Customizable AI for Depression Care: Improving the User Experience of Large Language Model-Driven Chatbots

Yi Li
School of Design
Hunan University
Changsha city, China
2012171@hnu.edu.cn

Xuanxuan Ding*
School of Design
Hunan University
ChangSha, China
dingxuanxuan@hnu.edu.cn

Yifan Chen
School of Design
Hunan University
Changsha, China
chenyifan001@hnu.edu.cn

Yeye Li
School of Design
Hunan University
Changsha, China
monicahaha@hnu.edu.cn

Nan Ma
School of Design
Hunan University
Changsha, China
mina@hnu.edu.cn

Abstract

Large Language Models (LLMs) demonstrate significant potential in the field of mental health. However, existing chatbots often lack personalized designs, which may limit their ability to fully address the complex needs of users with depression. This study builds upon the previously developed CloudEcho system, a mental health management application that integrates emotion monitoring and psychological support functions, to explore the impact of role customization features on user trust and customization experience. Using a mixed-methods approach, this study compares the differences between the system with role customization features and its original version. Quantitative results indicate that role customization can enhance user trust, showcasing high usability and satisfaction. Qualitative interviews further reveal the strengths and limitations of this feature and suggest directions for optimization. Together, these findings highlight the potential value of chatbot role customization in mental health support and offer theoretical and practical guidance for future LLM-driven personalized design and optimization in mental health contexts.

CCS Concepts

• **Human-centered computing** → **Empirical studies in interaction design**.

Keywords

Personalized Chatbots, Large Language Models, Mental Health Support, Persona Customization, User Trust

*Corresponding author.

DIS '25, Funchal, Portugal

ACM ISBN 979-8-4007-1485-6/25/07
https://doi.org/10.1145/3715336.3735795

ACM Reference Format:
Yi Li, Xuanxuan Ding, Yifan Chen, Yeye Li, and Nan Ma. 2025. Customizable AI for Depression Care: Improving the User Experience of Large Language Model-Driven Chatbots. In *Designing Interactive Systems Conference (DIS '25), July 05–09, 2025, Funchal, Portugal.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3715336.3735795

1 Introduction

Mental health issues have become increasingly severe worldwide, with approximately one in eight people suffering from mental disorders, such as anxiety and depression [116]. In China, about 95 million individuals are diagnosed with depression [20]. However, due to stigma [119] and limited healthcare resources [95], the actual number of cases is likely even higher. Emotional fluctuations and behavioral changes in individuals with depression often pose complex challenges to their social relationships and support networks, complicating interaction and assistance [32, 40]. Moreover, depression is strongly associated with high mortality rates, including those resulting from suicide [80], cardiovascular diseases [110] and accidents involving violence [66].

In recent years, conversational artificial intelligence (AI) tools powered by large language models (LLMs) have significantly transformed mental health support [6, 50], providing a stigma-free and accessible platform for psychological assistance [76, 77]. These tools also empower patients to engage more actively in self-care and decision-making [18, 44]. While prior research suggests that users generally enjoy interacting with chatbots [9], this does not necessarily imply meaningful trust, especially in high-stakes contexts like mental health [114]. In these scenarios, initial usage may stem from curiosity or novelty, but sustained engagement and effective psychological support require a deeper form of trust that enables emotional disclosure and relational alignment [14, 51]. Studies have shown that a good match between patients and therapists significantly enhances therapeutic outcomes, especially for individuals with high severity or complex condition [15]. This highlights that "liking" a system is not sufficient—the system must be able to foster identification-based trust to truly serve as a credible and supportive mental health partner.

To further explore the potential of LLMs in depression care, our research team previously developed CloudEcho, a mental health management tool driven by LLMs, designed to offer emotional support to users with depression. While CloudEcho demonstrated notable usability advantages, initial studies revealed that its lack of understanding of personalized user needs hindered trust-building and reduced user engagement with the system. Moreover, clinicians highlighted that chatbots' failure to address specific traumatic experiences of users with depression might lead to latent risks. Previous studies suggest that higher levels of personalization can make users feel that their individual needs are being considered, thereby increasing their trust and likelihood of using these services [35, 45, 57]. This underscores the importance of enhancing user trust and fostering sustained interaction through refined personalization in depression care.

Current personalization mechanisms in conversational agents within mental health contexts primarily rely on AI algorithms to analyze user behavior and preferences [69], dynamically adjusting interaction styles and therapeutic approaches [25, 47]. However, these mechanisms often depend on automated data analysis or case uploads, lacking direct user involvement. Additionally, their opaque processes raise privacy and ethical concerns, undermining users' trust in the system. In response to these challenges, this study aims to address the following research questions:

- **RQ1:** What is the impact of customizable chatbot roles on user trust in the context of depression care?
- **RQ2:** How do users customize LLM-driven chatbot roles in the context of depression care, and how can the customization process be optimized to enhance the user experience?

To address these questions, we enhanced the existing CloudEcho system by developing a user-customizable interface, enabling users with depression to define chatbot roles, language styles, and trigger conditions. This design aims to increase system transparency, reduce interaction complexity, and enhance users' sense of control and trust. Additionally, the system incorporates the participation of caregivers to ensure user safety and continuous support. We conducted an exploratory study comparing customizable and non-customizable chatbot roles in terms of user trust and employed both quantitative and qualitative analyses to examine the user experience of the customization process. The primary contributions of this study are as follows:

- This study provides empirical evidence that chatbot customization can enhance user trust in depression care contexts, suggesting its potential to foster sustained engagement and indirectly reduce caregiver workload over time.
- Design implications for chatbot customization interfaces in depression care, providing actionable insights on reducing operational complexity, accurately reflecting user intentions, and meeting the unique demands of this context.

The subsequent sections review relevant literature, detail the design of the customizable chatbot interface, outline the research methodology and experimental setup, and present a comprehensive analysis of the findings and their implications.

2 Related Work

Our review of the literature encompasses three primary areas of research: (1) trust in AI, (2) the application and development of LLMs-driven chatbots in mental health, and (3) recent advances in LLMs-driven chatbot role customization.

2.1 Trust in AI

Trust has emerged as a critical topic in contemporary human-computer interaction (HCI) research, particularly in understanding interactions between humans and virtual agents [11, 112]. Mayer et al. define trust as "the willingness of a user to rely on technology and its outputs" [62], establishing a foundational theoretical framework for exploring trust in technology [8]. Early efforts to measure trust can be traced back to the work of Muir and Moray, who developed a scale for assessing operator trust in industrial automation [72]. More recently, Gulati et al. [34] introduced a trust scale tailored to modern technological contexts, which serves as a significant tool for evaluating user trust in LLMs [26, 91].

In sensitive domains such as healthcare, trust in AI is paramount [49, 75]. This trust extends beyond cognitive dimensions, encompassing emotional elements of trust-building [85]. Research indicates that high levels of competence and benevolence are essential attributes for fostering trust in AI-powered healthcare providers [44]. Additionally, endowing AI systems with clearly defined roles or consistent personality traits has been shown to enhance user trust and acceptance of the technology [52, 71]. Personalization plays a particularly crucial role in fostering trust. Studies by Qin et al. highlight that personalized designs in healthcare AI not only boost cognitive and emotional trust but also encourage users to adopt AI-generated recommendations [85].

Although trust is widely recognized as crucial in AI-driven systems, it should not be seen as a static perception but rather as a developmental psychological structure. Lewicki and Bunker describe trust as evolving through three stages: deterrence-based (based on rules and constraints), knowledge-based (from cumulative interactions), and identification-based trust (rooted in shared values and emotional alignment) [53]. Similarly, Corritore et al. suggest that trust in human-computer interaction evolves from initial trust to mature trust with sustained engagement [23].Users' initial trust in chatbots typically arises from widespread technological adoption [4], reputable platform endorsements [74], or unrecognized privacy risks [9]. However, this initial trust is inherently fragile. In sensitive contexts like mental health, personalization aims precisely to transition such fragile initial trust toward stable, identification-based trust.

2.2 Applications of LLMs in Mental Health

The application of artificial intelligence in mental health spans a wide spectrum, from early identification of mental disorders to personalized treatment plans and the development of virtual therapists [76, 81]. Tools such as virtual therapists and chatbots have significantly improved the accessibility of mental health interventions, becoming essential channels for users to obtain psychological support [6, 82]. In recent years, large language model (LLM)-driven conversational AI systems have grown increasingly sophisticated, offering emotional support and cognitive behavioral interventions through empathetic interactions [76]. These systems provide a low-barrier, stigma-free platform for mental health management [77]. For instance, applications such as Lecter [96],VHope [12],Replika

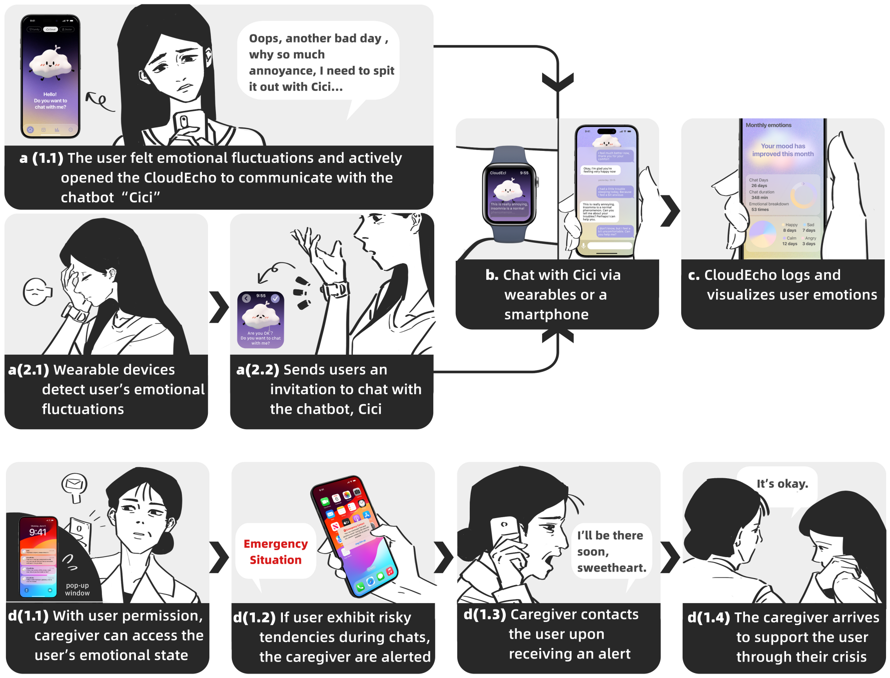


Figure 1: Overview of the CloudEcho usage process: (a) CloudEcho enables users to initiate chats with the chatbot “Cici” either proactively or when wearable devices detect emotional fluctuations. (b) Users can interact with Cici via the wearable device or a mobile application. (c) After the conversation, the application logs and visualizes emotional data for future reference. (d) With user consent, caregivers can receive updates on the user's emotional state; in critical situations, CloudEcho alerts caregivers to facilitate timely intervention and support.

[106], and HealMe [117] leverage mindfulness training and cognitive behavioral therapy to effectively alleviate anxiety and depressive symptoms while improving users' overall mental well-being [1, 55, 56, 61].Prior work has explored how chatbots function as intermediaries in mental health contexts. MindfulDiary employs a structured journaling interface that assists patients in recording emotional experiences for clinical use, offering limited conversational reciprocity and relatively weak social cues [42]. Building on this, Lee et al. investigate how chatbots may facilitate trust transfer to human therapists through self-disclosure mechanisms, marking a shift toward more socially engaged mediation [51]. However, the system adopts a generalized persona and also does not support dynamic tailoring to individual users. Wester et al. further examine the influence of perceived moral agency on users' trust and safety perceptions, highlighting both the relational potential and the ethical risks of heightened anthropomorphism [114]. Together, these studies reflect a progression from task-oriented assistance to socially responsive design.

Existing systems face significant limitations in their ability to deliver personalized experiences. Most rely on static, pre-defined roles and lack the flexibility to dynamically adapt to diverse user needs [5, 16, 83]. For example, models trained on specific datasets often exhibit inadequate adaptability and fairness when addressing users from varied cultural backgrounds or with complex symptom profiles, potentially leading to insensitivity and associated risks [83]. In contrast, personalized designs that dynamically adjust based on user behavior can enhance response accuracy [37] and enable real-time monitoring of high-risk situations, thereby notifying caregivers or clinicians to take proactive measures in mitigating relapse risks [76]. Moreover, personalized interventions should prioritize user autonomy and emphasize human-AI collaboration [90, 107]. Allowing users to customize chatbot roles enables them to directly participate in the personalization process, ensuring interventions align with their preferences and values while significantly enhancing trust and engagement [90].

To further explore the potential of LLMs in depression care, our team previously developed CloudEcho, an LLM-driven mental health management tool designed to provide emotional support and personalized collaboration. The system allows users to incorporate optional caregivers, such as family members, friends, or mental health professionals, based on their specific needs. In combination with wearable devices, CloudEcho continuously monitors users' emotional fluctuations. It initiates interactions to guide users in expressing emotions and sharing experiences while logging and visualizing emotional data for future management. When detecting potential crises, such as self-harm or suicidal tendencies, the system responds according to user-defined permissions. Figure 1 illustrates the typical workflow of CloudEcho, which includes emotion monitoring, interactive guidance, and crisis response, with a strong emphasis on privacy protection and adaptive responsiveness at every stage. However, initial studies revealed that despite its effectiveness in delivering basic emotional support, CloudEcho's "one-size-fits-all" role and interaction design struggle to meet the diverse needs of users, particularly in fostering long-term trust.

Against this backdrop, this study aims to investigate the application of chatbot role customization in mental health support. Focusing on the personalized needs of depression care scenarios, it seeks to provide both theoretical insights and practical guidance for enhancing user experience and delivering higher-quality mental health support.

2.3 Customizing Chatbot Personas

Since the HCI field began emphasizing the importance of avoiding "one-size-fits-all" interactions, personalized chatbot design has become a research focus [3]. Personalization is achieved by constructing user models and analyzing usage contexts to optimize specific interaction goals [7]. User models are typically generated through two approaches: direct input, where users actively provide personalized information, and indirect input, where systems dynamically adapt based on long-term interaction data [68]. For individuals with depression, the ability to customize a chatbot persona during the first use is particularly crucial, as it can quickly address their psychological support needs.

Key elements of persona customization include personality traits, demographic attributes, appearance, and linguistic style. For instance, the Five-Factor Model (FFM) has been widely used to simulate human personality traits [108] such as openness, conscientiousness, and extraversion [87]. Studies show that users with high agreeableness tend to prefer similarly agreeable AI personas [109], and consistency between a user's personality and a chatbot's interaction style significantly enhances positive evaluations of the technology [88]. Regarding linguistic style, users generally prefer chatbots with a sense of humor, but humor design must dynamically adapt to interactions to avoid discomfort caused by subjective differences [101]. In terms of demographic attributes, age and gender significantly influence initial impressions and acceptance of suggestions [103], but using the likenesses of celebrities or familiar individuals without permission may raise ethical concerns [36].

Despite the demonstrated value of personalized personas, direct user involvement in the customization process remains underexplored [84]. Current customization methods primarily fall into two categories: text-input-based models (e.g., ChatGPT [78]), where users are prompted to adjust model behavior and style, but may struggle to articulate their needs clearly; and form-based models (e.g., CloChat [36]), which guide users through customization via predefined options. While the latter reduces operational complexity, an abundance of options can confuse first-time users. These methods require emotional sensitivity, privacy protection, and adherence to ethical standards, particularly when applied in mental health scenarios [13].

This study focuses on the depression care context to analyze the core needs of users customizing chatbot personas. Since individuals experiencing depression often face cognitive symptoms such as concentration difficulties, decision-making fatigue, and reduced motivation [46, 64], the design emphasizes minimizing cognitive load to reduce operational barriers and support sustained engagement. By avoiding undue emotional strain, this research aims to advance the practical application of LLM-driven chatbots in personalized mental health support.

3 User Interface Design for Customizable Chatbots

To address the research questions, we developed a user-customizable interface based on the previously established CloudEcho system. This interface allows users with depression to define the identity traits, conversational style, and activation conditions of the chatbot, "Cici." It serves as an empirical exploration of the effects of customization on user trust and how the customization process can be enhanced to improve user experience.

3.1 Design Goals

This study aims to deliver a unique conversational experience where users can personalize the identity traits, conversational style, and activation conditions of the chatbot while incorporating caregivers as needed to enhance safety. Based on literature reviews and alignment with the research questions, the following design goals were established:

G1: Reduce Operational Complexity. Simplifying the customization process is critical to improving user experience, especially for individuals with depression. Existing customization settings often rely on intricate multiple-choice formats or complex text prompts, which are already challenging for general users and particularly so for users with depression. These cumbersome procedures may increase psychological strain and trigger frustration or resistance, undermining users' reliance on and trust in AI support systems [93]. This study seeks to optimize the customization process by streamlining interactions, enabling users with depression to easily complete personalization tasks, thereby improving their experience and fostering trust in the chatbot.

G2: Ensure Accurate Representation of User Intentions. He customization process must ensure that the system accurately understands and reflects user intentions, avoiding ambiguities or misinterpretations. Following the "Use Guidelines for Human-AI Interaction" by Microsoft, systems should maintain clarity in available actions and their implications, ensuring users understand how their inputs impact system behavior [67]. To this end, the design should incorporate real-time feedback mechanisms during the customization process, enabling users to comprehend the meaning and effects of each customization option, thereby enhancing the precision of intent expression.

G3: Adapt to Depression Care Scenarios. To ensure the AI system aligns with the unique needs of depression care, this goal emphasizes meticulous design and ethical considerations, building on the six design principles for Personalized Adaptive Conversational Agents (PACAs) proposed by Rangina Ahmad et al. [5]. We aim to enhance the transparency of customization settings, establish user-authorized sensitive data handling mechanisms, and introduce 24/7 proactive support. This ensures safety, transparency, and responsiveness in interactions. Moreover, the interface design adheres to the ethical design elements for depression-related applications outlined by Dionne Bowie-Dabreo et al. [16], emphasizing simple, intuitive interfaces and clear interaction flows to reduce cognitive load while minimizing psychological risks. These considerations collectively aim to provide users with depression a safe and inclusive personalization experience, fostering long-term mental health and emotional stability through daily interactions.

3.2 System Design

This section outlines the customization process design, which allows users to personalize Cici via conversational and form-based interactions. The overall process is illustrated in Figure 2. Upon first interaction, Cici prompts the user to initiate the customization setup, which can also be triggered manually at any time. Once the user completes the setup, the configuration is synchronized to the caregiver's interface, provided authorization is granted. The caregiver interface manages data in tabular form, reviews the preference part settings to ensure safety, and directly supplements the protective part to maintain reliability and privacy in information management. The following sections elaborate on the customizable elements and the design and implementation of the user and caregiver interfaces.

3.2.1 Theoretical Foundations and Classification of Customizable Content. To establish a design space for customizable role attributes, we conducted a systematic review of the literature, focusing on articles published in SIGCHI-related conferences (e.g., CHI, CSCW, and DIS). Using keywords such as "personalization," "conversational agents," "chatbots," and "mental health," we identified and manually curated 30 highly relevant studies through iterative screening and citation network exploration. These studies served as the theoretical foundation for developing our customization framework. Two researchers independently coded the characteristics extracted from the literature and reached a consensus through iterative discussions, resulting in nine classifications: identity attributes, language style, voice style, non-verbal behavior, conversational context, emotion-trigger mechanisms, data sharing, conversational topics, and sensitive events or keywords. These classifications provide a clear structure for subsequent design. Details of these classifications are provided in Appendix A.

3.2.2 Design of Customizable Content and User Safety Consideration. The customization process carefully balances user safety and personalization needs by categorizing user participation levels for different settings (Figure 3). This ensures user comfort and interaction transparency while addressing the unique requirements of depression care scenarios (G3).

Style part. This component enables users to fully customize elements such as role identity, language style, voice preferences, visual preferences, activation conditions, and data sharing. Its openness ensures user autonomy, allowing users to decide which information caregivers can access, thus safeguarding privacy and maintaining control. Once users input their preferences for role, language style, and voice style, these entries are automatically reviewed by the LLM's built-in content moderation mechanisms or predefined safety filtering modules. The review criteria include: Whether the language contains discriminatory, violent, offensive, self-harming, or otherwise inappropriate content, conflicts with the objectives of mental health support, or introduces ethical risks, such as power imbalances or unhealthy interpersonal interactions.

Preference part. Covering topics of interest, activity preferences during emotional fluctuations, and sensitive topics to avoid, this section requires caregiver validation to ensure accuracy and mitigate biases stemming from depressive symptoms. For instance, users with borderline personality disorder might exaggerate distress to seek attention [42]. Caregiver oversight prevents such inputs from undermining the system's effectiveness.

Protective part. Configured exclusively by caregivers, this section includes sensitive keywords, events, and crisis response protocols. When dangerous language is detected, the system triggers alerts for real-time intervention. To prevent sensitive terms from exacerbating the user's emotional distress, this section remains entirely hidden from the user, minimizing the risk of triggering adverse emotional reactions.

3.2.3 Interface Design for Users with Depression. This section focuses on interaction and visual aspects to enhance users' customization experience through simplified processes and dynamic visual feedback (Figure 4).

Interaction Design. The system combines conversational and form-based interactions to reduce the complexity of initial customization (G1). During the initial setup, users can define key elements such as role identity, language style, and visual preferences through voice or text interactions. The system employs guided conversations to present pre-set options incrementally (b), alleviating users' cognitive load. Users can seamlessly switch to a structured form (a) to review or modify settings, enhancing the intuitiveness and convenience of the maintenance process. At first use, the system automatically presents a setup prompt, whereas subsequent customizations can be initiated by issuing the "customize Cici" command to the chatbot. To accommodate the psychological state of users with depression (G3), the system provides clear progress indicators and allows users to pause the process at any time (c). If a user halts customization due to emotional fluctuations, the system saves their progress, enabling them to resume later when emotionally stable. Additionally, leveraging real-time monitoring via wearable devices, the system can proactively pause customization to safeguard users' mental well-being. In conversational mode, "Cici" provides immediate feedback after each key step (d), ensuring the alignment of customization with users' intentions (G2) and minimizing frustration caused by misinterpretations or errors.

Visual Design. Visually, the system employs dynamic character adjustments to deliver personalized feedback, strengthening role recognition and user trust (G2). For example, based on the selected role type, "Cici's" appearance dynamically changes: a mentor role might include glasses (e), a father role could feature a mustache, and a pet role might display relevant accessories. This nuanced feedback creates a direct link between the character's presentation and user expectations, enhancing the anthropomorphic effect and role distinguish ability [57]. By balancing engagement with consistency, dynamic visual feedback boosts immersion, reduces operational stress, and increases the enjoyment of the customization process, all while avoiding sensory overload.

3.2.4 Interface Design for Caregivers. Human supervision plays a crucial role in enhancing the safety and ethical reliability of artificial intelligence in mental health contexts [6]. Caregivers provide ongoing, dynamic insights into user needs and manage sensitive

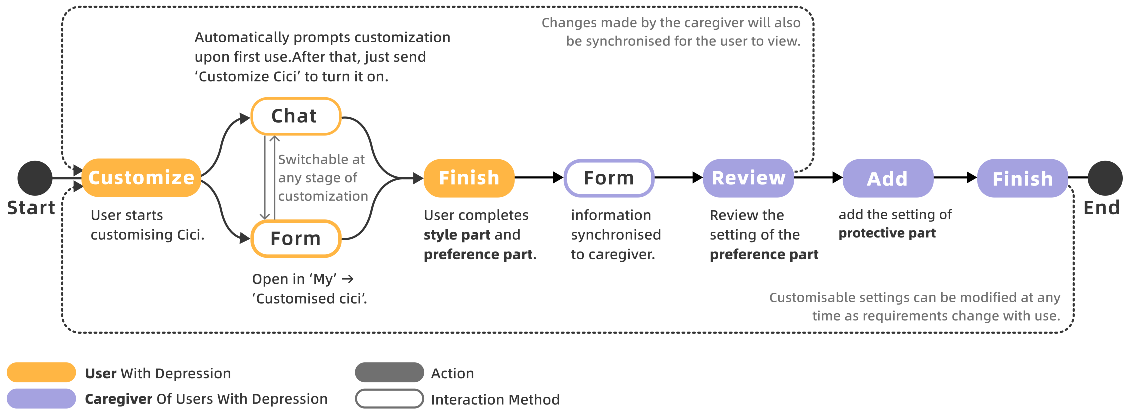


Figure 2: Customization process for AI chatbot roles involving both users with depression and their caregivers.

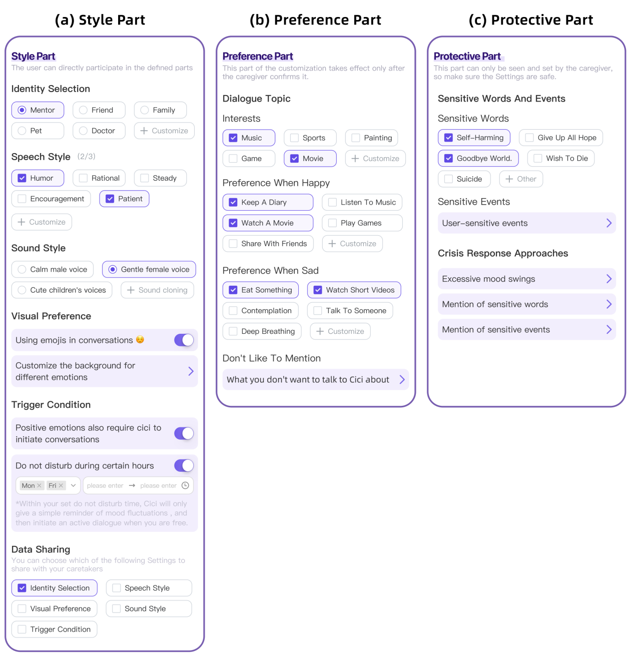


Figure 3: The design of customizable content is divided into three parts. (a) The Style part is independently set by the user, including role identity, language, voice, and visual styles, activation conditions, and data sharing, ensuring user autonomy and privacy protection. (b) The Preference part is collaboratively completed by the user and caregiver, covering topics of interest, emotional preferences, and taboo topics. This section requires caregiver approval to prevent subjective bias from misleading the system. (c) The Protective part is exclusively completed by the caregiver, addressing crisis-related sensitive keywords, events, and response plans, ensuring real-time intervention and emotional safeguarding.

content based on real-time feedback, thereby offering targeted protection (G3). This sustained involvement addresses the limitations of episodic clinical support, while bolstering users' emotional security and trust in the system (Figure 5).

The caregiver interface adopts a table-based interaction design to improve the efficiency and accuracy of information management. In the customization settings, only user-authorized content from the style part (b) is displayed. The preference part (c), including topics of interest and activity preferences set by users, requires caregiver review and approval before activation, ensuring that inappropriate inputs from users with depression do not compromise system performance. The protective part (d), designed exclusively for caregivers, manages crisis-related keywords, event records, and intervention strategies. To avoid triggering discomfort or emotional distress, this section is entirely hidden from users with depression. This multi-tiered permission management mechanism effectively balances personalized care with risk management, establishing a robust foundation for comprehensive protection and efficient collaboration in mental health support scenarios.

3.3 Design Implementation

In this section, we detail the technical aspects of the customization functionality (Figure 6). The CloudEcho system is built upon OpenAI's GPT-4o, integrated into a local service framework via the official API. Prior to this study, the system had been developed by the project team with built-in capabilities tailored for depression support scenarios. To ensure contextual appropriateness and ethical compliance in mental health dialogues, a structured system prompt was configured to guide the model's generation behavior during initial interactions. This system prompt comprises three components: first, it explicitly defines the conversational context as supporting users with depression, activating the model's internal safety constraints and ethical generation strategies; second, it incorporates established principles from mental health literature and clinical expert interviews, including empathetic responses, support for emotional labeling, and encouragement to communicate feelings with real-world caregivers; third, it defines the default persona settings for the chatbot "Cici," including core personality traits, linguistic tone, and voice preferences, establishing a gentle, patient, and trustworthy interactional identity. In addition, GPT-4o's base

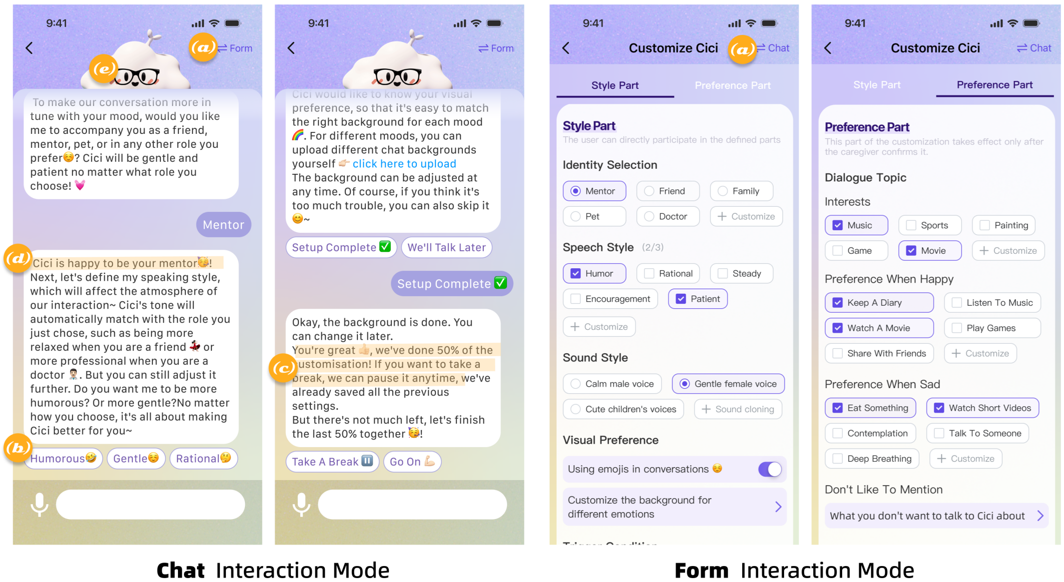


Figure 4: Main interface for the customization settings of users with depression. This interface integrates conversational and form-based interaction modes, allowing users to switch between them (a). In the conversational mode, tooltip suggestions (b) assist users in making choices, reducing operational complexity. Considering the unique context of depression care, the customization process provides progress updates and can be paused at any time (c). To ensure accurate reflection of user intentions, the system confirms choices at each critical step during the conversation (d), with dynamic visual adaptations of Cici, such as changes in appearance based on role selection (e).

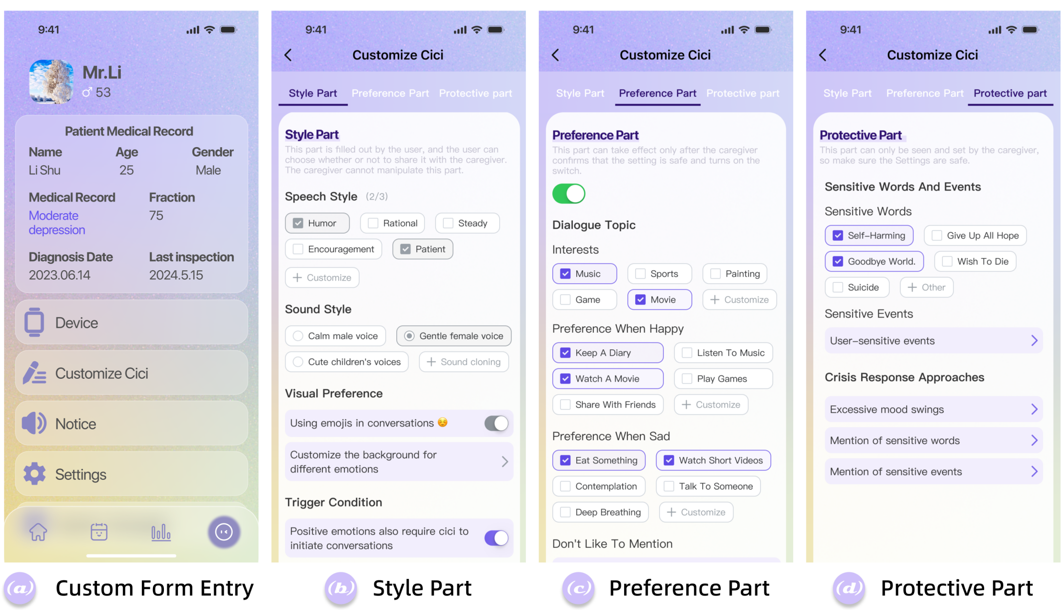


Figure 5: Main interface for customization settings on the caregiver's end. Caregivers can access the customization form via "My" > "Customize Cici" (a). The form includes the Style part (b), Preference part (c), and Protective part (d), with content from the protective part visible exclusively to caregivers.

Figure 6: Technical architecture of the CloudEcho customization process. The system integrates inputs from users with depression and caregivers to define user roles and preferences through chat-based or form-based modes (step 1). User-defined settings are transformed into structured JSON data (step 2) and incorporated into GPT-4o prompts (step 3) to enable personalized interactions. Visual feedback and dynamic role-based elements enhance the user experience, while sensitive word detection and contextual analysis ensure safety and caregiver notifications (step 4).

model includes a built-in content moderation mechanism capable of detecting high-risk expressions related to violence, self-harm, and discrimination, providing preliminary safeguards within mental health contexts [29, 113].

On top of this foundation, user- and caregiver-defined personalization inputs are appended to the model prompt in a structured format. These inputs adjust persona-related elements such as personality attributes, language style, and interactional preferences. They selectively override only the corresponding default persona fields, while preserving the original ethical constraints and behavioral boundaries of the system, ensuring that the customization process remains within a semantically controllable and contextually appropriate scope. The customization functionality offers two interaction modes: a form-based approach and a conversational interface. Users can either complete role-setting information through an intuitive form or describe their personalized needs in natural language. These settings are stored as structured data and communicated via a RESTful API to enable efficient parsing and synchronization. The personalized role information is incorporated into the model's input, enabling responses tailored to individual user needs. From an implementation perspective, the front-end interface is built using React, while the back-end leverages the Flask framework to handle business logic and interact with a MySQL database for managing user data, personalization settings, and interaction logs. All data transmissions are secured using HTTPS encryption protocols to ensure privacy and security. A visual feedback mechanism updates the chatbot's appearance in real time by invoking pre-configured dynamic image resources as users complete their role settings. For security management, the system integrates a keyword- and context-based detection technology to identify potentially high-risk content. If high-risk language, such as self-harm or suicidal expressions, is detected, the system immediately sends alerts to the caregiver for timely intervention. The sensitive keyword database allows caregivers to expand the vocabulary based on user-specific needs. By incorporating contextual semantic analysis, the system significantly enhances the accuracy of high-risk content detection while minimizing false positives.

4 USER STUDY

This study conducted a controlled experiment comparing CloudEcho systems with and without customization features to investigate their effects on user trust and the customization experience in the context of depression care (see Section 1 for research questions). The experiment was conducted in a controlled laboratory environment under full researcher supervision and was approved by the Ethics Committee of the School of Design, Hunan University, ensuring compliance with ethical standards.

4.1 Participants

To ensure the reliability and applicability of the experimental data, participants with depression and their caregivers were recruited through advertisements on local community health platforms in China. Eligible participants were required to have a confirmed diagnosis of depression, with no recent history of self-harm or suicidal tendencies. Caregivers needed to have an established, long-term relationship with the participants and be familiar with their behaviors and emotions. Both participants and caregivers were also required to possess basic smartphone proficiency and be able to visit the laboratory in person. Additionally, all participants were required to complete the Self-Rating Depression Scale (SDS) [120] and their suitability for the study was confirmed through psychiatric evaluations. To enhance the representativeness of the results, the recruitment process ensured a balanced gender and age distribution among the participants. In total, 16 pairs of participants and their caregivers took part in the experiment. Detailed demographic information is summarized in Table 1. Each participant received RMB 200 as compensation, an amount calibrated to local economic standards and approved by the ethics committee.

4.2 Study Setup and Procedure

Experimental environment. The study was conducted in a controlled laboratory with three designated zones: participant interaction, caregiver tasks, and research monitoring. Dedicated recording devices and cameras were used to document interactions while adhering to strict privacy and data security protocols.

Pre-Experiment Preparations. Participants completed the SDS scale [120] to confirm eligibility (scores between 53–62) for safety and applicability (Figure 7a). The research team provided detailed experiment instructions and obtained informed consent (Figure 7b). Prior to interaction, participants received a standardized introduction to the application's interface, features, and interaction modes to ensure baseline familiarity. Pre-experiment interviews (Figure 7c) guided participants to recall three recent emotional events, reconstructing scenarios to enhance realism and immersion.

Interacting with the Chatbot. Participants engaged with two versions of the chatbot "Cici": one with customization functionality (Figure 7d) and one without (Figure 7e). Participants were randomly divided into groups to ensure balanced interaction sequences. Text-based interactions were chosen for simplicity and accuracy, focusing on one of three emotional events recalled during interviews. Under the customization condition, participants completed role configurations before interactions began (Figure 7d-1). This process included customizing the style part and preference part, followed by caregiver safety reviews of the preference part and additions to the protective part. These measures ensured the accuracy and safety of participant configurations while maintaining ethical compliance. Each complete session—including onboarding, interaction with both chatbot versions, questionnaires, and interviews—lasted approximately 90 minutes.

Post-Experiment Questionnaires and Interviews. After interacting with the customizable and non-customizable versions of Cici, participants completed the Human-Computer Trust Scale (HCTS) developed by Gulati et al. [34] to quantify and compare trust levels under both conditions. HCTS is specifically designed to assess user trust in modern AI systems, making it suitable for evaluating trust in LLM-powered chatbots [26, 91]. Those using the customizable Cici also completed the USE Questionnaire [59] to evaluate the customization process. This questionnaire includes four dimensions—usefulness, ease of use, ease of learning, and satisfaction—which provide a comprehensive measure of the user experience with customizable interfaces, and directly align with our design goals G1 (reducing operational complexity) and G2 (reflecting user intentions). Semi-structured interviews explored their subjective perceptions of the system's functionality and emotional support quality, focusing on how customization affected interaction experiences and emotional support depth. Interview data were systematically coded and analyzed using thematic analysis to identify opportunities for design improvements and theoretical contributions to human-computer interaction research.

4.3 Data analysis

This study employed a mixed-methods approach, integrating quantitative and qualitative data to address the research questions comprehensively. In the quantitative phase, the HCTS [34] was used to measure trust under two interaction conditions. The 12-item

Table 1: Participants

Label	Gender	Age	SDS score	Frequency of use of LLM	Identity of caregivers
P1	Female	22	53	4 or more times per week	Brother
P2	Female	30	55	Everyday	Husband
P3	Female	25	61	Everyday	Mother
P4	Male	26	57	4 or more times per week	Girlfriend
P5	Male	32	56	1-2 times per month	Wife
P6	Male	28	55	4 or more times per week	Friend
P7	Female	23	59	1-2 times per month	Sister
P8	Female	21	53	Everyday	Father
P9	Female	49	55	4 or more per month	Husband
P10	Male	21	58	4 or more per month	Mother
P11	Male	37	60	Everyday	Wife
P12	Male	47	54	1-2 times per week	Son
P13	Female	45	56	1-2 times per month	Daughter
P14	Female	19	56	Everyday	Mother
P15	Male	21	55	1-2 times per week	Sister
P16	Male	31	57	4 or more times per week	Wife

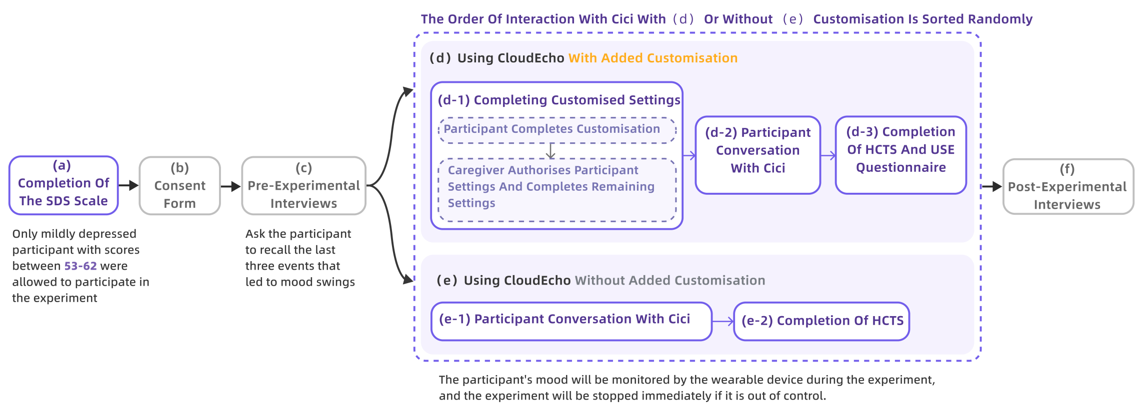


Figure 7: Experimental Procedure. Before the experiment, participants were required to complete the SDS scale to ensure suitability and safety (a). After signing the consent form (b) and participating in the preliminary interview (c), participants interacted with CloudEcho with (d) and without (e) customizable functionality through the chatbot "Cici." As shown, half of the participants interacted with the customizable Cici first, followed by the non-customizable version, while the other half followed the reverse order. The study concluded with a post-experiment interview (f).

scale evaluates five dimensions: perceived risk, competence, benevolence, reciprocity, and trust, rated on a 5-point Likert scale (1 = "strongly disagree," 5 = "strongly agree"). The USE Questionnaire [59] assessed the customization process across four dimensions: usefulness, ease of use, ease of learning, and satisfaction, rated on a 7-point scale (1 = "strongly disagree," 7 = "strongly agree"). Ease of use and ease of learning validated G1 (reducing operational complexity), while usefulness and satisfaction assessed G2 (reflecting user intentions). G3 (adapting to depression care scenarios) was further examined through qualitative analysis. In the qualitative phase, semi-structured interviews explored the customization feature's impact on trust, user experience, and future expectations. Interview recordings were analyzed using MAXQDA with bottom-up thematic coding. These insights complemented the quantitative findings, highlighting opportunities and challenges for long-term use and informing future design improvements.

4.4 Ethical Considerations for Users with Depression

To protect the mental health and rights of participants with depression, the study followed strict protocols. Screening was conducted using the SDS scale, inviting only those scoring 53–62, with caregivers consulted to confirm the absence of recent self-harm or suicidal intent. Participants repeated the SDS assessment before the experiment to ensure suitability. During the study, emotional monitoring devices tracked participants' states, with a psychologist available for crises. Participants retained the right to withdraw from the study at any point, respecting their autonomy and minimizing additional psychological stress caused by the experiment. Data was securely encrypted to maintain confidentiality, adhering to ethical guidelines and data protection standards.

5 Result

In this section, we address the two core research questions by integrating quantitative and qualitative data, delving into the impact of customization features on user trust and experience within depression care scenarios. Additionally, the analysis identifies opportunities to improve the customization process and role design, offering guidance for future applications of LLMs in mental health design.

5.1 Customization Features Enhance User Trust in Chatbots

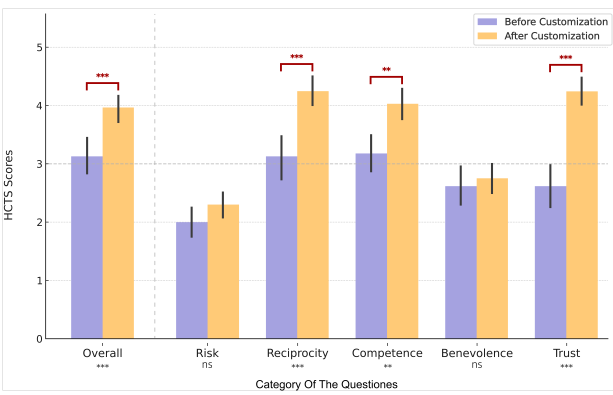


Figure 8: Comparison of mean HCTS overall and dimension-specific scores before and after customization (Section 5.1.1). Asterisks indicate the statistical significance levels based on paired-sample t-tests (*: p < .001, **: p < .01, *: p < .05). As shown, overall scores and three dimensions—Reciprocity, Competence, and Trust—showed significant improvements after customization (p < .05), while Risk and Benevolence showed no significant changes.**

5.1.1 *Changes in Trust Toward Chatbots During the Study.*

To evaluate the impact of customization features on user trust, we conducted paired-sample t-tests to compare the HCTS scores before and after customization. As shown in Figure 8, overall trust scores significantly increased from 3.71 (SE = 0.09) to 3.97 (SE = 0.06) post-customization ($p < 0.001$). Specifically, scores for Reciprocity ($p < 0.001$), Competence ($p < 0.01$), and Trust ($p < 0.001$) showed significant improvements, while Risk and Benevolence dimensions did not exhibit statistically significant changes.

Reciprocity. The significant improvement in Reciprocity ($p < 0.001$) highlights the perceived mutuality between users and the chatbot. Interview data revealed that users experienced a sense of accomplishment as they gradually refined their settings. P2 shared, *"As I progressively added settings, Cici became closer to the character I envisioned, which made me trust it more."* This iterative process not only fostered functional alignment but also reinforced a positive feedback loop between user effort and system response. P5 and P14 emphasized forming a *"connection"* with Cici after investing time and emotions, further strengthening the sense of reciprocity.

Competence. The significant improvement in Competence scores ($p < 0.01$) suggests that customization positively influenced users' perception of Cici's capabilities. Participants frequently noted that the customized Cici better aligned with their personal interests and needs, particularly in content recommendations and personalized language. P7 highlighted, *"The music Cici recommended matched my preferences perfectly, even though I didn't explicitly set them."* Similarly, P14 remarked, *"When I mentioned my breakup, it quoted my favorite movie line to comfort me, which was a delightful surprise."* Such implicit personalization enhanced trust and engagement. However, participants also reported limitations in emotionally complex scenarios. For instance, P2 mentioned that Cici's use of humor during moments of emotional distress could have a counterproductive effect.

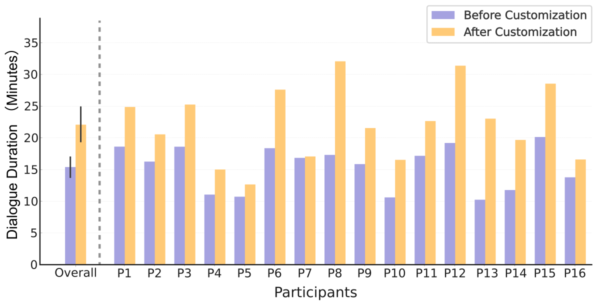
Trust. Trust scores significantly increased ($p < 0.001$), indicating the effectiveness of customization in building user trust. Participants described feeling reassured and emotionally understood by the customized Cici. P1 noted, *"Customized Cici is no longer a generic, one-size-fits-all responder but feels more tailored to my needs."* Similarly, P8 observed that customized Cici demonstrated greater care for emotional needs, thereby reducing interpersonal distance and increasing willingness to confide. However, some participants identified limitations in Cici's reactive nature. P4 and P8 suggested that its current interaction model risked reinforcing the perception of Cici as merely a *"negative emotion management tool."* P9 noted that trust was easier to establish when Cici consistently met user expectations.

Risk. Changes in the Risk dimension scores were not statistically significant ($p > 0.05$). During interviews, some users mentioned that the customization feature slightly enhanced their sense of control over the system. P1 noted, *"It outputs content according to my preferences, which makes me feel I have more control over it."* However, concerns about data privacy remained widespread. P13 commented, *"Although Cici offers more personalized settings, I still worry about data security."* This concern likely stems from a general lack of trust in the transparency of AI data usage. Additionally, P8 expressed apprehension about long-term data storage.

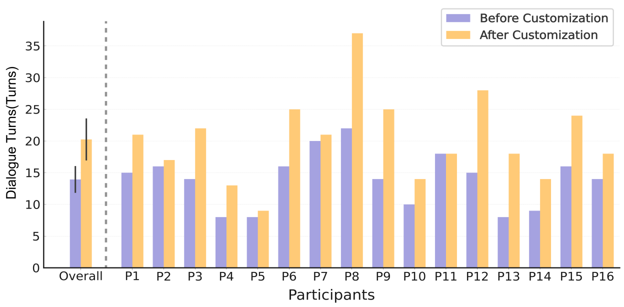
Benevolence. The Benevolence dimension scores before and after customization showed no significant difference ($p > 0.05$). While most participants felt the customized Cici better met their needs, its impact on perceived benevolence was minimal. P14 noted, *"The original Cici was already very kind; customization mainly refined its role attributes without significantly increasing its benevolence."* P4 observed that setting roles like mentor or doctor made Cici's responses feel more distant. Some participants wanted Cici to show more proactive care, with P8 saying, *"I wish Cici would engage with me more frequently, not just when detecting emotional fluctuations."*

5.1.2 *Changes in Dialogue Duration and Turns During the Study.*

Beyond the significant findings from HCTS, notable trends in participants' interaction behaviors were observed. After using customization settings, both the average dialogue duration and the



(a) Comparison of user **dialogue duration** before and after customization



(b) Comparison of user **dialogue turns** before and after customization

Figure 9: Comparison of Dialogue Duration (a) and Dialogue Turns (b) Before and After Customization (Section 5.1.2). Overall, both the Dialogue Duration and Dialogue Turns significantly increased after customization, reflecting deeper interactions between participants and the chatbot.

number of dialogue turns with Cici increased. While these observations were not statistically tested, they suggest that enhanced trust and personalization may encourage longer and more engaging interactions.

As shown in Figure 9, the average Dialogue Duration rose from 15.4 minutes (SE = 0.87) before customization to 22.17 minutes (SE = 1.46) afterward. Similarly, the average Dialogue Turns increased from 13.94 (SE = 1.08) to 20.25 (SE = 1.69). These changes indicate that customization contributed to deeper interactions. Interview feedback supports this: P8 shared, *"After customization, I felt the chatbot understood me better, which encouraged me to share more personal experiences."* P12 added, *"The customized prompts were more precise, making it easier for me to express myself and unintentionally extending the conversation."*

However, not all participants exhibited significant changes. Individual differences and goals likely influenced these variations. For example, P5 noted, *"As a rational user, I prefer clear advice."* Meanwhile, P11 mentioned, *"Customized responses were more detailed, requiring extra time to read and reflect."* Similarly, P16 said, *"The longer replies with role-specific information increased my reading time significantly."* Some participants highlighted unique reasons for extended interactions. P15 stated, *"I tested the customization feature's adjustments, which increased my conversation time but wasn't directly related to the topic."* P3 observed, *"Knowing customization was available, I shared more to help the chatbot understand me faster."* These findings highlight diverse impacts of customization on interaction patterns, suggesting opportunities for further refinement.

5.1.3 Participants' Preferences for Customized Chatbot Identities. examines participants' choices of chatbot identities to explore the factors driving trust. Although primarily based on interviews without quantitative validation, the findings offer valuable insights into trust-building mechanisms.

As shown in Table 2, participants' chosen identities were categorized into five groups: Real-life Figures, Fictional Human Characters, Animals, Anime Characters, and Non-living Entities. The most selected categories were Fictional Human Characters and Animals, indicating a preference for animated or living entities. Only one participant selected a deceased relative, and almost no one chose a currently living person.

Participants avoided real-life figures for various reasons. P8 said, *"I don't have someone I rely on, so I'm unsure who to choose."* Privacy concerns and dynamic relationships also played a role. P3 remarked, *"If I choose someone I know, I might equate them with Cici and worry about them sharing my secrets."* P12 added, *"If Cici represents someone close to me, real-life conflicts could reduce my trust in Cici."* P2 noted that discrepancies between Cici and the real-life individual could create a sense of disconnection. Conversely, deceased relatives provided comfort. P6 explained, *"Because they're no longer here, everything is just my imagination, so I feel more comfortable sharing."* P10 chose to set Cici as "my second self," stating, *"Only I can fully understand myself."*

Animal characters were highly favored for their low-pressure interaction. P8 noted, *"Animal characters allow for more imagination."* However, some found animal or non-living entities unnatural. P10 said, *"It feels odd when a cat answers my questions."* Compatibility with Cici's visual design also influenced preferences. P11 observed, *"Knowing Cici is a cloud-themed IP, I didn't associate it with anyone I know."*

Identity choices also impacted satisfaction. Emotional support roles (e.g., "An understanding elder sister") scored higher than professional guidance roles (e.g., "A psychologist"), likely due to their ability to meet emotional needs. P4 noted, *"Even with a stern professor role, I still hope it conveys warmth."* Animal roles also garnered high satisfaction, with many citing improved performance post-customization. However, unclear preferences led to lower satisfaction for some users.

5.2 User Experience of the Chatbot Customization Process

This section explores user experience based on USE Questionnaire data and interview feedback, focusing on operational complexity, functional precision, scenario adaptability, and interaction modes. The aim is to validate the design objectives and identify opportunities for optimization, providing theoretical guidance for applications in depression care contexts.

5.2.1 Streamlined and Efficient Customization Process (G1). This section evaluates the operational experience of chatbot customization through the Ease of Use and Ease of Learning dimensions

Table 2: Participant selection of identities

Category	Identities	Participant	Satisfaction rating (1-5)
Real-life figures	A deceased grandfather	P6	4
Fictional human characters	An understanding elder sister	P1	4
	A psychologist	P2	2
	A close and empathetic friend	P3	3
	A professor of literature	P4	3
	An authoritative psychologist	P5	3
	A life mentor	P9	3
	Alter Ego	P10	4
	A benevolent and affectionate grandmother	P13	5
Animals	A cute rabbit	P7	4
	An affectionate ragdoll cat	P8	3
	A friendly golden retriever	P12	5
	A deceased pet dog	P14	4
Anime characters	Totoro	P16	4
Non-living entities	A cloud	P11	4
	A magical spirit from a fairy tale	P15	3

of the USE Questionnaire, complemented by user interviews. According to USE data (Figure 10), the overall mean score for the customization feature was 5.69 (SE = 0.16), reflecting positive user perceptions. The Ease of Learning dimension scored the highest at 6.25 (SE = 0.10), while Ease of Use achieved an average of 5.79 (SE = 0.07), both significantly above the midpoint of 4. These results highlight the system's effectiveness in reducing the learning curve and simplifying operational workflows.

Interview feedback further supports these findings. Many participants noted that guided prompts and keyword recommendations effectively streamlined the customization process, enabling them to complete the setup more easily. For instance, P16 remarked, *"It guides me step by step, making it very simple."* Similarly, P13 shared, *"It's hard for me to come up with it on my own without the Keywords."* P10 also highlighted that the tooltip feature helped alleviate decision-making pressure.

Nevertheless, some areas for improvement were identified. Differences in user backgrounds influenced their operational experiences. For tech-savvy participants, the process was perceived as smooth and user-friendly. However, less familiar users required additional guidance; P7 suggested, *"More detailed tooltips and guidance might help users who are not familiar with the process."* Additionally, while the streamlined process facilitated quick setup, it did not always translate into deeper trust in the chatbot. P9 observed, *"Customization felt like completing a checklist. It seemed like Cici's understanding of me was superficial, lacking a sense of natural accumulation."*

5.2.2 Accurate Reflection of User Intentions (G2). This section examines how well the customization feature reflects user intentions using the "Usability" and "Satisfaction" dimensions of the USE questionnaire and interview feedback. As shown in Figure 10, the Usability score averaged 4.95 (SE = 0.12), while Satisfaction reached 6.07 (SE = 0.08), both above the median. The notably high Satisfaction score reflects strong user approval of the customization experience.

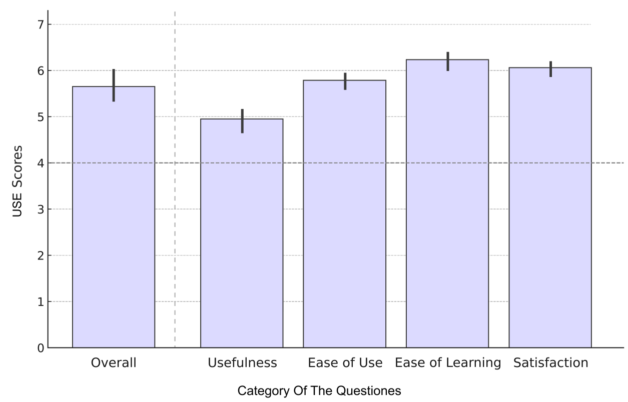


Figure 10: Comparison of overall and dimension-specific USE Questionnaire scores (Section 5.2.1, 5.2.2). The results indicate consistently positive feedback, with all dimensions scoring above 4 on a 7-point Likert scale.

Interview feedback supports these findings. Users appreciated the clear process and effective feedback mechanisms. P16 said, *"The setup is simple and clear, with system guidance on remaining steps, making the process feel manageable."*

Similarly, P13 noted: *"The system confirms each step, which is reassuring."* However, some areas for improvement were identified. P7 suggested adding engaging visual feedback: *"Progress could be visualized, like making the cloud IP more transparent as customization advances, for a more enjoyable experience."* Others recommended more specific feedback. P2 said: *"Concrete examples of settings, like humor levels, would better align with user needs."* Decision-making challenges also surfaced, as P8 shared: *"Choosing identity and style was sometimes overwhelming."*

5.2.3 Adaptation to Depression Care Scenarios (G3). The sensitive word alert mechanism was widely recognized as essential for safeguarding user privacy and preventing unintended emotional harm. As a caregiver of P3 noted, *"Because these settings are private and vary by individual. Cici might unintentionally trigger discomfort otherwise."* P10 also emphasized that discussing certain taboo topics could lead users to abandon the application entirely.

Interestingly, no sensitive word alerts were triggered during the experiment. P7 reflected that, within a lab environment, participants tended to stay emotionally restrained and carefully chose their words: *"Even if I felt something, I wouldn't express it too directly in this setting."* This suggests that emotional inhibition under observation may have reduced the likelihood of triggering such alerts—highlighting a possible discrepancy between lab behavior and real-world risk expression.

Meanwhile, several caregivers expressed difficulty in configuring the sensitive word list. P2's caregiver noted, *"It's hard to think of specific phrases within a short time, even though I know it's important."* P8's caregiver added, *"I worry that if I miss something, risks might be overlooked—but if I include too much, constant alerts could cause resistance from my child."* As P15 also pointed out, *"Not all sensitive words mean danger. Phrases like 'I want to die' could be habitual, and automatic caregiver alerts might feel intrusive and discourage use."* These findings suggest the need for more nuanced support during sensitive content configuration and raise important questions about the ecological validity of keyword-based alert mechanisms in emotionally complex, real-world scenarios.

In terms of interaction design, participants found the guided process effective in reducing cognitive load and aiding expression. P7 stated, *"Step-by-step guidance helped me express myself."* P15 suggested pre-set templates for users with limited energy, offering quicker options instead of relying on manual customization. Stigma and social pressure were notable challenges. P6 highlighted, *"Cici's appearance is too obvious. If its IP image appears when my wearable detects I'm upset, it might reveal I'm using this app, increasing stigma."* The permissions design also drew criticism for potentially reducing user autonomy. P8 commented, *"Requiring caregiver approval for preference settings makes users feel incapable of managing themselves."*

5.2.4 Combined Chat and Form-Based Customization Enhances User Experience. The integration of chat and form-based customization was widely appreciated by participants for balancing interaction flexibility and structured efficiency. P6 noted, *"This approach is more engaging and avoids monotony."* P7 added, *"Chat excels in guidance and helps me realize what I want through prompts."* Similarly, P8 stated, *"Forms can feel restrictive, while chats are more casual. However, forms are faster and irreplaceable for modifications."* P5 highlighted the humanized nature of chat, stating it *"feels like conversing with a person."*

Opinions varied on text length. P15 found long texts overwhelming, while P13 remarked, *"Longer texts make me feel cared for, but voice interaction might be better suited for such scenarios."* Many participants, including P13, showed strong interest in voice interaction, citing its unique advantages in reducing cognitive load and enhancing emotional expression. P11 observed, *"Voice input is easier than typing as it doesn't require careful wording. Additionally, it conveys emotional nuances."* P1 further emphasized, *"Voice interaction subtly encourages sharing more detailed content."*

6 Discussion

This study explored the impact of chatbot persona customization on user trust (RQ1) and examined user experience and potential improvements during the customization process (RQ2) in depression care. Findings showed that customization significantly enhanced emotional connection, trust, engagement, and satisfaction. These results validated customization as an effective strategy for building human-AI trust while identifying limitations and design challenges in the process.

This chapter discusses the role of customization in trust-building and its broader implications (Section 6.1) while reflecting on interaction simplification and ethical considerations (Section 6.2). It then provides a critical examination of the broader implications of personalization in LLM-based mental health systems (Section 6.3), and proposing directions for ethically grounded, clinically informed design.

6.1 Designing for Evolving Trust in Mental Health Chatbots

Trust in AI chatbots is not static but evolves through ongoing interaction. This study shows that allowing users to customize chatbot personas significantly enhances trust in depression care contexts (Section 5.1). These findings support prior work suggesting that systems attuned to users' individual needs and emotional preferences can foster greater trust [85, 105]. In what follows, we examine how personalization shapes different layers of trust and discuss design directions to support its development.

6.1.1 How Customization Enhances User Trust. Customization was a key factor in enhancing user trust, as confirmed by HCTS data, particularly in the improved reciprocity, competence, and trust dimensions (Section 5.1.1). First, customization strengthened users' sense of control and agency. By defining chatbot traits, users constructed a personalized entity, with trust growing over time through sustained interaction [30]. This participatory process fostered emotional connection and reciprocal relationships. Second, customization enhanced system adaptability, enabling users to feel more accurately understood by the chatbot. In mental health contexts, users expect tailored responses. Adjusting persona traits [89] increased personalization, enhancing recognition of the system's competence.Third, customization promoted safer interactions. Sensitive word and event settings helped prevent emotional triggers, while transparent customization pathways reduced concerns about the "black-box" nature of LLMs, boosting cognitive and emotional trust.

Persona selection also shaped trust-building (Section 5.1.3). Prior research indicates that users prefer chatbots exhibiting high conscientiousness [19] and extroversion [109], with helpful and friendly traits significantly increasing their appeal [70]. This observation is validated by the results of the present study. Emotional personas were favored for their warmth, while professional personas met expertise needs but lacked emotional connection due to rational communication styles. Balancing functionality and emotional expression is essential for holistic interactions. the virtual and abstract

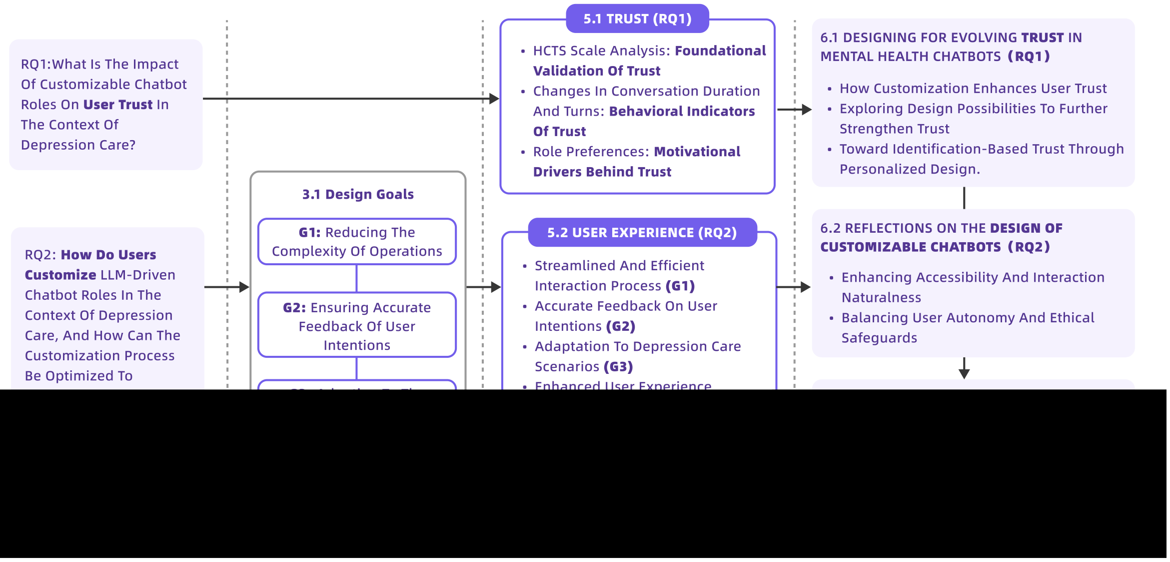


Figure 11: Research Framework.This figure illustrates the overall framework of the study, focusing on the impact of chatbot persona customization on user trust (RQ1) and the user experience during the customization process (RQ2). The framework begins with the formulation of research questions (Section 1, 2), proceeds with the design of a customizable chatbot interface based on defined design goals (Section 3), followed by experimental validation (Section 4, 5), and concludes with an analysis and synthesis of findings in the discussion (Section 6).

persona designs, like Cici's "cloud" identity, reduced psychological burdens. Participants noted that virtual personas avoided direct links to real-world individuals, alleviating privacy concerns and trust barriers from inconsistent portrayals. This aligns with ethical considerations in mental health support [36], where real-life personas risk causing discomfort or stress if chatbot behavior deviates from expectations.

The increased dialogue duration and turns provide further insight into trust dynamics (Section 5.1.2). Longer and deeper interactions suggest that personalization encourages users to build sustained connections with the system. However, extended conversations also pose efficiency challenges. Future designs should focus on emotionally rich yet efficient interactions, avoiding user fatigue from lengthy or repetitive exchanges, and balancing trust-building with interaction efficiency.

6.1.2 Exploring Design Opportunities to Enhance Trust. While customization improved user trust overall, the lack of significant changes in the Risk and Benevolence dimensions highlights areas for refinement.

The Risk dimension reflects persistent concerns about data privacy and security (Section 5.1.1). The personalization-privacy paradox suggests that greater personalization can heighten fears of privacy breaches [35]. Healthcare design must balance personalization and privacy by reducing reliance on user data, enhancing transparency, and improving user control [92]. Future designs should integrate emotionally resonant, transparent features to address these concerns effectively.

The Benevolence dimension reveals limitations in emotional care within the current interaction model (Section 5.1.1). Cici's proactive interactions focus mainly on negative emotional triggers, leading users to perceive it as a "crisis management tool" rather than a holistic companion. Expanding Cici's role to include interactions based on user interests or routines could balance negative emotion management with positive engagement, fostering greater intimacy and trust while reducing the perception of a "problem-oriented" approach.

6.1.3 Toward Identification-Based Trust through Personalized Design. In this study, personalization features were found to significantly support the formation of knowledge-based trust by enhancing users' sense of control, improving the interactional appropriateness of system responses, and enabling user-defined configuration processes. However, to facilitate the evolution of trust into a deeper, more enduring form—identification-based trust—more targeted matching strategies are required to strengthen users' perception of long-term relational alignment with the system.

On the one hand, capability-based matching has been shown to enhance users' perception of system competence and reliability. Prior research indicates that users are more likely to trust and cooperate with professionals who have demonstrated experience with similar issues [15]. Platforms such as Meela Health implement refined matching algorithms that go beyond professional background, aligning users with therapists based on therapeutic style, communication preferences, and shared values [65]. These capability-oriented strategies reinforce the role of AI systems as reliable cognitive assistants, thereby strengthening knowledge-based trust.

On the other hand, achieving identification-based trust necessitates attention to emotional resonance and value congruence between the user and the system. Studies have shown that when users perceive an AI agent to exhibit empathy or stylistic compatibility in communication, they are more likely to develop affective trust grounded in emotional alignment [97]. Lee et al. further demonstrated that in contexts where chatbots function as intermediaries for psychological support, appropriate self-disclosure not only increases trust in the chatbot itself but also facilitates trust transfer, encouraging users to open up to subsequent human counselors [51].

We therefore suggest that future personalization designs should not only focus on capability and interest alignment, but also explore affective expression strategies, self-characterization mechanisms, and the system's potential as a social mediator. These directions can help shift trust from a function-oriented to a relationship-oriented paradigm.

6.2 Reflections on the Design of Customizable Chatbots

This section examines the user experience of customizable chatbots from a design perspective, focusing on strategies to enhance accessibility, interaction naturalness, and the balance between user autonomy and ethical safety in mental health contexts.

6.2.1 Enhancing accessibility and interaction naturalness. Findings (Section 5.2.1,5.2.2) reveal that users often struggle to achieve desired settings due to limited understanding of LLMs or inadequate guidance. This can lead to misaligned outcomes and reduced trust in the chatbot. Future designs should include multi-layered support systems, such as preset templates for quick selection, paired with gradual adjustments during interactions. Additionally, analyzing user-uploaded medical records or interaction histories to generate tailored initial settings could balance usability for low-tech users with flexibility for more experienced ones [43].

While the current hybrid chat-and-form model improves usability (Section 5.2.4), it could better support human-like interaction and natural expression. Voice interaction offers potential advantages: it allows freer expression than text and conveys emotional cues like tone, helping models interpret users' states more effectively [48].

A further challenge is achieving a natural interaction rhythm. As highlighted in Section 5.2.1, trust develops gradually, and overly rapid personalization risks disrupting this progression [21], particularly for older users [115]. Accelerated customization may break the natural flow, hindering long-term trust. Future designs should adopt incremental personalization strategies to better align with the natural pace of trust-building.

6.2.2 Balancing user autonomy and ethical safeguards. This study (Section 5.2.3) highlights that while safety mechanisms protect the mental health of users with depression, they may also challenge their sense of autonomy. For instance, the preference part requires caregiver approval to mitigate potential behaviors like exaggerated expressions of distress [28]. However, this design made some users feel their rights were overlooked. Prior research indicates that individuals with depression are often stigmatized as "weak" or "incapable" [19] and excessive intervention risks reinforcing this stereotype, exacerbating discomfort.

Building on these observations, we propose a dynamic weighting mechanism as a future design direction for preference part modules. Specifically, both users and caregivers would contribute input to the preference configuration, while the system would implicitly adjust the relative weight of each party's input based on factors such as the user's clinical stability, behavioral history, and linguistic coherence. For instance, if a user is assessed to be in a stable condition with no signs of emotional escalation, their input could receive a higher adoption ratio; conversely, caregiver input would carry more weight in contexts suggesting elevated risk. This backend weighting strategy maintains necessary safeguards while avoiding the explicit display of hierarchical control, thereby better balancing user dignity, expressive autonomy, and ethical system governance.

Furthermore, keyword-triggered mechanisms remain inadequate in capturing the semantic complexity of language associated with suicide risk. Expressions like "I want to see my mother again", particularly in the context of bereavement, may implicitly signal risk but are often overlooked due to their non-obvious lexical form. Conversely, seemingly direct phrases may trigger false positives despite being emotionally neutral in context—thereby exacerbating users' psychological burden. These limitations stem from static, decontextualized word matching that lacks situational awareness.

To address this, future systems should integrate historical dialogue and user behavioral data to enable dynamic intent recognition rather than relying on isolated keyword detection. Incorporating multimodal signals—such as heart rate variability or electrodermal activity from wearable devices—could also serve as cross-modal validation for linguistic input. Additionally, when caregivers configure the protective part, systems should provide scenario-based guidance (e.g., "Has the user experienced a recent bereavement?") to support more targeted risk profiling. Such mechanisms not only improve detection accuracy but also mitigate users' perception of being monitored.

6.3 Reflections and Prospects on Personalization in Mental Health LLMs

This study, alongside prior research, underscores the potential of large language models (LLMs) in mental health. LLMs provide innovative ways to overcome geographic, economic, and stigma-related barriers, creating opportunities for personalized depression care [22, 33]. In this study, we focused specifically on the role of persona customization in shaping user trust and experience. Our findings show that well-designed LLMs enhance user trust and engagement in mental health scenarios.

While the findings of this study demonstrate that persona customization can enhance user trust (Section 5.1.1), extend the duration of human-AI interactions (Section 5.1.2), and improve overall user experience (Section 5.2), these indicators primarily reflect user engagement rather than direct therapeutic outcomes. The Affective Use Study released by OpenAI similarly revealed that most interactions with affective chatbots center around lightweight social needs—such as casual conversation or companionship—rather than structured psychological intervention [79]. Our qualitative interviews echoed this trend: several participants deviated from emotionally expressive dialogue, perceiving the system more as a daily companion than as a tool for mental health support (Section 5.1.2). These findings caution against equating enhanced engagement through personalization with clinically meaningful improvements in psychological well-being.

Given this observation, we argue that the therapeutic potential of persona customization requires further empirical validation. Future studies should extend evaluation to longer-term deployment in naturalistic settings, incorporating standardized mental health scales and behavioral metrics to systematically assess the intervention's efficacy and user trajectory over time. However, even with positive outcomes in more ecologically valid contexts, personalization alone cannot overcome the structural limitations inherent to mental health technology. Psychological conditions are highly individualized and dynamically evolving, making them ill-suited to one-dimensional adaptation strategies. Prior research has shown that while affective chatbots may provide short-term emotional comfort, they also risk eliciting artificial emotional attachment, disrupting real-life support networks, or exacerbating feelings of loneliness and dependency [114].

Given these challenges, future research should move beyond surface-level engagement metrics and adopt evaluation frameworks that align more closely with therapeutic outcomes. This includes incorporating standardized psychological assessments, behavioral health indicators, and longitudinal tracking mechanisms to systematically examine the actual impact of personalization mechanisms on users' mental health trajectories. From a design perspective, personalization should evolve from static user preference settings toward more adaptive, context-aware systems. By integrating historical interaction data, emotional fluctuation patterns, and multimodal signals—such as vocal tone, physiological responses, and data from wearable devices—systems can continuously optimize personalization in response to users' dynamic mental states. In parallel, ethical safeguards must be reinforced to prevent users from developing excessive emotional dependence on chatbots. Systems should provide clear identity signaling and user education to mitigate anthropomorphic misunderstandings, while also implementing adjustable intervention boundaries and configurable levels of emotional engagement. Personalization should serve to enhance the system's assistive value—not to blur the line between AI and human support networks.

6.4 Limitations and Future Work

This study examined the impact of role customization on user trust and the user experience in depression care, though several limitations warrant further exploration. First, the sample size was relatively small and drawn from specific cultural and age groups, potentially limiting the generalizability of the findings. Future research should broaden the participant base to include more diverse cultural and age demographics, enhancing the applicability of the results. Additionally, the short-term experimental design did not capture the long-term effects of role customization on sustained trust and interaction behaviors. Longitudinal studies are needed to evaluate its utility over extended periods and provide a deeper understanding of its dynamic influence on user experience. Moreover, while customization led to improved trust scores and longer engagement in this short-term setting, our findings do not imply that personalized chatbots are inherently superior in addressing mental health needs. Some users reported that increased engagement stemmed from curiosity or surface-level enjoyment rather than genuine therapeutic value. Future work should further examine how personalization affects deeper psychological outcomes, beyond interaction metrics.

Second, the current interaction design for role customization leaves room for improvement. This study primarily employed a combination of forms and dialogues for role creation, without exploring the potential of multimodal interactions. Future work could integrate multimodal sensing technologies, such as emotion recognition, vocal cues, and biosensor data, to better capture users' emotional states and enhance the accuracy and naturalness of role customization. Furthermore, privacy protection and ethical considerations require more attention. Specifically, data collection and sensitive word monitoring mechanisms should prioritize transparency and user control to mitigate privacy concerns and build trust. It is also worth noting that no sensitive word alerts were triggered during our experiment. Participants reported restrained emotional expression in the lab setting, and some caregivers found it difficult to predefine meaningful risk-related expressions. These limitations suggest that future research should investigate sensitive word detection in real-life, longitudinal contexts where emotional disclosure is more spontaneous and nuanced.

Finally, localized adaptation is crucial for developing mental health chatbots. Although this study was designed with a focus on the Chinese context, the impact of cultural adaptability on user role preferences and interaction behaviors remains underexplored. Future research should investigate cultural differences in user needs and use these insights to refine designs for greater inclusivity and adaptability, thereby maximizing the applicability of LLMs in mental health scenarios.

7 Conclusion

This study explored the potential of the role customization feature's potential in depression care, focusing on its impact on user trust and the customization process experience. Using the CloudEcho system, role customization allowed users to personalize chatbot settings, addressing diverse needs. Results showed significant trust improvements, particularly in reciprocity, competence, and trust dimensions, driven by enhanced control, adaptability, and transparency, fostering emotional connection and security. However, concerns about privacy risks and viewing chatbots as limited to managing negative emotions were noted.

The customization process was seen as efficient, with high usability and ease of learning scores. Still, some users struggled to achieve desired results due to unclear needs or a lack of preview features. Satisfaction with role types varied: emotionally supportive roles were favored for their empathy, while professionally oriented roles lacked emotional connection. This highlights the need to balance expertise and emotional resonance in role design for mental health scenarios.

Taken together, these findings suggest the promise—but also the limitations—of chatbot role customization in supporting trust and user engagement in sensitive care contexts. Future designs should offer diverse role templates, enhance real-time adjustments, and integrate multimodal data to improve the customization experience. This study provides empirical evidence for conversational AI in mental health and offers insights for advancing personalized human-AI interactions.

Acknowledgments

We extend our gratitude to the experimental participants for their valuable insights and to the anonymous reviewers for their substantial contributions in enhancing this paper.

References

[1] 2019. Use of Social Robots in Mental Health and Well-Being Research: Systematic Review. *Journal of Medical Internet Research* 21, 7 (July 2019). doi:10.2196/13322

[2] 2021. Making conversations with chatbots more personalized. *Computers in Human Behavior* 117 (April 2021), 106627. doi:10.1016/j.chb.2020.106627 Publisher: Pergamon.

[3] Alaa A. Abd-Alrazaq, Mohannad Alajlani, Nashva Ali, Kerstin Denecke, Bridgette M. Bewick, and Mowafa Househ. 2021. Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review. *Journal of Medical Internet Research* 23, 1 (Jan. 2021), e17828. doi:10.2196/17828

[4] Saleh Afroogh, Ali Akbari, Emmie Malone, Mohammadali Kargar, and Hananeh Alambeigi. 2024. Trust in AI: progress, challenges, and future directions. *Humanities and Social Sciences Communications* 11, 1 (Nov. 2024), 1–30. doi:10.1057/s41599-024-04044-8 Publisher: Palgrave.

[5] Rangina Ahmad, Dominik Siemon, Ulrich Gnewuch, and Susanne Robra-Bissantz. 2022. Designing Personality-Adaptive Conversational Agents for Mental Health Care. *Information Systems Frontiers* 24, 3 (June 2022), 923–943. doi:10.1007/s10796-022-10254-9

[6] Ahmed M Alhuwaydi. 2024. Understanding the conflicting demands of family caregivers caring for depressed family members. *Risk Management and Healthcare Policy* 17 (Dec. 2024), 1339–1348. doi:10.2147/RMHP.S461562 Publisher: Dove Medical Press _eprint: https://www.tandfonline.com/doi/pdf/10.2147/RMHP.S461562.

[7] Muhammad Asif and John Krogstie. 2012. Taxonomy of personalization in mobile services. 343–350. https://iadisportal.org/digital-library/taxonomy-of-personalization-in-mobile-services

[8] Tita Alissa Bach, Amna Khan, Harry Hallock, Gabriela Beltrão, and Sonia Sousa. 2024. A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective. *International Journal of Human–Computer Interaction* 40, 5 (March 2024), 1251–1266. doi:10.1080/10447318.2022.2138826

[9] Petter Bae Bae Brandtzæg, Marita Skjuve, Kim Kristoffer Kristoffer Dysthe, and Asbjørn Følstad. 2021. When the Social Becomes Non-Human: Young People's Perception of Social Support in Chatbots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3411764.3445318

[10] Austin Beattie, Autumn P. Edwards, and Chad Edwards. 2020. A Bot and a Smile: Interpersonal Impressions of Chatbots and Humans Using Emoji in Computer-mediated Communication. *Communication Studies* 71, 3 (May 2020), 409–427. doi:10.1080/10510974.2020.1725082 Publisher: Routledge _eprint: https://doi.org/10.1080/10510974.2020.1725082.

[11] Izak Benbasat and Weiquan Wang. 2005. Trust In and Adoption of Online Recommendation Agents. *Journal of the Association for Information Systems* 6, 3 (March 2005). doi:10.17705/1jais.00065

[12] Jackylyn L. Beredo and Ethel C. Ong. 2022. A Hybrid Response Generation Model for an Empathetic Conversational Agent. In *2022 International Conference on Asian Language Processing (IALP)*. 300–305. doi:10.1109/IALP57159.2022.9961311

[13] Shalini Bhatia. 2016. Multimodal sensing of affect intensity. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16)*. Association for Computing Machinery, New York, NY, USA, 567–571. doi:10.1145/2993148.2997622

[14] Timothy W. Bickmore and Rosalind W. Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.* 12, 2 (June 2005), 293–327. doi:10.1145/1067860.1067867

[15] James F. Boswell, Michael J. Constantino, Alice E. Coyne, and David R. Kraus. 2022. For whom does a match matter most? Patient-level moderators of evidence-based patient–therapist matching. *Journal of Consulting and Clinical Psychology* 90, 1 (Jan. 2022), 61–74. doi:10.1037/ccp0000644

[16] Dionne Bowie-DaBreo, Corina Sas, Heather Iles-Smith, and Sandra Sünram-Lea. 2022. User Perspectives and Ethical Experiences of Apps for Depression: A Qualitative Analysis of User Reviews. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–24. doi:10.1145/3491102.3517498

[17] Kyle Boyd, Courtney Potts, Raymond Bond, Maurice Mulvenna, Thomas Broderick, Con Burns, Andrea Bickerdike, Mike Mctear, Catrine Kostenius, Alex Vakaloudis, Indika Dhanapala, Edel Ennis, and Fred Booth. 2022. Usability testing and trust analysis of a mental health and wellbeing chatbot. In *Proceedings of the 33rd European Conference on Cognitive Ergonomics (ECCE '22)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3552327.3552348

[18] Paulina Chametka, Sana Maqsood, and Sonia Chiasson. 2023. Security and Privacy Perceptions of Mental Health Chatbots. In *2023 20th Annual International Conference on Privacy, Security and Trust (PST)*. 1–7. doi:10.1109/PST58708.2023.10320174 ISSN: 2643-4202.

[19] Ana Paula Chaves and Marco Aurelio Gerosa. 2021. How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. *International Journal of Human–Computer Interaction* 37, 8 (May 2021), 729–758. doi:10.1080/10447318.2020.1841438 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2020.1841438.

[20] China National Narcotic Drugs Association and Haoxinqing. 2023. *The 2023 Blue Book of China's Mental Health*. Technical Report. Haoxinqing Internet Hospital, Beijing, China. https://www.haoxinqing.cn/article/detail31087.html

[21] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300705

[22] Marco Colizzi, Antonio Lasalvia, and Mirella Ruggeri. 2020. Prevention and early intervention in youth mental health: is it time for a multidisciplinary and trans-diagnostic model for care? *International Journal of Mental Health Systems* 14, 1 (March 2020), 23. doi:10.1186/s13033-020-00356-9

[23] Cynthia L. Corritore, Beverly Kracher, and Susan Wiedenbeck. 2003. On-line trust: concepts, evolving themes, a model. *International Journal of Human-Computer Studies* 58, 6 (June 2003), 737–758. doi:10.1016/S1071-5819(03)00041-7

[24] Samuel Rhys Cox and Wei Tsang Ooi. 2022. Does Chatbot Language Formality Affect Users' Self-Disclosure?. In *Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3543829.3543831

[25] Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma, and Sandeep Kaushik. 2019. Big data in healthcare: management, analysis and future prospects. *Journal of Big Data* 6, 1 (June 2019), 54. doi:10.1186/s40537-019-0217-0

[26] Julian De Freitas, Ahmet Kaan Uğuralp, Zeliha Oğuz-Uğuralp, and Stefano Puntoni. 2024. Chatbots and mental health: Insights into the safety of generative AI. *Journal of Consumer Psychology* 34, 3 (2024), 481–491. doi:10.1002/jcpy.1393 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcpy.1393.

[27] Kerstin Denecke, Sayan Vaaheesan, and Aaganya Arulnathan. 2021. A Mental Health Chatbot for Regulating Emotions (SERMO) - Concept and Usability Test. *IEEE Transactions on Emerging Topics in Computing* 9, 3 (July 2021), 1170–1182. doi:10.1109/TETC.2020.2974478

[28] Patrícia Deud Guimarães and Vânia Paula De Almeida Neris. 2024. A Tool-Supported Approach to Adapt Web User Interfaces Considering the Emotional State of the User. In *Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems (IHC '23)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3638067.3638115

[29] Danica Dillion, Debanjan Mondal, Niket Tandon, and Kurt Gray. 2024. Large Language Models as Moral Experts? GPT-4o Outperforms Expert Ethicist in Providing Moral Guidance. doi:10.31234/osf.io/w7236

[30] Aaron C. Elkins and Douglas C. Derrick. 2013. The Sound of Trust: Voice as a Measurement of Trust During Interactions with Embodied Conversational Agents. *Group Decision and Negotiation* 22, 5 (Sept. 2013), 897–913. doi:10.1007/s10726-012-9339-x

[31] Ahmed Fadhil, Gianluca Schiavo, Yunlong Wang, and Bereket A. Yilma. 2018. The Effect of Emojis when interacting with Conversational Interface Assisted Health Coaching System. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '18)*. Association for Computing Machinery, New York, NY, USA, 378–383. doi:10.1145/3240925.3240965

[32] Ian H. Gotlib. 1992. Interpersonal and cognitive aspects of depression. *Current Directions in Psychological Science* 1, 5 (1992), 149–154. doi:10.1111/1467-8721.ep11510319 Place: United Kingdom Publisher: Blackwell Publishing.

[33] J. P. Grodniewicz and Mateusz Hohol. 2023. Waiting for a digital therapist: three challenges on the path to psychotherapy delivered by artificial intelligence. *Frontiers in Psychiatry* 14 (June 2023). doi:10.3389/fpsyt.2023.1190084 Publisher: Frontiers.

[34] Siddharth Gulati, Sonia Sousa, and David Lamas. 2019. Design, development and evaluation of a human-computer trust scale. *Behaviour & Information Technology* 38 (Aug. 2019), 1–12. doi:10.1080/0144929X.2019.1656779

[35] Xitong Guo, Xiaofei Zhang, and Yongqiang Sun. 2016. The privacy–personalization paradox in mHealth services acceptance of different age groups. *Electronic Commerce Research and Applications* 16 (March 2016), 55–65. doi:10.1016/j.elerap.2015.11.001

[36] Juhye Ha, Hyeon Jeon, Daeun Han, Jinwook Seo, and Changhoon Oh. 2024. CloChat: Understanding How People Customize, Interact, and Experience Personas in Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–24. doi:10.1145/3613904.3642472

[37] Chen-Wei Hsieh and Sherry Chen. 2016. A Cognitive Style Perspective to Handheld Devices: Customization vs. Personalization. *The International Review of Research in Open and Distributed Learning* 17 (Feb. 2016). doi:10.19173/irrodl.v17i1.2168

[38] Hyu Jung Huh, Kwangyeol Baek, Jae-Hyung Kwon, Jaeseung Jeong, and Jeong-Ho Chae. 2016. Impact of childhood trauma and cognitive emotion regulation strategies on risk-aversive and loss-aversive patterns of decision-making in patients with depression. *Cognitive Neuropsychiatry* 21, 6 (Nov. 2016), 447–461. doi:10.1080/13546805.2016.1230053 Publisher: Routledge _eprint: https://doi.org/10.1080/13546805.2016.1230053.

[39] Alison KY Hui, Paul WC Wong, and King-wa Fu. 2014. Building a model for encouraging help-seeking for depression: a qualitative study in a Chinese society. *BMC Psychology* 2, 1 (April 2014), 9. doi:10.1186/2050-7283-2-9

[40] Thomas E. Joiner Jr. and Jennifer Katz. 1999. Contagion of depressive symptoms and mood: Meta-analytic review and explanations from cognitive, behavioral, and interpersonal viewpoints. *Clinical Psychology: Science and Practice* 6, 2 (1999), 149–164. doi:10.1093/clipsy.6.2.149 Place: United Kingdom Publisher: Blackwell Publishing.

[41] Inchan Jung, Hankyung Kim, and Youn-kyung Lim. 2021. Understanding How Users Experience the Physiological Expression of Non-humanoid Voice-based Conversational Agent in Healthcare Services. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference (DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1433–1446. doi:10.1145/3461778.3462082

[42] Taewan Kim, Seolyeong Bae, Hyun Ah Kim, Su-woo Lee, Hwajung Hong, Chanmo Yang, and Young-Ho Kim. 2024. MindfulDiary: Harnessing Large Language Model to Support Psychiatric Patients' Journaling. doi:10.1145/3613904.3642937 arXiv:2310.05231 [cs].

[43] Bart P. Knijnenburg, Martijn C. Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4 (Oct. 2012), 441–504. doi:10.1007/s11257-011-9118-4

[44] Naja Kathrine Kollerup, Joel Wester, Mikael B. Skov, and Niels van Berkel. 2024. How Can I Signal You To Trust Me: Investigating AI Trust Signalling in Clinical Self-Assessments. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 525–540. doi:10.1145/3643834.3661612

[45] Sherrie Komiak and Izak Benbasat. 2006. The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. *MIS Quarterly* 30 (Dec. 2006), 941–960. doi:10.2307/25148760

[46] Brage Kraft, Ragnhild Bø, Rune Jonassen, Alexandre Heeren, Vidar Sandsaunet Ulset, Tore C. Stiles, and Nils Inge Landrø. 2023. The association between depression symptoms and reduced executive functioning is primarily linked by fatigue. *Psychiatry Research Communications* 3, 2 (June 2023), 100120. doi:10.1016/j.psycom.2023.100120

[47] Yogesh Kumar, Apeksha Koul, Ruchi Singla, and Muhammad Fazal Ijaz. 2023. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *Journal of Ambient Intelligence and Humanized Computing* 14, 7 (2023), 8459–8486. doi:10.1007/s12652-021-03612-z

[48] Craig Kuziemsky, Anthony J. Maeder, Oommen John, Shashi B. Gogia, Arindam Basu, Sushil Meher, and Marcia Ito. 2019. Role of Artificial Intelligence within the Telehealth Domain. *Yearbook of Medical Informatics* 28, 1 (Aug. 2019), 35–40. doi:10.1055/s-0039-1677897

[49] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie Y S Lau, and Enrico Coiera. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association* 25, 9 (Sept. 2018), 1248–1258. doi:10.1093/jamia/ocy072

[50] Ellen E. Lee, John Torous, Munmun De Choudhury, Colin A. Depp, Sarah A. Graham, Ho-Cheol Kim, Martin P. Paulus, John H. Krystal, and Dilip V. Jeste. 2021. Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 6, 9 (Sept. 2021), 856–864. doi:10.1016/j.bpsc.2021.02.001

[51] Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1 (May 2020), 31:1–31:27. doi:10.1145/3392836

[52] Nadine Lessio and Alexis Morris. 2020. Toward Design Archetypes for Conversational Agent Personality. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 3221–3228. doi:10.1109/SMC42975.2020.9283254 ISSN: 2577-1655.

[53] Roy J. Lewicki and Barbara Benedict Bunker. 1996. Developing and Maintaining Trust in Work Relationships. In *TRUST IN ORGANIZATIONS*. SAGE Publications, Inc, United States, 114–139. doi:10.4135/9781452243610.n7

[54] Brenna Li, Ofek Gross, Noah Crampton, Mamta Kapoor, Saba Tauseef, Mohit Jain, Khai N. Truong, and Alex Mariakakis. 2024. Beyond the Waiting Room: Patient's Perspectives on the Conversational Nuances of Pre-Consultation Chatbots. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–24. doi:10.1145/3613904.3641913

[55] Han Li. 2023. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. (2023).

[56] Shi Min Lim, Chyi Wey Claudine Shiau, Ling Jie Cheng, and Ying Lau. 2022. Chatbot-Delivered Psychotherapy for Adults With Depressive and Anxiety Symptoms: A Systematic Review and Meta-Regression. *Behavior Therapy* 53, 2 (March 2022), 334–347. doi:10.1016/j.beth.2021.09.007

[57] Kaifeng Liu and Da Tao. 2022. The roles of trust, personalization, loss of privacy, and anthropomorphism in public acceptance of smart healthcare services. *Computers in Human Behavior* 127 (Feb. 2022), 107026. doi:10.1016/j.chb.2021.107026

[58] Xiao-Xin Liu, Cheng-Yue Yin, and Meng-Ran Li. 2024. The power of voice! The impact of robot receptionists' voice pitch and communication style on customer value cocreation intention. *International Journal of Hospitality Management* 122 (Sept. 2024), 103819. doi:10.1016/j.ijhm.2024.103819

[59] Arnold Lund. 2001. Measuring Usability with the USE Questionnaire. *Usability and User Experience Newsletter of the STC Usability SIG* 8 (Jan. 2001).

[60] Diana Löffler, Lennart Giron, and Jörn Hurtienne. 2017. Night Mode, Dark Thoughts: Background Color Influences the Perceived Sentiment of Chat Messages. In *Human-Computer Interaction - INTERACT 2017*, Regina Bernhaupt, Girish Dalvi, Anirudha Joshi, Devanuj K. Balkrishan, Jacki O'Neill, and Marco Winckler (Eds.). Springer International Publishing, Cham, 184–201. doi:10.1007/978-3-319-67684-5_12

[61] Bethanie Maples, Merve Cerit, Aditya Vishwanath, and Roy Pea. 2024. Loneliness and suicide mitigation for students using GPT3-enabled chatbots. *NPJ Mental Health Research* 3 (Jan. 2024), 4. doi:10.1038/s44184-023-00047-6

[62] Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An Integrative Model of Organizational Trust. *The Academy of Management Review* 20, 3 (1995), 709–734. doi:10.2307/258792 Publisher: Academy of Management.

[63] Marilyn Rose McGee-Lennon, Maria Klara Wolters, and Stephen Brewster. 2011. User-centred multimodal reminders for assistive living. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2105–2114. doi:10.1145/1978942.1979248

[64] Roger S. McIntyre, Danielle S. Cha, Joanna K. Soczynska, Hanna O. Woldeyohannes, Laura Ashley Gallaugher, Paul Kudlow, Mohammad Alsuwaidan, and Anusha Baskaran. 2013. Cognitive Deficits and Functional Outcomes in Major Depressive Disorder: Determinants, Substrates, and Treatment Interventions. *Depression and Anxiety* 30, 6 (2013), 515–527. doi:10.1002/da.22063 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/da.22063.

[65] Meela. [n. d.]. Meela - Match with your best-fit therapist. https://meelahealth.com/uk/en/

[66] Ana Paula Souto Melo, Ilse N. Dippenaar, Sarah Charlotte Johnson, Nicole Davis Weaver, Francisco de Assis Acurcio, Deborah Carvalho Malta, Antônio Luiz P. Ribeiro, Augusto Afonso Guerra Júnior, Eve E. Wool, Mohsen Naghavi, and Mariangela Leal Cherchiglia. 2022. All-cause and cause-specific mortality among people with severe mental illness in Brazil's public health system, 2000–15: a retrospective study. *The Lancet Psychiatry* 9, 10 (Oct. 2022), 771–781. doi:10.1016/S2215-0366(22)00237-1 Publisher: Elsevier.

[67] Microsoft. 2023. Use guidelines for human-AI interaction. https://learn.microsoft.com/en-us/training/modules/introduction-to-microsofts-responsible-ai-approach/3-use-guidelines-for-human-ai-interaction

[68] Bamshad Mobasher, Robert Cooley, and Jaideep Srivastava. 2000. Automatic personalization based on Web usage mining. *Commun. ACM* 43, 8 (Aug. 2000), 142–151. doi:10.1145/345124.345169

[69] Syed Naveed Mohsin, Abubakar Gapizov, Chukwuyem Ekhator, Noor U Ain, Saeed Ahmad, Mavra Khan, Chad Barker, Muqaddas Hussain, Jahnavi Malineni, Afif Ramadhan, and Raghu Halappa Nagaraj. [n. d.]. The Role of Artificial Intelligence in Prediction, Risk Stratification, and Personalized Treatment Planning for Congenital Heart Diseases. *Cureus* 15, 8 ([n. d.]), e44374. doi:10.7759/cureus.44374

[70] Joonas Moilanen, Aku Visuri, Sharadhi Alape Suryanarayana, Andy Alorwu, Koji Yatani, and Simo Hosio. 2022. Measuring the Effect of Mental Health Chatbot Personality on User Engagement. In *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia (MUM '22)*. Association for

Computing Machinery, New York, NY, USA, 138–150. doi:10.1145/3568444.3568464
[71] Sara Moussawi and Raquel Benbunan-Fich. 2021. The effect of voice and humour on users' perceptions of personal intelligent agents. *Behaviour & Information Technology* 40, 15 (Nov. 2021), 1603–1626. doi:10.1080/0144929X.2020.1772368 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/0144929X.2020.1772368.
[72] Bonnie M. Muir and Neville Moray. 1996. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39, 3 (March 1996), 429–460. doi:10.1080/00140139608964474
[73] Marcia Nißen, Dominik Rüegger, Mirjam Stieger, Christoph Flückiger, Mathias Allemand, Florian V Wangenheim, and Tobias Kowatsch. 2022. The Effects of Health Care Chatbot Personas With Different Social Roles on the Client-Chatbot Bond and Usage Intentions: Development of a Design Codebook and Web-Based Study. *Journal of Medical Internet Research* 24, 4 (April 2022), e32630. doi:10.2196/32630
[74] Cecilie Bertinussen Nordheim, Asbjørn Følstad, and Cato Alexander Bjørkli. 2019. An Initial Model of Trust in Chatbots for Customer Service—Findings from a Questionnaire Study. *Interacting with Computers* 31, 3 (May 2019), 317–335. doi:10.1093/iwc/iwz022
[75] Shantanu Nundy, Tara Montgomery, and Robert M. Wachter. 2019. Promoting Trust Between Patients and Physicians in the Era of Artificial Intelligence. *JAMA* 322, 6 (Aug. 2019), 497–498. doi:10.1001/jama.2018.20563
[76] David B. Olawade, Ojima Z. Wada, Aderonke Odetayo, Aanuoluwapo Clement David-Olawade, Fiyinfoluwa Asaolu, and Judith Eberhardt. 2024. Enhancing mental health with Artificial Intelligence: Current trends and future prospects. *Journal of Medicine, Surgery, and Public Health* 3 (Aug. 2024), 100099. doi:10.1016/j.glmedi.2024.100099
[77] Batyrkhan Omarov, Zhandos Zhumanov, Aidana Gumar, and Leilya Kuntunova. 2023. Artificial Intelligence Enabled Mobile Chatbot Psychologist using AIML and Cognitive Behavioral Therapy. *International Journal of Advanced Computer Science and Applications* 14, 6 (2023). doi:10.14569/IJACSA.2023.0140616
[78] OpenAI. 2023. Custom Instructions for ChatGPT.
[79] OpenAI and MIT Media Lab Research. 2025. Early methods for studying affective use and emotional well-being on ChatGPT. https://openai.com/index/affective-use-study/
[80] Laura Orsolini, Roberto Latini, Maurizio Pompili, Gianluca Serafini, Umberto Volpe, Federica Vellante, Michele Fornaro, Alessandro Valchera, Carmine Tomasetti, Silvia Fraticelli, Marco Alessandrini, Raffaella La Rovere, Sabatino Trotta, Giovanni Martinotti, Massimo Di Giannantonio, and Domenico De Berardis. 2020. Understanding the Complex of Suicide in Depression: from Research to Clinics. *Psychiatry Investigation* 17, 3 (March 2020), 207–221. doi:10.30773/pi.2019.0171
[81] SoHyun Park, Anja Thieme, Jeongyun Han, Sungwoo Lee, Wonjong Rhee, and Bongwon Suh. 2021. "I wrote as if I were telling a story to someone I knew.": Designing Chatbot Interactions for Expressive Writing in Mental Health. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference (DIS '21)*. Association for Computing Machinery, New York, NY, USA, 926–941. doi:10.1145/3461778.3462143
[82] Danika Passler Bates, Skyla Y. Dudek, James M. Berzuk, Adriana Lorena González, and James E. Young. 2024. SnuggleBot the Companion: Exploring In-Home Robot Interaction Strategies to Support Coping With Loneliness. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 2972–2986. doi:10.1145/3643834.3660702
[83] Adrianos Pavlopoulos, Theodoros Rachiotis, and Ilias Maglogiannis. 2024. An Overview of Tools and Technologies for Anxiety and Depression Management Using AI. *Applied Sciences* 14, 19 (Jan. 2024), 9068. doi:10.3390/app14199068 Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.
[84] Alisha Pradhan and Amanda Lazar. 2021. Hey Google, Do You Have a Personality? Designing Personality and Personas for Conversational Agents. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*. ACM, Bilbao (online) Spain, 1–4. doi:10.1145/3469595.3469607
[85] Hongyi Qin, Yifan Zhu, Yan Jiang, Siqi Luo, and Cui Huang. 2024. Examining the impact of personalization and carefulness in AI-generated health advice: Trust, adoption, and insights in online healthcare consultations experiments. *Technology in Society* 79 (Dec. 2024), 102726. doi:10.1016/j.techsoc.2024.102726
[86] Hua Xuan Qin, Shan Jin, Ze Gao, Mingming Fan, and Pan Hui. 2024. CharacterMeet: Supporting Creative Writers' Entire Story Character Construction Processes Through Conversation with LLM-Powered Chatbot Avatars. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3613904.3642105
[87] Boele Raad. 2000. *The Big Five Personality Factors: The psycholexical approach to personality.*
[88] Peter A. M. Ruijten. 2021. The similarity-attraction paradigm in persuasive technology: effects of system and user personality on evaluations and persuasiveness of an interactive system. *Behaviour & Information Technology* 40, 8 (June 2021), 734–746. doi:10.1080/0144929X.2020.1723701
[89] Seyed Saadati and Seyed Saadati. 2023. The Role of Chatbots in Mental Health Interventions: User Experiences. *AI and Tech in Behavioral and Social Sciences* 1 (April 2023), 19–25. doi:10.61838/kman.aitech.1.2.4
[90] Hamid Reza Saeidnia, Seyed Ghasem Hashemi Fotami, Brady Lund, and Nasrin Ghiasi. 2024. Ethical Considerations in Artificial Intelligence Interventions for Mental Health and Well-Being: Ensuring Responsible Implementation and Impact. *Social Sciences* 13, 7 (July 2024), 381. doi:10.3390/socsci13070381 Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.
[91] Mohammed Salah, Hussam Alhalbusi, Maria Mohd Ismail, and Fadi Abdelfattah. 2024. Chatting with ChatGPT: decoding the mind of Chatbot users and unveiling the intricate connections between user perception, trust and stereotype perception on self-esteem and psychological well-being. *Current Psychology* 43, 9 (March 2024), 7843–7858. doi:10.1007/s12144-023-04989-0
[92] Pedro Sanches, Axel Janson, Pavel Karpashevich, Camille Nadal, Chengcheng Qu, Claudia Daudén Roquet, Muhammad Umair, Charles Windlin, Gavin Doherty, Kristina Höök, and Corina Sas. 2019. HCI and Affective Health: Taking stock of a decade of studies and charting future research directions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3290605.3300475
[93] Subhro Sarkar, Sumedha Chauhan, and Arpita Khare. 2020. A meta-analysis of antecedents and consequences of trust in mobile commerce. *International Journal of Information Management* 50 (Feb. 2020), 286–301. doi:10.1016/j.ijinfomgt.2019.08.008
[94] Yuto Sawa, Julia Keckeis, and Katie Seaborn. 2023. Right for the Job or Opposites Attract? Exploring Cross-Generational User Experiences with "Younger" and "Older" Voice Assistants. In *Companion Publication of the 2023 ACM Designing Interactive Systems Conference (DIS '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 160–163. doi:10.1145/3563703.3596642
[95] Shekhar Saxena, Graham Thornicroft, Martin Knapp, and Harvey Whiteford. 2007. Resources for mental health: scarcity, inequity, and inefficiency. *The Lancet* 370, 9590 (Sept. 2007), 878–889. doi:10.1016/S0140-6736(07)61239-2 Publisher: Elsevier.
[96] R. Sham Sundhar, T. Shivavardhini, C. A. Daphine Desona Clemency, and M. S. Roobini. 2024. Lecter - A Large Language Model Chatbot for Cognitive Behavioral Therapy. In *Computational Intelligence in Data Science*, Mieczyslaw Lech Owoc, Felix Enigo Varghese Sicily, Kanchana Rajaram, and Prabavathy Balasundaram (Eds.). Springer Nature Switzerland, Cham, 41–53. doi:10.1007/978-3-031-69982-5_4
[97] Ruoxi Shang, Gary Hsieh, and Chirag Shah. 2025. Trusting Your AI Agent Emotionally and Cognitively: Development and Validation of a Semantic Differential Scale for AI Trust. In *Proceedings of the 2024 AAAI/ACM Conference on AI, Ethics, and Society (AIES '24)*. AAAI Press, San Jose, California, USA, 1343–1356.
[98] Wai Man Si, Michael Backes, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, Savvas Zannettou, and Yang Zhang. 2022. Why So Toxic? Measuring and Triggering Toxic Behavior in Open-Domain Chatbots. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22)*. Association for Computing Machinery, New York, NY, USA, 2659–2673. doi:10.1145/3548606.3560599
[99] Kellie Yu Hui Sim, Kohleen Tijing Fortuno, and Kenny Tsu Wei Choo. 2024. Towards Understanding Emotions for Engaged Mental Health Conversations. In *Companion Publication of the 2024 ACM Designing Interactive Systems Conference (DIS '24 Companion)*. Association for Computing Machinery, New York, NY, USA, 176–180. doi:10.1145/3656156.3663694
[100] Logan Stapleton, Sunniva Liu, Cindy Liu, Irene Hong, Stevie Chancellor, Robert E Kraut, and Haiyi Zhu. 2024. "If This Person is Suicidal, What Do I Do?": Designing Computational Approaches to Help Online Volunteers Respond to Suicidality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–21. doi:10.1145/3613904.3641922
[101] Xin Sun, Isabelle Teljeur, Zhuying Li, and Jos A. Bosch. 2024. Can a Funny Chatbot Make a Difference? Infusing Humor into Conversational Agent for Behavioral Intervention. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (CUI '24)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3640794.3665555
[102] Ekaterina Svikhnushina, Alexandru Placinta, and Pearl Pu. 2021. User Expectations of Conversational Chatbots Based on Online Reviews. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference (DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1481–1491. doi:10.1145/3461778.3462125
[103] Silke ter Stal, Monique Tabak, Harm op den Akker, Tessa Beinema, and Hermie Hermens. 2020. Who Do You Prefer? The Effect of Age, Gender and Role on Users' First Impressions of Embodied Conversational Agents in eHealth. *International Journal of Human–Computer Interaction* 36, 9 (May 2020), 881–892. doi:10.1080/10447318.2019.1699744 Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/10447318.2019.1699744.

[104] Wensheng Tian, Yifan Lu, Jinhao Yu, Jiafeng Fan, Panpan Tang, and Lei Zhang. 2022. A Privacy-Preserving Framework for Mental Health Chatbots Based on Confidential Computing. In *2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta)*. 1119–1124. doi:10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00160

[105] Indrit Troshani, Sally Rao Hill, Claire Sherman, and Damien Arthur. 2021. Do We Trust in AI? Role of Anthropomorphism and Intelligence. *Journal of Computer Information Systems* 61, 5 (Sept. 2021), 481–491. doi:10.1080/08874417.2020.1788473 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/08874417.2020.1788473.

[106] Use of Social Robots in Mental Health and Well-Being Research:, Systematic Review., and Luka, Inc. [n. d.]. Replika. https://replika.com/

[107] Usman Ahmad Usmani, Ari Happonen, and Junzo Watada. 2023. Human-Centered Artificial Intelligence: Designing for User Empowerment and Ethical Considerations. In *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. 1–7. doi:10.1109/HORA58378.2023.10156761

[108] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R. Cowan, and Heinrich Hussmann. 2021. Eliciting and Analysing Users' Envisioned Dialogues with Perfect Voice Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–15. doi:10.1145/3411764.3445536

[109] Sarah Theres Völkel and Lale Kaya. 2021. Examining User Preference for Agreeableness in Chatbots. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*. ACM, Bilbao (online) Spain, 1–6. doi:10.1145/3469595.3469633

[110] Jifei Wang, Zhenping Zhao, Jing Yang, Limin Wang, Mei Zhang, and Maigeng Zhou. 2024. The Association Between Depression and All-Cause, Cause-Specific Mortality in the Chinese Population — China,2010–2022. *China CDC Weekly* 6, 40 (Oct. 2024). doi:10.46234/ccdcw2024.212

[111] Shihan Wang, Chao Zhang, Ben Kröse, and Herke van Hoof. 2021. Optimizing Adaptive Notifications in Mobile Health Interventions Systems: Reinforcement Learning from a Data-driven Behavioral Simulator. *Journal of Medical Systems* 45, 12 (Oct. 2021), 102. doi:10.1007/s10916-021-01773-0

[112] Weiquan Wang, Lingyun Qiu, Dongmin Kim, and Izak Benbasat. 2016. Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. *Decision Support Systems* 86 (June 2016), 48–60. doi:10.1016/j.dss.2016.03.007

[113] Ruoyu Wen, Stephanie Elena Crowe, Kunal Gupta, Xinyue Li, Mark Billinghurst, Simon Hoermann, Dwain Allan, Alaeddin Nassani, and Thammathip Piumsomboon. 2024. Large Language Models for Automatic Detection of Sensitive Topics. doi:10.48550/arXiv.2409.00940 arXiv:2409.00940 [cs].

[114] Joel Wester, Henning Pohl, Simo Hosio, and Niels van Berkel. 2024. "This Chatbot Would Never...": Perceived Moral Agency of Mental Health Chatbots. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 133:1–133:28. doi:10.1145/3637410

[115] Novia Wong, Sooyeon Jeong, Madhu Reddy, Caitlin A. Stamatis, Emily G. Lattie, and Maia Jacobs. 2024. Voice Assistants for Mental Health Services: Designing Dialogues with Homebound Older Adults. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 844–858. doi:10.1145/3643834.3661536

[116] World Health Organization. 2022. *World mental health report: Transforming mental health for all.* Technical Report ISBN 9789240049338. World Health Organization, Geneva, Switzerland. https://www.who.int/publications/i/item/9789240049338

[117] Mengxi Xiao, Qianqian Xie, Ziyan Kuang, Zhicheng Liu, Kailai Yang, Min Peng, Weiguang Han, and Jimin Huang. 2024. HealMe: Harnessing Cognitive Reframing in Large Language Models for Psychotherapy. 1707–1725. doi:10.18653/v1/2024.acl-long.93

[118] Anna Xygkou, Panote Siriaraya, Alexandra Covaci, Holly Gwen Prigerson, Robert Neimeyer, Chee Siang Ang, and Wan-Jou She. 2023. The "Conversation" about Loss: Understanding How Chatbot Technology was Used in Supporting People in Grief.. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3544548.3581154

[119] Fang Yang, Bing Xiang Yang, Teresa E. Stone, Xiao Qin Wang, Yang Zhou, Juan Zhang, and Shu Fen Jiao. 2020. Stigma towards depression in a community-based sample in China. *Comprehensive Psychiatry* 97 (Feb. 2020), 152152. doi:10.1016/j.comppsych.2019.152152

[120] WILLIAM W. K. ZUNG. 1965. A Self-Rating Depression Scale. *Archives of General Psychiatry* 12, 1 (Jan. 1965), 63–70. doi:10.1001/archpsyc.1965.01720310065008

A Customizable Content Framework for Chatbot Personas

In Section 3.2.1, we provide a detailed overview of the classification framework and interactive design examples employed in the chatbot customization process. This framework is grounded in a systematic literature review, which identified nine key categories of customizable role attributes. These categories form the foundation for designing chat-based interactions that align with user needs and preferences. Table 3 summarizes these categories, including their theoretical references and interaction text examples, demonstrating how these concepts are practically implemented.

Table 3: User-customizable content and interactive text in chat mode

User	Part	Category	Interaction Text in Chat Mode	References
User with depression	**Style Part**	Identity Selection	To make our conversation more in tune with your mood, would you like me to accompany you as a friend, mentor, pet, or in any other role you prefer? Cici will be gentle and patient no matter what role you choose!	[73, 86, 118]
		Speech Style	Cici is happy to be your XXX! Next, let's define my speaking style, which will affect the atmosphere of our interaction. Cici's tone will automatically match with the role you just chose, such as being more relaxed when you are a friend or more professional when you are a doctor. Do you want me to be more humorous? Or more gentle? No matter how you choose, it's all about making Cici better for you.	[24, 54, 101, 115]
		Sound Style	Ok, Cici will adapt to your preferences! Do you have any preference for Cici's voice? I can be a calm male voice, a gentle female voice, and a fun child's voice. Of course, Cici also supports voice cloning. You can become any voice you like.	[41, 54, 58, 94]
		Emoji preference	Okay , I hope you'll like Cici's new voice!Cici would also like to know, would you like us to use emoji when we chat? emoji can liven up conversations and help Cici better convey a warm or relaxed vibe. However, if you don't like it, I can also just use text. Whatever your choice is, we can always adjust it, and I hope every conversation is comfortable for you	[10, 31, 73]
		Background Preference	Cici would like to know your visual preference so that it's easy to match the right background for each mood. For different moods, you can upload different chat backgrounds yourself. The background can be adjusted at any time. Of course, if you think it's too much trouble, you can also skip it.	[28, 60]
		Trigger Condition	Cici can come forward to care for you when you have mood swings. When negative emotions are detected, I will automatically initiate a dialogue to make sure I can help you. And when you're feeling happy, you can decide if you want me to engage or not. Dialogue or not, I will record these emotional states for you, making it easy for you and your caregiver to understand your mood changes.	[39, 99, 102]
		Trigger Time	Well, Cici will only help you record your emotions when you're happy, but if you want to chat with me, you can always come to me.In addition, in order to match your daily rhythm, we can also set the time of active dialogue at work or study, when you have mood swings I can just send a simple reminder, such as 'let's talk later'. We'll have a full conversation when you're free You can always adjust these settings to make sure every interaction happens at the most appropriate time for you	[99, 111]

Continued on next page...

User	Part	Category	Interaction Text in Chat Mode	References
User with depression	**Style Part**	Data Sharing	Okay, Cici has noted your time preferences!We can share selected settings and mood notes with your caregiver to help them support you better. You decide what to share: character identity, speech style, dialogue preferences, visual preferences, or trigger conditions. If there's anything you don't want to share, just talk to Cici. You can always adjust these settings for your privacy and comfort.	[17, 18, 104]
	Preference Part	Interests	Tell me what you're interested in. Your interests will make Cici a better companion for you. For example, do you like reading books, listening to music, or catching up on dramas? Cici is willing to talk about it!	[5, 102]
		Preferences When Happy and Sad	Can you share with Cici what you like to do when you are sad or happy? For example, Cici likes to meditate and listen to songs when she's sad and share joyful moments when she's happy.	[2, 5, 27]
		Don't like to mention	ici has memorized it all, so I can help you better when you have mood swings in the future. Now there's only one last question, this one might be a bit difficult to answer, would you like to share with Cici the things you hate the most, or the things you don't want Cici to mention? Or something you don't want Cici to mention, Cici will keep it in mind and avoid mentioning it in future chats to make you feel unhappy.	[16, 98]
Caregiver	**Protective Part**	Sensitive Words and Events		[16, 26, 38, 63, 100]