Re-imagining Mental Health Access: The Role of Human, AI and Design

Zainab Iftikhar
zainab_iftikhar@brown.edu
Computer Science
Brown University
Providence, Rhode Island, USA

ABSTRACT

Access to care is one of the major barriers to mental health treatment, leading many individuals to turn to unproven methods for support, such as large language models (LLMs). This dissertation investigates how humans, design, and AI can collectively support mental health needs without overburdening practitioners. Through user studies and mixed-method analysis of a text-based peer support dialogue dataset, my research has examined the impact of training peer support providers in Cognitive Behavioral Therapy (CBT) techniques and explores how redesigning text-based platforms with typing indicators can increase users' perceived social presence. These studies have highlighted how CBT training improves users' perceived empathy and how platform design can affect human connection. However, my work reveals a disconnect between the nuanced human experience and the often limited machine interpretation of empathy in mental health care. This finding paves the way for my future research, which will focus on developing new evaluation metrics and benchmarks to evaluate and improve the interpretability of LLM outputs in mental health care.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing**; • **Computing methodologies** → **Discourse, dialogue and pragmatics**; • **Applied computing** → *Health informatics*.

KEYWORDS

technology-assisted mental health; social computing; interaction design; large language models; responsible AI

ACM Reference Format:
Zainab Iftikhar. 2024. Re-imagining Mental Health Access: The Role of Human, AI and Design. In *Companion of the 2024 Computer-Supported Cooperative Work and Social Computing (CSCW Companion '24), November 9–13, 2024, San Jose, Costa Rica.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3678884.3682056

CSCW Companion '24, November 9–13, 2024, San Jose, Costa Rica

ACM ISBN 979-8-4007-1114-5/24/11
https://doi.org/10.1145/3678884.3682056

1 INTRODUCTION

One in four Americans prefer talking to AI instead of a therapist, with 80% of them considering it an effective alternative [8]. The United States has an average of thirty psychologists per a hundred thousand people [10], leading to issues of inaccessibility, high costs, and complexity, leaving more than half of U.S. adults with mental illness without adequate care [16]. Improving the patient-practitioner ratio solely through training additional professionals is unlikely to meet the current healthcare needs. Instead, new scalable approaches are emerging to expand access to care, including peer support platforms [11, 20] and recent advancements in AI, also known as large language models (LLMs) [13]. From a language generation system to "therapist", LLMs have enabled some people immediate access to care through fully automated means [15].

However, using LLMs for support when they were originally designed for natural language processing raises questions about their effectiveness and reliability in providing mental health care due to their lack of high-quality training data, inherent societal biases, and concerns regarding data privacy and confidentiality [12]. Despite these concerns, an overwhelming number of online users attest to their effectiveness. However, informal and personal evaluations of these tools are subjective and influenced by individual circumstances. Those seeking support may not fully recognize the risks, as LLMs offer immediate support, albeit limited in scope. This gap has led to ongoing debates between users and experts. The rising public interest in using LLMs for mental health, coupled with expert reservations, calls for a more rigorous and objective evaluation.

This thesis proposal will evaluate LLM-based support in mental health through clinical metrics. Pre-existing text-based session dialog datasets from human peers will be reconstructed using a CBT-prompted LLM. Through a mixed-methods analysis rooted in CBT literature, licensed psychologists will annotate and rate both human and LLM sessions for quality of support using the Cognitive Therapy Rating Scale (CTRS) and qualitative comments. By doing this, this proposal aims to understand the risks and benefits of both support methods and explore how effective collaboration between human and AI peers can be facilitated through safe and responsible AI.

Disclaimer: This research will involve a controlled study where LLM-generated sessions will be reconstructed from publicly available peer support sessions. An artificial experiment design is chosen to avoid an experimental method that may cause harm.

2 RELATED WORK

This section covers the current cost-effective alternatives like peer support platforms and AI-mediated health care that have emerged as accessible solutions in digital mental health.

2.1 Technological Modalities in Mental Health: From Conversational Agents to Large Language Models

Administering AI-Mediated Therapeutic Interventions: Researchers in digital mental health have focused on innovative care strategies, such as using self-disclosing chatbots to foster meaningful user interactions[6], integration of large-scale counseling data with natural language processing to enable quality improvement [7] and LLMs for mental health applications. These applications range from prompt design to treatment and care evaluation and have demonstrated AI's efficacy in non-judgmental and accessible mood management [5]. While current work has also examined the risks and benefits of interacting with these models, this investigation is user-subjective, needing an objective understanding of how LLMs operate and behave from a clinical point of view [2].

Human-AI Comparison: Replace or Augment? With the rise of AI-mediated care, there is an increased interest in comparing the agent of care (AI) with its traditional counterpart: the human. While some research demonstrates that AI chatbots can produce more empathetic and high-quality responses than human physicians [1], other studies have found that users prefer empathetic responses written by their human peers over AI [7]. This also aligns with previous work highlighting the gap between human and machine understanding of empathy in peer support sessions [17]. Although individuals in online peer support platforms rate their interactions high in perceived empathy, current state-of-the-art deep learning methods score these interactions low, primarily because they focus on sentence structure with a bias towards verbose, redundant statements over genuine emotional connection.

2.2 Scalable Mental Health — Designing Text-based Peer Support Platforms

Apart from AI-mediated care, one way to overcome challenges associated with the growing demand for mental health professionals is to provide training and mentoring to build the capacity of peers to provide support. By removing the need for one-on-one time with a trained professional, peer support platforms can provide therapeutic support, minimize wait times, reduce treatment costs, and mitigate stigma [11, 14]. Research indicates that helping others also positively impacts peers' mood and alleviates depressive symptoms [3].

Recently, text-based online platforms have become increasingly popular because certain aspects of text can help reduce feelings of social anxiety and inhibition [4, 7]. Nevertheless, compared to richer formats like video and audio, expressing empathy through text can be challenging, as it lacks tone, facial expressions, body language, and social presence—essential components for conveying empathy [4].

3 ESTABLISHING THE GROUNDWORK: RESEARCH AND DEVELOPMENTS ANTECEDENT TO CURRENT STUDY

Our recent study, accepted at CHI'24, conducted in collaboration with an online peer support service, aimed to investigate effective methods of peer training. We analyzed 116 text-based peer support sessions, focusing on factors that increase users' perceived empathy. The sessions were categorized into low and high empathy sessions and analyzed for user experience reflections, peer supporter strengths, and areas for improvement. Our results indicated that techniques from CBT training, such as active listening and reflective restatements, significantly increased users' perceived level of empathy. However, rigid adherence to training and overuse of techniques could hinder connection and result in low perceived empathy [17].

One key challenge with text-based peer support platforms is their design, as texting lacks the social cues present in face-to-face interactions [9]. To address this, we explored social presence communication techniques and evaluated their implementations in a 24-person study [4]. We designed two interactive typing indicators: *masked-typing* and *live-typing*. Our goal was to create two different indicators to represent when someone is typing.

For *masked-typing*, characters appear as they are typed, but instead of seeing the actual characters, the recipient sees a '#' symbol for each character. This allows the recipient to understand that the sender is actively typing and to gauge the speed and length of the message as it is composed or edited, but the actual content of the message remains hidden until it is sent. Conversely, the *live-typing* indicator shows the recipient the characters in real-time as they are being typed, including any edits or pauses, essentially replicating the experience of watching the sender's screen.

The library and code for each indicator are open-source and can be easily incorporated into existing web applications. The software is hosted at brownhci/live-typing[1]. We found that *live-typing* increased participants' collaboration through validation, active listening, communication, and connection—key concepts of effective peer support. Participants indicated that these indicators are highly suitable for mental health support.

Having covered the two main components — the role of humans and design — in improving access to support, the final stage of this thesis considers how technology can complement these efforts to further increase mental health accessibility.

4 RESEARCH DESIGN AND METHODS

So far, through peer training and platform design, my research has discovered that text-based peer support platforms enable users to connect and experience high empathy [4, 17]. However, a discrepancy has been observed where AI approaches do not align with user experiences on these platforms. This leads to pressing questions: considering that many people now use and review LLM chatbots for peer support and self-therapy online [8], there is a need for a clinical investigation into the quality of support provided by LLMs and how it differs from human counterparts. This investigation will help to understand the strengths and limitations of LLM-provided

[1] https://github.com/brownhci/live-typing

care and to evaluate if and how they can be integrated with human peers to enhance access to care.

In collaboration with a non-commercial company with extensive experience training human peers for CBT, the next stage of this dissertation examines how LLM-based peer supporters compare to human counterparts. To ensure a controlled environment, LLM and human peers are trained using the same techniques by the same team.

The initial phase of the study will involve an analysis of two distinct sets of therapeutic sessions ($n_1 = 27$; $n_2 = 27$), where n_1 represents public sessions conducted with a trained human peer counselor (dataset available at [2]) and n_2 represents the sessions that will be recreated using HELPERT, a detailed prompt run on the GPT-4 large language model (Figure 1). Sessions will be anonymized to reduce bias and then independently evaluated by five clinical psychologists specializing in CBT for quality assessment. The psychologists will complete their evaluations independently. Each psychologist will evaluate each session quantitatively using the Cognitive Therapy Rating Scale (CTRS), an observer-rated tool designed to assess a counselor's proficiency in administering CBT [19]. The scale (Table B) is composed of 11 items that are each evaluated on a 7-point Likert-type scale; this ranges from poor (0) to excellent (6). Therefore, session scores can vary from 0 to a maximum of 66. The sessions will also be accompanied by qualitative commentary by each psychologist.

Using thematic analysis on coded session transcripts and psychologists' qualitative opinions, we will understand the risks and challenges associated with LLMs and human peer supporters. The mixed-method evaluation process will indicate the areas that require further technological advancement and ethical considerations, especially when dealing with vulnerable populations. The next step is to develop newer benchmarks for evaluating and improving the interpretability of LLM outputs in mental health. This is imperative given that people – especially those who are vulnerable or in dire need of support – will invariably interact with these models [8].

5 CONTRIBUTIONS AND FUTURE WORK

The results of this proposal will inform future research in multiple ways. Firstly, note that this investigation is synthetic and does not interact with human participants. Instead, LLM sessions are artificially reconstructed using pre-existing, consented peer support sessions. Through our study, researchers who wish to conduct a study involving human subjects and compare AI with human peer supporters in a live setting will be aware of the potential risks. Simultaneously, our analysis will inform us how to implement AI technical measures to improve the interpretability of LLM outputs and mitigate the associated risks identified in the study. Lastly, we might be able to identify opportunities, if any, for AI to enhance, not substitute, human peer supporters' capabilities, leading to more effective support. Human-AI collaboration can make the therapeutic process safer without losing the authenticity of human interactions or the efficacy of an advanced artificial intelligence model [18].

[2] https://cheeseburgertherapy.org/

6 ACKNOWLEDGMENTS

My deepest gratitude to my Ph.D. advisor, Dr. Jeff Huang, and my mentors, Dr. Sean Ransom and Dr. Nicole Nugent, my collaborators at Cheeseburger Therapy and the Brown HCI Lab.

REFERENCES

[1] John W Ayers, Adam Poliak, Mark Dredze, Eric C Leas, Zechariah Zhu, Jessica B Kelley, Dennis J Faix, Aaron M Goodman, Christopher A Longhurst, Michael Hogarth, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine*, 2023.

[2] Charlotte Blease and John Torous. Chatgpt and mental healthcare: balancing benefits with risks of harms. *BMJ Ment Health*, 26(1), 2023.

[3] Bruce P Doré, Robert R Morris, Daisy A Burr, Rosalind W Picard, and Kevin N Ochsner. Helping others regulate emotion predicts increased regulation of one's own emotions and decreased symptoms of depression. *Personality and Social Psychology Bulletin*, 43(5):729–739, 2017.

[4] Zainab Iftikhar, Yumeng Ma, and Jeff Huang. "together but not together": Evaluating typing indicators for interaction-rich communication. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 2023.

[5] Harsh Kumar, Ilya Musabirov, Jiakai Shi, Adele Lauzon, Kwan Kiu Choy, Ofek Gross, Dana Kulzhabayeva, and Joseph Jay Williams. Exploring the design of prompts for applying gpt-3 based chatbots: A mental wellbeing case study on mechanical turk. arXiv preprint arXiv:2209.11344, 2022.

[6] Silja Martikainen, Laura Kohonen-Aho, Niina Seittenranta, Emilia Makkonen, Mari Falcon, Valtteri Wikström, and Katri Saarikivi. Text-based patient–doctor discourse online and patients' experiences of empathy. *Computer Supported Cooperative Work (CSCW)*, pages 1–25, 2023.

[7] Robert R Morris, Kareem Kouddous, Rohan Kshirsagar, and Stephen M Schueller. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. *Journal of medical Internet research*, 20(6):e10148, 2018.

[8] Jesse Noyes. Perceptions of ai in healthcare: What professionals and the public think. The Intake. Available at: https://www.tebra.com/theintake/medical-deep-dives/tips-and-trends/research-perceptions-of-ai-in-healthcare, 2023.

[9] Kathleen O'Leary, Arpita Bhattacharya, Sean A Munson, Jacob O Wobbrock, and Wanda Pratt. Design opportunities for mental health peer support technologies. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 1470–1484, 2017.

[10] World Health Organization. Psychologists working in mental health sector (per 100,000). 2024.

[11] Aisling Ann O'kane, Sun Young Park, Helena Mentis, Ann Blandford, and Yunan Chen. Turning to peers: integrating understanding of the self, the condition, and others' experiences in making sense of complex chronic conditions. *Computer Supported Cooperative Work (CSCW)*, 25:477–501, 2016.

[12] Sun Young Park, Pei-Yi Kuo, Andrea Barbarin, Elizabeth Kaziunas, Astrid Chow, Karandeep Singh, Lauren Wilcox, and Walter S Lasecki. Identifying challenges and opportunities in human-ai collaboration in healthcare. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing*, pages 506–510, 2019.

[13] Aditya kumar Purohit, Aditya Upadhyaya, and Adrian Holzer. Chatgpt in healthcare: Exploring ai chatbot for spontaneous word retrieval in aphasia. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, pages 1–5, 2023.

[14] Amnon Ribak, Michal Jacovi, and Vladimir Soroka. " ask before you search" peer support and community building with reachout. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, 2002.

[15] Elizabeth C Stade, Shannon Wiltsey Stirman, Lyle H Ungar, Cody L Boland, H Andrew Schwartz, David B Yaden, João Sedoc, Robert J DeRubeis, Robb Willer, and Johannes C Eichstaedt. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *npj Mental Health Research*, 3(1):12, 2024.

[16] Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the united states: results from the 2020 national survey on drug use and health. HHS Publication No. PEP21-07-01-003, 2021.

[17] Sara Syed, Zainab Iftikhar, Amy Wei Xiao, and Jeff Huang. Machine and human understanding of empathy in online peer support: A cognitive behavioral approach. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2024.

[18] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. From human-human collaboration to human-ai collaboration: Designing ai systems that can work together with people. In *Extended abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2020.

[19] JE Young and Aaron T Beck. Cognitive therapy scale. Unpublished manuscript, University of Pennsylvania, 1980.

[20] Renwen Zhang, Jordan Eschler, and Madhu Reddy. Online support groups for depression in china: Culturally shaped interactions and motivations. *Computer Supported Cooperative Work (CSCW)*, 27(3):327–354, 2018.

A PUBLIC AND REPRODUCED LLM SESSION

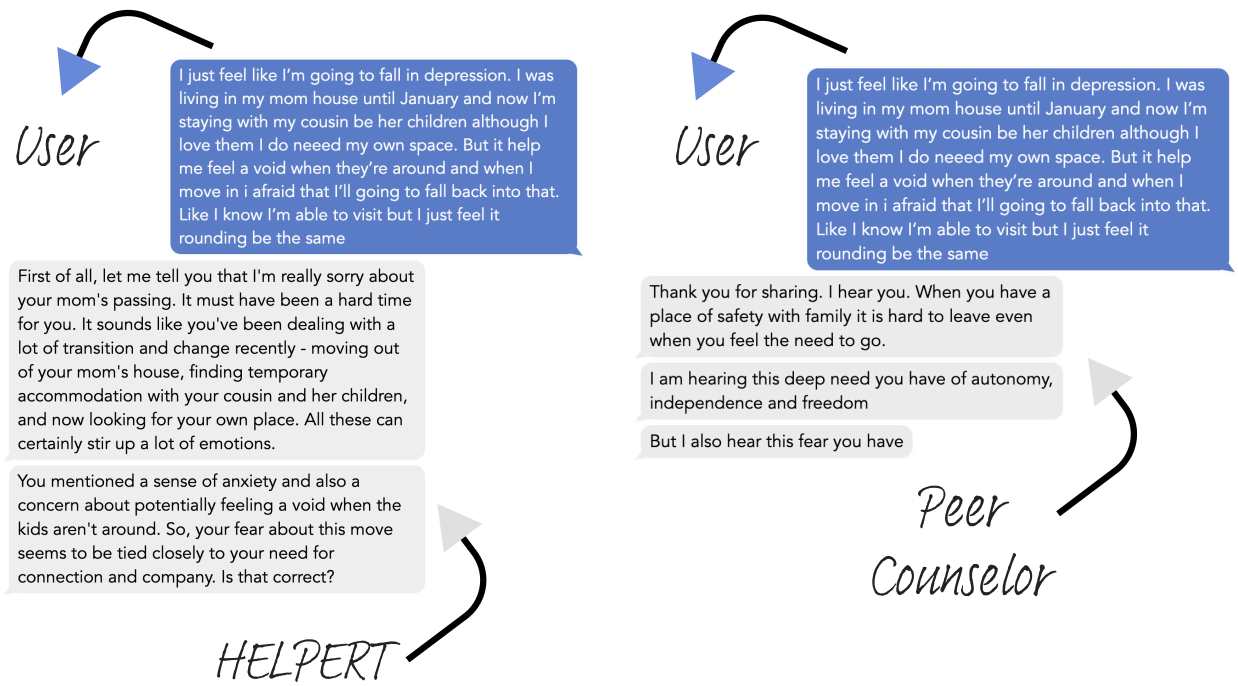

Figure 1: Example public session by a peer counselor (right) reproduced using CBT-prompted large language model-based system called HELPERT (left).

B COGNITIVE THERAPY RATING SCALE (CTRS)

Table 1: Cognitive Therapy Rating Scale (CTRS). Items on the CTRS are divided into two sub-groups: General Therapeutic Skills and Conceptualization, Strategy, and Technique

General Therapeutic Skills	
Agenda	Ability to collaboratively define an appropriate plan for the session, factoring time constraints and prioritization
Feedback	Extent to which the counselor applies patient feedback to ensure understanding and satisfaction within the session
Understanding	Capacity to comprehend both explicit and nonverbal communication, demonstrating listening skills and empathy
Interpersonal Effectiveness	Expression of positive interpersonal traits such as warmth, sincerity, confidence, and relationship quality
Collaboration	Efforts to foster a collaborative work relationship, active involvement in the process
Pacing & Efficient Use of Time	Ability to effectively manage and structure the session time to ensure progress
Conceptualization, Strategy, and Technique	
Guided Discovery	Skill in promoting self-discovery through measured questioning instead of persuasive tactics
Strategy for Change	Capacity to develop and follow a consistent strategy for change, using appropriate CBT techniques
Focusing on Key Cognitions or Behaviors	Ability to target essential thoughts or behaviors relevant to the seeker's problems
Application of CBT Techniques	Level of skill and resourcefulness exhibited in applying CBT
Homework	Ability to assign, explain, review and use tailored homework as an active element in the CBT process