Domain Specific Improvement On Psychotherapy Chatbot Using Assistant

Cheng Kang[1], Katerina Urbanova[2], Yuqing Cheng[3], Yong Hu[4], Daniel Novak[1]

[1] Czech Technical University in Prague
[2] National Mental Health Institute
[3] Shenzhen Mental Health Centre
[4] The University of Hong Kong

Check out our project page here (with demos)

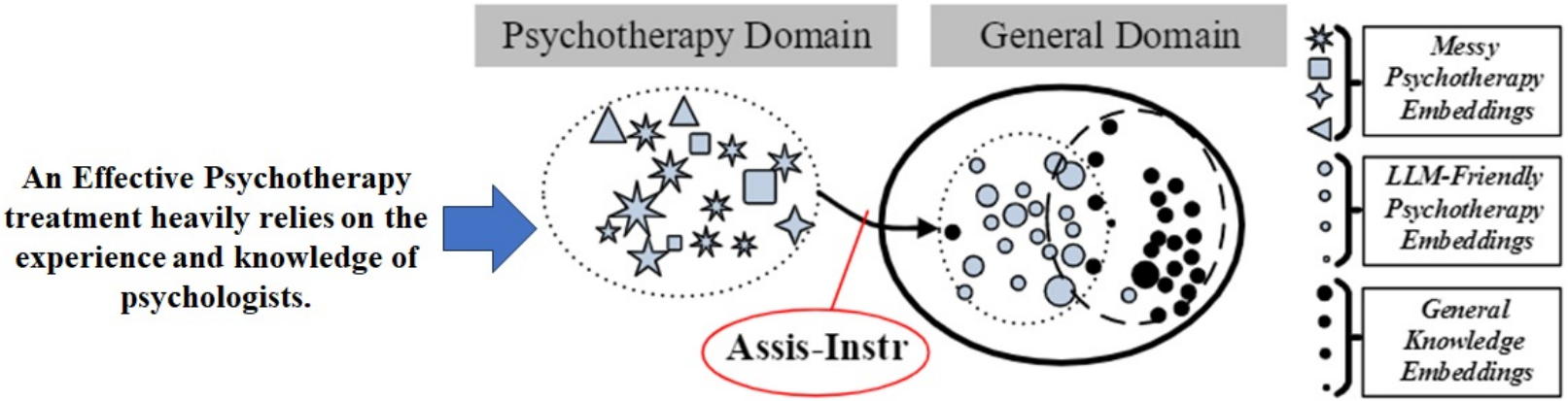
Check out our project page here (with demos)

Introduction

Background

- Effective Psychotherapy Treatments;
- Psychiatric Chatbot Using Large Language Models (LLMs);
- The lack of psychotherapy knowledge of LLMs;

A successful model is expected to use the provided instructions (including task- and domain-definition examples) to output professional instances. The **Assistant** can be: Psychologists (Experts), Machines, or Mixture use.



Previous Solutions

Two main technical ways have been developed for extending and augmenting the domain-specific data:

1. Using human-annotated data on a wide range of tasks

- Reinforcement Learning on Human Feedback (RLHF) [4];
- Human annotated prompts [6].

2. Using datasets augmented with manually or automatically generated instructions

- Machine generated instruction following data [3];
- Self-Instruct tuning [5].

Our Method

The Assistant-Instruction (semi-self-instruction) in psychotherapy domains.

- Data Cleaning and Information Extracting;
- Task Identification;
- Data revision and knowledge expansion;
- Evaluation and decision of Acceptability.

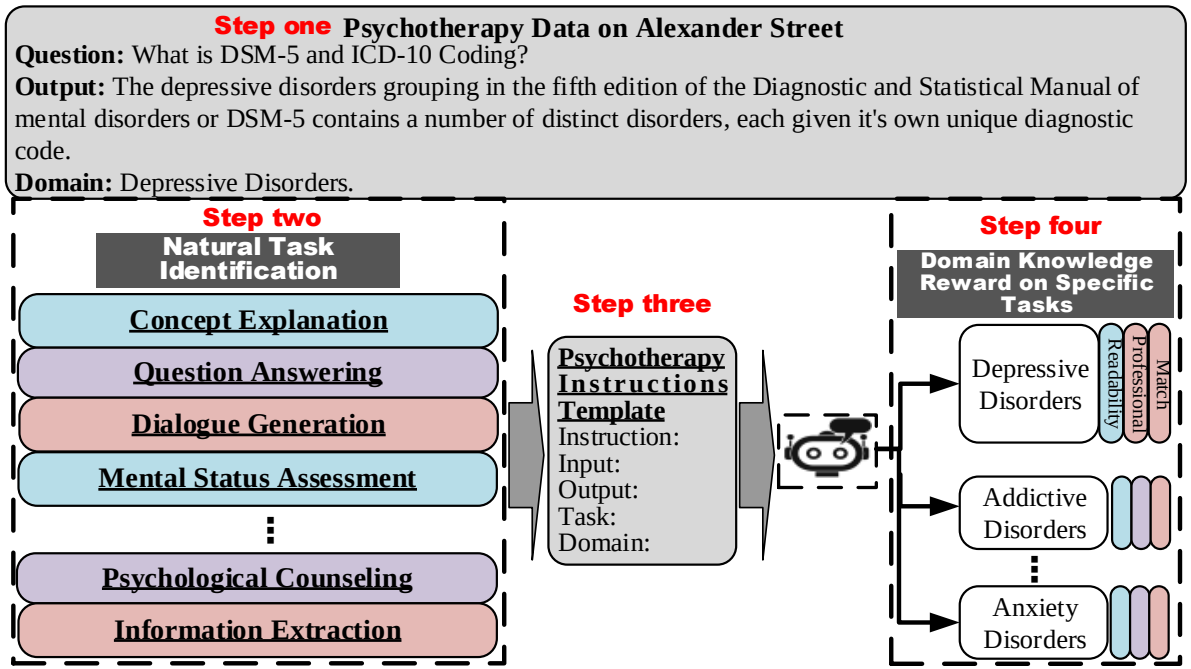


Figure: Steps to generate the Assistant-Instruction Data.

Data Collection and Models

Using GPT-4 as an assistant, and applying inhibited Low Rank Adaption (LoRA) [2] and Self Retrieval Augmented Generation (Self-RAG) [1], Assistant-Instruction fine-tuned LLMs are evaluated under two main metrics: Automatic evaluation and Human evaluation.

- Alexander Street counselling transcripts;
- GPT-4 as an assistant;
- LoRA fine-tuning and RAG;
- Evaluation of Assistants and Experts;

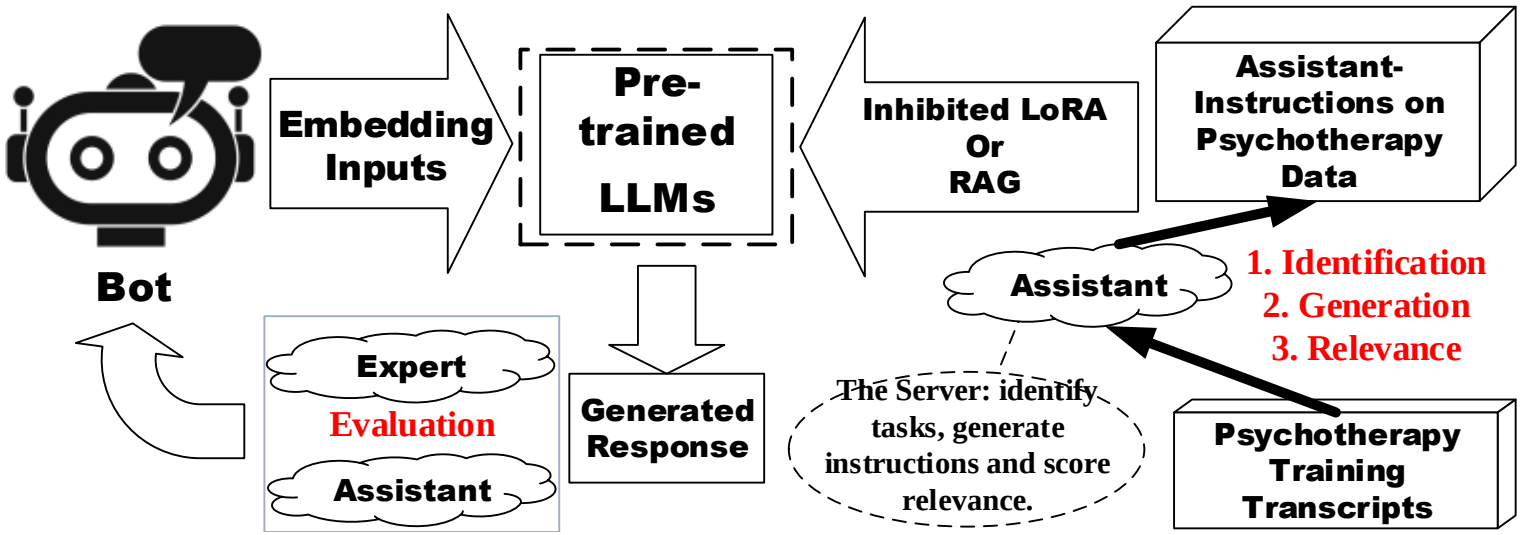


Figure: The schematic of our system.

The Performance of Assistant-Instruction

Table: Based on Llama2-7B, we illustrate the performance of Zero-Shot, inhibited LoRA Tuned and RAG methods on Psychotherapy data.

Instruction: Please provide professional suggestions or comments to effectively address and alleviate [Autism]. **Input:** We are discussing [Autism].
Output from Zero-Shot: Autism is a developmental disorder that can cause significant social, communication and behavioral challenges. There is no known single cause for autism, but increased awareness and funding support research into possible causes and better ways to treat the disorder.
Output from Inhibited LoRA: Autism is a developmental disorder that can cause significant social, communication and behavioral challenges. It is a spectrum disorder, which means that symptoms and severity vary widely among individuals.
Output from Inhibited LoRA + RAG: Autism is a developmental disorder that can cause significant social, communication and behavioral challenges. It is a spectrum disorder, which means that symptoms can range from mild to severe. Some people with autism may also have other conditions, such as attention deficit hyperactivity disorder (ADHD), anxiety disorders, or epilepsy. Applied Behavior Analysis (ABA) a widely recognized and evidence-based therapy for Autism. ABA techniques focus on promoting positive behaviors and reducing challenging behaviors. Speech and language therapy can help improve communication skills, such as speech articulation, expressive language, and social communication.

Evaluation Results

- 1. Automatic Evaluation: rouge (higher is better) and fluency (lower is better, perplexity based on GPT-2);
- 2. Human Evaluation: Readability, Professional, Match. (**Rule:** *Extremely Bad (1), Bad (2), Neutral (3), Acceptable (4), Good (5) and Very Good (6)*)

Table: For evaluating the performance of LLMs on psychotherapy domain, two methods - inhibited LoRA and RAG - were used on two pre-trained LLMs.

Inhibited LoRA Finetuning (without / with Asisstant-Instruction)					
Pretrained	Automatic		Human Evaluation		
LLM	Rouge ↑	Fluency ↓	Read	Prof	Match
ChatGLM2-6B	24.3/27.1	49.4/48.7	4.8/4.9	2.9/3.3	2.1/2.5
Llama2-7B	15.1/16.9	20.9/20.5	5.0/5.2	3.0/3.2	1.9/2.3
Retravel Augmented Generation (without / with Asisstant-Instruction)					
Pretrained	Automatic		Human Evaluation		
LLM	Rouge ↑	Fluency ↓	Read	Prof	Match
ChatGLM2-6B	25.1/32.8	56.4/46.7	4.6/5.3	3.9/4.2	2.9/3.3
Llama2-7B	15.4/22.4	30.3/20.7	4.8/5.2	3.7/4.1	3.0/3.4

Conclusion

We propose a novel method called ASSISTANT-INSTRUCT to improve the instruction-following ability of LLM in the psychotherapy domains. This method

- can provide additional knowledge to LLMs, but avoid heavily manual work.
- combines common knowledge and professional psychotherapy knowledge to generate instruction data with the help of LLM assistants.
- retains general knowledge and incorporates specific psychotherapy knowledge of LLM from Assistant-revised instructions.

References

Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi.

Self-rag: Learning to retrieve, generate, and critique through self-reflection.

arXiv preprint arXiv:2310.11511, 2023.

Cheng Kang, Jindich Prokop, Lei Tong, Huiyu Zhou, Yong Hu, and Daniel Novak.

Ina: Inhibition adaption on pre-trained language models.

Available at SSRN 4551993.

Baolin Peng, Chunyuan Li, Pengcheng He, Michel Galley, and Jianfeng Gao.

Instruction tuning with gpt-4.

arXiv preprint arXiv:2304.03277, 2023.

Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting.

Large language models in medicine.

Nature medicine, 29(8):1930–1940, 2023.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi.

Self-instruct: Aligning language models with self-generated instructions.

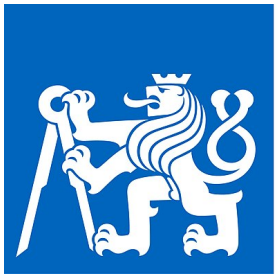
arXiv preprint arXiv:2212.10560, 2022.

Linyao Yang, Hongyang Chen, Zhao Li, Xiao Ding, and Xindong Wu.

Chatgpt is not enough: Enhancing large language models with knowledge graphs for fact-aware language modeling.

arXiv preprint arXiv:2306.11489, 2023.

Thank You!



Check out our project page here (with demos)

Check out our project page here (with demos)