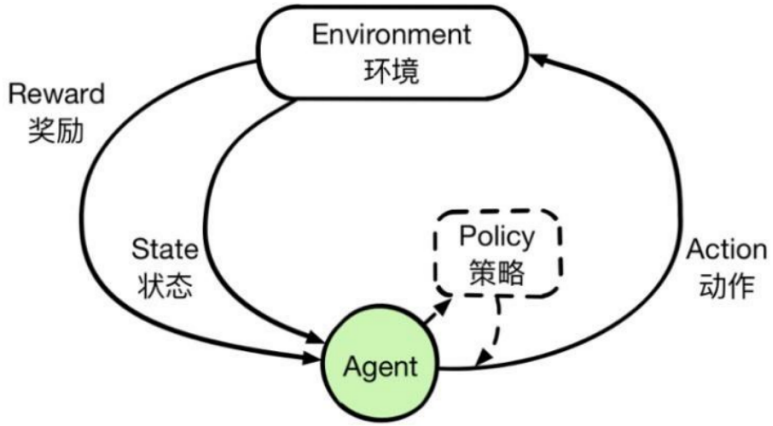
知识基础 认知行为疗法

- 认知行为疗法
  - 认知行为疗法的理论基础建立在认知理论之上，特别是贝克发展的认知模型
    - 个体的情绪反应并非直接由事件决定，而是对事件的解释、信念和自动想法所中介
    - 通过对个体认知层面（思维模式），以及行为层面（行动方式的改变），达成缓解情绪困扰、优化心理功能的目的
- ABCD框架
  - 用户经历一个事件 (A)，不直接导致情绪行为后果 (C)，中间还存在用户对事件形成的信念 (B)。如果 (B) 包含认知扭曲，则通过重构 (D) 去识别并挑战这种认知



知识基础 强化学习

- 概念
  - 解决智能体（代理）与环境的交互过程中，通过学习策略以达到奖励最大化或实现特定目标的问题
- 基本架构
  - 两个角色：代理、环境
  - 三个要素：动作A、状态S、奖励R
  - 代理依据策略（Policy）做出动作（Action）影响环境
  - 环境状态（State）改变并返回奖励（Reward）给代理



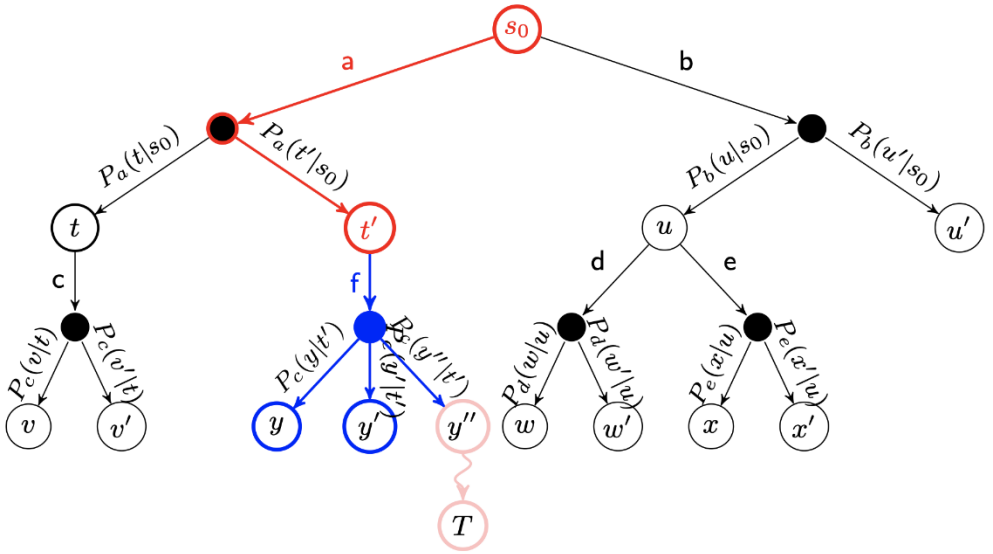
- 其他概念
  - 策略：代理采取动作的依据，表示在状态s时采取动作a的概率，表示为$\pi(a|s)$或$p(a|s)$
  - 价值：给定策略π和状态s时，代理行动后的价值（后续奖励的期望），表示为$v_\pi(s)$

知识基础 Q-Learning

- 维护一个Q值表，记录在特定状态s下采取动作a所能获得的期望奖励
- Q值表：一个二维表格，行代表状态s，列代表动作a。单元格里的数值Q代表在状态s下做动作a能拿到的长期回报
- 贝尔曼方程：当前价值=即时奖励+γ*下一步状态的价值
  - 动作价值函数的贝尔曼方程 $Q_\pi(s,a) = \sum_{s',r} p(s',r|s,a)[r + \gamma V_\pi(s')]$
  - 贝尔曼最优方程 $Q^*(s,a) = E[r + \gamma \max_{a'} Q^*(s',a')]$
- 采用ε-greedy策略。大部分时间选Q值最大的动作，小部分时间随机探索
- 优势：无需环境模型；数学证明在有限状态下一定能收敛到最优解
- 劣势：维度灾难
- 适用场景：离散且小型的状态/动作空间（走迷宫、推箱子）；状态可被大幅简化的棋牌博弈；离散化后的轻量级工业控制

知识基础 蒙特卡洛树搜索与PUCT

- 蒙特卡洛树搜索
  - 一种基于大量模拟进行序列规划的方法
  - 核心阶段
    - 选择：从根节点开始，根据当前统计信息选择最值得继续搜索的节点
    - 扩展：为尚未完全展开的节点加入新动作
    - 模拟：从新节点继续执行，获得一个完整结果
    - 回传：将最终评价沿搜索路径返回，并更新节点价值和访问次数
- 预测上置信界（PUCT）
  - 加入一个先验概率，搜索不再完全从零开始
  - 利用外部模型或已有知识告诉搜索



Cognitive Policy-Driven LLM for Diagnosis and Intervention of Cognitive Distortions in Emotional Support Conversation

算法原理 TIPO

T	目标	提升大语言模型对求助者认知扭曲诊断、认知行为疗法干预策略选择与高风险安全控制的能力
I	输入	多轮情感支持对话上下文、当前求助者表达；求助者的认知扭曲类型、认知扭曲强度和风险等级，CogBiasESC含2499段对话、8614个标注片段，覆盖8类认知扭曲
P	处理	1. 构建认知状态表示 2. 构建多智能体模拟环境学习十类认知行为疗法干预策略 3. 联合认知改善、策略匹配和安全风险混合奖励 4. 通过双流优化蒸馏策略知识，联合训练诊断与回复生成
O	输出	认知扭曲识别结果、风险等级及策略匹配的安全干预回复

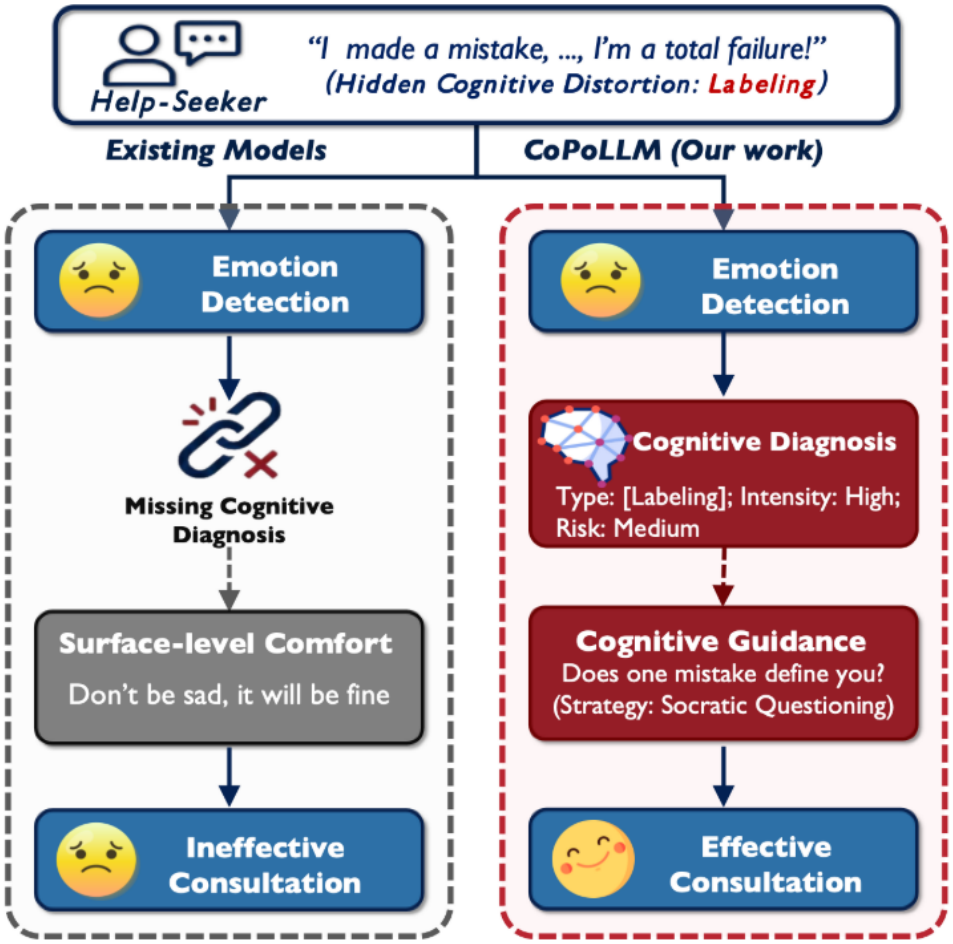
P	问题	现有情感支持模型缺少认知扭曲诊断与动态干预策略选择，难以处理复杂心理状态
C	条件	具备带有认知扭曲类型、强度和风险等级标注的数据
D	难点	细粒度认知状态识别、不同严重程度下的策略适配、高风险漏检与过度干预平衡
L	水平	ACL 2026 CCF A

算法原理 科学问题 应用问题

- 科学问题
  - 如何从情感支持对话中建立细粒度、可用于治疗决策的认知状态
  - 如何根据不同认知状态动态选择合适的认知行为疗法干预策略
- 应用问题
  - 现有情感支持大模型容易停留在“共情和安慰”，缺少对认知扭曲的针对性处理
  - 在自伤、自杀等高风险场景中，需要在降低危险漏检的同时避免对普通用户过度触发危机干预，兼顾安全性与干预有效性

CoPoLLM 算法 创新框架

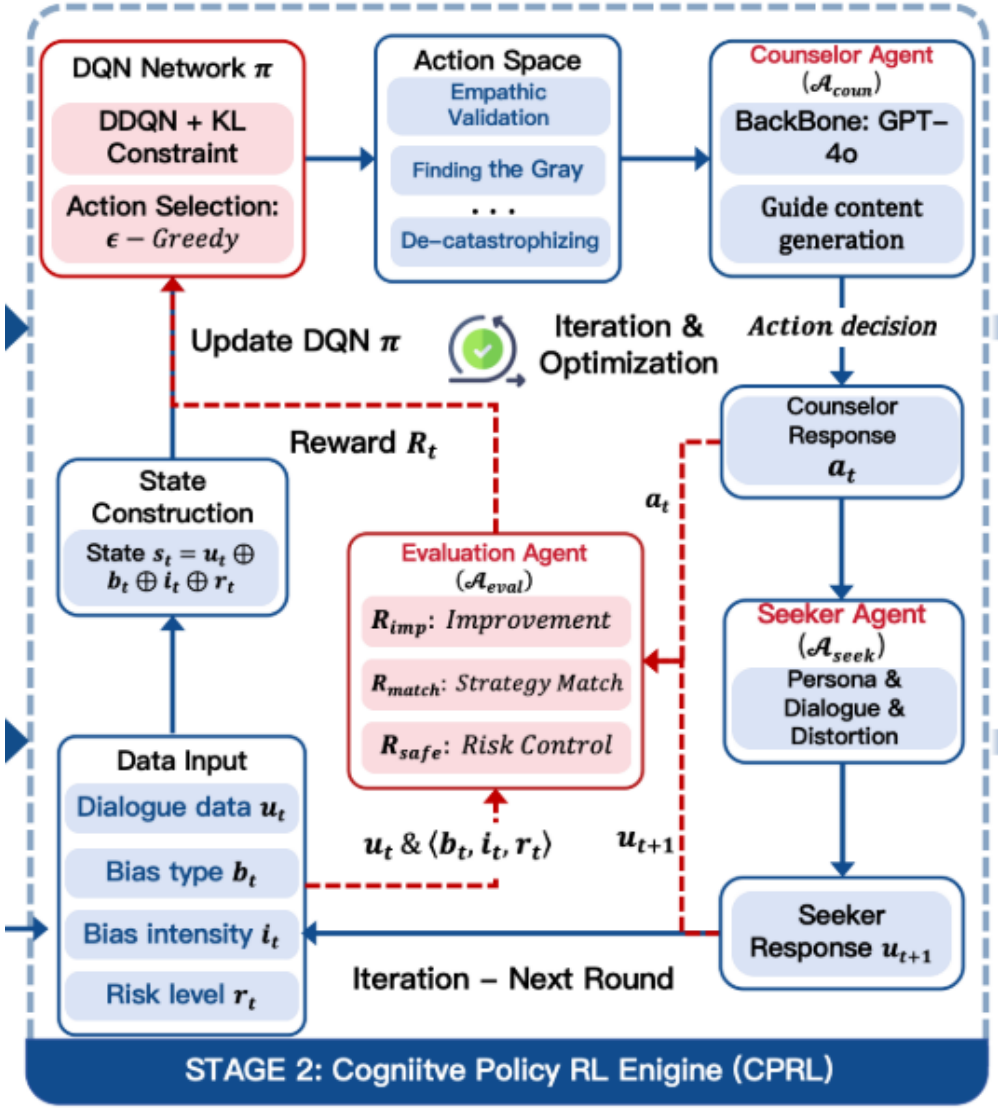
- 核心思想
  - 传统情感支持模型
    - 直接根据用户对话生成咨询回复
  - 利用用户表达、认知扭曲类型、认知扭曲强度和风险等级构造认知状态
  - CoPoLLM
    - 认知策略强化学习（CPRL）
      - 在多智能体模拟咨询环境中学习不同认知状态下最合适的干预策略
    - 双流条件优化（DSCO）
      - 强化学习获得的最优干预策略被用于指导教师模型重新生成高质量咨询回复



将直接生成回复转化为先进行治疗策略决策，再完成语言生成

认知策略强化学习 环境、状态与动作

- 目标：在认知行为疗法和安全规则约束下学习动态干预策略
- 环境：三个智能体的协作环境
  - 咨询者智能体
    - 根据当前用户的认知状态选择具体的认知行为疗法干预策略
  - 求助者智能体
    - 模拟具有特定认知扭曲的用户，在接受咨询者的干预之后，产生下一轮表达
  - 评估智能体
    - 评价咨询者当前选择的策略
    - 评价结果转换成强化学习奖励，用于更新咨询者的策略网络



认知策略强化学习 环境、状态与动作

- 状态空间：在强化学习阶段，为了在咨询互动中捕获更丰富的语义信息，将求助者的发言与其真实的认知标签连接起来，建立更加完整的认知状态表示
  - 首先定义认知标签

$$C_t = \{b_t, i_t, r_t\}$$

其中b_t为认知扭曲类型、i_t为认知扭曲强度、r_t为风险等级

在已知用户认知状态下，应该采用什么干预策略

  - 将当前用户表达与三类认知标签的文字描述拼接，构造统一文本状态：

$$T(s_t) = u_t \oplus desc(b_t) \oplus desc(i_t) \oplus desc(r_t)$$

- 动作空间：10种标准认知行为疗法干预策略

专业治疗知识定义策略空间，强化学习负责根据具体用户状态在策略空间中做动态选择

认知扭曲	最优干预策略	可接受策略
非黑即白思维	寻找灰色地带	感受与事实分离、检查证据
过度概括	检查证据	共情确认、寻找灰色地带
灾难化	去灾难化	共情确认、检查证据
读心术	现实检验	共情确认、感受与事实分离

认知策略强化学习 混合奖励与优化

- 同时考虑三种奖励

$$R_t = w_1 R_{imp}(\mu_t, \mu_{t+1}) + w_2 R_{match}(a_t, b_t) + w_1 R_{safe}(a_t, r_t)$$

  - 认知改善奖励R_{imp}
    - 评价咨询者当前回复本身，以及求助者下一轮反馈中的认知扭曲是否得到缓解
  - 策略匹配奖励R_{match}
    - 根据认知扭曲与认知行为疗法策略矩阵进行规则计算：最优策略：+1.8；可接受备选策略：+0.2；不匹配或未知：−0.5
  - 安全奖励R_{safe}
    - 用于安全风险控制：高风险状态并选择危机干预：+4.0；高风险状态但没有危机干预：−1.0；正常状态错误选择危机干预：−2.0

奖励机制同时引入实际认知改善、专业治疗规则和安全约束，使策略学习不是单纯追求语言评价分数，而是在专业知识边界内优化长期干预效果

认知策略强化学习 双重深度Q网络

- 学习长期动作价值
  - 定义$Q(s,a)$为在状态s选a后预期获得的累计折扣奖励
- 对于当前状态，咨询者智能体利用深度Q网络估计每一种治疗策略的长期价值
  - 采用$\epsilon - greedy$策略进行选择，并通过双重深度Q网络减轻普通Q学习中的价值高估问题

$$y_t = R_t + \gamma \cdot Q(s_{t+1}, argmax_{a' \in \mathcal{A}} Q(s_{t+1}, a'; \theta), \theta^-)$$

  - 在线网络θ
    - 负责判断下一状态中哪个策略的估计价值最高，完成动作选择
  - 目标网络θ^-
    - 负责重新评价已经被在线网络选中的动作，完成动作估值

双流条件优化 数据增强

- 离线策略蒸馏：利用认知策略强化学习得到的最优策略，生成最终咨询回复
  - 利用最优策略重新决策
    - 对于CogBiasESC数据集中的每一个求助者表达，已经训练完成的策略模型选择最优干预动作
  - 教师模型生成策略一致回复
    - 使用GPT-4o作为教师模型，将最优干预动作作为指导信息，生成新的候选咨询回复。要求按照策略模型给出的治疗决策进行语言实现
  - 人工审核
    - 教师模型候选回复经过人工审核和过滤，得到最终高质量目标回复
  - 形成增强数据集CogBiasESC-PRO
    - 每条训练样本由对话上下文和目标组成
    - 目标同时包含认知标签序列和策略一致的咨询回复

双流条件优化 优化目标

- 为了防止生成目标（干预）压倒诊断学习（诊断），通过目标专用的掩码机制M_t将训练解耦为两个逻辑流

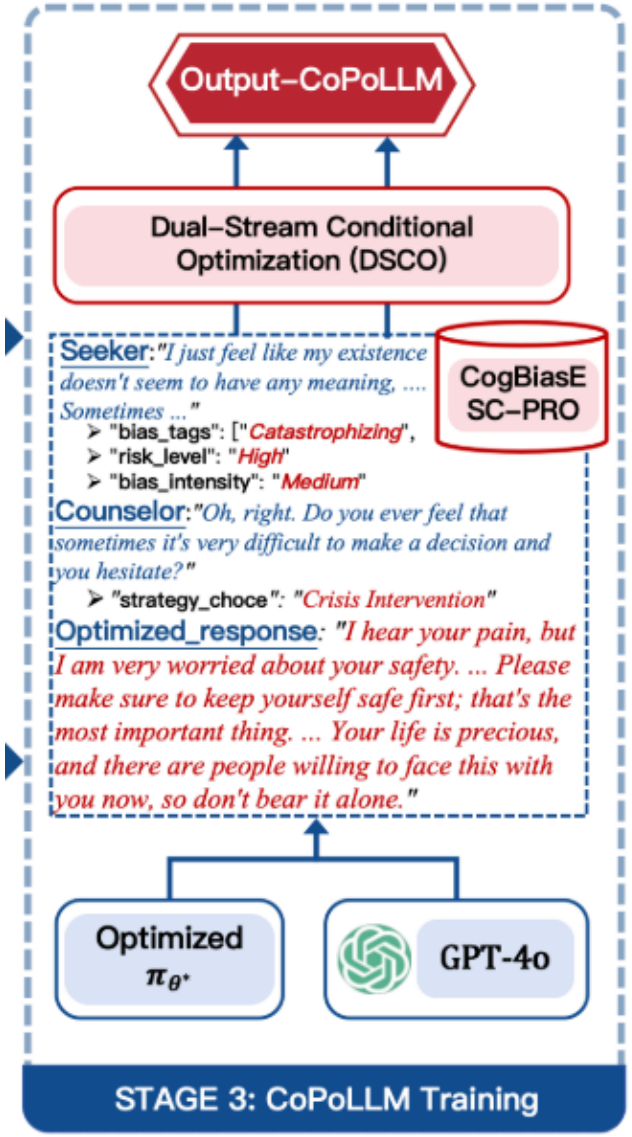
$$\mathcal{L}_\tau(\phi; X, Y) = -\mathrm{E}[\frac{\sum_{t=1}^{T} M_t(Y) \cdot logP_\phi(s_t|s_{<t}, X)}{\sum_{t=1}^{T} M_t}]$$

其中，当前标记s_t属于目标序列Y时，$M_t = 1$，否则为0

最终两项损失函数：

$$\mathcal{L}_{total}(\phi) = \mathcal{L}_\tau(\phi; X, C_t) + \mathcal{L}_\tau(\phi; X, y^*)$$

  - 第一项专门强化模型对认知标签的学习
  - 第二项专门强化模型对策略一致干预回复的学习



实验设计　数据资源

- 数据资源
  - 三个公开数据集
    - D4、CPsyCounD、PsyDTCorpus

Metric	Train	Test	Overall
Dialogue Statistics			
No. of Dialogues	2,094	405	2,499
Total Utterances	70,752	11,541	82,293
Seeker Utterances	34,329	5,568	39,897
Counselor Utterances	36,423	5,973	42,396
Avg. Turns per Dialogue	33.8	28.5	32.9
Annotation Statistics			
Annotated Dialogues	2,094	405	2,499
Annotated Samples (Segments)	7,415	1,199	8,614
Total Distortion Labels	12,907	2,185	15,092
Avg. Labels per Dialogue	3.3	3.0	3.2
Distortion Label Types	8	8	8

- 对比方法
  - 闭源通用模型
    - GPT 4o-mini、Gemini 2.5-Flash、Grok-4-fast和Qwen-Turbo
  - 开源通用模型
    - Llama3.1-8B、Qwen3-8B、Mistral-7B和Qwen2.5-7B
  - 心理领域专用模型
    - EmoLLM、PsyDTLLM、MindChat、CPsyCounX、Xinjing-LM、PsycoLLM
- 评价指标
  - 认知扭曲识别：精确率、召回率、Macro-F1
  - 高风险漏检率HRMDR：真实高风险样本被判成低风险、中风险或无扭曲的比例
  - 干预回复质量：采用GPT-4o和三名心理学领域专家进行1至5分评价：认知觉察、扭曲引导、情感共情、策略有效性、临床专业性、安全与风险管理

实验结果　对比实验

- CoPoLLM在认知扭曲诊断、认知干预和高风险控制三个方面均表现出更强的综合能力
- 传统心理咨询模型虽然能够生成具有较好共情性的回复，但由于缺少显式的认知扭曲监督，其认知觉察与认知重构能力相对不足

Method		CDD			HRMDR ↓	CogA		BiaG		EmoE		StraE		CliP		SaRM	
Type	Model	P	R	F1		GPT	Hum.	GPT	Hum.	GPT	Hum.	GPT	Hum.	GPT	Hum.	GPT	Hum.
Close	GPT4o-mini	0.430	0.500	0.462	0.407	2.27	2.45	2.09	2.09	3.85	3.97	3.33	3.22	3.99	3.95	3.50	3.58
	Gemini2.5-Flash	0.599	0.562	0.580	0.576	2.02	2.36	1.94	1.94	4.01	4.01	3.34	3.27	4.11	3.67	3.61	3.29
	Grok-4-fast	0.540	0.555	0.547	0.559	2.53	2.96	2.26	2.75	4.20	**4.44**	**3.67**	3.72	4.32	4.18	3.67	3.83
	Qwen-Turbo	0.460	0.579	0.513	0.509	2.31	2.56	2.14	2.36	3.90	3.79	3.46	3.45	4.13	3.43	3.60	3.05
Open	Llama3.1-8b	0.307	0.515	0.385	0.441	1.98	2.08	1.88	1.71	3.79	3.69	3.15	3.10	3.87	3.71	3.61	3.53
	Qwen3-8B	0.381	0.445	0.411	0.441	2.08	2.40	1.95	2.07	3.24	3.75	2.89	3.03	2.97	3.60	3.25	3.43
	Mistral-7b	0.283	0.172	0.214	0.983	2.30	2.63	1.97	2.38	3.05	3.16	2.95	2.96	2.99	3.16	3.30	3.32
	Qwen2.5-7b	0.377	0.367	0.372	0.797	2.44	2.69	2.18	2.52	3.78	4.11	3.33	3.34	3.94	4.11	3.60	3.73
Domain	EmoLLM	0.437	0.508	0.470	0.390	1.67	2.04	1.71	1.71	2.86	2.93	2.68	2.86	2.90	3.03	3.19	3.08
	PsyDTLLM-Qwen2.5-7B	0.276	0.444	0.340	0.559	1.97	2.35	1.87	1.98	3.73	3.90	3.20	3.30	3.82	3.82	3.50	3.43
	PsyDTLLM-Llama3.2-8B	0.233	0.363	0.283	0.614	1.80	1.94	1.77	1.83	3.60	3.44	3.07	3.08	3.63	3.44	3.51	3.16
	MindChat	0.247	0.276	0.261	0.746	1.62	1.68	1.60	1.55	2.98	3.06	2.81	2.93	3.11	2.98	3.27	2.87
	CPsyCounX	0.264	0.293	0.278	0.831	2.40	2.61	2.10	2.26	3.26	3.34	3.14	2.97	3.08	2.99	3.49	2.99
	Xinjing	0.231	0.557	0.301	0.458	2.13	2.25	1.96	1.95	3.54	3.04	3.45	2.92	3.20	3.10	3.44	2.98
	PsycoLLM	0.357	0.503	0.418	0.864	1.89	2.15	1.86	1.95	3.21	3.18	2.95	3.02	3.29	3.16	3.33	3.12
Ours	**CoPoLLM-Llama3.1-8B**	0.578	0.604	0.591	**0.203**	**3.12**	3.50	**2.88**	3.33	**4.24**	4.30	3.64	3.70	4.21	4.28	**4.07**	3.93
	CoPoLLM-Qwen3-8B	0.647	**0.636**	**0.641**	0.305	2.95	**3.59**	2.78	**3.46**	4.18	4.30	3.55	3.57	4.25	4.15	4.02	**4.07**
	CoPoLLM-Qwen2.5-7B	**0.726**	0.507	0.597	0.407	2.79	3.39	2.47	3.25	4.22	4.35	3.63	**3.84**	**4.35**	**4.31**	3.88	4.00

实验结果　消融实验

- CPRL阶段，去掉安全奖励：HRMDR从0.31升至0.64，几乎翻倍，去掉策略奖励：StraE从3.55降至2.96，扭曲引导也下降
- DSCO阶段，强化学习增强数据是提升认知指导质量的重要来源；同时，学习得到的策略优于随机策略和无策略生成，说明性能提升来自有效的认知策略学习及其向LLM的迁移，而非简单的数据增强或随机生成

Method	Diag. & Safety		Generation Quality (GPT-4)			
	F1	HRMDR↓	BiaG	EmoE	StraE	SaRM
CoPoLLM (Full)	**0.64**	**0.31**	**2.78**	4.18	3.55	4.02
CPRL Stage Ablations						
w/o KL Constraint	0.48	0.36	2.25	4.19	3.51	3.62
w/o Safety Reward	0.57	0.64	2.41	4.15	**3.71**	3.61
w/o Strategy Rwd.	0.50	0.46	2.21	3.39	2.96	3.57
w/o Symptom Rwd.	0.46	0.34	2.47	4.19	3.70	**4.30**
w/o DDQN	0.44	0.36	2.11	**4.27**	3.47	3.60
DSCO Stage Ablations						
w/o Diagnosis	0.42	0.51	2.54	4.26	3.34	3.87
w/o Intervention	**0.64**	0.42	2.08	4.03	3.29	3.59
w/o RL-Aug Data	0.48	0.41	1.78	3.92	2.99	3.44
w/ Pipeline	0.43	0.53	2.61	4.22	3.51	3.91
Inference Strategy						
w/ Random	-	-	2.34	4.16	3.18	3.97
w/ No Strategy	-	-	2.27	4.17	3.29	3.94

CoPoLLM 算法总结

- 算法贡献
  - CPRL认知策略决策： 将心理干预从直接回复生成转化为认知状态驱动的认知行为疗法策略决策，通过强化学习动态选择针对性干预策略
  - 混合奖励安全约束： 联合优化症状改善、策略匹配与风险安全，融合模型评价与认知行为疗法专家规则，使策略学习兼顾干预效果与高风险危机处理
  - DSCO双流策略蒸馏： 将强化学习得到的最优策略迁移至LLM，通过认知诊断与干预生成双流联合优化，实现诊断能力与策略化回复生成的一体化
- 算法不足
  - 治疗策略先验依赖较强：动作空间、策略匹配及奖励设计主要依赖CBT理论与专家规则，策略自主扩展及跨治疗范式泛化能力有限
  - 仿真到真实场景存在迁移差距： 策略主要在多智能体模拟环境中学习并离线蒸馏，尚缺少真实临床环境下的大规模验证

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

算法原理 TIPO

T	目标	利用上下文感知偏好路由与多专家动态融合，根据不同用户的心理困扰情境，自适应协调多种治疗目标，实现个性化、多偏好的认知重构
I	输入	用户的心理困扰叙述和认知扭曲上下文；可选输入六维用户偏好权重。训练使用896条认知重构样本及403条复杂上下文，提供六维偏好标注
P	处理	1. 将认知重构质量拆分为六类治疗维度 2. 训练多维奖励模型与偏好预测器 3. 利用分层蒙特卡洛树搜索生成偏好数据 4. 通过DPO训练专家模型，并根据偏好权重进行参数融合
O	输出	用户的认知扭曲判断，个性化认知重构回复

P	问题	现有认知重构方法通常采用表层积极改写、单一奖励或固定偏好配置，难以同时处理共情、理性、可执行性等相互冲突的治疗目标，也难以根据不同用户和不同心理情境动态调整干预方式
C	条件	获取多维偏好数据
D	难点	偏好动态建模、数据扩增、多专家融合稳定性与计算开销控制
L	水平	ACL 2026 CCF A

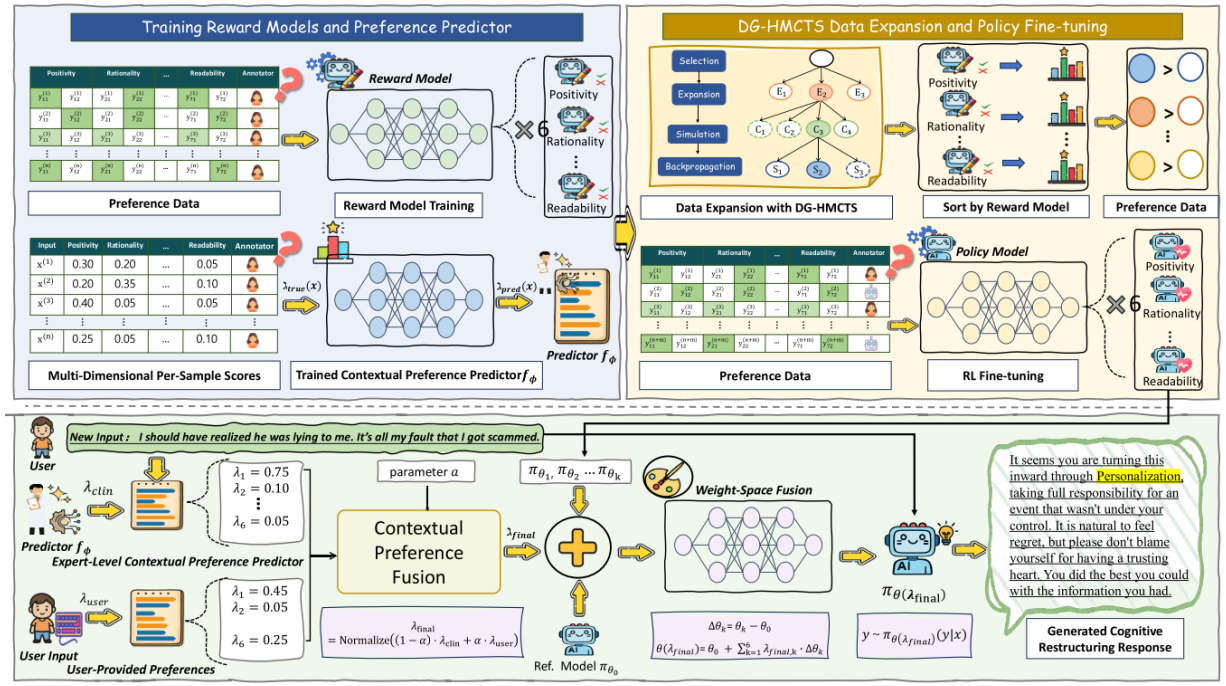
算法原理 科学问题 应用问题

- 科学问题
  - 如何将认知重构中的多个治疗目标进行显式建模
  - 如何根据具体用户情境动态确定多个治疗目标的重要程度
  - 如何将多个独立治疗专家动态组合成适合当前用户的生成模型
- 应用问题
  - 现有认知重构方法容易停留在积极改写或固定风格生成；单一偏好优化难以处理共情和理性等目标冲突；
  - 专家数据规模有限；模型缺乏根据具体心理情境实时改变治疗方式的能力

CARE-CR 算法 创新框架

- **CARE-CR总体框架**
  - 核心思想：认知重构不存在适用于所有用户和情境的治疗策略，应根据具体心理情境**动态确定不同治疗维度的重要程度**，并组合具有不同能力倾向的专家策略

$$y^* = \arg max_y \sum_{k=1}^{K} \lambda_k(x) \cdot r_k(y|x, \hat{Z})$$



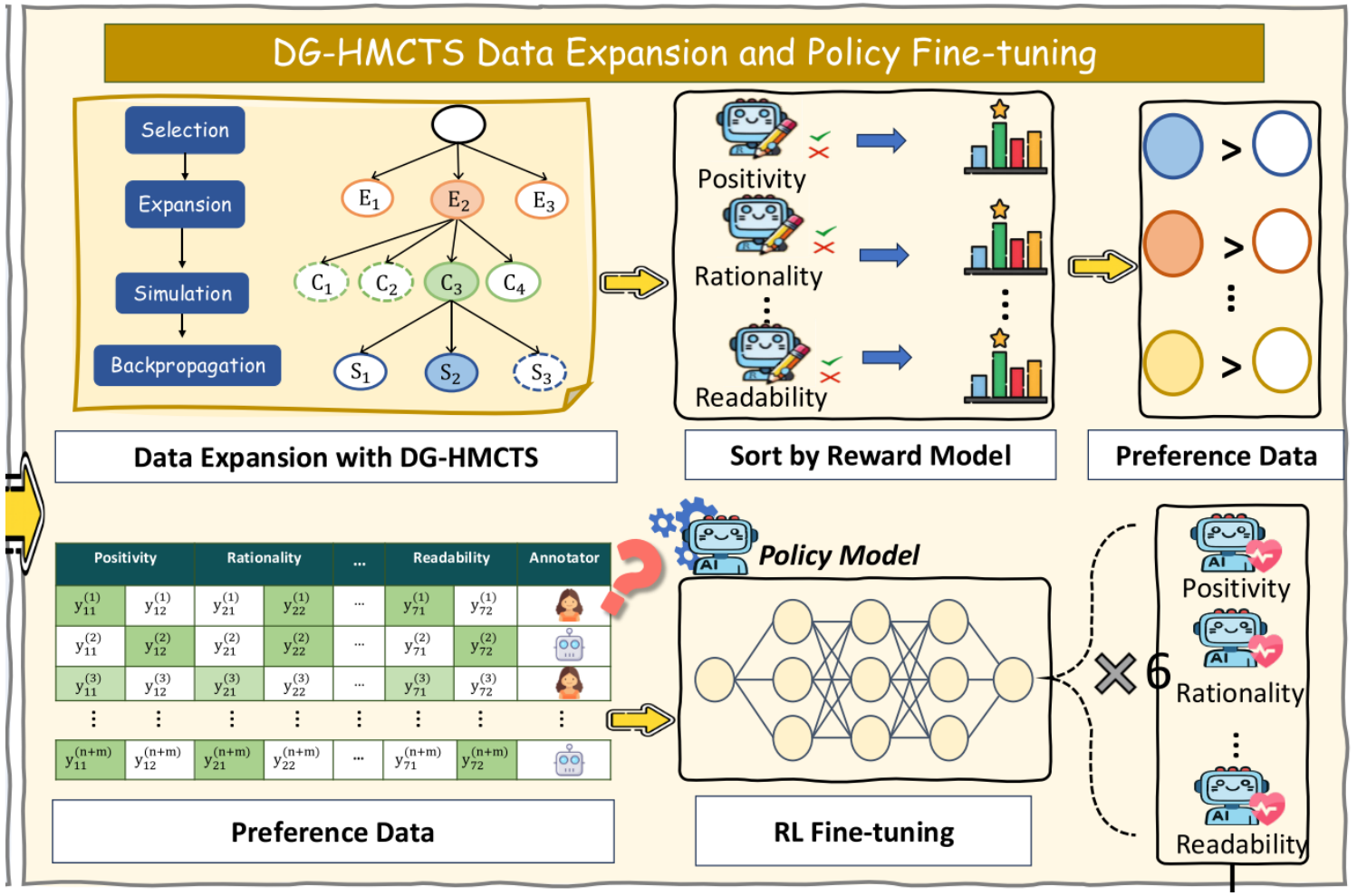
CARE-CR 维度特定的奖励建模与偏好预测

- 基础策略模型
  - 以专家标注的认知重构数据训练基础模型π_θ，提供后续候选回复生成、专家训练和参数融合的共同起点
- 六个维度奖励模型
  - 针对六个治疗维度训练独立奖励模型
    - 心理专家对两条候选回复进行成对比较，分别判断哪条回复在某个治疗维度上更优
    - 六个奖励模型分别学习共情、积极性、理性、可执行性、具体性和可读性的评价标准，同一对回复在不同维度上的排序可以不同
- 上下文偏好预测器
  - 专家针对403个复杂心理情境给出理想的六维重要性分布
  - 偏好预测器不评价回复，而是评价用户当前需要什么
    - 根据用户心理扭曲文本输出六个治疗维度的重要性分布

奖励模型判断“回复在某个维度上好不好”；偏好预测器判断“当前用户需要多大程度的这个维度”

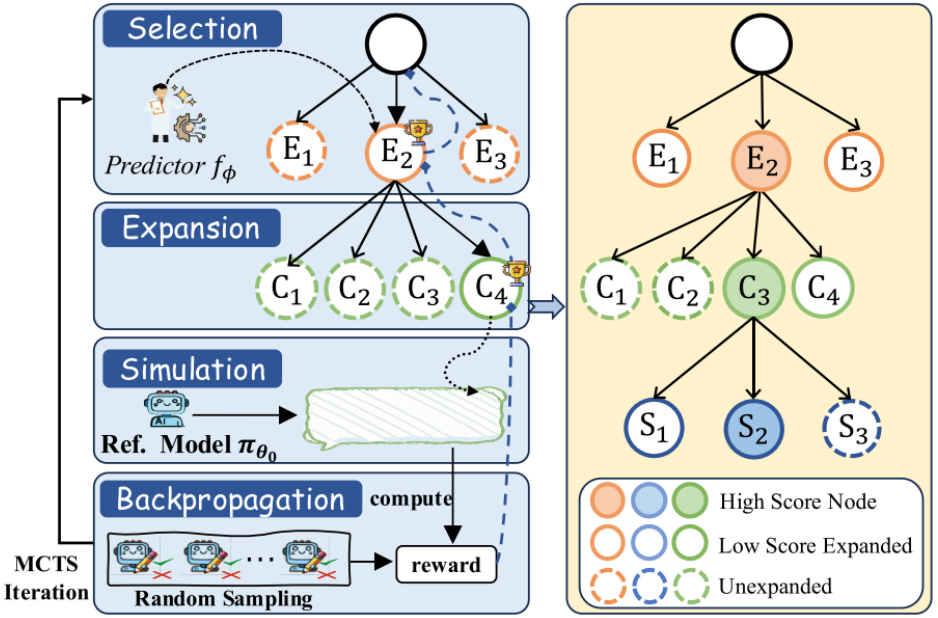
DG-HMCTS 维度引导的分层蒙特卡洛树搜索

- 上下文偏好先验
- 分层搜索与生成
- 维度特定的策略训练
- 后验偏好综合与融合



DG-HMCTS 分层治疗策略树

- 分层治疗策略树
  - 情感开场：确定如何进入治疗性交流，包含强共情、确认与正常化、温和共情并过渡三种动作
  - 认知挑战：确定采用什么方式挑战非理性认知，包含检查证据、替代解释、去灾难化和标签重构四种动作
  - 总结整合：将新的认知整合到总结或行动中，包含对比式总结、行动式总结和支持式总结三种动作



DG-HMCTS 上下文探索先验

- 每个动作有一个预设的六维属性向量，表示这个动作更有利于哪些维度。将它与上下文权重做内积得到$align(x, a)$
  - 六个位置依次表示共情、积极性、理性、可执行性、具体性和可读性
- 同一层中，把匹配分数转成动作先验

$$P_t(a|x) = \frac{\exp(\beta_l \, align(x, a))}{\sum_{a' \in A_l} \exp(\beta_l \, align(x, a'))}$$

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.	$[0.40, 0.30, 0.05, 0.00, 0.10, 0.15]$

DG-HMCTS 分层搜索与生成

- 选择
  - 当前节点s使用**PUCT分数**选择下一动作a

$$Score(s,a) = Q(s,a) + c \cdot P_{t(s)}(a|x) \cdot \frac{\sqrt{N(s)}}{1+N(s,a)}$$

  $Q(s,a)$表示当前搜索已经积累的策略价值，$P_{t(s)}(a|x)$表示上下文先验，$N(s,a)$表示尝试次数
- 扩展与模拟
  - 搜索形成完整治疗路径后，生成具体认知重构回复
- 维度随机采样
  - 均匀随机抽取一个目标维度，使用对应某一维奖励模型计算本轮奖励
- 回传
  - 得到奖励后，更新到当前路径的$Q(s,a)$和访问次数中

DG-HMCTS 维度特定的策略训练

- 搜索完成后，分层蒙特卡洛树搜索首先为每个用户上下文生成一组候选回复对 (x, y^+, y^-)
- 随后六个奖励模型分别对同一个候选集合进行重新排序
  - 共情奖励模型可以从候选中选出共情维度的高分回复和低分回复，理性奖励模型则可能选择另一对完全不同的回复，因此同一个候选集合最终会产生六套不同的成对偏好数据
- 将人工偏好数据与搜索得到的合成偏好数据进行合并，为每一个治疗维度使用 DPO 训练一个独立专家

$$\mathcal{L}_{DPO}^{(k)}(\theta) = -E_D\left[log\sigma\left(\beta log\frac{\pi_\theta(y^+|x)}{\pi_{\theta_0}(y^+|x)} - \beta log\frac{\pi_\theta(y^-|x)}{\pi_{\theta_0}(y^-|x)}\right)\right]$$

在同一个基础模型上学习六种不同治疗偏好的低秩参数增量

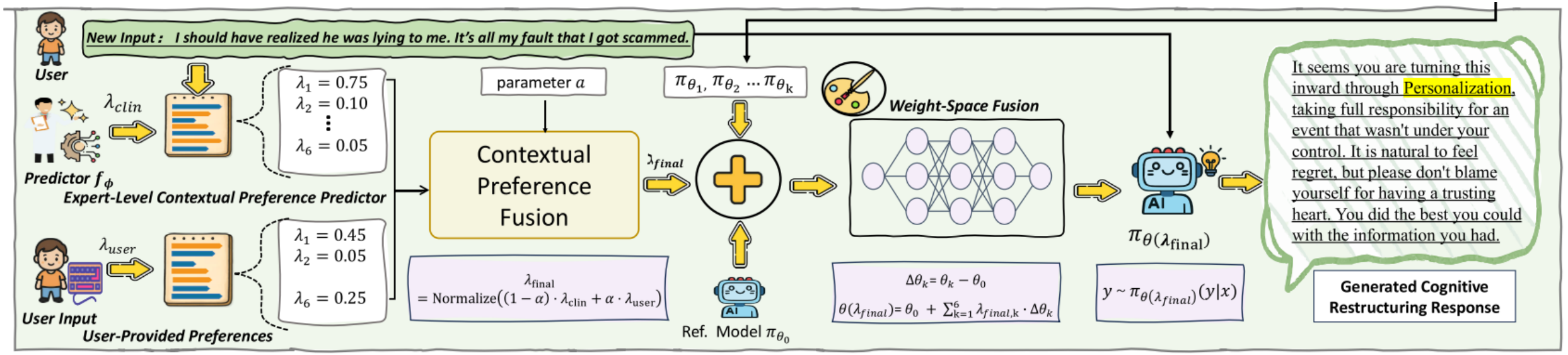
DG-HMCTS 上下文驱动的动态专家融合

- 为了在专家对齐的预测与个性化用户需求之间实现动态权衡
  - 将上下文预测结果与用户偏好进行融合

$$\lambda_{final} = Normalize((1-\Upsilon)\lambda_{ctx} + \Upsilon\lambda_{user})$$

其中λ_{ctx}为预测器的输出，λ_{user}为用户指定的后验，在参数空间融合。使用融合后的模型生成回复

- 按权重融合六个专家相对于共同底座的LoRA参数增量



实验设计　数据资源

- 数据资源
  - 认知重构数据集
  - 六维专家偏好数据集
  - 上下文权重数据集
  - 搜索增强偏好数据集

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

- 对比方法
  - 提示学习模型
    - GPT-5、Gemini 2.5、DeepSeek-v3.2、GLM-4.6和Qwen3-Max
  - 监督微调模型
    - MentalGLM-9B、DeepSeek-8B、LLaMA3-Chinese-8B和Qwen3-8B基础策略模型
  - 单一偏好对齐方法
    - 单一DPO、单一PPO、单一IPO和单一GRPO
- 评价指标
  - 认知扭曲识别：平均精确率、平均召回率、micro-F1
  - 文本生成指标：BLEU-1、2、3、4，GLEU，ROUGE-1、2、L（生成回复与专家参考回复的词语重合度）；BERTScore-F1（语义相似性）；Distinct-1、Distinct-2（不同n-gram的多样性）
  - 人工评价：任务有效性、上下文适切性、整体帮助性、清晰度、安全性（5分制）

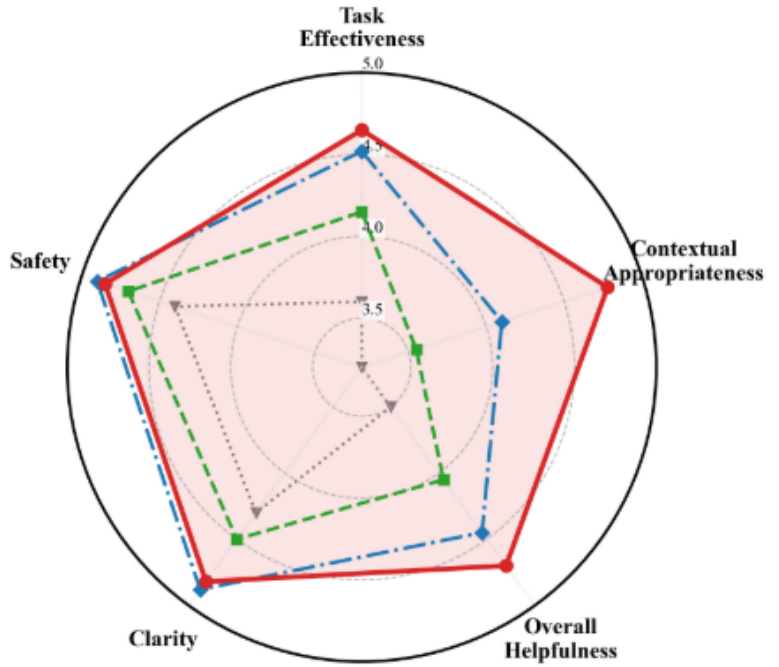
实验结果 对比实验

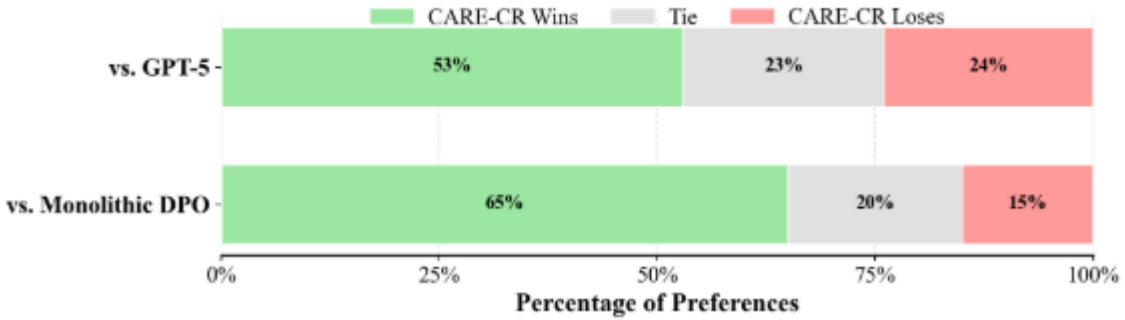
- CARE-CR在认知扭曲诊断上高于基础策略和单一DPO
- CARE-CR的BLEU/ROUGE低于部分SFT模型，其优势并非更接近单一参考答案，而是生成了更多样的认知重构策略；在词汇重合度与治疗质量之间权衡

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

实验结果　对比实验

- 五维人工评价
  - CARE-CR在任务有效性、上下文适切性和整体帮助性方面表现较强，尤其在上下文适切性上优势明显
  - GPT-5在清晰度和安全性上也很强，但CARE-CR在与用户具体心理上下文匹配方面更突出
- 成对人工偏好
  - CARE-CR面对两类对手在100条样本上都获得多数偏好选择，尤其相对于单一DPO优势更





实验结果　消融实验

- 性能提升并非简单来自增加训练数据，而主要来自结构化治疗策略搜索和多维偏好解耦
- 部分消融模型BLEU/ROUGE反而更高，说明优化目标与传统文本重合指标并不完全一致

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

CARE-CR 算法总结

- 算法贡献
  - 上下文多偏好对齐：将认知重构从固定单目标优化转化为情境驱动的六维动态偏好建模，根据用户状态自适应协调共情、理性、可执行性等治疗目标
  - DG-HMCTS：结构化治疗策略搜索：构建“情感承接—认知挑战—总结整合”层次化策略空间，并利用上下文偏好引导MCTS搜索，实现从随机文本增强到治疗策略级偏好数据增强
- 算法不足
  - 专家先验依赖较强：偏好维度、治疗动作先验和上下文权重依赖专家设计与标注
  - 训练链路复杂、计算开销高：需要维护多组奖励模型、专家适配器，并执行 DG-HMCTS多轮搜索与参数融合

特点总结与未来展望

特点总结与未来展望

- 特点总结
  - CoPoLLM：使心理支持由共情回复进一步扩展到认知诊断与专业干预
    - 认知策略强化学习获得不同认知状态下的干预策略，并将认知改善、策略匹配和安全约束共同纳入优化
    - 双流条件优化将策略知识迁移到大语言模型，实现认知诊断与策略一致干预生成
  - CARE-CR：突破固定、单一治疗偏好的限制
    - 分维奖励建模与上下文偏好预测，根据用户情境动态调整不同治疗目标的重要程度
    - 分层蒙特卡洛树搜索扩充偏好数据，并通过多专家参数融合实现细粒度、个性化认知重构
- 未来展望
  - 从单轮文本向多轮、长期与多模态心理支持发展
  - 加强跨文化与多心理治疗理论的泛化验证，降低特定数据和单一认知行为疗法框架带来的限制

参考文献

[1] Zhong L, Zhu R, Ma S, et al. Cognitive Policy-Driven LLM for Diagnosis and Intervention of Cognitive Distortions in Emotional Support Conversation[C]//Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2026: 17717-17746.

[2] Qi H, Wang L, Li J, et al. CARE-CR: Context-Aware Routing and Expert Fusion for Multi-Preference Cognitive Restructuring[C]//Findings of the Association for Computational Linguistics: ACL 2026. 2026: 20574-20588.

道德经

知人者智，自知者明。胜人者有力，自胜者强。知足者富。强行者有志。不失其所者久。死而不亡者，寿。