

## Original Paper

# Development and Preliminary Evaluation of a Conversational Agent Delivering Problem-Solving Therapy for Family Caregivers of Children With a Chronic Health Condition: Multiphase Mixed Methods Study

Weichao Yuwen<sup>1</sup>, PhD, RN; Liying Wang<sup>2</sup>, PhD; Serena Jinchen Xie<sup>3</sup>, PhD; Myra Divina<sup>3</sup>, MS; Xuehong Fan<sup>3</sup>; William Kearns<sup>1</sup>, PhD; Teresa M Ward<sup>3</sup>, PhD, RN

<sup>1</sup>School of Nursing & Healthcare Leadership, University of Washington Tacoma, Tacoma, WA, United States

<sup>2</sup>Florida State University, Tallahassee, FL, United States

<sup>3</sup>University of Washington, Seattle, WA, United States

**Corresponding Author:**

Weichao Yuwen, PhD, RN  
School of Nursing & Healthcare Leadership  
University of Washington Tacoma  
1900 Commerce Street, Campus Box 358421  
Tacoma, WA 98402  
United States  
Email: [wyuwen@uw.edu](mailto:wyuwen@uw.edu)

## Abstract

**Background:** Family caregivers of children with chronic health conditions experience substantial physical and mental health burdens, including burnout, anxiety, depression, fatigue, and sleep disturbances. Despite this need, validated digital mental health tools tailored to family caregivers remain limited. AI-powered conversational agents offer a promising approach for delivering on-demand, personalized mental health support, yet development and evaluation frameworks for this population are lacking.

**Objective:** This paper describes the iterative development and formative evaluation of COCO (Caring of Caregivers Online), a conversational agent designed for family caregivers of children with chronic health conditions. COCO integrates problem-solving therapy (PST) and motivational interviewing (MI) within a human-in-the-loop development framework that progressed from rule-based interactions to a large language model (LLM)-powered conversational agent.

**Methods:** COCO was developed across four phases: (1) caregiver persona and dialogue development based on PST and MI; (2) usability testing of a low-fidelity prototype with standardized patients in a single session of PST; (3) usability testing of a high-fidelity prototype with caregivers in a single session of PST (n=38); (4) integration of an LLM into COCO. The Wizard-of-Oz method was used across phases 2 and 3 to collect naturalistic dialogues and refine COCO's conversational design. In phase 3, usability of COCO was assessed using the System Usability Scale (SUS). Caregiver emotions were measured before and after the session using 6 subscales of the PANAS-X. In phase 4, GPT-4 was integrated into COCO with few-shot learning and evaluated by research team members using the caregiver personas. Descriptive statistics were used to summarize quantitative measures. The MI principles and techniques used by COCO across the 4 phases were coded using the *Motivational Interviewing Treatment Integrity Coding Manual*.

**Results:** In phase 1, 4 gold-standard dialogues were developed using caregiver personas. In phase 2, standardized patients described COCO as validating and identified its problem-solving and on-demand support as helpful for caregivers. In phase 3, COCO-Wizard-of-Oz achieved a mean SUS score of 75.6% (SD 12.9%), reflecting acceptable usability. Participants demonstrated significant improvement in negative affect, sadness, guilt, and fatigue following PST sessions ( $P<.05$ ). In phase 4, an LLM-powered COCO was developed and demonstrated promising initial conversational capabilities. Across all phases, conversational quality showed progressively improved, with LLM-powered COCO achieving the highest density of MI techniques per turn (2.56) and greater balance across MI strategy types, particularly in seeking collaboration and reflection.

**Conclusions:** COCO demonstrated feasibility and usability as a conversational agent for delivering protocolized therapeutic support to family caregivers of children with chronic conditions. The iterative, human-in-the-loop approach supported the

development of empathetic and therapeutically grounded responses. More broadly, this study provides a structured framework for systematically integrating and refining evidence-based therapeutic approaches through iterative testing before LLM deployment.

*J Med Internet Res* 2026;28:e94175; doi: [10.2196/94175](https://doi.org/10.2196/94175)

**Keywords:** health-wellbeing; family caregiver; usability study; mental health; conversational agent

## Introduction

In 2025, an estimated 63 million family caregivers provided unpaid care to adults and children living with chronic health conditions [1]. The roles and responsibilities of family caregivers often involve working full-time while attending to the needs (eg, medication administration, mobility, and attending health care provider appointments) of their loved one and their family's needs, which often results in burn-out and symptoms of anxiety, depression, sleep disturbances, fatigue, and distress [2-4]. Family caregivers need support and resources to assist with caregiving, specifically in managing stress [5]. These factors have contributed to a critical demand for effective ways to deliver self-management interventions that promote better self-care in family caregivers [6]. While various consumer-friendly informatics tools are available to patients and caregivers, barriers to access and lack of personalized design to meet the specific needs of caregiving populations have limited the technological adoption and efficacy of existing solutions. Similar to patients, family caregivers use technology for information and social support [7]. A recent systematic review reported on the lack of efficacy of existing health informatics tools for family caregivers, specifically the use of informatics tools to support family caregivers and their needs [8].

Studies have shown that digital mental health tools positively improve caregivers' mental and physical health [9, 10]. Caregivers shared that coping skills, emotion regulation, skill building, and education are important aspects of digital mental health tools [11]. Another study used a quantitative survey to measure pediatric blood marrow transplant caregivers' activation after using the blood marrow transplant Roadmap IT tool. The results indicated that the caregiver's anxiety has decreased, their quality of life has improved, and growth was found in activation compared to the baseline [12]. AI is one of the core technologies behind many smartphone apps, robotics, and medical technology. Leveraging AI can help rapidly scan data and provide tailored information based on individual needs. Although there are multiple positive impacts of health informatics tools found in several studies, validated AI-enhanced tools for family caregivers of children with chronic conditions are limited [8].

Conversational agents provide a promising opportunity for self-management support in health care. A conversational agent is a software program that leverages AI to simulate human-like conversations with the user through written text or spoken language [13,14]. The first health conversational agent dates back to 1966 with the introduction of ELIZA, which used memory and grammar rules to mimic a Rogerian and nondirective psychotherapist and constructed responses

based on user input [15]. Since then, numerous studies have reported the development of conversational agents using language modeling techniques, and several systematic reviews have described the scope of the use of conversational agents in health care [16,17]. Recent advancements in language modeling, most notably in large language models (LLM), have opened new doors for AI technology to provide self-management support through natural language conversations. Generative AI models, such as the GPT series, Claude, and the open-sourced LLaMA, have set new standards for conversational agents, increasing user expectations and significantly improving the ability of these agents to provide support. For example, a recent systematic review reported the promising potential of mental health conversational agents showing strong efficacy in mitigating distress and depression symptoms [18]. In another study, researchers found that users can establish a therapeutic alliance with the conversational agent, which is a critical factor in successful therapy [19]. However, despite these potentials, there are risks in adopting AI-based conversational agents for mental health, including the lack of evidence-based data on accuracy, reliability, and interpretability [20-22]. The majority of AI-based systems have focused on delivering psychotherapy and psychoeducational content for the general population, and less attention is given to particular populations with specific needs and stress, such as family caregivers of children with chronic health conditions [18]. The National Alliance for Caregiving reported that family caregivers juggle several tasks simultaneously [23-25]. While conversational agents developed for the general population could offer much-needed mental health support, their focus is only a small fragment of the needs of family caregivers. Wang et al [26] explored the use of conversational agents to provide therapy for caregivers, identified both promises and limitations, and highlighted the need for more attention on developing conversational agents for caregivers and improving the quality of emotional response of these agents [26-28]. Recognizing the need, potential, and risk of conversational agents for family caregivers of children with a chronic health condition, the current study aimed to develop and evaluate a human-in-the-loop conversational agent designed to deliver problem-solving therapy (PST) for caregivers of children with chronic conditions specifically.

To provide services that offer targeted care and help relieve their physical and mental pressure, we developed Caring of Caregivers Online (COCO). COCO is a mobile application with an embedded chatbot providing family caregivers—particularly those with limited resources and high stress and burnout—with on-demand caregiving support and interactive self- and family-management skill development in English. The primary goal of COCO is to provide

on-demand, personalized, and emotionally intelligent support and health solutions to reduce caregiver burnout, thereby promoting their health and well-being. Features and functions include evidence-based intervention components such as daily check-ins [29], weekly PST sessions [30], automated reminders for self-care [31], symptom self-tracking [24], on-demand health and caregiving question-and-answer capabilities [14], and tailored resource recommendations [32]. In this paper, we described the design, development, and preliminary testing of COCO's key feature of delivering AI-enhanced PST to family caregivers.

## Methods

The development and iterative testing of COCO went through 4 phases outlined below. Phases 1 and 2 described the initial development of the rule-based conversational agent, phase 3 was the main usability study with prechat and postchat caregiver emotions evaluation, and phase 4 was an

exploratory technical demonstration of an updated COCO using an LLM.

### *Phase 1: Gold Standard Dialogue Development Informed by PST and Motivational Interviewing*

#### Theoretical Foundation of COCO

COCO intervention was guided by PST to help caregivers reach self-management and adaptive coping of stressful life events associated with caregiving. PST is an evidence-based treatment approach that provides a framework for COCO to systematically assess the problem (ie, caregiving symptoms such as stress, worry, and sleep disturbances) at hand, set specific and realistic goals, and brainstorm solutions with users to generate adaptive coping strategies to resolve the problem and reach the goals [11]. The 7-step guideline of PST was used to develop the flow of the therapeutic dialogues underlying COCO (Textbox 1).

**Textbox 1.** Problem-solving therapy (PST): 7-step guideline for Caring of Caregivers Online (COCO) dialogue.

- Step 1: Build a therapeutic relationship (eg, introduce COCO and the team behind COCO, and get to know the caregiver and their family).
- Step 2: Explain the structure (eg, the PST process and short-term and long-term goals of the intervention).
- Step 3: Identify and assess a caregiving symptom (eg, cause, frequency, context, and severity).
- Step 4: Set goals to reduce the symptoms.
- Step 5: Generate solutions to achieve the goals.
- Step 6: Implement solutions.
- Step 7: Evaluate solutions (eg, satisfaction, learning anything new about the symptom, and what can be done differently).

Additionally, we incorporated motivational interviewing (MI) principles into COCO to enhance the user's willingness and readiness to change (Textbox 2). According to the stages of change model, individuals move through incremental processes as new behavior is cultivated, from not

considering change to feeling ambivalent to getting prepared for change [33]. MI is both principle-oriented and technique-based, which together catalyze motivation enhancement and ambivalence resolution.

**Textbox 2.** Motivational interviewing (MI) principles and techniques.

#### **MI Principles**

- Cultivate change talk
- Soften sustained talk
- Strengthen partnerships
- Soften sustained talk
- Empathy

#### **MI techniques**

- Giving information
- Questioning
- Reflection (simple or complex)
- Affirmation
- Persuading with permission
- Seeking collaboration
- Emphasizing autonomy
- Normalizing (added by the authors)

## Persona and Conversation Development

Persona-based prototyping is an established user-centered design method used to represent target users' goals, behaviors, and needs in order to guide the iterative development

of health technologies [34]. Personas serve as practical design tools that ground development decisions in the lived experiences of the intended population, helping ensure the technology is relevant, acceptable, and appropriately tailored before formal testing [35]. We developed 4 family caregiver

personas through a systematic process integrating (1) a review of the empirical literature on caregiver burden and needs; (2) findings from our prior research with this population; and (3) participatory design sessions conducted with family caregivers of children with chronic health conditions and community-based experts working with these families [17, 33,36-39]. Each persona was constructed to reflect common psychosocial symptom profiles documented in the literature, including worry, anxiety, fatigue, sleep deprivation, and tiredness, ensuring that COCO's design addressed clinically and experientially relevant challenges faced by this caregiver population. Based on the personas developed, the clinicians (n=4) on the team (nurses and clinical psychologists) then created dialogues between COCO and caregiver personas, generating a total of 4 dialogues. In writing the dialogue, we incorporated the PST structure and MI principles, focusing on empathetic responses. Using the set of dialogues curated by the team, we created the low-fidelity prototype of COCO.

## ***Phase 2: Usability Testing of Low-Fidelity Prototype With Standardized Patients***

### **Standard Patients**

Standardized patients (SPs) are trained actors widely used in medical education and health technology research to simulate realistic clinical encounters in a controlled, reproducible manner [40]. We recruited 2 actors and trained them to role-play as family caregivers using the 4 personas developed in phase 1. Each SP was briefed on the psychosocial profile, symptom presentation, and caregiving context of their assigned persona prior to testing. This approach allowed for systematic and reproducible usability testing across personas while ensuring that interactions reflected clinically relevant caregiver experiences.

### **Wizard-of-Oz**

The Wizard-of-Oz (WoZ) method is an established technique in conversational agent development that enables the collection of naturalistic user-system dialogues prior to full automation [41]. In WoZ testing, a user interacts with a mock interface believing it to be a functioning system, while a human "wizard" operating behind the scenes generates the system's responses in real time. This approach allows researchers to evaluate interaction quality and refine dialogue design before committing to full technical implementation, making it particularly well-suited for early-stage conversational agent development [42]. For this study, 2 research team members with clinical experience served as COCO wizards, simulating responses that COCO was designed to deliver. The COCO-WoZ system architecture and technical setup have been described in detail in our previous publication [17]. SPs interactions with the mock interface were not informed of the wizard's involvement, ensuring that their behavior and responses reflected naturalistic interaction with a conversational agent.

### **Usability Testing Procedure**

Usability testing sessions were conducted via videoconferencing. Each of the 4 caregiver personas was tested in a 1-hour

session, during which the SPs completed 2 PST sessions within that persona's caregiving context. Following each session, usability data were collected through five open-ended questions addressing: (1) COCO-WoZ's perceived personality, (2) system functionality, and (3) overall design quality.

## ***Phase 3: Usability Testing and Preliminary Effects of COCO-WoZ on Caregiver Positive and Negative Emotions***

In this phase, we tested COCO-WoZ with family caregivers of children with chronic health conditions, examining system usability and changes in caregiver emotions in a pre-post study without a control.

### **Participants**

We recruited participants using a snowball sampling method through flyers shared with online forums, clinics, and hospitals. Eligible participants must be over the age of 18 years, currently taking care of a child with a chronic condition, speak and understand English, have a computer or laptop, and have access to stable internet. Given the formative and exploratory nature of this study, no formal sample size or power calculation was performed. The study was designed to provide preliminary estimates of usability and short-term changes in affect and was not powered to demonstrate clinical effectiveness.

### **Procedure**

We used the WoZ interface for COCO (COCO-WoZ) to deliver PST sessions. Participants reviewed study information, provided informed consent, and scheduled sessions with COCO-WoZ. Participants received an email reminder the day prior to their scheduled session. Before each session, the participants completed a brief survey rating their emotions at the moment. After completing the sessions, participants filled out a survey about their emotions and additional questions about their experience with COCO-WoZ.

### **Measures**

We used the Positive and Negative Affect Schedule-Expanded Scale (PANAS-X) to measure changes in current emotions in users' presessions and postsessions with COCO-WoZ [43]. We selected 6 subscales that included 34 items from PANAS-X to reduce survey burden. We selected 6 affective states: overall negative affect, specific negative affect (guilt, sadness, and fatigue), and specific positive affect (joviality and serenity). These were selected based on common caregiver symptoms and our hypothesis that these emotions would likely be affected by the therapy session. Each item was rated on a 5-point Likert scale of 1 (not at all) to 5 (very much). Each subscale consisted of 3 to 10 emotional items, respectively, and the subscale score was calculated by adding the individual item scores. Higher scores mean more intense emotions. User experience was measured using the System Usability Scale (SUS) [44]. The SUS is a 5-point Likert scale (1="strongly agree" to 5="strongly

disagree”) with 10 items that assess system usability and learnability [45]. The total score of all items was converted to a 0-100 scale, with a higher score suggesting better usability. A score of 70 or greater indicates acceptable usability [45].

### Phase 4: Development and Usability Testing of LLM-Powered COCO

Based on phase 3’s results, we continued improving COCO. In this final phase, we developed a GPT-4-based COCO delivering PST with integrated MI techniques and used persona-based methods to examine its preliminary performance. The details of the chatbot development process and testing are published in a separate manuscript [46].

#### Development

To develop GPT-4-based COCO, we used the emerging prompting methods at the time and created several versions of COCO and selected the best-performing model through human evaluation using a persona-based approach similar to phase 2. Details of the development and evaluation were reported elsewhere [46]. In the current study, we examined the dialogues generated from the best-performing model, which used few-shot learning with a brief description of COCO. The few-shot conversation examples were carefully crafted by our clinical team members using dialogues from phases 1-3, each demonstrating one or more MI techniques to effectively engage users in behavior change while maintaining therapeutic alliance. The LLM-powered agent was deployed using Streamlit (Snowflake Inc), which enables users to interact with LLM-powered COCO through a web app.

#### Evaluation

Three clinicians in the team (2 nurses and 1 clinical psychologist) conversed with the LLM-powered COCO using caregiver personas and generated 12 dialogues for the current evaluation of coding MI techniques.

#### Ethical Considerations

Phases 2 and 3 were reviewed by the University of Washington Institutional Review Board (IRB) and received exempt status (STUDY00007247). Phases 1 and 4 were not considered human subject research, and thus no IRB approval was needed. Participants eligible for the study provided informed consent before proceeding with study procedures.

#### Data Analysis

Descriptive statistics were conducted to summarize quantitative measures in phase 3. We examined the distributions

of the prechat and postchat PANAS-X 6 subscale scores, and paired-sample 2-tailed *t* tests were conducted to detect changes in emotions. To account for multiple comparisons across the 6 PANAS-X subscales, *P*-values were adjusted using the Holm procedure to control the family-wise error rate at  $\alpha=.05$ . Effect sizes were calculated using Cohen *d* for paired samples, defined as the mean of the within-participant pre-post difference scores divided by the SD of the difference scores. Occasionally, a participant omitted 1 item in the PANAS-X. In this case, we did not compute the subscale score containing that emotion, and this participant was not included in the specific subscale *t* test.

We used the *Motivational Interviewing Treatment Integrity Coding Manual* developed by Moyers et al [23] to code MI principles and techniques reflected in participants’ dialogues with COCO across the 4 phases mentioned above (see [Textbox 2](#) for MI techniques and principles). Initially, the research team members reviewed the phase 1 conversations and discussed coding. When a response did not fit an existing MITI category, a study-specific category was developed through discussion. For example, simple grounding was added to capture brief grounding or orienting responses (eg, “I see”) that were observed in the dialogue. The definitions and coding rules for these study-specific categories were discussed and agreed upon before the final coding was conducted. After that, one trained coder coded all COCO-generated responses. A second reviewer then examined the coding for accuracy and consistency. Any discrepancies were discussed between the coder and the reviewer until consensus was reached.

## Results

### Phase 1: Dialogue Development Informed by PST and MI

The conversation generally starts with COCO greeting the caregiver. After the caregiver generates a response, COCO will respond with an empathetic response such as “There seems to be a lot on your plate,” followed by a therapeutic response appropriate to the PST step in the conversation. The way the conversation is structured supports the design of COCO to be tailored to family caregivers ([Textbox 3](#)).

#### Textbox 3. Example dialogues created in phase 1.

##### Example dialogue 1

COCO: Earlier you mentioned that your worrying is causing you to feel on edge and anxious.

COCO: Is this correct?

Caregiver: Yes, that’s correct.

COCO: Okay.

COCO: You also mentioned that what is causing you to feel this way is worrying about Thomas.

COCO: Is this also correct?

Caregiver: Yep.

COCO: Is it safe to say that your worrying, the feeling of putting you on edge, and your anxiousness are all related?

*Example dialogue 2*

COCO: What would be a personal goal for you to address the sleep deprivation?

Caregiver: I want to get some sleep. But I can't do that if I'm worried about my kid.

COCO: I understand.

COCO: It's difficult to focus on taking care of yourself when your child is not well.

COCO: In the future, I'd like us to explore the feeling of worry and work toward reducing that worry.

COCO: But for now, let's talk about setting one goal to help you with your sleep deprivation.

## Phase 2: Usability Testing of Low-Fidelity Prototype With SPs

Participants overall found their experience affirmed by COCO-WoZ, felt supported, and perceived the chatbot as pleasant to interact with. Regarding COCO-WoZ's perceived communication style, participants stated that the chatbot was helpful in providing actionable suggestions. They also liked the colloquial language the chatbot used, which was informal and conversational. Regarding current functions, participants expressed a preference for receiving solutions for their goals instead of being asked to generate their own, stating, "I want somebody to throw a couple [of] things, and we can together and figure out what sticks." Additionally, participants noted that it was unclear whether to respond to the chatbot's statement-like response (vs questions). For overall design quality, participants enjoyed the features of COCO-WoZ, such as providing vetted websites to offer information about online chatrooms to find emotional support or a community of caregivers. They particularly liked the

flexibility of on-demand support that COCO-WoZ offered, stating, "It's nice that it's available. Anytime that would fit anybody's schedule."

## Phase 3: Usability Testing and Preliminary Effects of COCO-WoZ on Caregiver Positive and Negative Emotions

### Participants

A total of 38 participants were enrolled in the study, with 6 participants having incomplete demographic data. Thirty-two participants were included in the analyses on usability testing and effects of COCO-WoZ on positive and negative emotions. The majority of the caregiver participants were female, above 36 years of age, with an annual income of US \$100,000 or more. The majority of the children under care were under 10 years old and presented with a range of diagnoses, such as autism, arthritis, and diabetes (Table 1).

**Table 1.** Demographic characteristics of caregivers and children in the sample.

| Characteristics                  | Participants, n (%) |
|----------------------------------|---------------------|
| Caregiver characteristics (n=32) |                     |
| Sex                              |                     |
| Male                             | 6 (18.75)           |
| Female                           | 26 (81.25)          |
| Age (y)                          |                     |
| 26-35                            | 8 (25)              |
| 36-45                            | 18 (56.25)          |
| 46-55                            | 6 (18.75)           |
| Income range (US \$)             |                     |
| Less than \$19,999               | 2 (6.25)            |
| \$20,000-\$39,999                | 1 (3.13)            |
| \$40,000-\$59,999                | 2 (6.25)            |
| \$60,000-\$79,999                | 3 (9.38)            |
| \$80,000-\$99,999                | 2 (6.25)            |
| ≥\$100,000                       | 10 (31.25)          |
| Do not know or want to share     | 12 (37.50)          |
| Child characteristics (n=32)     |                     |
| Sex                              |                     |
| Male                             | 16 (50)             |
| Female                           | 16 (50)             |

| Characteristics | Participants, n (%) |
|-----------------|---------------------|
| Age (y)         |                     |
| 1-5             | 11 (34.38)          |
| 6-10            | 13 (40.63)          |
| 11-15           | 6 (18.75)           |
| >16             | 2 (6.25)            |
| Diagnosis       |                     |
| Anxiety         | 1 (3.13)            |
| Arthritis       | 7 (21.88)           |
| Asthma          | 2 (6.25)            |
| Autism          | 8 (25)              |
| Depression      | 1 (3.13)            |
| Diabetes        | 3 (9.38)            |
| Other           | 10 (31.25)          |

## Usability Testing Feedback

Common feedback surrounded the response and conversation flow functionality, experience chatting with COCO-WoZ, and interface design. Some feedback from the participants' initial session found the chatbot "helpful, but the response speed is quite slow" and "didn't flow like a conversation." Other participants also described their experience chatting with COCO-WoZ as a comforting and calming experience. One participant noted, "Using the bot, even now, was calming to me. It felt like it gave me some direction and hope." Another shared, "Sharing my experience releases a lot of stress for me. Even just having someone listen." We also gained feedback on improving the interface design. As one suggested, "I would appreciate it if there were some interesting photos embedded in the conversation." The mean score for SUS was 75.6% (SD 12.9%). This gives COCO-WoZ a "B" level, implying that there are some problems with the chatbot that

need to be addressed, even though the users were generally satisfied with the product [47]. The item-by-item analysis of the SUS matches the qualitative feedback: caregivers generally thought COCO-WoZ was easy to use, would like to use it, and felt confident in using it (average scores above 3 out of 5 for these items). Learnability was also high. One specific item that received a relatively lower usability rating was "I found the system very cumbersome to use," which received an average of 2.21 out of 5. We thought this was mainly related to the slow response time in the WoZ system.

## Changes in Positive and Negative Emotions

Compared to the presession, the participants showed significant improvements in the postsession of negative emotions with categories of negative affect, sadness, guilt, and fatigue ( $P<.05$ ; Table 2).

**Table 2.** Results of negative and positive emotions.

| Emotions          | Mean (SD)    | <i>t</i> test ( <i>df</i> ) | Holm-adjusted <i>P</i> value | Cohen <i>d</i> |
|-------------------|--------------|-----------------------------|------------------------------|----------------|
| Negative emotions |              |                             |                              |                |
| Negative affect   | -6.80 (6.59) | 6.09 (34)                   | <.001                        | 1.03           |
| Guilt             | -0.74 (1.09) | 4.02 (34)                   | .001                         | 0.68           |
| Sadness           | -2.53 (3.62) | 3.95 (31)                   | .001                         | 0.70           |
| Fatigue           | -1.94 (3.88) | 2.92 (34)                   | .02                          | 0.49           |
| Positive emotions |              |                             |                              |                |
| Joviality         | +1.09 (6.24) | 1.00 (32)                   | .05                          | 0.17           |
| Serenity          | +1.21 (3.02) | -2.33 (33)                  | .33                          | 0.40           |

## Summary of MI Techniques Across All Phases

The results show a progressive and varied use of MI techniques across the 4 phases.

In phase 1, a total of 4 sessions were coded for MI techniques, with a total of MI instances being 202. The top

3 MI techniques used were simple grounding ( $n=58$ , 28.71%), question ( $n=44$ , 21.78%), and reflection ( $n=30$ , 14.85%). Seeking collaboration and reflection also had significant usage, indicating a balance between engaging caregivers and ensuring empathetic responses (Table 3).

**Table 3.** Types of MI techniques used across each phase.

| MI <sup>a</sup> code     | Phase 1                 | Phase 2                 | Phase 3                  | Phase 4                 |
|--------------------------|-------------------------|-------------------------|--------------------------|-------------------------|
|                          | MI count (n=202), n (%) | MI count (n=132), n (%) | MI count (n=2970), n (%) | MI count (n=385), n (%) |
| Seeking collaboration    | 25 (12.38)              | 21 (15.91)              | 396 (13.33)              | 80 (20.78)              |
| Question                 | 44 (21.78)              | 36 (27.27)              | 661 (22.26)              | 61 (15.84)              |
| Reflection               | 30 (14.85)              | 22 (16.67)              | 266 (8.96)               | 58 (15.06)              |
| Normalizing              | 5 (2.48)                | 3 (2.27)                | 167 (5.62)               | 53 (13.77)              |
| Affirm                   | 10 (4.95)               | 7 (5.3)                 | 500 (16.84)              | 49 (12.73)              |
| Persuade with permission | 7 (3.47)                | 13 (9.85)               | 359 (12.09)              | 29 (7.53)               |
| Giving information       | 21 (10.4)               | 10 (7.58)               | 167 (5.62)               | 28 (7.27)               |
| Simple grounding         | 58 (28.71)              | 19 (14.39)              | 446 (15.02)              | 24 (6.23)               |
| Emphasizing autonomy     | 2 (0.99)                | 1 (0.76)                | 8 (0.27)                 | 3 (0.78)                |

<sup>a</sup>MI: motivational interviewing.

In phase 2, with 132 MI instances in 4 sessions, the use of questions increased to 27.27% (n=36), while seeking collaboration (n=21, 15.91%) and reflection (n=22, 16.67%) remained consistent. Interestingly, simple grounding decreased, reflecting a shift in conversational style as real-time interactions with SPs were tested (Table 3).

By phase 3, with a total of 2970 MI instances in 38 sessions, questions (n=661, 22.26%) and affirming (n=500, 16.84%) were among the most used MI strategies, showcasing an emphasis on helping caregivers clarify and reinforce positive behaviors. Seeking collaboration (n=396, 13.33%) and simple grounding (n=446, 15.02%) were also prominent, indicating COCO's effectiveness in creating a collaborative environment (Table 3).

In phase 4, the LLM-powered version of COCO showed a more balanced distribution of MI techniques across

its 385 instances in 12 sessions. Seeking collaboration (n=80, 20.78%) became the most frequently used technique, followed by questions (n=61, 15.84%) and reflection (n=58, 15.06%). This phase emphasized collaboration and empathetic understanding more than earlier phases (Table 3).

The average number of turns per session and MI techniques used per turn varied across phases. Phase 1 had 22.25 turns and 2.27 MI techniques per turn, while phase 3 had a similar 22.11 turns but a slightly higher MI-per-turn ratio (2.33), reflecting the robust integration of MI principles with caregivers. Phase 4 showcased COCO's enhanced performance with the GPT model, which had fewer turns per session (12.55) but the highest MI-per-turn rate (2.56), illustrating its ability to deliver concentrated motivational support in fewer interactions (Table 4).

**Table 4.** The average number of motivational interviewing (MI) techniques per turn of conversations across 4 phases.

| Phase | COCO <sup>a</sup>                       | Caregiver                 | Average number of turns | Average number of MI <sup>b</sup> | Average number of MI or turn |
|-------|-----------------------------------------|---------------------------|-------------------------|-----------------------------------|------------------------------|
| 1     | Team member                             | Persona-based team member | 22.25                   | 50.5                              | 2.27                         |
| 2     | WoZ <sup>c</sup> -supported team member | Persona-based SP          | 25.75                   | 33                                | 1.28                         |
| 3     | WoZ-supported team member               | Family caregiver          | 22.11                   | 51.6                              | 2.33                         |
| 4     | GPT <sup>d</sup>                        | Persona-based team member | 12.55                   | 32.08                             | 2.56                         |

<sup>a</sup>COCO: Caring of Caregivers Online.

<sup>b</sup>MI: motivational interviewing

<sup>c</sup>WoZ: Wizard-of-Oz.

<sup>d</sup>GPT: generative pre-trained transformer

## Discussion

This study aimed to develop and evaluate COCO, a conversational agent designed to deliver PST with MI techniques to family caregivers of children with chronic health conditions. The research unfolded in 4 distinct phases, each contributing to the refinement and evaluation of COCO. We started with laying foundational work for a theoretically sound intervention tailored to the unique needs of family caregivers. The subsequent phases used a mixed methods design and WoZ testing to assess the usability of COCO and its preliminary impact on user affective state in a

prechat and postchat design. The last exploratory phase enhanced COCO's performance using an LLM, using a similar number of MI techniques as humans, while exhibiting higher variability within the MI techniques used.

The evolution of COCO across the 4 phases demonstrates a clear progression in its ability to deliver PST with MI techniques, particularly as it transitioned from human-guided interactions to a fully LLM-powered system in phase 4. Initially, COCO relied heavily on foundational MI strategies, such as simple grounding and open-ended questioning, to engage caregivers. With further iterations of COCO, the integration of MI principles became more

refined with increased emphasis on collaboration, reflection, and affirmation, creating a more balanced and empathetic dialogue. Using human-curated examples from the first 3 phases of the study, we developed an LLM-powered COCO. In the person-based testing environment in phase 4, this version of COCO was able to deliver an increased variety of MI techniques, fostering deeper engagement in fewer conversational turns. This shift indicates that as COCO's conversational sophistication improved, it was able to more effectively respond to caregivers' needs, providing a richer, more supportive environment. GPT-4's ability to maintain consistency in MI use while enhancing collaboration and reflection highlights the potential of AI-driven agents to meaningfully replicate therapeutic techniques in a scalable and responsive manner.

The implementation of PST and MI through a chatbot platform like COCO presents unique opportunities and challenges. COCO offers on-demand support, potentially reaching caregivers with limited access to traditional therapy due to time constraints, geographic limitations, or financial barriers. However, personalizing COCO to capture the nuances of individual circumstances and emotional states is a significant hurdle, as is the challenge of establishing rapport—an essential aspect of PST and MI—through a chatbot interface [48,49]. Continued refinement of COCO's conversational abilities and empathetic responses is needed to address these challenges. Ethical considerations also play a critical role, including ensuring user privacy and data security, and establishing appropriate escalation protocols for crisis situations.

We used the PANAX-S scale to measure changes in short-term affective state in phase 3. In the prechat and postchat design, caregivers showed significant reductions in negative emotions and an increase in the feeling of serenity. Although there is evidence showing the relationship between short-term affect change and long-term mental health, this study is only a first step toward examining the long-term therapeutic impacts of a conversational agent-delivered protocolized therapy [50,51]. In addition, we took the approach of WOZ testing in this phase, rather than a fully automated conversational agent. Due to the high-stakes nature of mental health support, we carefully chose this approach in our first pilot study with actual caregivers. The changes in emotions prechat and postchat were likely affected by a combination of the protocolized therapy delivered through COCO and the skilled human touches through the WoZ interface.

While the preliminary evaluation of the LLM-based conversational agent showed promise in adopting a range of MI techniques, challenges related to response quality and controllability remain. For example, we observed instances

in which COCO repeated information shared by the caregiver rather than advancing the therapeutic conversation. This highlights the need to ensure that AI-supported interventions are not only responsive but also clinically appropriate, purposeful, and consistent with therapeutic goals [52]. Future work should prioritize clinical oversight and quality assurance throughout development, evaluation, and deployment. This may include expert-developed therapy guidance, structured review points to evaluate session quality, and mechanisms for clinicians to identify and correct problematic responses [53,54]. Technical approaches such as retrieval-augmented generation or multiagent systems [21,55] may support these goals, but their value should be evaluated in terms of whether they improve personalization, safety, and fidelity to PST and MI principles.

This study has limitations that should be considered. First, 3 out of the 4 phases of testing used persona-based approaches, limiting the variability of possible dialogues. Second, the initial COCO intervention was developed with a rule-based chatbot design, limiting its therapeutic scope. With the recent advancements in LLMs, we have started expanding COCO's therapeutic scope while carefully examining its performance. Third, our analysis of MI techniques was exploratory in nature. The frequency of MI techniques and MI techniques per conversational turn were used as descriptive measures to compare conversational patterns across development phases and should not be interpreted as validated indicators of therapeutic quality or intervention effectiveness. Future studies could examine the relationship between MI technique use and clinically meaningful caregiver outcomes in chatbots. Finally, the pilot evaluation in phase 3 included a small convenience sample of predominantly well-educated, higher-income caregivers recruited from an urban university neighborhood, limiting the generalizability of the results. A larger definitive trial with a control condition and long-term clinical outcome evaluation is warranted.

In conclusion, COCO shows promise as an accessible approach for delivering PST and MI-informed support to family caregivers of children with chronic conditions. The human-in-the-loop development process helped refine COCO's conversational design and improve the relevance and responsiveness of its interactions. The LLM-powered version of COCO exhibited a broader range of MI techniques and greater flexibility in applying these techniques during conversations. While these findings are encouraging, they are exploratory and do not establish therapeutic effectiveness. Further research with larger and more diverse caregiver populations is needed to evaluate COCO's clinical impact, optimize personalization, and improve the controllability and safety of LLM-based interventions.

## Acknowledgments

We gratefully acknowledge the participants and families who generously gave their time and shared their experiences to make this work possible. We also thank the research staff and community partners whose contributions were essential to the completion of this study. No generative AI tools were used at any stage of the preparation of this manuscript.

## Funding

This research was, in part, funded by the National Institutes of Health (NIH) Agreement Number 1OT2OD032581 and R21NR020634. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the NIH.

## Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

## Authors' Contributions

WY served as principal investigator and led the conceptualization of the study and manuscript, providing critical revisions and substantive feedback throughout the writing process. LW contributed to study conceptualization and data analysis and coordinated with other co-authors and drafted the manuscript. SJX and WK contributed to the technical components of the study, including model development. MD coordinated the project and managed overall study operations. XF participated in data collection and analysis. TMW provided senior technical expertise and guidance across the project.

## Conflicts of Interest

None declared.

## References

1. Resendez J, Choula RB, Cantor K, Caldera S, Frank L, Raimondi A, et al. Caregiving in the US 2025. AARP; 2025. URL: <https://www.aarp.org/content/dam/aarp/ppi/topics/ltss/family-caregiving/caregiving-in-us-2025.doi.10.26419-2fppi.00373.001.pdf> [Accessed 2026-09-05]
2. Javalkar K, Rak E, Phillips A, Haberman C, Ferris M, Van Tilburg M. Predictors of caregiver burden among mothers of children with chronic conditions. *Children (Basel)*. May 16, 2017;4(5):39. [doi: [10.3390/children4050039](https://doi.org/10.3390/children4050039)] [Medline: [28509853](https://pubmed.ncbi.nlm.nih.gov/28509853/)]
3. Meltzer LJ, Sanchez-Ortuno MJ, Edinger JD, Avis KT. Sleep patterns, sleep instability, and health related quality of life in parents of ventilator-assisted children. *J Clin Sleep Med*. Mar 15, 2015;11(3):251-258. [doi: [10.5664/jcsm.4538](https://doi.org/10.5664/jcsm.4538)] [Medline: [25515280](https://pubmed.ncbi.nlm.nih.gov/25515280/)]
4. Schulz R, Sherwood PR. Physical and mental health effects of family caregiving. *Am J Nurs*. Sep 2008;108(9 Suppl):23-27. [doi: [10.1097/01.NAJ.0000336406.45248.4c](https://doi.org/10.1097/01.NAJ.0000336406.45248.4c)] [Medline: [18797217](https://pubmed.ncbi.nlm.nih.gov/18797217/)]
5. National Alliance for Caregiving. 2024. URL: <https://www.caregiving.org/> [Accessed 2026-02-20]
6. Dionne-Odom JN, Azuero A, Lyons KD, et al. Family caregiver depressive symptom and grief outcomes from the ENABLE III randomized controlled trial. *J Pain Symptom Manage*. Sep 2016;52(3):378-385. [doi: [10.1016/j.jpainsymman.2016.03.014](https://doi.org/10.1016/j.jpainsymman.2016.03.014)] [Medline: [27265814](https://pubmed.ncbi.nlm.nih.gov/27265814/)]
7. Suzuki LK, Kato PM. Psychosocial support for patients in pediatric oncology: the influences of parents, schools, peers, and technology. *J Pediatr Oncol Nurs*. 2003;20(4):159-174. [doi: [10.1177/1043454203254039](https://doi.org/10.1177/1043454203254039)] [Medline: [14567564](https://pubmed.ncbi.nlm.nih.gov/14567564/)]
8. Zhai S, Chu F, Tan M, Chi NC, Ward T, Yuwen W. Digital health interventions to support family caregivers: an updated systematic review. *Digit Health*. 2023;9:20552076231171967. [doi: [10.1177/20552076231171967](https://doi.org/10.1177/20552076231171967)] [Medline: [37223775](https://pubmed.ncbi.nlm.nih.gov/37223775/)]
9. Shin JY, Chaar D, Kedroske J, et al. Harnessing mobile health technology to support long-term chronic illness management: exploring family caregiver support needs in the outpatient setting. *JAMIA Open*. Dec 2020;3(4):593-601. [doi: [10.1093/jamiaopen/ooaa053](https://doi.org/10.1093/jamiaopen/ooaa053)] [Medline: [33758797](https://pubmed.ncbi.nlm.nih.gov/33758797/)]
10. Smeallie E, Rosenthal L, Johnson A, Roslin C, Hassett AL, Choi SW. Enhancing resilience in family caregivers using an mHealth app. *Appl Clin Inform*. Oct 2022;13(5):1194-1206. [doi: [10.1055/a-1967-8721](https://doi.org/10.1055/a-1967-8721)] [Medline: [36283418](https://pubmed.ncbi.nlm.nih.gov/36283418/)]
11. Petrovic M, Gaggioli A. Digital mental health tools for caregivers of older adults-a scoping review. *Front Public Health*. 2020;8:128. [doi: [10.3389/fpubh.2020.00128](https://doi.org/10.3389/fpubh.2020.00128)] [Medline: [32411643](https://pubmed.ncbi.nlm.nih.gov/32411643/)]
12. Runaas L, Hoodin F, Munaco A, et al. Novel health information technology tool use by adult patients undergoing allogeneic hematopoietic cell transplantation: longitudinal quantitative and qualitative patient-reported outcomes. *JCO Clin Cancer Inform*. Dec 2018;2:1-12. [doi: [10.1200/CCI.17.00110](https://doi.org/10.1200/CCI.17.00110)] [Medline: [30652535](https://pubmed.ncbi.nlm.nih.gov/30652535/)]
13. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. *JMIR Ment Health*. Oct 18, 2019;6(10):e14166. [doi: [10.2196/14166](https://doi.org/10.2196/14166)] [Medline: [31628789](https://pubmed.ncbi.nlm.nih.gov/31628789/)]
14. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc*. Sep 1, 2018;25(9):1248-1258. [doi: [10.1093/jamia/ocy072](https://doi.org/10.1093/jamia/ocy072)] [Medline: [30010941](https://pubmed.ncbi.nlm.nih.gov/30010941/)]
15. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun ACM*. Jan 1966;9(1):36-45. [doi: [10.1145/365153.365168](https://doi.org/10.1145/365153.365168)]
16. de Cock C, Milne-Ives M, van Velthoven MH, Alturkistani A, Lam C, Meinert E. Effectiveness of conversational agents (virtual assistants) in health care: protocol for a systematic review. *JMIR Res Protoc*. Mar 9, 2020;9(3):e16934. [doi: [10.2196/16934](https://doi.org/10.2196/16934)] [Medline: [32149717](https://pubmed.ncbi.nlm.nih.gov/32149717/)]

17. Kearns WR, Kaura N, Divina M, et al. A wizard-of-oz interface and persona-based methodology for collecting health counseling dialog. *CHI Conf Hum Factors Comput Syst*. Apr 25, 2020. [doi: [10.1145/3334480.3382902](https://doi.org/10.1145/3334480.3382902)]
18. Li H, Zhang R, Lee YC, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *NPJ Digit Med*. Dec 19, 2023;6(1):236. [doi: [10.1038/s41746-023-00979-5](https://doi.org/10.1038/s41746-023-00979-5)] [Medline: [38114588](https://pubmed.ncbi.nlm.nih.gov/38114588/)]
19. Beatty C, Malik T, Meheli S, Sinha C. Evaluating the therapeutic alliance with a free-text CBT conversational agent (Wysa): a mixed-methods study. *Front Digit Health*. 2022;4:847991. [doi: [10.3389/fdgth.2022.847991](https://doi.org/10.3389/fdgth.2022.847991)] [Medline: [35480848](https://pubmed.ncbi.nlm.nih.gov/35480848/)]
20. Chung NC, Dyer G, Brocki L. Challenges of large language models for mental health counseling. *arXiv*. Preprint posted online on Nov 23, 2023. [doi: [10.48550/arXiv.2311.13857](https://doi.org/10.48550/arXiv.2311.13857)]
21. Guo T, Chen X, Wang Y, Chang R, Pei S, Chawla NV, et al. Large language model based multi-agents: a survey of progress and challenges. In: Larson K, editor. *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*. IJCAI Organization; 2024:8048-8057. [doi: [10.24963/ijcai.2024/890](https://doi.org/10.24963/ijcai.2024/890)]
22. Lawrence HR, Schneider RA, Rubin SB, Matarić MJ, McDuff DJ, Jones Bell M. The opportunities and risks of large language models in mental health. *JMIR Ment Health*. Jul 29, 2024;11:e59479. [doi: [10.2196/59479](https://doi.org/10.2196/59479)] [Medline: [39105570](https://pubmed.ncbi.nlm.nih.gov/39105570/)]
23. Moyers TB, Rowell LN, Manuel JK, Ernst D, Houck JM. The Motivational Interviewing Treatment Integrity Code (MITI 4): rationale, preliminary reliability and validity. *J Subst Abuse Treat*. Jun 2016;65:36-42. [doi: [10.1016/j.jsat.2016.01.001](https://doi.org/10.1016/j.jsat.2016.01.001)] [Medline: [26874558](https://pubmed.ncbi.nlm.nih.gov/26874558/)]
24. Murnane EL, Cosley D, Chang P, et al. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. *J Am Med Inform Assoc*. May 2016;23(3):477-484. [doi: [10.1093/jamia/ocv165](https://doi.org/10.1093/jamia/ocv165)] [Medline: [26911822](https://pubmed.ncbi.nlm.nih.gov/26911822/)]
25. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. *Am J Health Promot*. 1997;12(1):38-48. [doi: [10.4278/0890-1171-12.1.38](https://doi.org/10.4278/0890-1171-12.1.38)] [Medline: [10170434](https://pubmed.ncbi.nlm.nih.gov/10170434/)]
26. Wang L, Mujib MI, Williams J, Demiris G, Huh-Yoo J. An evaluation of generative pre-training model-based therapy chatbot for caregivers. *arXiv*. Preprint posted online on Jul 28, 2021. [doi: [10.48550/arXiv.2107.13115](https://doi.org/10.48550/arXiv.2107.13115)]
27. Shankar R, Bunde A, Mukhopadhyay A. Effectiveness of Chatbot interventions for reducing caregiver burden: protocol for a systematic review and meta-analysis. *MethodsX*. Jun 2025;14:103272. [doi: [10.1016/j.mex.2025.103272](https://doi.org/10.1016/j.mex.2025.103272)] [Medline: [40201159](https://pubmed.ncbi.nlm.nih.gov/40201159/)]
28. Cheng ST, Ng PHF. The PDC30 chatbot-development of a psychoeducational resource on dementia caregiving among family caregivers: mixed methods acceptability study. *JMIR Aging*. Jan 6, 2025;8:e63715. [doi: [10.2196/63715](https://doi.org/10.2196/63715)] [Medline: [39761547](https://pubmed.ncbi.nlm.nih.gov/39761547/)]
29. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Ann Behav Med*. May 18, 2018;52(6):446-462. [doi: [10.1007/s12160-016-9830-8](https://doi.org/10.1007/s12160-016-9830-8)] [Medline: [27663578](https://pubmed.ncbi.nlm.nih.gov/27663578/)]
30. Malouff JM, Thorsteinsson EB, Schutte NS. The efficacy of problem solving therapy in reducing mental and physical health problems: a meta-analysis. *Clin Psychol Rev*. Jan 2007;27(1):46-57. [doi: [10.1016/j.cpr.2005.12.005](https://doi.org/10.1016/j.cpr.2005.12.005)] [Medline: [16480801](https://pubmed.ncbi.nlm.nih.gov/16480801/)]
31. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. *J Med Internet Res*. May 14, 2009;11(2):e16. [doi: [10.2196/jmir.1138](https://doi.org/10.2196/jmir.1138)] [Medline: [19632970](https://pubmed.ncbi.nlm.nih.gov/19632970/)]
32. Cohn WF, Lyman J, Broshek DK, et al. Tailored educational approaches for consumer health: a model to address health promotion in an era of personalized medicine. *Am J Health Promot*. Jan 2018;32(1):188-197. [doi: [10.1177/0890117116671082](https://doi.org/10.1177/0890117116671082)] [Medline: [29214832](https://pubmed.ncbi.nlm.nih.gov/29214832/)]
33. Bickmore T, Schulman D, Yin L. Maintaining engagement in long-term interventions with relational agents. *Appl Artif Intell*. Jul 1, 2010;24(6):648-666. [doi: [10.1080/08839514.2010.492259](https://doi.org/10.1080/08839514.2010.492259)] [Medline: [21318052](https://pubmed.ncbi.nlm.nih.gov/21318052/)]
34. Ten Klooster I, Wentzel J, Sieverink F, Linssen G, Wesselink R, van Gemert-Pijnen L. Personas for better targeted eHealth technologies: user-centered design approach. *JMIR Hum Factors*. Mar 15, 2022;9(1):e24172. [doi: [10.2196/24172](https://doi.org/10.2196/24172)] [Medline: [35289759](https://pubmed.ncbi.nlm.nih.gov/35289759/)]
35. De Vito Dabbs A, Myers BA, Mc Curry KR, et al. User-centered design and interactive health technologies for patients. *Comput Inform Nurs*. 2009;27(3):175-183. [doi: [10.1097/NCN.0b013e31819f7c7c](https://doi.org/10.1097/NCN.0b013e31819f7c7c)] [Medline: [19411947](https://pubmed.ncbi.nlm.nih.gov/19411947/)]
36. Fagnano M, Berkman E, Wiesensthal E, Butz A, Halterman JS. Depression among caregivers of children with asthma and its impact on communication with health care providers. *Public Health*. Dec 2012;126(12):1051-1057. [doi: [10.1016/j.puhe.2012.08.007](https://doi.org/10.1016/j.puhe.2012.08.007)] [Medline: [23102501](https://pubmed.ncbi.nlm.nih.gov/23102501/)]
37. Laster N, Holsey CN, Shendell DG, Mccarty FA, Celano M. Barriers to asthma management among urban families: caregiver and child perspectives. *J Asthma*. Sep 2009;46(7):731-739. [doi: [10.1080/02770900903082571](https://doi.org/10.1080/02770900903082571)] [Medline: [19728215](https://pubmed.ncbi.nlm.nih.gov/19728215/)]

38. Yuwen W, Chen ML, Cain KC, Ringold S, Wallace CA, Ward TM. Daily sleep patterns, sleep quality, and sleep hygiene among parent-child dyads of young children newly diagnosed with juvenile idiopathic arthritis and typically developing children. *J Pediatr Psychol*. Jul 2016;41(6):651-660. [doi: [10.1093/jpepsy/jsw007](https://doi.org/10.1093/jpepsy/jsw007)] [Medline: [26994855](https://pubmed.ncbi.nlm.nih.gov/26994855/)]
39. Yuwen W, Lewis FM, Walker AJ, Ward TM. Struggling in the dark to help my child: parents' experience in caring for a young child with juvenile idiopathic arthritis. *J Pediatr Nurs*. 2017;37:e23-e29. [doi: [10.1016/j.pedn.2017.07.007](https://doi.org/10.1016/j.pedn.2017.07.007)] [Medline: [28778610](https://pubmed.ncbi.nlm.nih.gov/28778610/)]
40. Hamilton A, Molzahn A, McLemore K. The evolution from standardized to virtual patients in medical education. *Cureus*. Oct 2024;16(10):e71224. [doi: [10.7759/cureus.71224](https://doi.org/10.7759/cureus.71224)] [Medline: [39525234](https://pubmed.ncbi.nlm.nih.gov/39525234/)]
41. Kelley JF. An iterative design methodology for user-friendly natural language office information applications. *ACM Trans Inf Syst*. Jan 1984;2(1):26-41. [doi: [10.1145/357417.357420](https://doi.org/10.1145/357417.357420)]
42. Joglekar C. WOzBot: a wizard of oz based method for chatbot response improvement [Master's Thesis]. Trinity College Dublin, University of Dublin; 2022. URL: <https://publications.scss.tcd.ie/theses/diss/2022/TCD-SCSS-DISSERTATION-2022-127.pdf> [Accessed 2026-09-05]
43. Watson D, Clark LA. The PANAS-X: manual for the positive and negative affect schedule - expanded form. The University of Iowa; 1994. URL: [https://iro.uiowa.edu/esploro/fulltext/other/The-PANAS-X-Manual-for-the-Positive/9983557488402771?repId=12674991360002771&mId=13675073940002771&institution=01IOWA\\_INST](https://iro.uiowa.edu/esploro/fulltext/other/The-PANAS-X-Manual-for-the-Positive/9983557488402771?repId=12674991360002771&mId=13675073940002771&institution=01IOWA_INST) [Accessed 2026-09-05]
44. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *Int J Hum Comput Interact*. Jan 1995;7(1):57-78. [doi: [10.1080/10447319509526110](https://doi.org/10.1080/10447319509526110)]
45. Bangor A, Kortum P, Miller JT. Determining what individual SUS scores mean: adding an adjective rating scale. *J Usability Stud*. 2009;4(3):114-123. URL: <https://dl.acm.org/doi/10.5555/2835587.2835589> [Accessed 2026-09-05]
46. Filienko D, Wang Y, Jazmi CE, et al. Toward large language models as a therapeutic tool: comparing prompting techniques to improve GPT-delivered problem-solving therapy. *AMIA Annu Symp Proc*. 2024;2024:417-426. [Medline: [40417590](https://pubmed.ncbi.nlm.nih.gov/40417590/)]
47. Sauro J. Measuring usability with the System Usability Scale (SUS). MeasuringU. Mar 19, 2011. URL: <https://measuringu.com/sus/> [Accessed 2026-09-05]
48. Kayeser Fatima J, Khan MI, Bahmannia S, Chatrath SK, Dale NF, Johns R. Rapport with a chatbot? The underlying role of anthropomorphism in socio-cognitive perceptions of rapport and e-word of mouth. *J Retail Consum Serv*. Mar 2024;77:103666. [doi: [10.1016/j.jretconser.2023.103666](https://doi.org/10.1016/j.jretconser.2023.103666)]
49. Lee J, Lee D, Lee JG. Influence of rapport and social presence with an AI psychotherapy chatbot on users' self-disclosure. *Int J Hum Comput Interact*. Apr 2, 2024;40(7):1620-1631. [doi: [10.1080/10447318.2022.2146227](https://doi.org/10.1080/10447318.2022.2146227)]
50. Charles ST, Piazza JR, Mogle J, Sliwinski MJ, Almeida DM. The wear and tear of daily stressors on mental health. *Psychol Sci*. May 2013;24(5):733-741. [doi: [10.1177/0956797612462222](https://doi.org/10.1177/0956797612462222)] [Medline: [23531486](https://pubmed.ncbi.nlm.nih.gov/23531486/)]
51. Akhter S, Kumar T, Shaheen M. Mapping the evidence base of positive psychology interventions: effectiveness, limitations, and future directions. *Front Psychol*. 2026;17. [doi: [10.3389/fpsyg.2026.1812638](https://doi.org/10.3389/fpsyg.2026.1812638)]
52. Mennella C, Maniscalco U, De Pietro G, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: a narrative review. *Heliyon*. Feb 29, 2024;10(4):e26297. [doi: [10.1016/j.heliyon.2024.e26297](https://doi.org/10.1016/j.heliyon.2024.e26297)] [Medline: [38384518](https://pubmed.ncbi.nlm.nih.gov/38384518/)]
53. AlMakinahR, Norcini-Pala A, DisneyL, CanbazMA. Enhancing mental health support through human-AI collaboration: toward secure and empathetic AI-enabled chatbots. In: 2025 IEEE Conference on Artificial Intelligence (CAI). IEEE; 2026:196-202. [doi: [10.1109/CAI64502.2025.00038](https://doi.org/10.1109/CAI64502.2025.00038)]
54. Olisaeloka L, Richardson CG, Wang AY, Munthali RJ, Vigo DV. Safety mechanisms and risk mitigation in generative AI mental health chatbots: a systematic scoping review. *Healthcare (Basel)*. May 20, 2026;14(10):1395. [doi: [10.3390/healthcare14101395](https://doi.org/10.3390/healthcare14101395)] [Medline: [42194487](https://pubmed.ncbi.nlm.nih.gov/42194487/)]
55. Hong S, Zhuge M, Chen J, Zheng X, Cheng Y, Zhang C, et al. MetaGPT: meta programming for a multi-agent collaborative framework. *arXiv*. Preprint posted online on Aug 1, 2023. [doi: [10.48550/arXiv.2308.00352](https://doi.org/10.48550/arXiv.2308.00352)]

## Abbreviations

**COCO:** Caring of Caregivers Online  
**IRB:** institutional review board  
**LLM:** large language model  
**MI:** motivational interviewing  
**PANAS-X:** Positive and Negative Affect Schedule - Expanded Scale  
**PST:** problem-solving therapy  
**SP:** standardized patient  
**SUS:** System Usability Scale

---

**WoZ:** Wizard-of-Oz

---

*Edited by Matthew Balcarras; peer-reviewed by Chiedozie Arum, Miloud Chakit, Zhao Liu; submitted 25.Feb.2026; final revised version received 17.Aug.2026; accepted 18.Aug.2026; published 06.Oct.2026*

*Please cite as:*

*Yuwen W, Wang L, Xie SJ, Divina M, Fan X, Kearns W, Ward TM*

*Development and Preliminary Evaluation of a Conversational Agent Delivering Problem-Solving Therapy for Family Caregivers of Children With a Chronic Health Condition: Multiphase Mixed Methods Study*

*J Med Internet Res 2026;28:e94175*

URL: <https://www.jmir.org/2026/1/e94175>

doi: [10.2196/94175](https://doi.org/10.2196/94175)

© Weichao Yuwen, Liying Wang, Serena Jinchen Xie, Myra Divina, Xuehong Fan, William Kearns, Teresa M Ward. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 06.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.