CARE: Empowering Peer Counselors via Automatic Suggestion Generation

ANONYMOUS

Millions of users come to online peer counseling platforms to seek support on diverse topics ranging from relationship stress to anxiety. However, studies show that online peer support groups are not always effective as expected largely due to users' negative experiences with unhelpful counselors. Peer counselors are key to the success of online peer counseling platforms, but most of them often do not have systematic ways to receive guidelines or supervision. In this work, we introduce CARE: an interactive AI-based tool to empower peer counselors through automatic suggestion generation. During the practical training stage, CARE helps diagnose which specific counseling strategies are most suitable in the given context and provides tailored example responses as suggestions. Counselors can choose to select, modify, or ignore any suggestion before replying to the support seeker. Building upon the Motivational Interviewing framework, CARE utilizes large-scale counseling conversation data together with advanced natural language generation techniques to achieve these functionalities. We demonstrate the efficacy of CARE by performing both quantitative evaluations and qualitative user studies through simulated chats and semi-structured interviews. We also find that CARE especially helps novice counselors respond better in challenging situations.

CCS Concepts: • **Human-centered computing → Collaborative and social computing systems and tools**.

Additional Key Words and Phrases: dialogue generation, natural language processing, intervention, human-AI collaboration, counseling

ACM Reference Format:

Anonymous. 2022. CARE: Empowering Peer Counselors via Automatic Suggestion Generation. In . ACM, New York, NY, USA, 26 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Millions of people (nearly 1 in 5 American adults) are touched by mental illness [22]. The majority of people affected never access care [25], due to structural barriers including the high cost of treatment and the lack of trained professionals to meet the demand [40, 41, 53]. One promising remedy for under-treatment and to meet the demand for mental health care is the use of online peer counseling platforms, because of the benefits of anonymity, empowerment, and access [35, 50, 63]. Online peer counseling platforms have been proven effective in reducing people's self-reported anxiety and depression and improving their quality of life [18, 23]. Nevertheless, studies show that online support groups are not always effective as expected [34], largely due to support seekers' negative experiences with unhelpful supporters [12, 13, 80]. Supporters are the key to the success of online peer counseling platforms [1], and the quality of the support that seekers receive within a community heavily depends on those providing support. Thus, finding ways to help peer counselors be more helpful and supportive remains especially important.

Evidence from volunteer-based support services indicates that even brief training may improve a supporter's ability to help seekers [70] and may ultimately lead to improved mental health outcomes in seekers [4]. Thus, there is a growing trend in scaling effective peer support training online. For instance, 7 Cups [56] provides users with training on active listening techniques [8], and *SAHAR*, an Israeli-based suicide-prevention initiative, trains its online counselors

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

Manuscript submitted to ACM

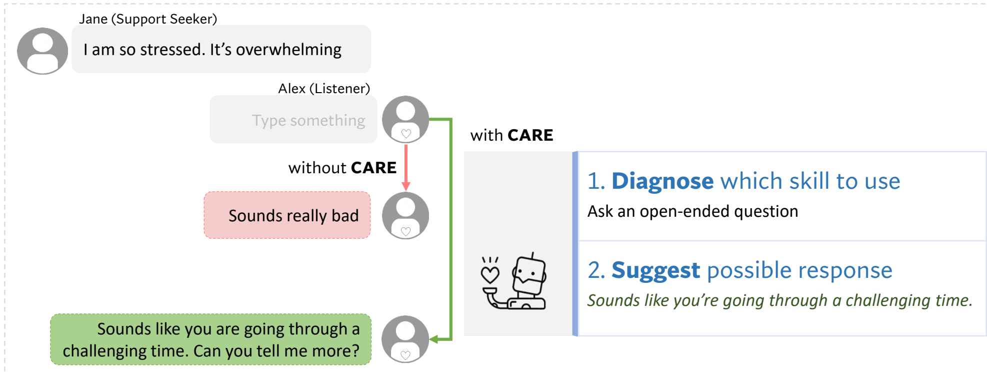


Fig. 1. Example interface of CARE to empower supporters by (1) diagnosing which skill to use and (2) suggesting responses.

extensively during weeks of in-person training to handle supportive chats [6]. However, supporters typically do not have the rigorous professional training of their offline counterparts [64], with very abbreviated training or no training at all for most online peer support groups [27, 85]. Thus, even if supporters want to help others in need they might lack the necessary set of skills or the knowledge to use evidence-based counseling strategies [83]. Moreover, unlike traditional counseling settings, online supporters often do not have systematic ways to receive supervision or guidance, especially during their conversations with seekers and when there are high levels of uncertainty about how to respond. Without appropriate guidance, supporters might develop biased or even inappropriate helping skills without being aware of it, based on their own experiences. This can lead to ineffective, or at worst, harmful, support which might prevent seekers from receiving other effective support options sooner [69, 81] and may also expose supporters themselves to self-doubts [83], increased stress and decreased self-efficacy [20, 71].

Existing mechanisms of training or scaffolding largely rely on human supervision, which requires an extensive amount of resources such as cost, time, labor, and expertise, making it hard to scale up to help a large number of supporters [5] who support millions of people in need of care in text-based online peer counseling platforms. Social computing research on online peer support groups mainly looked at how supporters' language use or counseling strategies are associated with better outcomes for seekers [2, 17, 57, 61, 62, 64]. In contrast, there have been relatively fewer studies on building scalable systems and tools to empower supporters to offer the most effective support in online peer counseling platforms, with a few exceptions [60, 76]. More importantly, the majority of research has specifically been about designing virtual counselors that emulate a human counselor in a chatbot-like setting [30, 31]. Different from previous work on chatbots around virtual counselors or client-clinician interaction [19], we propose to build an automatic *assistant* that can augment or empower supporters in terms of a wide range of counseling strategies.

Concretely, we propose to design an interactive training agent CARE to empower peer counselors[1] in text-based, online peer-to-peer counseling platforms. Here, we use 7 Cups as a research site and build contextualized language generation approaches to provide tailored assistance for supporters by highlighting which counseling strategies are needed in a given situation and suggesting example responses. Building upon psychotherapy theories and empirical studies on Motivational Interviewing (MI) [54, 74], we select a set of counseling strategies from MI to be our main focus

[1]In this work, we use *peer counselors* (or sometimes supporters, or support providers) to refer to the individuals providing help and *support seekers* (seekers) to refer to people seeking or receiving support.

for empowering volunteer counselors. We built machine learning classifiers on top of the large pretrained language model BERT [21] to predict the most suitable counseling strategies in a given context; we then develop language generation approaches via a large pretrained generation model DialoGPT [89] for generating example responses that counselors could further use and edit. As shown in Figure 1, we integrate different components together into an interactive interface for CARE to empower supporters at scale.

Furthermore, we use iterative, user-centered design principles in the development process of CARE by working closely with stakeholders from 7 Cups. To evaluate whether the proposed system CARE works, we perform both quantitative and qualitative evaluations, including system log analyses, questionnaires, and semi-structured interviews, to demonstrate the effectiveness of CARE. We also find that CARE especially helps beginner and novice counselors to better deal with these challenging situations. To sum up, we make the following contributions:

- Develop contextualized language generation techniques that provide tailored assistance for peer counselors by highlighting which counseling strategies are needed in a given situation and suggesting example responses.
- Create an interactive tool CARE to empower peer counselors in real-time on an online peer counseling platform.
- Evaluate CARE via both quantitative system log analyses and qualitative user studies to demonstrate the efficacy of CARE on a set of representative counseling scenarios.

2 RELATED WORK

2.1 Peer Support and Online Peer Counseling Platforms

Peer support has been shown effective in a wide range of mental health services [68], and is a major motivation for people who go online for mental health support [63]. Prior work has provided some evidence for the effectiveness of online peer support groups. For instance, a recent survey on 7 Cups indicated that users reported higher satisfaction with the support provided by 7 Cups counselors, and users who indicated receiving psychotherapy in the past marked the peer counselors' support to be as helpful as psychotherapy [7]. Most studies on online text-based peer counseling context have focused on the automatic modeling of behaviors from peer counselors and seekers and their correlations with successful outcomes [2, 42, 48, 87] or user engagement [3, 76]. For instance, [14] used n-grams to find the similarity between counselor and client speech to detect counselors' reflective listening skills automatically. Sharma and de Choudhury [77] classify messages into two categories of support: emotional and instructional, to find the relationship between linguistic accommodation and social support provision. Sharma et al. [75] develop a theoretically-grounded framework for empathy measurement on peer counseling platforms.

A growing amount of automatic text-based analytical work has also been conducted to identify behavioral codes derived from Motivational Interviewing (MI) Skill Codes [36, 52, 61, 62, 73, 82]. For instance, Perez-Rosas et al. [61] build a model to predict the MI strategies of therapists using clinical audio transcripts. Shah et al. [74] examine the relationship between the use of MI strategies and the satisfaction ratings from support seekers in *online* counseling sessions. Chen et al. [16] explore challenges faced and MI strategies employed by novice peer counselors via qualitative user studies and discuss design implications to better prepare for the skill development of novice therapists. Building upon these prior works, we focus on a subset of widely-used MI strategies and leverage these theory-driven counseling strategies to guide us in designing an interactive tool that can empower peer counselors.

2.2 Systems and Interventions for Peer Counseling

Interventions for support are generally valued by people with mental health concerns [57]. For example, people with schizophrenia showed a positive attitude toward using technology for care [26]. Supporters show a substantial interest in learning helping skills and using web-disseminated peer support interventions to help one another online [10]. Supportive chats guided by prompts in a design study were found to be associated with reduced anxiety, perceived as "deeper" with solutions to problems and new perspectives [58] compared to unguided ones, calling for natural language processing empowered systems to provide appropriate scaffolds at scale. This is in line with a recent study on the impact of technology in psychotherapy [37], which has identified the development of machine learning technologies for counselor's training and feedback as important needs where technology has a significant impact. To address this, our work will leverage advanced state-of-the-art natural language processing methods to provide peer counselors with contextualized training to help them acquire or improve their counseling skills.

There has been increased attention towards building automatic tools for online health communities [29, 60], however, the majority of research has specifically been about designing virtual counselors that emulate a human counselor in a chatbot-like setting. For instance, [30, 31] introduced a counseling dialogue system that interacts with users by recognizing what the users say and generating replies that utilize language templates to incorporate basic counseling techniques. Shen et al.[78] look at the generation of reflections using GPT-like models, and [72] focus on empathic rewriting. In one recent work, Sharma et al. [76] provide HAILEY, an AI-in-the-loop agent that provides just-in-time feedback to help participants who provide support respond more empathically to support seekers. Different from previous work on chatbots around virtual counselors or the facilitation of one single counseling strategy such as showing empathy, our work builds an automatic *assistant* that empowers counselors to help better train them with a wider range of theory-driven counseling strategies.

3 THE DESIGN OF CARE

Realistic practice and tailored feedback are key processes for training individuals with therapy skills [9, 43, 67]. As we discussed earlier, existing mechanisms of providing feedback largely depend on human supervision. It often takes well-trained human raters up to ten times as long as the duration of the session itself to finish labeling that session on the involved counseling strategies [5], let alone providing detailed feedback, making it difficult to scale up to help a large number of volunteer counselors who use online communities. Our work aims to build upon advances from natural language processing to automatically provide suggestions to assist volunteer counselors especially novices at scale [33].

Research Site. Our design of CARE will use 7 Cups as a case study. 7 Cups is an online peer support service, where support seekers with a variety of mental health problems participate in text-based chats with peer counselors who have completed active listening and other therapeutic training. As of February 2021, 7 Cups had over **200,000 trained peer counselors**, supporting over a million people a month.

Scope of CARE*.* Our work is mainly designed to be used in **a training or simulated environment** where counselors can practice their skills in a safe environment without harming real people, and where they can receive feedback or assistance. CARE allows for real-time human oversight and control, and we require counselors to review the suggested content before using it. Note that CARE will be mainly used by counselors who have already received very abbreviated default training on 7 Cups, **not** seekers or the vulnerable members who suffer from mental health issues; thus, there is no direct communication between seekers and CARE. In other words, counselors have full control and can decide whether and to what extent they would use the suggested content; counselors can easily turn on or turn off this feature.

Overview of CARE*.* CARE is an interactive tool that works in a private text-based session consisting of two speakers: a support seeker and a peer counselor. CARE aids users by recommending up to three possible responses to counselors in

Table 1. **Selected Motivational Interviewing strategies, their number of annotations from [74], and results of the most suitable counseling strategy prediction.** Using the human-annotated 7 Cups dataset from [74], to follow the distribution of counseling strategies on 7 Cups, we selected all 8 MI-consistent strategies from the 10 most frequent strategies on 7 Cups. Strategy, Description, and # Instances are distilled from Table 1 and 2 of [74]; Accuracy and F1 Score are the evaluation results of the Strategy Suggestion models trained on 7C-HQ, reported on 7C-MI in Section 3.2.1.

Strategy	Description	# Instances	Acc.	F1
Open Question	Open-ended questions that leave room for a response.	2507 (15.29%)	0.632	0.686
Closed Question	Questions with short specific answer.	1954 (11.91%)	0.612	0.672
Persuade with Permission	Counselor explicitly tries to change member's opinions, attitudes, or behavior based on logical arguments and facts. Counselor asks for permission first or emphasizes collaboration.	1918 (11.69%)	0.692	0.719
Reflection	Counselor captures the implicit meaning and feelings of client statements and returns it to the client through rephrases.	1697 (10.35%)	0.648	0.695
Support	Sympathetic, compassionate, or understanding comments encouraging client behavior.	1493 (9.10%)	0.672	0.705
Introduction/ Greeting	Counselor and seeker greet each other, exchange names etc.	1260 (7.68%)	0.867	0.874
Grounding	Counselor facilitates conversation through acknowledgements.	1027 (6.26%)	0.721	0.745
Affirm	Counselor compliments the seeker.	346 (2.11%)	0.722	0.747

a given situation. We choose to display 3 responses on the basis of the interface length and to allow for a good amount of flexibility for counselors. These suggestions are built upon psychotherapy literature on Motivational Interviewing [54, 74], which is a client-centered counseling style for eliciting positive behavior change by helping support seekers to explore and resolve ambivalence. We design the tool interface to be the same as the UI of 7 Cups [56] to simulate real-life user studies and obtain genuine user feedback. CARE comprises a backend system with suggestion-generation modules and a front-end user interface. They together form a full stack system. Note that our backend system is designed to be independent of the frontend client for generalizable deployment. The same backend system can be connected to other front-end clients in a plug-and-play fashion.

3.1 Motivational Interviewing Strategies

Motivational Interviewing (MI) Strategies are a set of client-centric techniques used by therapists for client betterment [54]. The underlying principle of MI techniques focuses on a therapeutic alliance between a patient and a therapist with an emphasis on patient autonomy. These techniques are suitable for online peer counseling platforms given their effectiveness in reducing drug and alcohol abuse [11, 32, 39, 49], decrease in smoking [28], and reduction of sexual risk behaviors [24, 44]. We used the corpus released by prior work on 7 Cups [74], which identifies and annotates 17 categories of counseling techniques on 14,797 utterances from 7 Cups chats (referred to as *7C-MI* hereafter). We selected a set of MI strategies based on a few criteria: (1) there are a relatively good amount of available annotations released by prior work so that we can build machine learning models with reasonable accuracy; and (2) these MI strategies are

Table 2. **Datasets used in this study.** We obtained permission and release from [74] for **7C-MI**. With the MI strategy classifiers from [74], we labeled **7C-HQ**. Finally, we ran the model selection experiments on **7C-HQ-small**, a 1-month subset. To evaluate CARE, we trained and fine-tuned models on **7C-HQ** and reported results by evaluating them on **7C-MI**.

Abbreviation	Description	MI Strategy Annotations	# Utterances	Usage
7C-MI	From [74]	Annotated by humans	14,797	Evaluate models
7C-HQ	Highly rated	Labeled by classifiers from [74]	20,445,517	Train models
7C-HQ-small	Highly rated; 1 month	No	344,335	Select models

widely used by prior work with a sufficient grounding in the MI literature. This led us to 8 strategies, as shown in 1. Future work can easily annotate other types of counseling strategies to expand this set.

3.1.1 Data Preparation. With written permission from [74], we assign MI strategy labels to our data with their released strategy classifiers. The purpose of assigning labels to our data instead of using their annotations directly is to obtain a greater number of MI strategy annotations on highly rated conversations to build our tool. For each strategy, we use the binary classifier of that strategy to assign labels to 20,445,517 utterances from 7 Cups conversations that receive a rating of 5 on a 5-point Likert scale from support seekers (*7C-HQ* hereafter). Lastly, we train and develop on 7C-HQ, setting aside 7C-MI as the ground-truth test set. We also created a 7C-HQ-small, a 1-month subset of high-quality conversations for response generation model selection (Section 3.2.2), which are explorative, computationally intensive experiments. The information about the datasets is presented in Table 2. We received a data-use agreement from 7 Cups for this research study and used dataset split train/dev/test = 8:1:1 throughout this study for model development.

3.2 Generating Suggestions

CARE suggests counseling strategies and potential responses for a conversation in real-time by using a multi-step process described as follows (Figure 2):

Step 1. For each of the 8 counseling strategies described in Table 1, CARE predicts the probability that which strategy should be used in the next peer counselor' response using the 5 latest utterances in the chat and a strategy-specific fine-tuned machine learning classifier.

Step 2. CARE generates a potential counselor response for each of the top three most probable counseling strategies. This generation is conditioned with the 5 latest utterances in the chat and uses a DialoGPT model fine-tuned with one of the 8 counseling strategies.

Step 3. CARE ensures another layer of support seeker safety by filtering all potentially inappropriate generations using a fine-tuned HateBERT [15].

Step 4. The de-duplicated responses are rendered on the front end in descending order of the probabilities from Step 1.

We now describe in detail the machine learning models involved in Step 1, Step 2, and Step 3.

3.2.1 Step 1: Predicting the Most Suitable Counseling Strategy. Prior research shows that large pretrained language models [21, 47] demonstrate state-of-the-art performances on a large set of language-based downstream classification tasks [84]. Thus, we use one of the most widely used pretrained language models Bidirectional Encoder Representations from Transformers (BERT) [21] to predict the most appropriate counseling strategy that a counselor should use in response to the utterance from the support seeker. Mathematically, we frame it as a multi-label classification task: given the current conversation between a support seeker and a peer counselor, and a set of possible counseling strategies,

Table 3. Comparison of different models conditioned on predicted strategies on 7C-HQ-small, and Results of strategy-conditioned next-utterance generation (trained on 7C-HQ, reported on 7C-MI). *Positive* indicates agreement of predictors by Shah et al. [74] on the strategy of the generated response. These BLEU scores are competitive to that of the related work [66, 75, 89].

Models	Total # Instances	Avg # Words	ROUGE-1	ROUGE-2	ROUGE-L	BERT Score	BLEU	Positive
BART	14797	7.710	0.128	**0.017**	**0.117**	0.845	0.074	NA
GPT-2	14797	9.728	0.111	0.013	0.100	**0.879**	**0.084**	NA
DialoGPT	14797	8.871	**0.132**	0.012	0.114	0.878	**0.084**	NA
Strategy	# Instances	Avg # Words	ROUGE-1	ROUGE-2	ROUGE-L	BERT Score	BLEU	Positive
Open Questions	1,853	7.991	0.164	0.034	0.155	0.876	0.188	0.930
Closed Questions	1,785	10.754	0.138	0.020	0.124	0.870	0.185	0.880
Persuade	1,883	13.809	0.111	0.010	0.093	0.861	0.177	0.680
Reflection	1,638	13.417	0.107	0.010	0.090	0.863	0.167	0.642
Support	1,364	11.042	0.172	0.037	0.156	0.872	0.167	0.809
Introduction	214	4.734	0.248	0.092	0.244	0.879	0.191	0.925
Grounding	960	1.984	0.143	0.035	0.143	0.883	0.098	0.948
Affirm	324	12.320	0.172	0.040	0.155	0.872	0.183	0.766

identify a set of strategies that are most suitable for the given context. Practically, we will first use our automated models of identifying MI strategies mentioned in Section 3.1, to identify (i.e., label) the strategies used in counselors' responses across all conversations. Then, we treat pairs of contexts and counselors' responses identified as utilizing a certain set of strategies as positive instances of those strategies and that *not* labeled with them as negative instances.

During the implementation stage, we transform this multi-label classification task into eight binary classification tasks, following the binary relevance paradigm for multi-label classification. We used the most recent five messages as the context for the current conversation, as suggested by prior work [74]. To deal with any class imbalance issue, we used downsampling by sampling the negative class size to be the same as the positive automatic annotations. The final classifiers are built by fine-tuning a pre-trained `BERT-base-uncased` checkpoint from Hugging Face [86] on 7C-HQ. The results are reported on 7C-MI, the human-annotated instances from [74] as the test set. We show the number of instances, model accuracy, and the F1 scores in Table 1. We found that our machine learning classifiers performed reasonably well, with an overall F1 score larger than 0.705. Given these reasonable predictions of the most appropriate counseling strategies, we then can move forward with the detailed feedback generation.

3.2.2 Step 2: Generating Example Responses. Supporters, especially beginners, might not be able to identify how to respond to a given situation [55], or they may not feel confident about providing an ideal response. To this end, we introduce the task of contextualized suggestion generation with humans in the loop. The goal here is not to replace counselors, but to **augment** them when they experience high uncertainty in conversation, as well as to these supporters in formulating their responses. We formulate this task as a conditional text generation problem: given an input (a seekers' post and the context) and a suitable strategy, we generate a statement that provides a sample response so that volunteer counselors can review and use—or edit and then use—in a given session. Similar to the setup in Section 3.1.1, we identified a set of pairs (post, response) from high-quality conversations filtered by satisfaction ratings as our corpus.

Auto-regressive models like GPT-2 [65] and BART [45] have shown great performance on generation tasks. Here, we choose a variant of GPT-2, DialoGPT [89] trained on multi-turn dialog conversations, as its domain shares lots of similarities with the counseling setting. Specifically, our model generates responses conditioned on (1) the counseling strategy, and (2) the most recent five utterances in the conversation. We do this by appending the predicted MI strategy

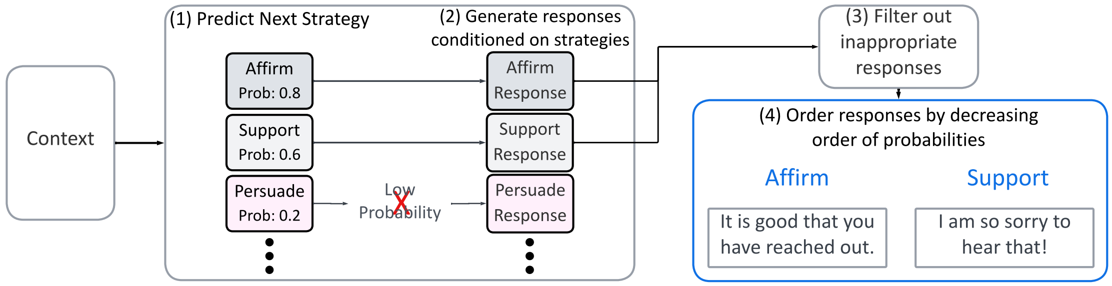


Fig. 2. CARE system architecture. CARE consists of 8 next-strategy predictors, one next-response generator, and one inappropriate response classifier. It predicts the probability distribution of the next counseling strategies and generates suggested responses for each confident strategy independently. In the end, the list of strategies and responses is filtered and ordered decreasingly by the confidence of the predictors.

token to the dialog history at both training and inference time. As shown in Tables 3, overall, DialoGPT-2[2] outperforms the GPT-2 and BART in terms of our automatic evaluation metrics [46, 59, 88]. We then use this best-performing model DialoGPT-2 to generate the example response when conditioning on a specific counseling strategy. The lower half of Table 3 shows the acceptable performances of our generation models, with an overall semantic similarity score between a generated text and the ground truth text larger than 0.87 (indicated by BERTScore). Furthermore, the column of *Positive*, which refers to whether the generated text exhibits the type of counseling strategy that it is supposed to have, demonstrates relatively good agreement with an overall score of 78%.

3.2.3 Step 3: Filtering Out Inappropriate Responses. As an extra layer of safety protection, CARE filters out inappropriate and undesired responses before presenting them to peer counselors. A response is defined as *inappropriate* if it consists of abusive behavior, the inquiry of unnecessary support seeker personal information/ personal identifiers, or contains unnecessary and excessive swear words. To avoid outputting inappropriate example responses, CARE employs the *inappropriate* classifier [74], which is a HateBERT model [15] finetuned on human annotations. The classifier is recall-heavy to ensure that CARE refrains from suggesting possibly inappropriate responses. It gives an F1 score of 84.21% on the human-annotated test set of 7C-MI.

3.3 System Development

Building upon the aforementioned steps, we design an interactive system CARE to empower volunteer counselors during their chats with support seekers. CARE integrates iteratively developed front-end and back-end components to optimize the user experience and runtime efficiency. Specifically, the system is hosted on an Amazon AWS server with an Ubuntu Operating system (32 Threads, 480GB RAM) having 8 Tesla k80 GPUs (12 GB RAM). These specifications enable CARE to interact seamlessly with humans. The back end is implemented with a Flask API that uses WebSockets to enable continuous bi-directional communication with the chat client. The functions provided by the back-end API fulfill the requirements of the system such as posting and pushing log-in credentials, posting chat logs, and pushing generations from the back-end models. Server-side event handlers are invoked through API function calls made by the front-end interface. All models (DialoGPT, BERT) are hosted on the same server to enable faster communication. The Server-side event handlers are flexible to connect with any type of front-end client. This allows CARE to be used with any 7 Cups-like

[2]Table 6 in Appendix provides a more detailed comparison in different settings for different models.

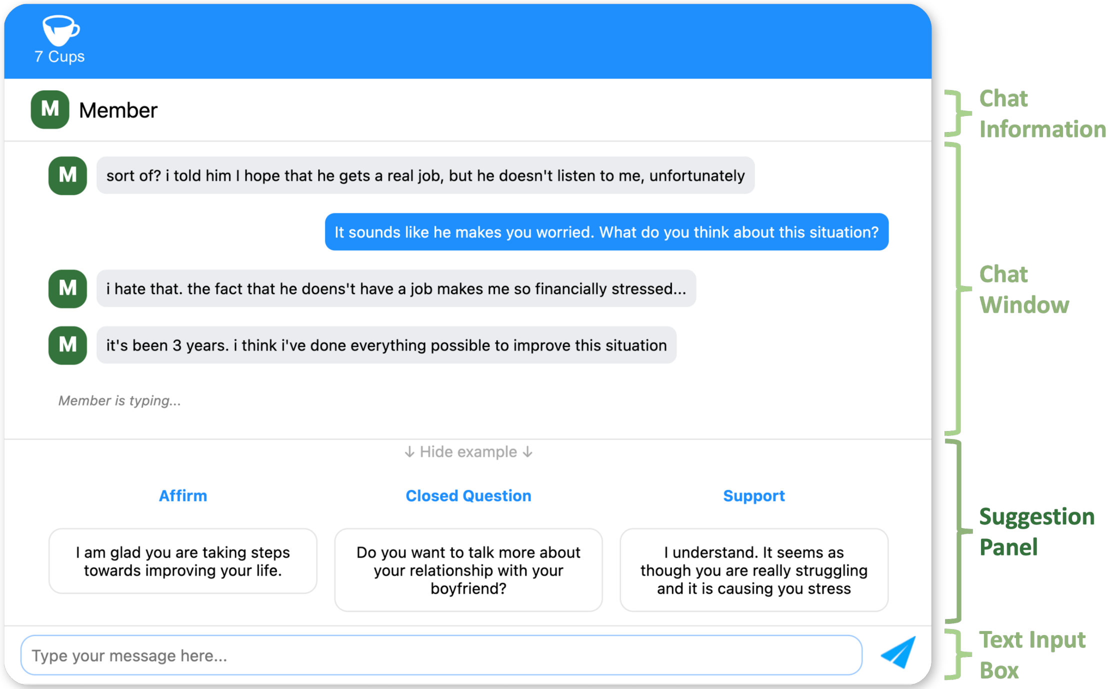


Fig. 3. The peer counselor side frontend view. We design CARE as a **Suggestion Panel**. The other components (Chat Information, Chat Window, and Text Input Box) mimic the frontend view of 7 Cups. CARE is comprised of three elements: 1) **The blue words**: Suggested MI counseling strategy to be used in the next peer counselor response. Users can hover over the strategy to see its 1-sentence description. 2) **Black sentences in boxes**: A generated next peer counselor response that utilized the MI strategy above it. Users can click the response to auto-fill the input text box for modification before sending. 3) **Hide/Show example button**: Toggle for collapsing and expanding Suggestion Panel.

online mental health platform and provides training to peer counselors in a platform-agnostic manner. A client-side API reference is used to facilitate the processing of function calls to the back-end API. This enables functionalities such as the initialization of new chat sessions, the addition of new messages to the chat log, and the clearing of sessions. With this architecture, the chat system has the ability to return suggestions and generations in real-time. We ensure that the client-side API is lightweight for quick loading of the system on the client end.

3.3.1 User-Centered Design of the System Front-End. The front end has two views, one for the support seeker and one for the peer counselor. The view for the peer counselor is given in Figure 3. The UI contains the four major components to replicate a chat environment like 7 Cups. These components are the chat information toolbar on the top, the chat window, the collapsible panel for suggestions and responses generated by CARE, and the text input box. The support seeker view of the platform is the same as the peer counselor view except for the absence of the collapsible panel. Feedback on the interface was obtained by working with domain users from the 7 Cups platform to simulate the look and feel of the platform. Key features added from the iterative development design through user feedback are (1) *a user is typing* indicator, and (2) the option to hide the tool window. When a peer counselor clicks on one of the potential

Table 4. **Participant Information**. **Tenure** denotes the length a participant has been a peer counselor on 7 Cups by the date they complete their user study session. These quantities are rounded to the closest integer in the largest non-zero unit. **CARE Chat** marks whether the participant see CARE's AI assistance in the first (1) or the second (2) chat in our user study session, which is randomly assigned. **Category** is the more familiar or comfortable category the participant chose from ***A****nxiety* and ***R****elationship Stress* for the mock chats. **Device(s) Used for 7 Cups** is collected because 7 Cups provides UIs for computers, tablets, and cellphones, but for consistency, we conducted our user studies with solely computers. **Most Recent Background** denotes the participant's most recent, external, related academic or professional background. Given the variety of responses and the sample size, we include all essential participant information here but do not analyze the relations between these variables and the participants' perceptions of CARE.

Participant ID	Tenure	CARE Chat	Category	Device(s) Used for 7 Cups	Most Recent Background
P1	3 months	2	R	Mostly computers	Degree in Sociology
P2	5 years	1	A	Always computers	No
P3	3 years	2	A	Mostly computers	Degree in Psychology
P4	7 years	2	R	Always computers	Degree in Counseling
P5	8 years	2	R	Half-and-half	Unknown
P6	4 years	1	A	Always cellphones/tablets	Student in Clinical Social Work
P7	2 years	1	R	Mostly computers	No
P8	4 years	1	R	Mostly computers	Student in Psychology
P9	8 years	2	A	Half-and-half	Crisis Counselor
P10	3 years	2	A	Half-and-half	No
P11	1 year	2	A	Mostly computers	Student in Psychology
P12	5 days	1	A	Always cellphones/tablets	No
P13	8 years	1	R	Always computers	Unknown
P14	1 month	2	R	Always cellphones/tablets	No
P15	3 months	1	R	Half-and-half	No

responses in the tool, the text input box gets populated with the response. The peer counselor can then choose to modify the response before sending it.

3.3.2 Logging of peer counselor actions. We log[3] peer counselor actions on the suggested responses to understand the impact of our tool in chats. For each utterance, we log the CARE suggestions, the peer counselor's click (choice) in the suggested responses, the peer counselor's click to show and hide the collapsible panel, and the final peer counselor response. We use these logs for our subsequent quantitative evaluation.

4 EVALUATION OF CARE VIA USER STUDIES

Evaluating the proposed system CARE requires rigorous examination and ethical consideration based on its sensitive and high-stakes cases. We use iterative, user-centered design principles in the development of the interfaces, by working closely with stakeholders from 7 Cups who best know the domain for interface development and by conducting user studies to test our interfaces so that our tool may be used widely in practice.

4.1 Recruitment of Participants

This study has been approved by the Institutional Review Board (IRB) at the first author's institution. By working with community moderators on 7 Cups, we recruit 15 peer counselors who are 7 Cups users located in the United States (11 females, 3 males, 1 unknown, age 18–69). All of the participants were recruited via notifications on the 7 Cups mobile/ web app and were compensated $20 for their time (the average time duration of the study was a little under

[3]During the user studies, we ask for written consent from the participants before doing so.

1 hour). Our recruitment criteria allowed only those participants who have completed at least 25 conversations on 7 Cups to ensure platform familiarity. All peer counselors in the study had a peer counselor rating of 4+ (out of 5). Lastly, we set an exclusion criterion for any peer counselor who may be previously blocked by a support seeker for researcher safety. This set of recruitment strategies led to 15 participant studies. The experience level of these recruited participants ranged from a tenure of 5 days to 8 years on the platform by the time of their interview. We aggregate their non-identifiable information in Table 4.

4.2 User Study Details

We test how a system like CARE empowers volunteer counselors by conducting user studies with two conditions: (1) no support, and (2) supported by CARE. We now describe in detail the different components involved in this user study.

4.2.1 Procedure and Method. For each user study session, a participant (peer counselor) first selects a chat category between *anxiety* and *relationship stress* based on their own familiarity and comfort. The participant then tests the system by interacting with a researcher who acts as a support seeker. Both of the systems mimic the chat interface of 7 Cups as described in Section 3.3.1. The order of the chats is randomly assigned using a computer randomizer: one with CARE, and the other without CARE.

4.2.2 Participants Onboarding with CARE*.* To get participants familiar with CARE, we briefly introduce the tool before they start playing around with the system. The participants are also asked to watch a tutorial video emphasizing the voluntariness of the use of CARE and explicitly teaching how to hide and show CARE's AI assistance. The video also describes the 8 counseling strategies along with their examples. We clarify all doubts and confusions a participant may have about the study. During the chat, when CARE shows AI assistance for the first time, we verbally instruct the participants again that they may show/hide the tool, ignore/modify generated responses, or use it as they prefer.

4.2.3 Conducting Simulated Chats. We select a larger set of representative seeker-supporter conversations from 7 Cups, remove any personal identifiers in these conversations, and manually paraphrase and fuse them, to obtain a smaller set of representative chats. In total, we devise four representative scenarios, two for each category and their corresponding scripts. In each simulated chat, the researchers act as a support seeker by sending messages modified from a pre-defined script. In detail, for our focused two categories: relationship stress and anxiety, the research team read roughly 30+ conversations from actual chats in 7 Cups. Researchers involved in the user study in our team have signed up as volunteer counselors and support seekers and conducted observations with more than 100 cumulative hours of counseling time on the platform. This enables us to be well-equipped in interacting with participants while conducting the studies. While the prepared scripts help maintain a coherent theme, we also allow for variability in the conversation due to the variability in the peer counselor's responses. The scenarios are summarized in Table 5.

4.3 Measure and Evaluation

We perform quantitative analyses on counselors' use of suggested counseling strategies when they interact with seekers, and qualitative analyses on the participants' feedback on what they liked and disliked about each session. To deeply understand the participant's perceptions of CARE, we design a post-task, anonymous questionnaire (Table 8 in Appendix), and also conduct a semi-structured interview. The questionnaire does not ask for any personal user identifiers to allow participants to freely disagree/ agree to the use of CARE and leave their opinions and comments. The questionnaire contains a mixture of Likert scales, checkboxes, and an open-ended response textbox, depending on the

Table 5. Summary of the mock chat scenarios.

Category	Order	Scenario Summary
Anxiety	1	A college student is worried about an upcoming admission exam and school-life balance.
	2	A person feels lonely and seeks to take their friendship to a deeper level.
Relationship Stress	1	A college freshman struggles with a newly long-distance relationship.
	2	A person feels financially stressed after their lover becomes unemployed for 3 years.

nature of the question. The participants complete the questionnaire during the session and may opt to unmute and ask questions whenever the questionnaire is unclear to them. As for the semi-structured interview, we prepare a list of questions and asked follow-up questions whenever we identify a potential theme.

4.4 Safety Protocol

We take careful steps to ensure the safety of our participants. First, CARE is designed for *training* and not for real-time uses; thus, supporters are involved in a simulated environment with little risk. We collect feedback and refine CARE in an interactive iterative process based on user insights to make sure it can work as expected. Second, at all times, the participants were free to leave the study (and will receive full payment) if they feel uncomfortable with the topics discussed in the chats. Third, our research team was involved in every support provision conversation transcript; if there are any signs of negative impacts (e.g., negative mood, frustration), our team will talk to participants immediately to discuss the issue and take appropriate actions such as referring them to the consulting clinician and sharing information about available national resources.

5 RESULTS

As described in Section 4.3, our participants fill out an anonymous questionnaire after they complete the first 30 utterances of each of the two system-testing chats. There are 30 system testing chats from the 15 user studies, which result in 926 utterances sent in those mock chats, where 470 utterances (50.76%) are sent by peer counselors, and the remaining 456 utterances (49.24%) are sent by us, the simulated help seekers. On top of the obtained numerical data from log analysis (Section 5.1) and Survey Responses (Section 5.2), we also conduct semi-structured interviews to understand the impact of CARE (Section 5.3). The quantitative and qualitative results of our study are given below.

5.1 Log Analysis

We analyze the system logs to understand whether CARE has effects on peer counselors and show that peer counselors **actually make use of** CARE when it is provided. We report the aggregated numbers at both, the utterance level and the chat level, to give a sense of the overall statistics and account for the difference between individual preferences.

5.1.1 CARE provides assistance 84% of the time. CARE's AI assistance is frequently provided to peer counselors during the experiment chats. CARE starts suggesting strategies and responses when a chat reaches at least 5 utterances. After that, CARE presents potential responses if the confidence score of the next strategy prediction is greater than a threshold of 0.5. Our analysis shows that across all utterances, CARE gives suggestions for 374 out of 467 utterances (80.09%) that were sent in chats, which is 84.38% of the time if we compute the median at the chat level.

5.1.2 *Peer counselors choose to see* CARE *93% of the time.*

Despite being told not necessarily to, instead of turning off CARE, most peer counselors almost always keep the CARE panel open during the mock chats. The peer counselors choose to turn on CARE for 340 utterances out of 374 utterances (90.91%) when CARE is available during the experimental chats, which takes up 93.20% of the time span on average across chats (std=23.376%, median=100%). Notably, 13/15 peer counselors keep the AI assistance panel shown throughout their experimental chat.

5.1.3 *Peer counselors check* CARE*'s suggestions before sending in around 47% of responses.*

We join system logs on clicks and messages together to analyze their usage of CARE. When it is the peer counselors' turn and they see CARE's suggestion, they click them before sending their response 75 out of 199 times, making the total click-through rate 37.69%. The median across the 15 experimental chats with CARE yields a 46.67% click-through rate.

5.1.4 *Peer counselors use a* CARE*'s response directly or edit it.*

We compare the text similarity between what CARE suggests and what peer counselors finally send out. From the utterances peer counselors sent after clicking responses suggested by CARE, we found that **most (45/75; 60%) of the responses were sent without modification from CARE's suggestion**, while the others were altered before being sent. For the 30/75 (40%) edited responses, we calculate the length of the Longest Common Subsequence (LCS) between the clicked suggested response and the actual utterance sent by the peer counselor. This calculation yields a median difference of 41.5 characters between the two strings. On one hand, if we divide the length of each LCS by the length of its corresponding generated response, we get the median 99.3% (mean=87.7%, std=19.6%). On the other hand, if each LCS length is instead divided by the length of the corresponding actual response peer counselors sent, the resulting median is 47.7% (mean=50.8%, std=20.6%). These two ratios together advise that **when peer counselors modify a CARE's suggestion, they usually build upon it by preserving almost the entire suggestion and adding more content to it.**

5.1.5 *Peer counselors send longer responses with* CARE.

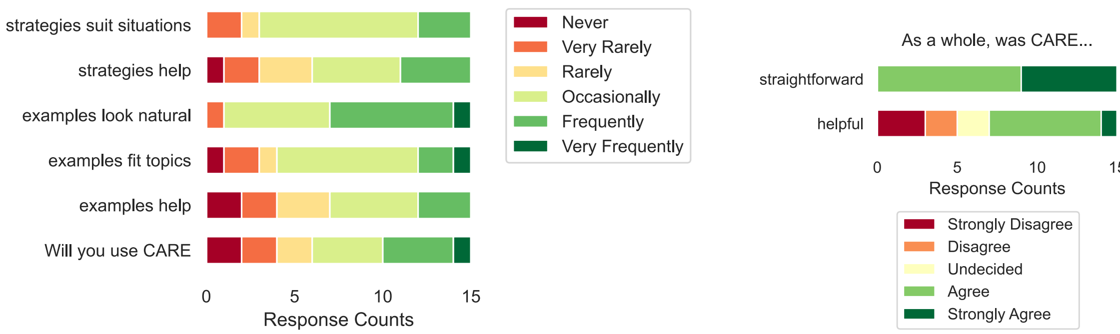
The median length of peer counselors' responses increases with CARE compared with its counterpart without CARE. Concretely, when peer counselors see CARE's suggestions, the median length of responses (77 characters) is significantly longer than the response length without CARE (61 characters) ($N = (199, 271)$, $p < 0.01$; Mann-Whitney U rank test)..

5.2 Survey Responses

This section summarizes the responses to the questionnaires in terms of Liker chart questions (Section 5.2.1) and check box questions (Section 5.2.2 and 5.2.3). Responses to the open-ended questions are rather free of form and highly relevant to the interview contents, so they are coded with interview transcripts in Section 5.3.

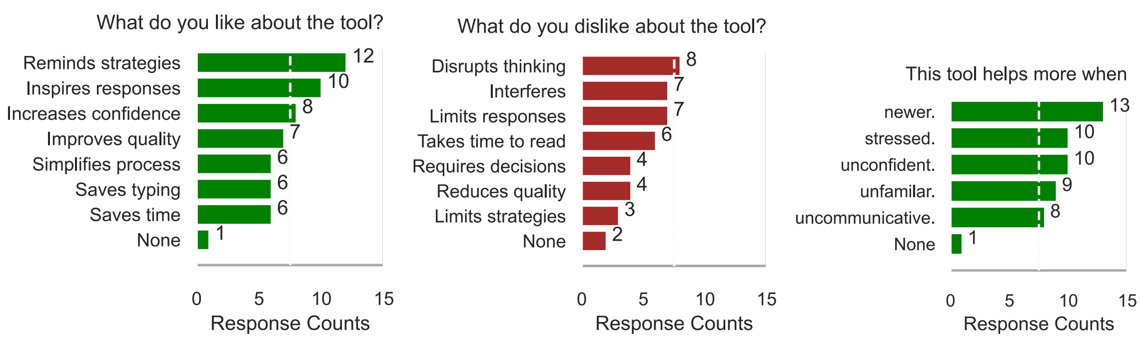
5.2.1 *Overall Perceptions.*

As shown in Figure 4a, the responses to the model perception questions show that for most participants, the strategy prediction and suggestion generation of CARE are leaning towards helpful in terms of whether the counseling strategies suit the situation, and whether the generated responses fit the conversation topic, and help peer counselors. The figure also shows that most participants feel that the model suggestions look natural for a 7 Cups-like peer counseling platform. As a whole, shown in Figure 4b, all participants agree that CARE is straightforward to use, while more than half of them think it is overall helpful. The median response to "Will you use CARE" (*If the tool is provided to you, how often do you think you will make use of it?* in the questionnaire) tilts towards positive. Lastly, responses to the question "Do the examples contain *harmful* message" are: Never (4/15), Very Rarely (6/15), Rarely (3/15), Occasionally (1/15), Frequently (1/15), and Very Frequently (0/15). This indicates that peer counselors generally think the example responses CARE suggested can hardly cause harm.



(a) Perceptions of Models and Future Usage (b) Perceptions of CARE

Fig. 4. The responses to perception questions. a) A bar chart of the participants' answers to Likert scale questions regarding their perception of the suggested counseling strategies (*strategies*), example responses (*examples*), and future usage of CARE. b) A bar chart of participants' overall perception of CARE. The dashed lines mark the medians of responses.



(a) Like Aspects (b) Dislike Aspects (c) Helpful Occasions

Fig. 5. The responses to multi-select checkbox questions regarding preferences. The dashed lines mark the medians of responses.

5.2.2 Like & Dislike. To deeply understand what features or aspects of CARE are more appreciated by peer counselors, we visualized their preference responses in Figure 5. We find that more than half of the participants agree that CARE reminds them of counseling strategies (12/15), inspires them to better responses (10/15), and increases their confidence (8/15). While about half of the participants think CARE can disrupt the thought process during conversations, we believe that the seamlessness of the tool increases with greater familiarity and use.

5.2.3 When Is CARE *Helpful?* As shown in Figure 5c, for the question of when CARE is helpful, we saw from the top three selections to our question "CARE helps more when...?" that, 13/15 said YES to *the peer counselor is new to listening chats*, 10/15 mentioned *the peer counselor feels more stressed*, and 10/15 emphasized when *the peer counselor is less confident in the chat's category*. This result confirms the effectiveness of CARE in helping especially novice peer counselors.

5.3 In-depth Interview Results

To provide an in-depth understanding of our findings from the log analyses and questionnaires, this section takes a deep dive into what participants shared with us during their semi-structured interviews.

5.3.1 *How peer counselors use* CARE.

Preference for and against showing CARE. Participant P7 attributed why they keep CARE open throughout the simulated chat to potential inspirations. This might explain the result about the major user preference of showing CARE more often than hiding it in Section 5.1.2.

> *"I think I would probably **have it on more often than not**, I'm just in case it suggested something that I hadn't that hadn't popped into my mind. " – P7*

Alternatively, P6 and P14 prefer to use CARE as a safety net and refer to CARE when there are uncertainties.

> *"Only if I'm really on, like, lost on what to say, I don't have any ideas or strategies to go, I would just go there [*CARE*'s suggestions]. Other than that, I prefer to go with my train of thought without looking at it." – P14.*

Importance of Optionality. Although most (13/15) participants chose to show CARE instead of hiding it, giving peer counselors an option to hide the tool is still essential. Some peer counselors (P3, P4, P6, P10) pointed out the need for this option to switch the tool on/off.

> *"You can switch it on and off as needed. If you pick up a chat with a subject you're not as familiar with, if you decide [that] **you want to put it on, you have that option**. If you don't want to, you don't have to." – P4*

Adopting a CARE *Response.* Section 5.1.3 shows that peer counselors click CARE's suggestion before sending in around 47% of responses. P7 demystifies this behavior by illustrating their usage of CARE:

> *"So you know, I might take a quick glance at it [CARE] and see if it's the same thing I was going to type anyway. And then **click on it** [the suggestion], that's definitely a **time saver**." – P7*

In terms of intentions behind the modification, P8 mentioned their rationale for doing so is because *"There were some open-ended questions that the tool suggested that I didn't think of that I thought is really good, and **sometimes I use exactly what was suggested**."*. At other times, P8 modifies the suggestion for **personalization**: *"The **personalization** part, I still think that the listener would have to be able to write the words out themselves."* Thus, when P8 wants to make the response conform to their style more, they *"fix up the wording to make it a bit warmer and less cold"* once in a while.

5.3.2 *Peer counselors think* CARE *is straightforward and helpful.*

Participants' interview comments on the straightforwardness and helpfulness of CARE are consistent with the survey response results in Section 5.2.1.

Straightforwardness of CARE. When being asked to comment on the UI/UX of CARE, P10 compliments on the UI design as they *"like how the suggestions are right above the chat bar."*. P12 describes the intuitive nature of CARE as *"Even if you didn't know about it [how to use the tool] before, you can [directly] use this technology here."*.

Helpfulness of CARE. Many participants (P1, P6, P7, P8, P9, P15) also agree that CARE is helpful and also mention the utility of the strategies. Others show satisfaction in the quality of CARE's suggested responses.

> *"I think it could be a **great tool**. I really do. I think it's something that we need in 7 Cups because we don't really train listeners very much. We it's very minimal training. So we kind of grab listeners once problems are occurring. So if we could put this at the front end, that would be phenomenal." – P1*

Risks of CARE. While CARE is perceived rarely harmful (Section 5.2.1), peer counselors who see CARE's mistakes think CARE may not actually cause negative consequences:

> *"I did notice just a couple of suggested responses that didn't quite fit. But um, I mean, I think the average listener would know not to click on them.* ***I don't think that they would just randomly start clicking.****" – P7*

5.3.3 *Interpretation of Likes and Dislikes.*

This part extends Section 5.2.2 from what to why to gauge the reason behind peer counselors' preference for and against CARE.

Reminds peer counselors of counseling strategies. The top questionnaire answer for what peer counselors like about CARE finds that it *reminds peer counselors of counseling strategies* (Section 5.2.2). 80% of peer counselors like the fact that CARE reminds them of MI strategies, indicating that the tool's ability to suggest relevant MI strategies is a helpful feature that supplements the short training these counselors receive. Peer counselors P6 and P12 mentioned in their interview how CARE helped them with learning MI strategies.

> *"The tool is using what the member [support seeker] is saying to come up with these prompts, and it's giving me open question options, it's giving me the different active listening skills that you're supposed to use when you're listening. And people often forget that as listeners. They get stuck on asking questions all the time. Or just validating all the time or just reflecting all the time. And I think this gives me the confidence* ***to explore the different skills*** *that you should have as a listener." – P6*

> *"I think there were a* ***lot of good [strategy] suggestions.*** *And it helped remind me of different coping mechanisms that were appropriate to like, mention in the use with the member [support seeker], for example, the whole challenging negative thought thing. The suggestions, you know, reminded me of that, because with affirmation and as being one tool, for example, it really helped me." – P12*

Inspire better responses. In Section 5.2.2, we bring out that 10/15 of the surveyed participants reported that they liked how CARE *inspires better responses.* CARE's dual-functioning of providing the MI strategy and potential response of what to say is a helpful feature. Even if peer counselors know which MI strategy they want to use, their training and experience may not be extensive enough to adequately apply the strategy to the situation. This is where CARE is able to support counselors by providing examples of using MI in the specific scenarios at hand. Even if counselors do not directly use the suggested utterance, it gives insight into what an alternative/better response might be. Peer counselor P6 mentioned multiple times how CARE **helped write better responses, even in situations where the generated responses are not a perfect fit**, they inspire the choice of response and the listener can then modify the suggestions, as P6 mentioned: *"[it] inspires me to use different strategies and* ***gives me ideas*** *and then I take that and modify it."*

Increases peer counselors' confidence. More than half of the participants mentioned that using CARE increases their confidence (8/15). During the interviews, many peer counselors (P1, P3, P6, P7, P8, P9, P10, P11) support that the tool would be greatly beneficial to novice counselors by empowering them in new situations and would help build confidence. Peer counselor P3, who often coaches peer counselors to be better, also mentioned how it helps in new situations. Peer counselor P10 stated that CARE would help provide better service to the support seekers.

> *"I liked it honestly, I do listener coaching and mentoring. And I liked that this really prompted I think a nice variety of responses depending on how comfortable the person is and really diving more deeply into whatever the issue that the member [support seeker] has. So I thought that as I was going through, I thought, this is a* ***great tool for listeners who are newer or are not feeling comfortable with the topic.****" – P3*

> *"I feel like it [deploying CARE on 7 Cups as part of the training program] would make them [new peer counselors] more knowledgeable. It would make it easier for them to **provide better service** in a way for the members [support seekers]." – P10*

Disrupts thinking. Some participants (8/15) pointed out that the addition of an interactive tool may distract a peer counselor from the actual conversation. The same peer counselor, P6, who said that the tool gave ideas, also thinks that it leads to an unconscious comparison between the peer counselor's response and the suggested response, which can lead to distractions—*"I am always **comparing what it was saying to what I was going to say**. And just like having another thing to see."* However, a few participants (P3, P8) disagree. This is supported by prior research that shows similar user feedback on the introduction of any new AI assistance tools: users feel that the tool disrupts thinking initially; over time, they feel that the tool is indispensable as they get used to it [38, 79]. P3 attributes this to friction during the process of getting used to AI assistance.

> *"Well, you know, the thing is, it's interesting that you say that because at first, I was thinking, Oh, this is going to be distracting. But then when I didn't have it I was like well, **I have to work a little harder because I have to come up with everything on my own.**" – P3*

5.3.4 CARE *helps both new and experienced peer counselors.*

CARE *helps inexperienced peer counselors* ***learn and grow****.* As most peer counselors (13/15), experienced and inexperienced, believe that CARE helps more *the peer counselor is newer to listening chats* as it *"gives them suggestions about what to say or how to respond"* (P1). *Aiding new peer counselors* is also a common theme mentioned in interviews and open-ended questionnaire responses.

> *"New people haven't really dealt with the situation before. And they're not listening and they don't really know exactly what to say yet. [CARE] can be quite a **guide**." – P2*

A couple of peer counselors, especially those who are involved in the new peer counselor training and mentoring efforts, think CARE will be **useful in training new peer counselors**. Some compare CARE with the current training modules and reach-out resources on 7 Cups, saying that **learning through real-time experience with CARE is more natural and practical than current static training modules**:

> *"With this tool, I think it'll be really helpful because this helps like when they are taking the chat instead of like help before or after that because this is **real-time help**. So it will be a lot more helpful for the new listeners to get familiar with the process." – P8*

> *"As most listeners, you know, they don't devote the time [on training modules]. Maybe not most, but some of them at least don't. So this way, you'll just, it's much **better to pick up through experience**." – P12*

Others even make recommendations on how to incorporate CARE into the peer counselor training process. P3 mentions how suggested responses give new peer counselors a sense of the expected language and the degree of formality on 7 Cups, making it a helpful tool for new peer counselors to refer to during the practice chats before they are verified. Still others, especially those who join 7 Cups newly, think other than helping them to learn, hypothetically, the fact that CARE is available on 7 Cups will itself increase their willingness to contribute more to chats. This can potentially help mitigate the high dropout rate or low engagement of peer counselors.

> *"It makes the whole process likely easier. It could only be helpful. So I guess it would **make me feel more open to doing more chats than I usually do**." – P15*

Peer counselors also foresee that new peer counselors' dependence on CARE can decrease over time since they gradually pick up what and how to use different strategies and respond with CARE, which meet our goal of helping new peer counselors to grow. One listener depicts their personal experience with CARE during the mock chats.

> *"When I started listening, it was helpful because when **when you first get into the listening, it's kind of hard to figure out immediately what you're going to say. So I did read the suggestion that was there**. But once I'm more comfortable, I was more comfortable listening and I kind of got the idea of how to be empathetic and how to ask open or closed questions or just sympathize." – P14*

CARE *helps experienced peer counselors **in difficult situations**.* In terms of the CARE's effect on experienced peer counselors, other than the aforementioned responses in the questionnaire, peer counselors think of other occasions where CARE can help, such as dealing with culture/ language barriers, unfamiliar topics, or context switches when being in multiple chats. By suggesting suitable responses, CARE can help mitigate the language barrier or age gap between a peer counselor and a help seeker.

> *"It [CARE] could be really helpful. So I know for some people on 7 Cups, they don't always have the same native language. Sometimes there are **language differences** or, like slang words that aren't always understood." – P1*

> *"Sometimes you have trouble **connecting with people who are not your age** or even closer I want to feel like a, like a colleague and not a mother. You know, so maybe the tool would help in that in that instance." – P4*

Peer counselors often specialize in a set of topics, yet they often have to converse on topics beyond that. A few participants describe difficulties in new topics which CARE helps mitigate: *"[listeners often] feel stuck in a chat, which happens quite often"* (P9). Specifically, P4 encountered this situation during the simulated chat and reflects that:

> *"You know what my specialty is grief. Relationship stress, [the category of the simulated chat], is not something that I'm as familiar with. I mean, I am but I'm better on certain topics. So I think **it would help me you know, with the category being as not as familiar** so it would help me like that." – P4*

Many peer counselors manage multiple chats concurrently, P3 talks about how CARE can help in such context switches of peer counselors by hinting according to their experience from peer counselor coaching.

> *"I have some people that I coach who take multiple chats and have a hard time managing them. **This [CARE] would allow them I think to manage them more properly** because they would have some kind of preset options to choose from and not have to come up with everything originally." – P3*

To sum up, via our system log analyses, questionnaires, and semi-structured interviews, we found that:

- **General Perception**: Overall, peer counselors think CARE is helpful and often use the tool when provided. The tool also leads to longer peer counselor responses on average. Peer counselors also agree that the tool is easy to use and these suggestions are natural. Analysis of system logs and surveys both showed that users choose to use CARE more often than not when given an option.
- **Situations where CARE helps**: Peer counselors think that CARE inspires better responses and reminds them of relevant counseling strategies. This helps novice counselors gain experience and confidence. Additionally, experienced peer counselors agree that the agent will help them with difficult and new topics. Semi-structured interviews also revealed that CARE holds the possibility to help bridge language, age, and cultural barriers.
- **Potential concerns towards CARE**: Some peer counselors report it to be sometimes distracting, therefore, including an option to show or hide CARE is important for better training.

6 CONCLUSION

This work develops an interactive training agent CARE for empowering volunteer counselors on online peer-to-peer counseling platforms For any given chat between a support seeker and a peer counselor, CARE automatically suggests counseling strategies and provides tailored responses based on the conversation context to assist counselors. We built upon the motivational interviewing framework and focus on a set of representative counseling strategies. We then develop contextualized classifiers to highlight which counseling strategies are needed in a given situation and scalable language generation approaches to generate example responses on the fly. We integrate different components together into an interactive CARE to enable a realistic user experience and adopt a user-centered design principle by working with domain users. Through both quantitative analysis and qualitative user studies, we demonstrate the efficacy of CARE on a set of representative counseling scenarios.

6.1 Implications and Design Recommendation

Theoretically, this work creates a novel synthesis of psychotherapy theories and advances in natural language processing to build an interactive assistant CARE that empowers peer counselors on an online peer counseling platform. Our contributions consist of operationalizing theory-driven motivational interviewing strategies, developing contextualized natural language models, designing an interactive HCI system, analyzing observational data from system logs and questionnaire results, as well as performing semi-structured interviews—all of these demonstrate a user-centered example paradigm on how to build both theory-driven and model-driven systems to augment human capabilities via AI.

One straightforward design recommendation is to use CARE to complement the current training programs on 7 Cups and many other similar online peer counseling platforms. As we discussed earlier, existing mechanisms of training or scaffolding largely rely on human supervision, which is costly and difficult to scale. Volunteer counselors often do not have the rigorous professional training as their offline counterparts and most online counseling platforms only provide very abbreviated training. By evaluating it with 15 stakeholders both quantitatively and qualitatively, we demonstrate the possibility of CARE as an initial and effective effort to improve these training mechanisms.

Our work sheds light on how CARE may be integrated into platforms such as 7 Cups to better support peer counselors. It has great potential to be deployed for *real time scaffolding* that supporters can turn to during their interactions with seekers, to obtain an actionable suggestion on which skill best fits in a given scenario, and on how to formulate what they are going to say, towards the long-term goal of increasing the number of people they can help and the effectiveness of their interactions in addressing the mental health needs of support seekers.

We design CARE to be primarily used as a training or simulated environment for support providers so that they can practice their skills in a safe environment, without harming real people. While the user studies and findings are grounded in the context of 7 Cups as the primary platform, findings on the impact of such an AI tool, design recommendations for the deployment of the tool, and the general perception of the tool are broadly applicable to other online health communities with similar setups of private support seeker peer counselor chat rooms. In other words, this system is generic and agnostic to any online peer counseling platforms such as MellowTalk [51], thus it can be used in many similar environments, not only specific to 7 Cups.

From a technical perspective, CARE consists of two main functions: highlighting which counseling strategies are most suitable in a given context, and suggesting example responses in real-time. Each function can be used as an individual sub-tool to help specific use cases in online support groups. For instance, the resulting machine learning classifiers in detecting the most suitable strategies can also be used to understand the relationship between different strategies uses and their influence on conversation success or the well-being of support seekers. This work also provides insights and raises questions for the field of natural language processing and machine learning on how to generate responsible and contextualized responses in high-skate scenarios.

6.2 Limitations and Future Work

This work is subject to a few limitations. First, given the sensitive nature of this study on online peer counseling platforms, any use of AI models in providing assistance might introduce concerns about user safety and the inherent biases of these AI models. While we have taken very careful steps such as implementing an inappropriate content filer as shown in Figure 2 and maintaining safety protocols as described in Section 4.4, there might still be cases that are potentially problematic and undesirable. Future work can help mitigate this concern by developing better algorithms for debiasing and content filtering as well as introducing more real-time human oversight and control.

Second, we design CARE as a training tool primarily for one-to-one conversations and evaluate it with relatively short conversations. However, some online support provisions might happen in multi-user chat forums like Reddit or public chat groups. Furthermore, conversations between support seekers and peer counselors can be very long in the real world; such long-term dependencies may not be captured well in the current setup when providing assistance. How to generalize CARE to multi-party and long conversations requires further investigation.

Third, although we utilized the current state-of-the-art models for recommending counseling strategies and example responses, some generation performances are still unsatisfactory as pointed out by participants. Such limited quality of generations may negatively affect the training or suggestion provision to peer counselors. Future work can further improve these models by performing extensive parameter searches and optimizing model architectures.

Fourth, as an initial effort to empower peer counselors, our work mainly focuses on the development and design of CARE, and the evaluation of it with domain users via interviews and system logs in a simulated one-time chat environment. Our work has not evaluated any *long-term* effects of CARE, such as its influence on certain skill acquisition of peer counselors or on the mental health outcome changes in support seekers. We plan to continue improving CARE and evaluate its long-term impact by working with more domain users in a longitudinal way.

Next, in terms of multiple design choices such as counseling strategies, we mainly focus on a subset of motivational interviewing strategies, which provide a solid foundation for CARE. However, there are other types of MI strategies or other prominent counseling techniques such as cognitive behavioral therapy which could be strong alternatives for empowering peer counselors. Similarly, for multiple thresholding choices such as the number of responses to show on the CARE interface, we chose them based on the researchers' expertise in design and feedback from stakeholders. We plan to iteratively refine them and perform more robustness checks for the next version of CARE.

Last but not least, our user studies involve only 15 peer counselors, who are located in the United States and who use English in their counseling sessions. During our user studies, we only asked peer counselors to work on two of the most popular topics: relationship stress and anxiety. However, in practice, topics discussed on online peer counseling platforms cover a wide variety of situations. Some examples of popular situations and topics are relationship stress, anxiety, depression, health, LGBTQ, dissociative identity, home life, self-improvement, etc. Future work can explore the efficacy of our system across more topics for better generalization. Extending this work to more participants with diverse peer counselor backgrounds, language use, locations, and expertise can bring in more nuanced findings.

7 ETHICAL CONSIDERATIONS

The Institutional Review Board has been approved for this study at the researchers' institution. The data for the study was collected in partnership with 7 Cups, following Health Insurance Portability and Accountability Act (HIPAA) and confidentiality data use agreements. To protect the privacy of the participants, the data has been anonymized and additional steps have been taken to ensure that it cannot be linked to any specific user. At all times, actual user data was present in secure servers, which could be only accessed by the researchers of this study. All the researchers involved in this study have completed CITI Program certifications on responsible code of conduct in research. All the participants in the user study are older than 18 and have signed a consent form stating their explicit consent. The researchers acknowledge and address the potential safety issues for support seekers and peer counselors when using such a system for training in the Safety Protocol in Section 4.4 and the Limitations in Section 6.2.

REFERENCES

[1] Kathina Ali, Louise Farrer, Amelia Gulliver, and Kathleen M Griffiths. 2015. Online peer-to-peer support for young people with mental health problems: a systematic review. *JMIR mental health* 2, 2 (2015), e4418.

[2] Tim Althoff, Kevin Clark, and Jure Leskovec. 2016. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics* 4 (2016), 463–476.

[3] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2018. Social support, reciprocity, and anonymity in responses to sexual abuse disclosures on social media. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 5 (2018), 1–35.

[4] Joe Armstrong. 2010. How effective are minimally trained/experienced volunteer mental health counsellors? Evaluation of CORE outcome data. *Counselling and Psychotherapy Research* 10, 1 (2010), 22–31.

[5] David C Atkins, Mark Steyvers, Zac E Imel, and Padhraic Smyth. 2014. Scaling up the evaluation of psychotherapy: evaluating motivational interviewing fidelity via statistical text classification. *Implementation Science* 9, 1 (2014), 1–11.

[6] Azy Barak. 2007. Emotional support and suicide prevention through the Internet: A field project report. *Computers in Human Behavior* 23, 2 (2007), 971–984.

[7] Amit Baumel. 2015. Online emotional support delivered by trained volunteers: users' satisfaction and their perception of the service compared to psychotherapy. *Journal of mental health* 24, 5 (2015), 313–320.

[8] Amit Baumel and Stephen M Schueller. 2016. Adjusting an available online peer support platform in a program to supplement the treatment of perinatal depression and anxiety. *JMIR mental health* 3, 1 (2016), e5335.

[9] James Bennett-Levy. 2006. Therapist skills: A cognitive model of their acquisition and refinement. *Behavioural and Cognitive Psychotherapy* 34, 1 (2006), 57–78.

[10] Samantha L Bernecker, Kaitlin Banschback, Gennarina D Santorelli, and Michael J Constantino. 2017. A web-disseminated self-help and peer support program could fill gaps in mental health care: Lessons from a consumer survey. *JMIR mental health* 4, 1 (2017), e4751.

[11] Brian L Burke, Hal Arkowitz, and Marisa Menchola. 2003. The efficacy of motivational interviewing: a meta-analysis of controlled clinical trials. *J Consult Clin Psychol* 71, 5 (Oct. 2003), 843–861.

[12] Moira Burke and Robert Kraut. 2008. Mind your Ps and Qs: the impact of politeness and rudeness in online communities. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. 281–284.

[13] Moira Burke, Robert Kraut, and Elisabeth Joyce. 2010. Membership claims and requests: Conversation-level newcomer socialization strategies in online groups. *Small group research* 41, 1 (2010), 4–40.

[14] Doğan Can, Panayiotis G Georgiou, David C Atkins, and Shrikanth S Narayanan. 2012. A case study: Detecting counselor reflections in psychotherapy for addictions using linguistic features. In *Thirteenth Annual Conference of the International Speech Communication Association*.

[15] Tommaso Caselli, Valerio Basile, Jelena Mitrović, and Michael Granitzer. 2021. HateBERT: Retraining BERT for Abusive Language Detection in English. In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*. 17–25.

[16] Tianying Chen, Kristy Zhang, Robert E. Kraut, and Laura Dabbish. 2021. Scaffolding the Online Peer-Support Experience: Novice Supporters' Strategies and Challenges. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 366 (oct 2021), 30 pages. https://doi.org/10.1145/3479510

[17] Prerna Chikersal, Danielle Belgrave, Gavin Doherty, Angel Enrique, Jorge E Palacios, Derek Richards, and Anja Thieme. 2020. Understanding client support strategies to improve clinical outcomes in an online mental health intervention. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–16.

[18] Greg Clarke, Donna Eubanks, Ed Reid, Chris Kelleher, Elizabeth O'connor, Lynn L DeBar, Frances Lynch, Sonia Nunley, and Christina Gullion. 2005. Overcoming Depression on the Internet (ODIN)(2): a randomized trial of a self-help depression skills program with reminders. *Journal of medical Internet research* 7, 2 (2005), e16.

[19] David Coyle, Conor Linehan, Karen Tang, and Sian Lindley. 2012. Interaction design and emotional wellbeing. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*. 2775–2778.

[20] Joe Curran, Glenys D Parry, Gillian E Hardy, Jennifer Darling, Ann-Marie Mason, and Eleni Chambers. 2019. How does therapy harm? A model of adverse process using task analysis in the meta-synthesis of service users' experience. *Frontiers in psychology* 10 (2019), 347.

[21] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR* abs/1810.04805 (2018). arXiv:1810.04805 http://arxiv.org/abs/1810.04805

[22] Joseph L Dieleman, Ranju Baral, Maxwell Birger, Anthony L Bui, Anne Bulchis, Abigail Chapin, Hannah Hamavid, Cody Horst, Elizabeth K Johnson, and Jonathan Joseph. 2016. US spending on personal health care and public health, 1996-2013. *Jama* 316, 24 (2016), 2627–2646.

[23] Gunther Eysenbach, John Powell, Marina Englesakis, Carlos Rizo, and Anita Stern. 2004. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *Bmj* 328, 7449 (2004), 1166.

[24] Jeffrey D Fisher, William A Fisher, Deborah H Cornman, Rivet K Amico, Angela Bryan, and Gerald H Friedland. 2006. Clinician-delivered intervention during routine clinical care reduces unprotected sexual behavior among HIV-infected patients. *J Acquir Immune Defic Syndr* 41, 1 (Jan. 2006), 44–52.

[25] Center for Behavioral Health Statistics and Quality. 2015. Behavioral health trends in the United States: results from the 2014 National Survey on Drug Use and Health. *HHS Publication No. SMA 15-4927, NSDUH Series H-50* (2015).

[26] Katrina Gay, John Torous, Adam Joseph, Anand Pandya, and Ken Duckworth. 2016. Digital technology use among individuals with schizophrenia: results of an online survey. *JMIR mental health* 3, 2 (2016), e5379.

[27] Madelyn S Gould, Wendi Cross, Anthony R Pisani, Jimmie Lou Munfakh, and Marjorie Kleinman. 2013. Impact of applied suicide intervention skills training on the national suicide prevention lifeline. *Suicide and Life-Threatening Behavior* 43, 6 (2013), 676–691.

[28] Emily Gray, Jim McCambridge, and John Strang. 2005. The effectiveness of motivational interviewing delivered by youth workers in reducing drinking, cigarette and cannabis smoking among young people: quasi-experimental pilot study. *Alcohol Alcohol* 40, 6 (Aug. 2005), 535–539.

[29] Stephanie Greer, Danielle E. Ramo, Yin-Juei Chang, Michael Fu, Judith Tedlie Moskowitz, and Jana Haritatos. 2019. Use of the Chatbot "Vivibot" to Deliver Positive Psychology Skills and Promote Well-Being Among Young People After Cancer Treatment: Randomized Controlled Feasibility Trial. *JMIR mHealth and uHealth* 7 (2019).

[30] Sangdo Han, Jeesoo Bang, Seonghan Ryu, and Gary Geunbae Lee. 2015. Exploiting knowledge base to generate responses for natural language dialog listening agents. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 129–133.

[31] Sangdo Han, Kyusong Lee, Donghyeon Lee, and Gary Geunbae Lee. 2013. Counseling dialog system with 5w1h extraction. In *Proceedings of the SIGDIAL 2013 Conference*. 349–353.

[32] Jennifer Hettema, Julie Steele, and William R Miller. 2005. Motivational interviewing. *Annu Rev Clin Psychol* 1 (2005), 91–111.

[33] Clara E Hill, Jessica Stahl, and Melissa Roffman. 2007. Training novice psychotherapists: Helping skills and beyond. *Psychotherapy: Theory, Research, Practice, Training* 44, 4 (2007), 364.

[34] Brenda E Hogan, Wolfgang Linden, and Bahman Najarian. 2002. Social support interventions: Do they work? *Clinical psychology review* 22, 3 (2002), 381–440.

[35] Thomas K Houston, Lisa A Cooper, and Daniel E Ford. 2002. Internet support groups for depression: a 1-year prospective cohort study. *American Journal of Psychiatry* 159, 12 (2002), 2062–2068.

[36] Xiaolei Huang, Lixing Liu, Kate Carey, Joshua Woolley, Stefan Scherer, and Brian Borsari. 2018. Modeling temporality of human intentions by domain adaptation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 696–701.

[37] Zac E Imel, Derek D Caperton, Michael Tanana, and David C Atkins. 2017. Technology-enhanced human interaction in psychotherapy. *Journal of counseling psychology* 64, 4 (2017), 385.

[38] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T. Hancock, and Mor Naaman. 2019. AI-Mediated Communication: How the Perception That Profile Text Was Written by AI Affects Trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300469

[39] Chad D Jensen, Christopher C Cushing, Brandon S Aylward, James T Craig, Danielle M Sorell, and Ric G Steele. 2011. Effectiveness of motivational interviewing interventions for adolescent substance use behavior change: a meta-analytic review. *J Consult Clin Psychol* 79, 4 (Aug. 2011), 433–440.

[40] Viktoria Kantor, Matthias Knefel, and Brigitte Lueger-Schuster. 2017. Perceived barriers and facilitators of mental health service utilization in adult trauma survivors: A systematic review. *Clinical Psychology Review* 52 (2017), 52–68.

[41] Alan E Kazdin and Stacey L Blase. 2011. Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on psychological science* 6, 1 (2011), 21–37.

[42] Florian E Klonek, Vicenç Quera, and Simone Kauffeld. 2015. Coding interactions in Motivational Interviewing with computer-software: What are the advantages for process researchers? *Computers in Human Behavior* 44 (2015), 284–292.

[43] Avraham N Kluger and Angelo DeNisi. 1996. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin* 119, 2 (1996), 254.

[44] Lisette Kuyper, John de Wit, Titia Heijman, Han Fennema, Jan van Bergen, and Ine Vanwesenbeeck. 2009. Influencing risk behavior of sexually transmitted infection clinic visitors: efficacy of a new methodology of motivational preventive counseling. *AIDS Patient Care STDS* 23, 6 (June 2009), 423–431.

[45] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 7871–7880.

[46] Chin-Yew Lin and Eduard Hovy. 2003. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*. 150–157.

[47] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR* abs/1907.11692 (2019). arXiv:1907.11692 http://arxiv.org/abs/1907.11692

[48] Sarah Peregrine Lord, Elisa Sheng, Zac E Imel, John Baer, and David C Atkins. 2015. More than reflections: empathy in motivational interviewing includes language style synchrony between therapist and client. *Behavior therapy* 46, 3 (2015), 296–303.

[49] Brad Lundahl, Chelsea Kunz, Cynthia Brownell, Derrik Tollefson, and Brian Burke. 2010. A Meta-Analysis of Motivational Interviewing: Twenty-Five Years of Empirical Studies. *Research on Social Work Practice - RES SOCIAL WORK PRAC* 20 (03 2010), 137–160. https://doi.org/10.1177/1049731509347850

[50] Belinda Melling and Terry Houguet-Pincham. 2011. Online peer support for individuals with depression: A summary of current research and future considerations. *Psychiatric rehabilitation journal* 34, 3 (2011), 252.

[51] MellowTalk. 2021. *MellowTalk*. https://mellowtalk.com/

[52] William R Miller. 1996. Motivational interviewing: research, practice, and puzzles. *Addictive behaviors* 21, 6 (1996), 835–842.

[53] Ramin Mojtabai, Mark Olfson, Nancy A Sampson, Robert Jin, Benjamin Druss, Philip S Wang, Kenneth B Wells, Harold A Pincus, and Ronald C Kessler. 2011. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. *Psychological medicine* 41, 8 (2011), 1751–1761.

[54] T.B. Moyers, J.K. Manuel, and D. Ernst. 2014. *Motivational Interviewing Treatment Integrity Coding Manual 4.1. Unpublished manual.* University of New Mexico Center on Alcoholism, Substance Abuse, and Addictions (CASAA).

[55] John A Naslund, Rahul Shidhaye, and Vikram Patel. 2019. Digital technology for building capacity of non-specialist health workers for task-sharing and scaling up mental health care globally. *Harvard review of psychiatry* 27, 3 (2019), 181.

[56] 7 Cups of Tea. 2021. *7cups*. https://www.7cups.com/

[57] Kathleen O'Leary, Arpita Bhattacharya, Sean A Munson, Jacob O Wobbrock, and Wanda Pratt. 2017. Design opportunities for mental health peer support technologies. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1470–1484.

[58] Kathleen O'Leary, Stephen M Schueller, Jacob O Wobbrock, and Wanda Pratt. 2018. "Suddenly, we got to become therapists for each other" Designing Peer Support Chats for Mental Health. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[59] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 311–318.

[60] Zhenhui Peng, Qingyu Guo, Ka Wing Tsang, and Xiaojuan Ma. 2020. Exploring the Effects of Technological Writing Assistance for Support Providers in Online Mental Health Community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[61] Verónica Pérez-Rosas, Rada Mihalcea, Kenneth Resnicow, Satinder Singh, Lawrence An, Kathy J Goggin, and Delwyn Catley. 2017. Predicting counselor behaviors in motivational interviewing encounters. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. 1128–1137.

[62] Verónica Pérez-Rosas, Xinyi Wu, Kenneth Resnicow, and Rada Mihalcea. 2019. What makes a good counselor? learning to distinguish between high-quality and low-quality counseling conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 926–935.

[63] John Powell, Aileen Clarke, et al. 2007. Investigating internet use by mental health service users: interview study. In *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems*. IOS Press, 1112.

[64] Yada Pruksachatkun, Sachin R Pendse, and Amit Sharma. 2019. Moments of change: Analyzing peer-based cognitive support in online mental health forums. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[65] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.

[66] Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 5370–5381.

[67] Ray Reagans, Linda Argote, and Daria Brooks. 2005. Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management science* 51, 6 (2005), 869–881.

[68] Julie Repper and Tim Carter. 2011. A review of the literature on peer support in mental health services. *Journal of mental health* 20, 4 (2011), 392–411.

[69] Julie S Reynolds and Nancy A Perrin. 2004. Mismatches in social support and psychosocial adjustment to breast cancer. *Health Psychology* 23, 4 (2004), 425.

[70] Philip L Rodgers. 2010. *Review of the applied suicide intervention skills training program (ASIST): rationale, evaluation results, and directions for future research*. LivingWorks Education Incorporated Calgary, Alberta, Canada.

[71] Alexander Rozental. 2016. *Negative effects of Internet-based cognitive behavior therapy: Monitoring and reporting deterioration and adverse and unwanted events*. Ph. D. Dissertation. Department of Psychology, Stockholm University.

[72] Tulika Saha, Vaibhav Gakhreja, Anindya Sundar Das, Souhitya Chakraborty, and Sriparna Saha. 2022. Towards Motivational and Empathetic Response Generation in Online Mental Health Support. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2650–2656.

[73] Craig S Schwalbe, Hans Y Oh, and Allen Zweben. 2014. Sustaining motivational interviewing: A meta-analysis of training studies. *Addiction* 109, 8 (2014), 1287–1294.

[74] Raj Sanjay Shah, Faye Holt, Shirley Anugrah Hayati, Aastha Agarwal, Yi-Chia Wang, Robert E Kraut, and Diyi Yang. 2022. Modeling Motivational Interviewing Strategies On An Online Peer-to-Peer Counseling Platform. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022),

1–24.
[75] Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. 2021. Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach. In *Proceedings of the Web Conference 2021*. 194–205.
[76] Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. 2022. Human-AI Collaboration Enables More Empathic Conversations in Text-based Peer-to-Peer Mental Health Support. *arXiv preprint arXiv:2203.15144* (2022).
[77] Eva Sharma and Munmun De Choudhury. 2018. *Mental Health Support and Its Relationship to Linguistic Accommodation in Online Communities*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174215
[78] Siqi Shen, Charles Welch, Rada Mihalcea, and Verónica Pérez-Rosas. 2020. Counseling-style reflection generation using generative pretrained transformers with augmented context. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 10–20.
[79] Swearingen Sinha, Kirsten Medhurst, and Rashmi Sinha. 2001. Beyond Algorithms: An HCI Perspective on Recommender Systems. (09 2001).
[80] Janet Smithson, Siobhan Sharkey, Elaine Hewis, Ray Jones, Tobit Emmens, Tamsin Ford, and Christabel Owens. 2011. Problem presentation and responses on an online forum for young people who self-harm. *Discourse studies* 13, 4 (2011), 487–501.
[81] Lene E Søvold, John A Naslund, Antonis A Kousoulis, Shekhar Saxena, M Walid Qoronfleh, Christoffel Grobler, and Lars Münter. 2021. Prioritizing the Mental Health and Well-Being of Healthcare Workers: An Urgent Global Public Health Priority. *Frontiers in public health* 9 (2021).
[82] Michael Tanana, Kevin A Hallgren, Zac E Imel, David C Atkins, and Vivek Srikumar. 2016. A comparison of natural language processing methods for automated coding of motivational interviewing. *Journal of substance abuse treatment* 65 (2016), 43–50.
[83] Anne Thériault, Nicola Gazzola, and Brian Richardson. 2009. Feelings of incompetence in novice therapists: Consequences, coping, and correctives. *Canadian Journal of Counselling and Psychotherapy* 43, 2 (2009).
[84] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2018. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. In *BlackboxNLP@EMNLP*.
[85] Yi-Chia Wang, Robert E Kraut, and John M Levine. 2015. Eliciting and receiving online support: using computer-aided content analysis to examine the dynamics of online social support. *Journal of medical Internet research* 17, 4 (2015), e99.
[86] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. HuggingFace's Transformers: State-of-the-art Natural Language Processing. arXiv:1910.03771 [cs.CL]
[87] Bo Xiao, Daniel Bone, Maarten Van Segbroeck, Zac E Imel, David C Atkins, Panayiotis G Georgiou, and Shrikanth S Narayanan. 2014. Modeling therapist empathy through prosody in drug addiction counseling. In *Fifteenth annual conference of the international speech communication association*.
[88] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. BERTScore: Evaluating Text Generation with BERT. In *International Conference on Learning Representations*.
[89] Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and William B Dolan. 2020. DIALOGPT: Large-Scale Generative Pre-training for Conversational Response Generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 270–278.

A APPENDIX

Table 6. Comparison of different models on the two conditioning scenarios on 7C-HQ-small.

Conditioned on MI Strategies	Model	ROUGE-1	ROUGE-2	ROUGE-L	BERT Score (F)	BLEU	Avg Length
No	BART	0.118	0.015	0.111	0.841	0.061	7.269
No	GPT-2	0.119	**0.018**	0.110	0.839	0.058	7.496
No	DialoGPT	0.119	0.015	0.111	0.841	0.060	7.127
Yes	BART	0.128	0.017	**0.117**	0.845	0.074	7.710
Yes	GPT-2	0.111	0.013	0.100	**0.879**	**0.084**	9.728
Yes	DialoGPT	**0.132**	0.012	0.114	0.878	**0.084**	8.871

Table 7. Results of the next utterance strategy predictions (trained on 7C-HQ, reported on 7C-MI).

Strategy	# Instances	Accuracy	Recall	Precision	F1 Score
Open Question (QUO)	3,586	0.632	0.802	0.599	0.686
Closed Question (QUC)	3,714	0.612	0.795	0.582	0.672
Persuade (PR)	3,778	0.692	0.788	0.661	0.719
Reflection (RF)	3,31	0.648	0.804	0.613	0.695
Support (SUP)	2,742	0.672	0.783	0.640	0.705
Introduction/Greeting (INT)	430	0.867	0.916	0.835	0.874
Grounding (GR)	1,926	0.721	0.817	0.685	0.745
Affirm (AF)	666	0.722	0.820	0.686	0.747
Overall	20,154	0.664	0.801	0.631	0.705

Table 8. Questionnaire presented to study participants. The participants completed this questionnaire during the user studies. In addition to the questions in this table, screenshots of individual UI components are also included in the questionnaire as a visual aid to help particiapnts identify which component of CARE a question refers to.

Questions	Question type
How often did the suggested strategies (blue hints) **suit the situation?**	Likert scale with six options Never, Very Rarely, Rarely, Occasionally, Frequently, Very Frequenty
How often did the suggested strategies (blue hints) **help?**	Likert scale with six options Never, Very Rarely, Rarely, Occasionally, Frequently, Very Frequenty
How often did the example responses (white buttons) **look natural on 7 Cups?**	Likert scale with six options Never, Very Rarely, Rarely, Occasionally, Frequently, Very Frequenty
How often did the example responses (white buttons) **fit the conversation topic?**	Likert scale with six options Never, Very Rarely, Rarely, Occasionally, Frequently, Very Frequenty
How often did the example responses (white buttons) **help?**	Likert scale with six options Never, Very Rarely, Rarely, Occasionally, Frequently, Very Frequenty
How often did the example responses (white buttons) **contain *harmful* message?**	Likert scale with six options Never, Very Rarely, Rarely, Occasionally, Frequently, Very Frequenty
As a whole, was the tool **straightforward to use?**	Likert scale with five options Strongly Disagree, Disagree, Undecided, Agree, Strongly Agree
As a whole, was the tool **helpful?**	Likert scale with five options Strongly Disagree, Disagree, Undecided, Agree, Strongly Agree
What do you like about the tool?	Check box with the following options Saves time Saves typing Simplifies decision process Increases listeners' confidence Reminds listeners of counseling strategies Inspires better responses Improves the overall quality of your responses None of the above Other (also has a text box)
What do you dislike about the tool?	Check box with the following options Takes more time to read examples Requires more decisions Disrupts thinking Limits the variety of counseling strategies Limits the diversity of responses Reduces the overall quality of your responses Interferes in listener-member communication None of the above Other (also has a text box)
This tool helps more when	Check box with the following options the Listener is newer to Listening chats. the Listener is less confident in the chat's category. the Listener is less familiar with the Member's chief complaint. the Member is less communicative. the Listener feels more stressed. None of the above Other (also has a text box)
If the tool is provided to you, how often do you think you will make use of it?	Likert scale with six options Never, Very Rarely, Rarely, Occasionally, Frequently, Very Frequenty
Any other comments?	Open-ended text box