

DEMENTIA CARE RESEARCH (RESEARCH PROJECTS; NONPHARMACOLOGICAL)

Revolutionizing Dementia Care: Enhancing Talk Therapy with
Fine-Tuned Large Language Models Using GPT Self-Generated
Data

Jia Xu | Tianyi Wei | Bojian Hou | George Demiris | Li Shen

University of Pennsylvania, Philadelphia, PA,
USA

Correspondence

Jia Xu, University of Pennsylvania,
Philadelphia, PA, USA.
Email: jiayu@alumni.upenn.edu

Abstract

Background: It has been validated that dementia can be alleviated by talk therapy. While general large language models (LLMs) have the potential for talk therapy and dementia treatment, customized LLMs resembling a mental-health therapist can be more professional and powerful. We introduced an LLM solution designed for tailored mental health support for future dementia care. By employing GPT self-generated data, encompassing a diverse range of mental health scenarios, we've significantly enhanced the model's ability to understand and engage with complex emotional and psychological queries from patients. This advancement in LLMs paves the way for more efficient and effective delivery of talk therapy, shedding light for future dementia care and treatment.

Method: We fine-tuned five open-source LLMs to enhance their capabilities in the context of mental health counseling. The fine-tuning process was conducted using the quantized Low-Rank Adapters (QLoRA) method on an A40 GPU. We generated 10,000 question-answer pairs as the training data using GPT-3.5-Turbo, designed to simulate interactions between mental health counselors and patients. This dataset encompassed a broad spectrum of mental health issues, including but not limited to depression, anxiety, and sleep disorders. Post fine-tuning, the performance of these models in test data gathered from the Mental Health Forum and Quora (Fig. 2) was judged comprehensively by GPT-4 Turbo. This evaluation was structured around six mental health metrics (Fig. 1) to ensure precise and relevant assessment in the context of mental health therapy.

Result: The top-performing model, "Mixtral-8x7B-v0.1-qlora," achieved a remarkable average score of 7.83 (Tab. 1), highlighting its superior capability over the base model in handling complex mental health interactions. Other fine-tuned models, such as

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Alzheimer's Association. *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

"Zephyr-7b-gpt" and "Vicuna-7b-gpt," also showcased exceptional performance over their base models, scoring 7.82 and 7.78 respectively. These results underscore the efficacy of fine-tuning in enhancing the empathetic and contextual responsiveness of LLMs in mental health settings, particularly in dementia care.

Conclusion: The study's results showcased the effectiveness of fine-tuned large language models (LLMs) suffixed by "QLoRA" in enhancing talk therapy. The leading model, "Mistral-7B-Instruct-v0.2-QLoRA," demonstrated superior empathy and understanding over the base model in complex dementia interactions by leveraging domain-specific data.

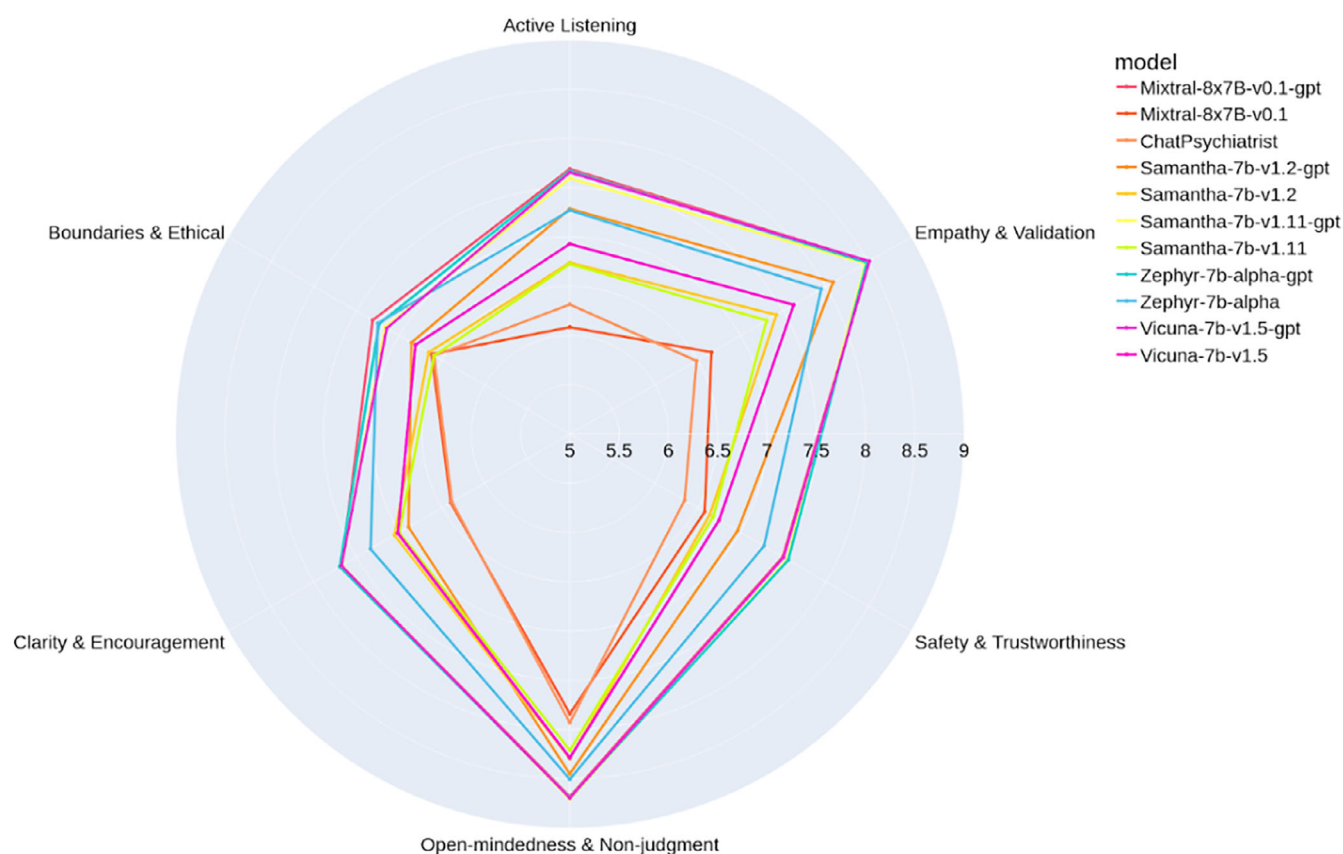


Fig. 1: GPT-4-Turbo evaluation in six mental health metrics: Active Listening, Empathy & Validation, Safety & Trustworthiness, Open-mindedness & Non-judgment, Clarity & Encouragement, Boundaries & Ethical.

	Model	Average Metrics Score	Version
1	Mixtral-8x7B-v0.1-QLoRA	7.83	Fine-tuned
2	Zephyr-7b-QLoRA	7.82	Fine-tuned
3	Vicuna-7b-QLoRA	7.78	Fine-tuned
4	Samantha-7b-v1.11-QLoRA	7.77	Fine-tuned
5	Zephyr-7b-alpha	7.48	Base
6	Samantha-7b-v1.2-QLoRA	7.37	Fine-tuned
7	Vicuna-7b-v1.5	7.08	Base
8	Samantha-7b-v1.2	7.03	Base
9	Samantha-7b-v1.11	6.98	Base
10	ChatPsychiatrist	6.48	Other
11	Mixtral-8x7B-Instruct-v0.1	5.85	Base

Tab. 1: Average scores of base and fine-tuned LLMs over six mental health metrics.