第 1 轮修改

审稿人一，意见 1，

引言部分

对于 CBT 对话系统的文献综述应补充国外 Woebot health,Wysa 团队的研究，这些团队使用了 CBT 的工作思路进行人机对话，和本文的思路有何异同之处。

回应 1，

感谢审稿专家的提醒，已补充了 Woebot Health 和 Wysa 团队的研究。

意见 2，

出现英文缩写的第一次应进行说明，方便读者阅读。

回应 2，

已改。

意见 3，

现有研究总结、不足之处建议进行充分梳理，目前的不足之处提出比较仓促

回应 3，

综合两位评审的建议，已重写相关工作、研究总结。

意见 4，

考虑将 2 相关工作部分重新整理，作为引言介绍的一部分。

回应 4，

综合两位评审的建议，以及参考本刊的写作风格，引言部分精简为开头的一小段。重写了相关工作，相关工作更加连贯、聚焦。

意见 5，

技术部分

多个智能体采用链式结果，如果第一步认知扭曲识别产生错误，是否会连带后续出错?如何规避或解决这个问题?

回应 5，

感谢审稿专家的宝贵建议。认知扭曲识别是一个很好的研究课题，我们参考了下面这篇 2024 发表在 ACL 上的论文：ERD。ERD 的推理思路也正是我们擅长的研究方法。而且 ERD 善于正确识别没有扭曲的病例，避免过度诊断认知扭曲的陷阱。 但我们没做正确率检测，我们考虑后续进行这方面的工作。

在文中 3.3，我们简单补充了说明。

Lim, S., Kim, Y., Choi, C.-H., Sohn, J., & Kim, B.-H. (2024). ERD: A Framework for Improving LLM Reasoning for Cognitive Distortion Classification. Proceedings of the 6th Clinical Natural Language Processing Workshop, 292–300.

意见 6，

建议比对不同提示策略的差异(比如思维链提示和少样本提示等)

回应 6，

我们补充了多个对比实验。在第 4.3 节 对比实验中，我们增加了对不同提示策略的对比分析，包括 Zero-Shot Prompting、Standard Prompting、以及 CBT-LLM Prompting。

意见 7，

建议提供模型选择和微调的更多技术细节，方便复现。

回应 7，

模型选择方面，我们在新版论文中已经对模型选择进行了详细说明。在第 4.2 节 CBT-CoT 小节中，我们明确了选取基座模型的过程。为了选择最适合 CBT-CoT 系统的基座模型，我们测试了国内三种主流模型：DeepSeek-Chat、Qwen-Plus 和 Doubao-1-5-pro，并进行了实验对比，以确定最适合的模型。

关于微调，初稿中，微调是构建 CBT-CoT 数据集之后用于大模型的微调。我们综合分析两位评审的意见，经讨论，认为构建 CBT-CoT 数据集与本文的标题和核心宗旨不符，决定删去构建 CBT-CoT 数据集这部分内容，所以本文已不涉及微调了。

意见 8，

实验设计评估部分

需要考虑自动评估未必非常可信，可能评估模型本身就有偏。另外，一般来说使用大模型进行评分的效果不一定精准，是否考虑采用多个大模型投票机制(即相对比较而非绝对打分)，也可以进一步输出评价的文字内容作为依据，目前只使用 deepseek 作为评分模型存在方法上的不足。

回应 8，

感谢审稿专家的宝贵建议。关于“自动评估的可信度”问题，我们在第 4.4.1 节 大模型自动评估中已经做了改进。为了避免单一模型评分可能带来的偏差，我们使用了三个大模型（DeepSeek-Chat、Qwen-Plus 和 Doubao-1-5-pro）充当裁判进行评估，并采用加权平均的方式来计算最终分数。这样可以通过多个模型的投票机制来减小个别模型可能存在的偏差。我们还有少量样本的人工评估环节，确保自动评估的可靠性。

意见 9，

摘要部分

目前摘要不符合学术论文规范，建议重新改写

回应 9，

已按本刊摘要的要求重写。

要求如下：摘要 研究报告与论文请附 300 字以内的中文摘要，内容应包括

科学问题(研究目的、研究背景)、方法(研究设计)、结果、结论。注意，不要把结果与结论混为一谈，研究方法(变量、材料、流程、被试的任务)在字数限制之内尽可能详细。请充分利用 300 字限制，让摘要独立成文。

意见 10，

模型试用部分

模型目前四个模块建议采用统一语言表述，在图部分示意不够清晰，建议重新做

回应 10，

我们重新梳理认知重构的工作流程，修改了图 1 CBT-CoT 框架结构示意图，使得工作流程更清晰。我们用 3.2 - 3.5 节详细介绍了 4 个智能体的工作细节。

我们把图 2 改为表 1，用一个简单的案例化繁为简地呈现模型的心理推理过程。

我们不仅重画了图和表，还重做了实验。

意见 11，

3.5 推理过程不清晰，建议重新梳理。

回应 11，

我们重新梳理认知重构的工作流程，修改了图 1 CBT-CoT 框架结构示意图，使得工作流程更清晰。我们用 3.2 - 3.5 节详细介绍了 4 个智能体的工作细节。为梳理各 Agent 交互逻辑关系，我们在 4.2 的式（1）中作出了定义。

意见 12，

最后，作为问答系统，希望作者能提供一个模型的 demo 让评审来试用，可以附上链接。

回应 12，

感谢审稿专家的宝贵建议。首先，原标题为：CBT-CoT：基于多智能体思维链的心理健康认知行为疗法问答系统。

我们经过讨论，我们把标题改为：CBT-CoT：基于多智能体思维链的认知行为疗法框架。

一是原标题太长，“心理健康”略显多余。二是 CBT-CoT 具有良好的可扩展性，如向对话系统、或者与其它心理治疗方法结合，因此改为框架更为合适。

我们已经将多智能体框架代码上传至 GitHub，并且提供了相关的 requirements.txt 文件，包含了所需的包和版本，以便于其他研究人员能够复现我们的工作。链接将在论文中提供。

关于“提供模型的 demo”问题，由于时间太紧，这段时间在忙于修改稿件，来不及完成。如果论文录用，我们争取在会议上呈现我们的 demo。

审稿人二，意见 1，

主要问题

四个智能体之间的工作逻辑没有特别强的理论依据。CBT 的工作一般是识别自动化思维(事实->想法->情绪三栏表，这和 ABC 也不一样，ABC 不是 Beck 提出来的)、对自动化思维进行评价 (想法 OR 事实，相信程度，歪曲类型)、重建自动化思维(苏格拉底式提问)、形成替代性思维。本研究的 CBT 大模型工作顺序跟一般的临床工作不一样，先进行认知检测，检测歪曲的自动化思维，再进行 ABC，这个在逻辑上无法说得通。ABC 本来就是用于识别自动化思维，识别之后才评价是否有歪曲。

回应 1，

感谢审稿专家的宝贵意见，关于“多智能体结构未能充分贴合 CBT 标准工作流”的问题，我们认真反思并对系统架构进行了调整，确保其更为精准地符合 CBT 治疗的经典工作流程。

我们作了如下修改，并重新实验：

重新定义了四个智能体的角色与功能，并严格遵循了 CBT 治疗的核心流程。系统中的四个智能体依次执行以下任务：

（1）自动化思维识别： 首先，系统通过“三栏记录表”（事件-自动化思维-情绪反应）结构，全面提取用户在特定情境下的初级认知与情绪反应，确保每个认知点都能得到准确的拆解。

（2）认知评估： 与初稿版本不同，我们对逻辑顺序进行了优化。首先构建三栏表，随后进入多维度认知评估环节。此阶段不仅判断自动化思维是否存在认知扭曲，还新增了两个重要评估指标：

1）事实与想法判断： 系统将自动化思维分为客观事实与主观解释/预测，确保认知扭曲分析聚焦于非事实性认知。

2）信念强度分析： 我们通过模拟评估用户对自动化思维的信任程度（0-100%），进一步确定是否需要介入认知重构。

（3）结构化认知推理与重构： 在评估的基础上，系统进入推理阶段，通过选择恰当的 CBT 技术（如苏格拉底式提问、去灾难化等），生成“替代思维”和“情绪预测”，并最终构建五栏记录表。

（4）个性化语言输出：最后，系统通过共情性反馈和个性化语言回应用户，提供认知转变与情绪缓解的具体建议。

相较于初稿版本的结构，改进后的 CBT-CoT 系统更为贴合 CBT 理论的标准工作流程。通过明确的任务分工和顺序衔接，系统不仅增强了各阶段的逻辑一致性，还优化了对用户认知偏差和情绪需求的精准应对。此改进提升了系统的专业性与临床适用性，使其更符合实际治疗过程中对结构化推理和认知评估的高要求。

意见 2，

前言部分

研究空白定位不够聚焦。文中提出了多个研究不足点(如未能实现系统化推理链、缺乏精细建模等)，但逻辑串联略显松散，建议强化主线，并围绕它展开。

回应 2，

感谢审稿专家的细致建议。我们十分重视，围绕主线重写了相关工作。

意见 3，

CBT 的认知有三级，技术也比较多，为什么只关注自动化思维的重新建构，认知重构在 CBT 技术里面占怎样的位置？为什么只把这一技术拎出来，需要说明。

回应 3，

感谢审稿专家提出的宝贵问题。关于为何专注于认知重构技术，我们选择这一技术是因为它在 CBT 中的核心地位，且具有结构化推理的特点，非常适合与大语言模型的思维链（CoT）技术结合。认知重构通过多步推理帮助个体识别和调整非理性思维，而 CoT 技术也依赖逐步推理，能够有效地模拟这一过程。

相比之下，CBT 中的行为技术，如暴露疗法和行为试验，依赖个体在实际情境中的互动与行为反应，这对于目前的计算机大模型来说实现起来较为困难。

因此，本研究聚焦于认知重构，因为它更符合当前技术的应用场景，能够通过推理链生成个性化的替代思维和情绪调整。

意见 4，

对认知重构的工作流程解释不够清晰。

回应 4，

我们重新梳理认知重构的工作流程，修改了图 1 CBT-CoT 框架结构示意图，使得工作流程更清晰。我们用 3.2 - 3.5 节详细介绍了 4 个智能体的工作细节。

我们把图 2 改为表 1，用一个简单的案例化繁为简地呈现模型的心理推理过程。

意见 5，

方法部分(系统设计)

四个智能体之间的信息流未详细描述。图 1 虽有概括，但文中缺少对各 Agent 交互逻辑的文字描述，建议增加一段描述各模块之间如何传递信息，是否存在状态维护、缓存机制或反馈机制。

回应 5，

感谢审稿专家的宝贵意见。正如您提到的，CBT-CoT 多智能体的信息交互相对简单，主要体现在三栏表、五栏表的信息读写共享，我们修改了图 1，让其能明显呈现。各 Agent 交互逻辑关系，我们在 4.2 的式（1）中作出了定义。

感谢审稿专家对状态维护、缓存机制和反馈机制的深刻见解。针对您的问题：

（1）状态维护

当前，CBT-CoT 系统通过结构化中间表（如三栏表和五栏表）实现隐式状态维护。这些中间表在智能体间传递信息，确保推理链的连续性。每个智能体的输入包括前序智能体的输出，并依赖大语言模型（如 DeepSeek-Chat）维护上下文（如其上下文窗口）。这种设计避免了显式状态机的复杂性，符合当前多智能体研究中常见的上下文传递范式（如 Park et al., 2023）。未来，我们计划引入显式状态追踪模块，进一步支持多轮对话的状态维护。

（2）缓存机制

目前，CBT-CoT 系统未实现动态缓存机制，所有智能体均实时调用基座模型进行推理。虽然以往的基于规则的聊天机器人（如 Rasa）通常依赖状态维护和缓存机制来跟踪对话历史，以保证多轮对话的连贯性和上下文信息的传递，但基于大语言模型（LLM）的问答系统，凭借其强大的上下文理解和推理能力，能够自动处理多轮对话中的信息流，因此不再依赖外部的缓存机制。我们计划在后续版本中引入基于 LRU（Least Recently Used）策略的缓存机制，以提高推理效率，特别是在高频场景下（如 Sharma et al., 2023 的优化方案）。

（3）反馈机制

目前，CBT-CoT 系统通过多智能体的协作实现内部反馈，特别是在认知评估阶段，系统会对用户提供的自动化思维进行初步筛选，识别并过滤掉不合理、极端或存在认知扭曲的思维模式，以确保后续的推理和认知重构基于更加理性和现实的思维框架。然而，当前系统缺乏显式的用户反馈收集模块，尚未针对用户的满意度或反馈进行闭环优化。未来，我们计划在系统中引入用户反馈收集模块，允许用户对生成的回复进行评分，并基于此反馈进行模型调整。这将有助于实现更加个性化和动态优化的反馈机制，并推动系统在实际应用中的持续改进。

这些机制在现有多智能体研究中较少涉及（如 Yang et al., 2024 的 PsyChoGAT），但您的建议将显著增强系统实用性。

意见 6，

图 2 缺少文字对照解释。虽然图 2 提到了推理过程示例，但缺乏具体文字介绍推理链条是如何展开的，建议将该示例进一步文字化，说明每一步如何处理用户输入与生成逻辑。

回应 6，

感谢审稿专家的宝贵意见。针对“图 2 缺少文字对照解释”的问题，我们已对图 1 进行了进一步的修改与完善工作流程图，明确展示了各智能体之间的信息流动和任务分配。为了更好地帮助读者理解每个步骤的推理过程，删除了图 2，将图 2 改为表 1，用一个简单的案例化繁为简地呈现模型的心理推理过程。

意见 7，

实验与结果部分

实验设计未展开。只提到“利用真实问答进行验证”，但没有说明实验对照组设置、评价指标(如 ROUGE、BLEU、人评指标、共情评分等)、样本数量等，建议系统性补充。

回应 7，

关于“实验设计未展开”的问题，我们对实验设计部分进行了详细补充。在第 4 节 CBT-CoT 实验及效果评估中，我们明确了实验对照组的设置，包括与现有的自动评估方法（如 DeepSeek）和人工评估进行对比。同时，实验中使用了六维度的评分体系，涵盖了共情、认知重构、推理深度、表达策略合理性、逻辑连贯性、个性化程度等多个维度，以全面评估 CBT-CoT 的表现。

此外，我们增加了样本数量的描述，共计使用了 360 个样本进行测试，并进行了大规模的验证以确保评估结果的稳定性和可靠性。

关于语言学指标（如 ROUGE、BLEU 等），需要有对应心理问题的标准答案，计算实验结果与标准答案的偏差。但数据集中，心理问题对应的壹心理的网友回复，专业水平参差不齐，作为标准答案的说服力不强。因此放弃这些指标。

参考多个同类研究，我们构建了 6 个维度的评估指标，这更符合 CBT 框架中的核心目标。

意见 8，

数据集构建过程模糊。CBT-COT 数据集如何构建?是否全部为模型生成，是否有人工校对?数据清洗或标注是否考虑伦理因素?这部分对数据的合法性、代表性、标注规范缺乏描述

回应 8，

感谢审稿专家的宝贵意见。我们分析了两位专家的意见，经讨论，认为构建 CBT-COT 数据集的工作与论文的标题和核心宗旨不符，为避免论文的焦点分散，决定从原文中删除。

意见 9，

结果不全面。如未报告模型性能指标或与其他系统的对比结果。建议补充实验表格或图表，至少展示几种模型在关键维度的差异

回应 9，

已补充实验。我们在第 4.3 节 对比实验中已增加了与其他方法的对比结果，并通过新增的实验表格和图表展示了几种模型在关键维度上的性能差异。我们在对比中涵盖了 Zero-shot Prompting、Standard Prompting、CBT-LLM Prompting 等方法，以及原始的人工回复，并展示了它们在共情、认知重构、推理深度等多个维度上的具体表现。

意见 10，

只用大模型评测，不够有说服力，可考虑增加人工评测。

回应 10，

已增加人工评估。我们在第 4.4.2 节中对人工评估进行了详细说明。

意见 11，

讨论部分

讨论略显欠缺，未见深入分析。当前未展开对结果的解释、限制与未来工作的展望。

回应 11，

已改。参考本刊的风格，增加了对实验结果的深入分析，并提出了未来工作的展望。

意见 12，

其他建议

文章写作需要优化，语病较多。

回应 12，

已逐行逐句进行反复修改。

意见 13，

文献引用格式需要统一，不够规范。

回应 13，

用专用的软件统一文献引用格式，确保符合规范。

第 2 轮修改

审稿人一，意见 1，

核心推理思路依然存在较大混乱，从三栏表-认知评估-认知推理-认知重构，这四个概念无法并列（三栏表是工具）

回应 1，

回应：感谢审稿人的专业指正。我们认同早期版本存在将“三栏表”与认知过程并列表述不当的问题。经与心理学教授讨论后，我们对推理结构进行了重新梳理和明确划分。

我们将 CBT-CoT 框架划分为两个核心阶段：认知评估阶段与认知重构阶段。分别由四个智能体依序协作完成：

1）认知评估阶段由两个智能体完成：

自动思维提取 Agent：将用户的问题拆分为事件、自动思维以及情绪和行为反应填写“三栏表”。

自动思维评估 Agent：基于三栏表中自动思维进行分析，通过事实或想法、相信程度、歪曲类型来进行判断，是否进行后续的结构化推理。

2）认知重构阶段由两个智能体完成：

CoT 推理 Agent：基于三栏表和歪曲类型，通过检验假设、检查证据、匹配认知重构技巧，生成五栏记录表和认知重构技巧。

回复生成 Agent：基于五栏表和技巧生成具有针对性和共情性的回复。值得强调的是，为了增强回复的接受性，专门设计了避免心理学专业术语的生成策略，确保回复语言自然易懂。

我们已在论文第 2 节“CBT-CoT 框架”与第 3 节“CBT-CoT 的心理推理”中对上述推理逻辑进行了修订，删除了将工具与过程并列的表达，确保结构逻辑清晰、术语使用严谨。

希望以上修改能够更准确地呈现本研究的推理机制，再次感谢您的悉心指导。

意见 2，

CBT 干预的核心需要充分呈现用户认知上的偏差到底是什么，然后才有可能去讨论个性化的替代思维。一次互动即生成替代思维（认知重构）可能还会产生干预的副作用，建议在帮助用户识别自己的自动思维部分尽可能做扎实，再考虑怎么去调整用户也许多年来形成的自动思维，勿操之过急、适得其反。

回应 2，

回应：感谢审稿人指出该关键问题。我们完全认同 CBT 干预中对认知偏差的识别应当作为认知重构的前提，且认知调整应在充分理解个体自动思维的基础上逐步推进。在现实实践中，一轮对话往往难以实现完整而深入的认知重构。

本研究所提出的 CBT-CoT 框架，确实属于初步探索性的尝试，旨在验证大

语言模型是否具备模拟 CBT 基本推理路径的能力，为未来多轮动态干预的建构奠定基础。我们并不认为一次性生成替代思维已能替代真实 CBT 过程，而是通过结构化推理链“提取—评估—推理—生成”模拟治疗流程，以期探索其在 AI 辅助心理咨询中的可能性。

我们也在实验中设置了人工评估环节以监测系统风险和质量。例如，在对 100 条模型标注的认知扭曲类别样本中，完全正确占比 69%，部分正确 27%，完全错误仅占 4%，说明在认知偏差识别方面，系统已具备较强稳定性。该部分人工标注的数据已发布于 GitHub 项目主页（https://github.com/dlq1998/CBT-COT/sampled_data.xlsx），以供同行验证。

此外，我们对回复内容也进行了人工评分，在六维心理支持能力评分体系下，CBT-CoT 平均得分为 17.18（满分 18），表明系统输出具有较高的语言质量与心理合理性。

针对审稿人建议，我们已在论文的“5 不足与展望”部分补充了引入多轮对话机制与阶段式意图建模的计划，以增强系统在复杂心理问题处理中的适应性与对话连贯性。该机制旨在实现对用户心理状态的动态追踪与连续性支持，从而提升 CBT-CoT 框架在实际咨询情境中的应用深度与稳定性。

意见 3，

人工评估需考虑评估者一致性等细节，呈现都不充分。

回应 3，

回应：感谢审稿人对人工评估方法严谨性的关注。本文在设计人工评分流程时，确实充分考虑了评分者之间的一致性问题，并已在第 4.5.2 节“人工评估”中作出具体说明。有关内容如下：

“... （1）随机选择 15 个用户问题，通过 CBT-CoT 为每个用户问题生成回复。

（2）各评分员根据附录表 S3 打分标准，独立对 CBT-CoT 回复进行评分。

（3）对于分差超过 2 分的回复，在心理咨询师的介入下进行讨论，达成一致。

（4）重复上述 3 个步骤，直到各评分员之间达到足够高的一致性水平（要求加权 Cohen’s Kappa 值大于 0.7）。正式评估时，在 CBT-CoT 方法的回复样本中随机抽取 100 条数据进行人工复核。

...”

意见 4，

参考文献存在多处引用混乱。

回应 4，

回应：我们已在多轮修改过程中对参考文献进行了全面校对和调整，统一了引用格式，逐条核实文献的标注位置与文后条目，确保引文与文献编号一致、来

源信息准确。

意见 5，

因无法试用该思维链，对其实际应用效果存疑。

回应 5，

回应：感谢审稿人对本研究应用效果的关注。基于计算机学术界的开放共享惯例，我们已将完整的代码与提示词实现细节公开发布于 GitHub，任何研究者均可免费下载、复现和使用。代码链接为：https://github.com/dlq1998/CBT-COT。

此外，为更直观展示系统推理链条与真实运行效果，我们还录制并上传了实际运行的视频演示，视频链接为（百度网盘）：https://pan.baidu.com/s/1aRal4u4ARmBOiNOIak1GmA?pwd=gwuf 提取码: gwuf。需要特别说明的是，为了体现系统在真实场景下的适用性与随机性，视频中所用样例均来自“壹心理”平台近期更新的问题，未经过人工筛选或特殊设计，力求还原系统的真实表现。（由于当前演示视频主要目的是展示系统的实际运行效果，因此未在视频中包含作者的个人身份信息。）

由于时间限制，尚未搭建在线 demo 平台，但若论文获录用，我们将优先完成交互式 demo 的开发与部署。

意见 6，

在重要研究方法局限上回应不足。例如，ERD 的推理思路也正是我们擅长的研究方法。而且 ERD 善于正确识别没有扭曲的病例，避免过度诊断认知扭曲的陷阱。 但我们没做正确率检测，我们考虑后续进行这方面的工作。

回应 6，

回应：感谢审稿人指出关于认知扭曲识别准确性的重要问题。针对模型在识别认知偏差方面的表现，本文已通过人工标注进行初步验证：我们邀请两位心理学研究生对系统输出的认知扭曲类型进行评估，在 100 条样本中，完全正确占比为 69%，部分正确为 27%，完全错误仅为 4%，表明系统在识别存在认知扭曲的情境下具有较高的稳定性。我们在修改后的论文中增加了以下的内容：

“为检验 CBT-CoT 框架中自动思维评估 Agent 在认知扭曲识别任务中的有效性，本文从构建的数据集中随机抽取 100 条样本，邀请两位心理学研究生分别对模型标注的认知扭曲类别进行人工评分。评分采用三等级标准：0 分表示完全不正确，1 分表示部分正确（识别到部分标签但不完整或存在干扰标签），2 分表示完全正确（与人工标注完全一致）。在评分过程中，对于存在分歧的样本，由两位评分者协商讨论，最终达成一致评分。...”（4.3 认知扭曲识别质量评估）

审稿人二，意见 1，

文中心理学词汇的表述不够规范，如“三栏表”和“五栏表”是常用技术，目的是认知识别和认知调整，与其他“认知评估-认知推理-认知重构”，无法并列，最好与其他词汇并列。

回应 1，

回应：感谢审稿人对术语规范性的专业指正。我们完全认同“三栏表”和“五栏表”在认知行为疗法（CBT）中属于临床常用的结构化技术工具，主要用于辅助认知识别与认知调整，不应与“认知评估”“认知推理”“认知重构”等心理过程并列表达。

我们将 CBT-CoT 框架划分为两个核心阶段：认知评估阶段与认知重构阶段。分别由四个智能体依序协作完成：

1）认知评估阶段由两个智能体完成：

自动思维提取 Agent：将用户的问题拆分为事件、自动思维以及情绪和行为反应填写“三栏表”。

自动思维评估 Agent：基于三栏表中自动思维进行分析，通过事实或想法、相信程度、歪曲类型来进行判断，是否进行后续的结构化推理。

2）认知重构阶段由两个智能体完成：

CoT 推理 Agent：基于三栏表和歪曲类型，通过检验假设、检查证据、匹配认知重构技巧，生成五栏记录表和认知重构技巧。

回复生成 Agent：基于五栏表和技巧生成具有针对性和共情性的回复。值得强调的是，为了增强回复的接受性，专门设计了避免心理学专业术语的生成策略，确保回复语言自然易懂。

我们已在论文第 2 节“CBT-CoT 框架”与第 3 节“CBT-CoT 的心理推理”中对上述推理逻辑进行了修订，删除了将工具与过程并列的表达，确保结构逻辑清晰、术语使用严谨。

希望以上修改能够更准确地呈现本研究的推理机制，再次感谢您的悉心指导。

意见 2，

讨论中建议补充 CBT-CoT 的伦理挑战与误用风险（例如误判认知扭曲的后果、共情失败等）将使论文更完整，尤其对临床心理学界更具说服力。

回应 2，

回应：感谢审稿人提出的宝贵建议。我们完全认同在 AI 心理支持系统的研究中，必须严肃对待潜在的伦理风险与误用隐患。

在模型实际表现方面，本文已对认知扭曲识别模块进行了人工校验，邀请两位心理学研究生对 100 条系统标注样本进行评分，完全正确占比 69%，部分正确为 27%，完全错误仅为 4%，初步验证了模型在识别典型认知偏差方面的稳定性。

该部分人工标注的数据已发布于 GitHub 项目主页（https://github.com/dlq1998/CBT-COT/sampled_data.xlsx），以供同行验证。我们在修改后的论文中增加了以下的内容：

“为检验 CBT-CoT 框架中自动思维评估 Agent 在认知扭曲识别任务中的有效性，本文从构建的数据集中随机抽取 100 条样本，邀请两位心理学研究生分别对模型标注的认知扭曲类别进行人工评分。评分采用三等级标准：0 分表示完全不正确，1 分表示部分正确（识别到部分标签但不完整或存在干扰标签），2 分表示完全正确（与人工标注完全一致）。在评分过程中，对于存在分歧的样本，由两位评分者协商讨论，最终达成一致评分。...”（4.3 认知扭曲识别质量评估）

然而，我们也充分认识到，任何形式的误判（如将合理陈述误判为扭曲）都可能对用户产生负面影响，尤其是在缺乏专业人工监督的情况下，风险不容忽视。我们将在后续研究中持续加强这一环节的防控设计，提升模型的判断准确性与心理安全性保障。

针对审稿人提出的建议，我们已在论文“5 不足与展望”部分补充说明了系统的伦理风险防控思路，并明确规划了两项关键优化方向：（1）引入多轮对话机制与阶段式意图建模，以实现对用户自动思维的反复确认与认知偏差的递进识别，避免认知重构过程中的仓促判断；（2）增强系统在连续性心理引导中的结构组织能力，以更好应对复杂问题场景下的咨询需求，提升对话连贯性与干预的心理接受度。

再次感谢您对本研究严谨性与社会责任感的关注。

意见 3，

文中仍存在多处写作语言不精炼，专业表述不够规范。

回应 3，

回应：感谢审稿人指出语言表达与专业表述方面的问题。针对该意见，我们已邀请多位心理学教授及学校心理咨询中心的资深咨询师对全文内容进行了审阅与修改，重点优化了术语使用的规范性和心理学表达的专业性。同时，我们也对冗余句式和不够精炼的描述进行了统一调整，提升语言的准确性与表达质量。感谢您的细致审阅与宝贵建议。

第 3 轮修改

审稿人一，意见 1，

建议让作者补充智能体的使用路径，评审可试用该智能体并进行判断。应用智能体能够有效应用才是关键。

回应 1，

回应：感谢审稿专家的宝贵建议。我们已在修订稿中补充智能体的使用路径，并提供在线试用入口：http://121.43.112.225:8502。

说明：该 Demo 为轻量原型，界面仅展示最终回答；论文所述的内部流程在后台完成、未在界面显式呈现。再次感谢您的指正。

审稿人二，意见 1，

文中引文和参考文献格式需要统一。

回应 1，

回应：感谢审稿专家的指正。我们已对全文进行统一规范处理：文内引文与参考文献格式按同一标准彻底统一，并完成逐条核对与交叉检查。针对相关问题，已按期刊规范逐条更正，例如：

“H. Zhang et al., 2024” 更正为： “Zhang et al., 2024”

“G. I. Clark & Egan, 2015” 更正为： “Clark & Egan, 2015”

多处同类问题已全文统一；文末参考文献表亦与文内引文一一对应。同时，我们已邀请一位应用心理学博士、副教授、国家二级心理咨询师对稿件进行挑剔性通读与校对，依据反馈完成相应修改，相关修订已体现在本次修订稿中。再次感谢您的审阅与指导。

编委意见:

请作者对文字和语言进行进一步提升，并请中级职称及以上同行进行挑剔性阅读。

回应：非常感谢编委专家和审稿人对本研究的宝贵意见和建议。我们已对全文逐段开展语言润色与行文优化，统一术语与表述；同时邀请一位应用心理学博士、副教授、国家二级心理咨询师对稿件进行挑剔性通读与校对，并据其反馈逐条修订，相关修改已体现在本次稿件中。

重点修改了第 1.1 节末段与第 1.2 节的专业表述，使逻辑更严谨、语句更流畅，以进一步提升论文的可读性与学术规范。

此外，全文各章节（含摘要、引言、方法、结果、讨论、结论、图表题注及附录等）均已按同一标准完成系统性修订与一致性校对，确保行文更为流畅、逻辑更为清晰、术语更为规范。