Psy-Copilot: Visual Chain of Thought for Counseling

Anonymous submission

Abstract

Large language models (LLMs) are becoming increasingly popular in the field of psychological counseling. However, when human therapists work with LLMs in therapy sessions, it is hard to understand how the model gives the answers. To address this, we have constructed Psy-COT, a graph designed to visualize the thought processes of LLMs during therapy sessions. The Psy-COT graph presents semi-structured counseling conversations alongside step-by-step annotations that capture the reasoning and insights of therapists. Moreover, we have developed Psy-Copilot, which is a conversational AI assistant designed to assist human psychological therapists in their consultations. It can offer traceable psycho-information based on retrieval, including response candidates, similar dialogue sessions, related strategies, and visual traces of results. We have also built an interactive platform for AI-assisted counseling. It has an interface that displays the relevant parts of the retrieval sub-graph. The Psy-Copilot is designed not to replace psychotherapists, but to foster collaboration between AI and human therapists, thereby promoting mental health development. Our code and demo are both open-sourced and available for use. Users can act as clients or therapists to interact with Psy-Copilot on this website [1].

Introduction

In recent years, empathetic conversational agents that can provide psychological support have become increasingly popular. However, most existing mental support dialogue agents are designed to interact directly with clients, lacking the ability to collaborate with psychotherapists in counseling (Lin et al. 2020; Sabour et al. 2022; Zhang et al. 2024). In psychological counseling, human therapists need a copilot to assist in consultations, rather than assigning patients directly to a black-box AI conversational agent. Psychological copilot should not only provide response candidates but, more importantly, offer traceable and reliable information for therapists.

To achieve this goal, we construct a Psy-COT graph and Psy-Copilot conversational agent. They constitute an AI assistant for psychological counseling tasks. The Psy-COT graph consists of 941 embedded dialogue sessions and a directed graph of events and strategies. Psy-COT is a multi-level, indexable graph, which is designed for visualizing and

retrieving the chain of thought progress in counseling.

As shown in Figure 1, Psy-Copilot is a conversation assistant that can aid therapists in counseling. It can leverage retrieval text to generate traceable counseling responses. With indexes in Psy-COT, Psy-Copilot can retrieve precise similar study cases and required strategies for reasoning. Psy-Copilot can enhance the performance of LLMs and provide better visualization of how the AI agent generates response candidates.

Psy-COT has similarities to knowledge graphs (Hogan et al. 2021). As shown in Figure 2, unlike knowledge graphs, which focus on inclusion relationships between entities (eg., "Disappointed"), Psy-COT emphasizes the logical chain of causality in events and the temporal evolution of strategies in counseling (eg., Expand Client's Client's relationship).

In our research, we focus on (1) Collecting counseling dialogue and interpretation data to build a graph that maps out event connections and strategic transitions; (2) Constructing an AI agent that can produce reliable, explainable, and traceable results; (3) Building an interactive platform to ensure that the AI agent can collaborate with human therapists, with outcomes that are visualized and easily comprehensible.

Figure 1: Screenshot of Psy-Copilot. The left side is the chat interface, and the right side is the visualized COT.
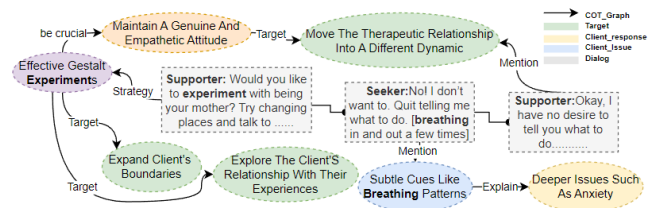

Figure 2: A small sub-graph of Psy-COT graph. The Psy-COT maps events and strategies to dialogue units, preserving causal and temporal relationships.
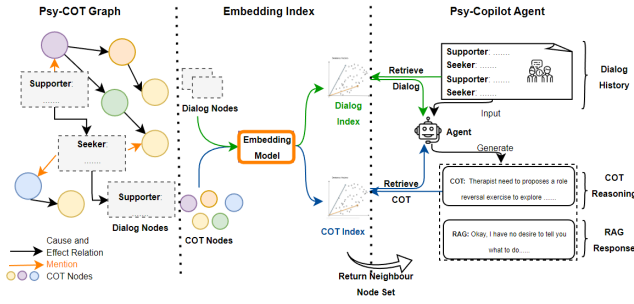

Figure 3: Overview of indexes in Psy-COT and retrieval progress in Psy-Copilot. Psy-COT has two vector indexes for dialog and COT reasoning content respectively. There is a two-stage retrieval argument generation in Psy-Copilot.

Technical Details

Psy-COT Graph Construction

Data Source We collect 941 psycho-counseling sessions from psychological blogs, each containing dialog and corresponding explanatory text.

Construction of COT Graph We design a multi-agent framework to automate the extraction of strategy descriptions from unstructured data. Based on psychological counseling research (Radcliffe, Comfort, and Fawcett 2015), we define the types of strategies and causal relationships. Using a Deepseek-powered Langchain Agent, we extract text fragments describing strategies and established causal relationships between nodes. Subsequently, we employ a sliding window technique to map these COT nodes to dialogue nodes, thereby establishing temporal relationships.

Multi-Level Index for Retrieval In contrast to graph-GAG (Edge et al. 2024), which employs embedded text chunks for document retrieval, we adopt a fine-grained approach that is designed for counseling dialog. As shown in the left side of Figure 3, we construct two index structures for COT nodes and dialogue nodes.

We have found the distinction between the content of the chain of thought and the dialogue. The COT content is derived from explanations in blogs, using specialized psychological terms and direct descriptions, such as "Effective Gestalt Experiments" or "Deeper Issues Such As Anxiety." However, the dialogue content is more euphemistic and, like, "Alright, I'm not here to tell you what to do." Given these differences, we have created separate indexes in Psy-COT for each content type, which makes vector retrieval more precise.

Psy-Copilot Agent

Base LLMs We utilize open-source models Qwen-2.5-7B (2023) and Deepseek-V2 (2024) to power our retrieval and chat agents, making Psy-Copilot easy to reproduce.

Retrieval Argument Generation Based on two types of embedding index in Psy-COT, we adopt a two-stage retrieval argument generation for Psy-Copilot. As shown on the right side of Figure 3, green arrows indicate dialogue retrieval pipeline for dialogue and blue arrows show retrieval progress for generated reasoning data.

In the first step marked by the green arrow, the agent uses the conversation history to find similar dialogues in the Psy-COT dialogue index by comparing them using cosine similarity. Next, the agent reasons step-by-step based on the conversational history, to analyze the user's case and determine counseling strategies (Wei et al. 2022). Based on the generated reasoning, the agent retrieves relevant counseling strategies in the COT index.

We filter the nodes and their neighboring nodes obtained from the results of COT retrieval and dialog retrieval, prioritizing overlapping nodes. We concatenate the content of dialogue nodes into few-shot examples, and COT contents into instructions. Finally, the model generates psychological counseling responses based on retrieval information.

Demo UI We have constructed an online website to provide an interactive experience with Psy-Copilot and showcase the Psy-COT graph. During each dialogue round, we dynamically retrieve and display related sub-graphs for the therapist. Sub-graphs make the logic chain behind generation easier to understand and trace.

Objective Evaluation Our agent can generate multi-turn dialogue. However, the interactive evaluation of multi-turn dialog is time-consuming and difficult to quantify. Inspired by Emobench (Sabour et al. 2024), we apply GLM4-9b (Du et al. 2022) LLM for evaluating the emotional intelligence of the Psy-Copilot.

As shown in Table 1, we assessed the multi-turn dialogues with 4 metrics: Fluency (Flu.), Helpfulness (Hel.), Naturalness (Nat.), and Comforting Effectiveness (Com.). Detailed definitions of metrics and evaluation codes are available on our paper's website. Our experiments show that Psy-Agent outperforms baseline models across all four metrics,

Model	Flu.	Hel.	Nat.	Com.
Deepseek-V2	8.2	7.2	6.7	7.6
Psy-Copilot-Dialog	8.5	7.4	**7.2**	**8.2**
Psy-Copilot-COT	**8.6**	**7.5**	7.0	7.9

Table 1: Human evaluation focuses on fluency, helpfulness, naturalness, and comforting effectiveness.

Conclusion

We demo Psy-COT and Psy-Copilot, a visual graph and a conversational agent aimed at providing traceable and reliable psychological counseling. Psy-Copilot can act as a helpful AI assistant for psychotherapists. We open-source the pipeline for graph construction and deploy an interactive counseling conversation platform. Our goal is to demonstrate the AI's response crafting process, cognitive patterns, and interaction techniques with patients, thereby fostering a sense of trust in the AI system among therapists. We aim to enhance the synergy between AI and therapists, creating a harmonious and balanced relationship among AI, therapists, and patients.

References

Bai, J.; Bai, S.; Chu, Y.; Cui, Z.; Dang, K.; Deng, X.; Fan, Y.; Ge, W.; Han, Y.; and et. al. 2023. Qwen Technical Report. arXiv:2309.16609.

DeepSeek-AI; Liu, A.; Feng, B.; Wang, B.; Wang, B.; Liu, B.; Zhao, C.; Dengr, C.; Ruan, C.; Dai, D.; Guo, D.; Yang, D.; and et. al. 2024. DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model. arXiv:2405.04434.

Du, Z.; Qian, Y.; Liu, X.; Ding, M.; Qiu, J.; Yang, Z.; and Tang, J. 2022. GLM: General Language Model Pretraining with Autoregressive Blank Infilling. In Muresan, S.; Nakov, P.; and Villavicencio, A., eds., *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 320–335. Dublin, Ireland: Association for Computational Linguistics.

Edge, D.; Trinh, H.; Cheng, N.; Bradley, J.; Chao, A.; Mody, A.; Truitt, S.; and Larson, J. 2024. From Local to Global: A Graph RAG Approach to Query-Focused Summarization. arXiv:2404.16130.

Hogan, A.; Blomqvist, E.; Cochez, M.; D'amato, C.; Melo, G. D.; Gutierrez, C.; Kirrane, S.; Gayo, J. E. L.; Navigli, R.; Neumaier, S.; Ngomo, A.-C. N.; Polleres, A.; Rashid, S. M.; Rula, A.; Schmelzeisen, L.; Sequeda, J.; Staab, S.; and Zimmermann, A. 2021. Knowledge Graphs. *ACM Computing Surveys*, 54(4): 1–37.

Lin, Z.; Xu, P.; Winata, G. I.; Siddique, F. B.; Liu, Z.; Shin, J.; and Fung, P. 2020. CAiRE: An End-to-End Empathetic Chatbot. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(09): 13622–13623.

Radcliffe, J. N.; Comfort, P.; and Fawcett, T. 2015. Psychological strategies included by strength and conditioning coaches in applied strength and conditioning. *The Journal of Strength & Conditioning Research*, 29(9): 2641–2654.

Sabour, S.; Liu, S.; Zhang, Z.; Liu, J. M.; Zhou, J.; Sunaryo, A. S.; Li, J.; Lee, T. M. C.; Mihalcea, R.; and Huang, M. 2024. EmoBench: Evaluating the Emotional Intelligence of Large Language Models.

Sabour, S.; Zhang, W.; Xiao, X.; Zhang, Y.; Zheng, Y.; Wen, J.; Zhao, J.; and Huang, M. 2022. Chatbots for Mental Health Support: Exploring the Impact of Emohaa on Reducing Mental Distress in China.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35: 24824–24837.

Zhang, C.; Li, R.; Tan, M.; Yang, M.; Zhu, J.; Yang, D.; Zhao, J.; Ye, G.; Li, C.; and Hu, X. 2024. CPsyCoun: A Report-based Multi-turn Dialogue Reconstruction and Evaluation Framework for Chinese Psychological Counseling. arXiv:2405.16433.