



OPEN ACCESS

EDITED BY

Jennifer Apolinário-Hagen,
Heinrich Heine University of Düsseldorf,
Germany

REVIEWED BY

Isabella Choi,
The University of Sydney, Australia,
Yunfei Long,
University of Essex, United Kingdom,
Izidor Mlakar,
University of Maribor, Slovenia

*CORRESPONDENCE

Sahand Sabour

✉ sahandfer@gmail.com

Minlie Huang

✉ aihuang@tsinghua.edu.cn

[†]These authors contributed equally to this work and share first authorship.

RECEIVED 29 December 2022

ACCEPTED 17 April 2023

PUBLISHED 04 May 2023

CITATION

Sabour S, Zhang W, Xiao X, Zhang Y, Zheng Y, Wen J, Zhao J and Huang M (2023) A chatbot for mental health support: exploring the impact of Emohaa on reducing mental distress in China.

Front. Digit. Health 5:1133987.

doi: 10.3389/fdgth.2023.1133987

COPYRIGHT

© 2023 Sabour, Zhang, Xiao, Zhang, Zheng, Wen, Zhao and Huang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A chatbot for mental health support: exploring the impact of Emohaa on reducing mental distress in China

Sahand Sabour^{1*†}, Wen Zhang^{2†}, Xiyao Xiao³, Yuwei Zhang³, Yinhe Zheng³, Jiaxin Wen¹, Jialu Zhao⁴ and Minlie Huang^{1,3*}

¹The CoAI Group, DCST, Institute for Artificial Intelligence, State Key Lab of Intelligent Technology and Systems, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, China, ²Department of Psychology, Beijing Normal University, Beijing, China, ³Department of Research and Development, Beijing Lingxin Intelligent Technology Co., Ltd, Beijing, China, ⁴Center for Counseling and Psychological Development Guidance Center, Tsinghua University, Beijing, China

Introduction: The growing demand for mental health support has highlighted the importance of conversational agents as human supporters worldwide and in China. These agents could increase availability and reduce the relative costs of mental health support. The provided support can be divided into two main types: cognitive and emotional. Existing work on this topic mainly focuses on constructing agents that adopt Cognitive Behavioral Therapy (CBT) principles. Such agents operate based on pre-defined templates and exercises to provide cognitive support. However, research on emotional support using such agents is limited. In addition, most of the constructed agents operate in English, highlighting the importance of conducting such studies in China. To this end, we introduce Emohaa, a conversational agent that provides cognitive support through CBT-Bot exercises and guided conversations. It also emotionally supports users through ES-Bot, enabling them to vent their emotional problems. In this study, we analyze the effectiveness of Emohaa in reducing symptoms of mental distress.

Methods and Results: Following the RCT design, the current study randomly assigned participants into three groups: Emohaa (CBT-Bot), Emohaa (Full), and control. With both Intention-To-Treat ($N = 247$) and PerProtocol ($N = 134$) analyses, the results demonstrated that compared to the control group, participants who used two types of Emohaa experienced considerably more significant improvements in symptoms of mental distress, including depression ($F[2, 244] = 6.26, p = 0.002$), negative affect ($F[2, 244] = 6.09, p = 0.003$), and insomnia ($F[2, 244] = 3.69, p = 0.026$).

Discussion: Based on the obtained results and participants' satisfaction with the platform, we concluded that Emohaa is a practical and effective tool for reducing mental distress.

KEYWORDS

Chatbots, conversational agents, emotional support, mental health support, cognitive behavioral therapy (CBT), deep learning

1. Introduction

Concerns regarding mental health are prevalent in the modern world due to the increasing morbidity of mental diseases (1). During the COVID-19 pandemic, depression, anxiety, and other mental health issues have increased significantly (2). Specifically, a

review by Lakhan et al. (2) highlighted a 20% and 35% rise in depression and anxiety, respectively, for 113,285 individuals across 16 studies. Additionally, an international study with a sample of 22,330 adults showed that about 17.4% of the participants met the criteria for a probable insomnia disorder (3). These mental health issues impact people's daily lives, leading to social dysfunction and risks of self-harm and suicide (4). Due to the rapidly increasing demands, mental health services worldwide face challenges regarding the lack of professional training and stigmatization of mental illness. These challenges can lead to low diagnosis accuracy and patient treatment delays (2).

Similarly, the prevalence of mental health diseases in China is increasing (5–7). According to the epidemiological survey of mental disorders in China (8), the lifetime prevalence rate of mental disorders in adults, excluding senile dementia, is 16.57%. Specifically, the prevalence of anxiety disorder was reported the highest in China, with a 12-month prevalence rate of 4.98% (8). While possessing one of the largest populations worldwide, the number of licensed psychiatrists, though gradually increasing, is extremely limited, with a recent estimate suggesting that China had only 36,610 psychiatrists (2.6 per 100,000 population) in 2018 (9, 10). Similar to the limited number of mental health services in China, the quality of such services is also inadequate (7, 8, 11). Additionally, recent research has also shown that stigmas related to mental health support in China and concerns regarding burdening others also affect an individual's willingness to seek support (12, 13). Due to these challenges, only a limited number of Chinese patients are receiving appropriate support and treatment. Hence, the invention of high-technology tools or treatments in China is essential as it can provide effective, available, and affordable support for improving individuals' mental health.

Advancements in Artificial Intelligence (AI) and the field of Natural Language Processing (NLP) have highlighted the potential of machines to serve as anthropomorphic conversational agents (11). One of the essential applications of such agents is health care, mainly for providing mental health support. Employing machines for such tasks increases availability while reducing the costs of seeking support, as these agents could be widely accessible and affordable through mobile devices (14). Previous work has shown that individuals are willing to self-disclose their emotional problems with machines (15–17), which is significant as users' self-disclosure is essential for providing support. It demonstrates user rapport with these agents and highlights their potential as practical and beneficial supporters (18), thus serving as a strong motivation for this study.

Extensive research has been conducted on the effects of machine-based support. As proposed by Rimé (19), there are two main types of support to reduce mental distress: cognitive and emotional. Cognitive support enables individuals to reassess their situation from a different perspective and realize a new way of thinking about their problem (20, 21). In contrast, emotional support includes providing validation and understanding to cause relief and improve emotional distress (20, 22, 23). Recent work has mainly focused on delivering cognitive support through conversational agents adopting Cognitive Behavioral Therapy (CBT) principles and has demonstrated the

efficacy of such interventions in reducing users' mental distress, mainly depression and anxiety (14, 24–26). While most of the existing research on this topic is in English (18, 27, 28), there have been attempts to create Chinese chatbots for CBT (26, 29–31), demonstrating the importance of employing such systems in China.

In addition, research on machine-based emotional support is comparatively limited. Liu et al. (23) constructed a dataset of emotional support conversations based on Hill's (22) helping skills and demonstrated the feasibility of machine-based emotional support. Their work facilitated the research in this direction, and several approaches have been proposed to improve machines' emotional support ability (32–37). These approaches achieved promising results on aspect-based human evaluation (e.g., fluency and coherence). However, their corresponding studies did not create prototype agents, nor did they conduct empirical studies of their effectiveness in reducing users' mental distress. In addition, all of the mentioned work was implemented in English, highlighting the lack of research and resources for Chinese machine-based emotional support.

To the best of our knowledge, Pauw et al. (21) presented the first and only study investigating the effects of different types of machine-based support, including emotional support. However, their proposed prototypes only produced a set of pre-defined statements (e.g., "I am sorry to hear that") rather than generating responses based on the users' messages. Therefore, with mental health being a rising issue in the Chinese community, existing high demands for available and affordable support in China, and the limited research in this area, we believe constructing and conducting a study on a conversational agent for support in Chinese is crucial.

This study investigates the efficacy of conversational agents for providing cognitive and emotional support. Specifically, it aims to study the effectiveness of agents providing different types of support in reducing mental distress and assess the acceptability and practicality of such interventions for mental health support. We introduce Emohaa, a hybrid system involving a platform based on CBT principles and exercises for cognitive support and a conversational platform for emotional support regarding various topics. We recruit participants from mainland China and hypothesize that using Emohaa, which includes completing daily exercises and emotional venting, would improve their symptoms of mental distress, specifically depression, anxiety, negative affect, and insomnia.

2. Materials and methods

2.1. Emohaa

Our proposed conversational agent consists of two platforms. First, a template-based platform that contains conversations with pre-defined options and exercises that assist participants in improving their mental distress based on CBT principles (CBT-Bot). Second, a generative dialogue platform that allows conversations regarding various emotional issues in an open-ended manner (i.e., without requiring the users to choose predefined conversational options) and provides emotional support (ES-Bot).

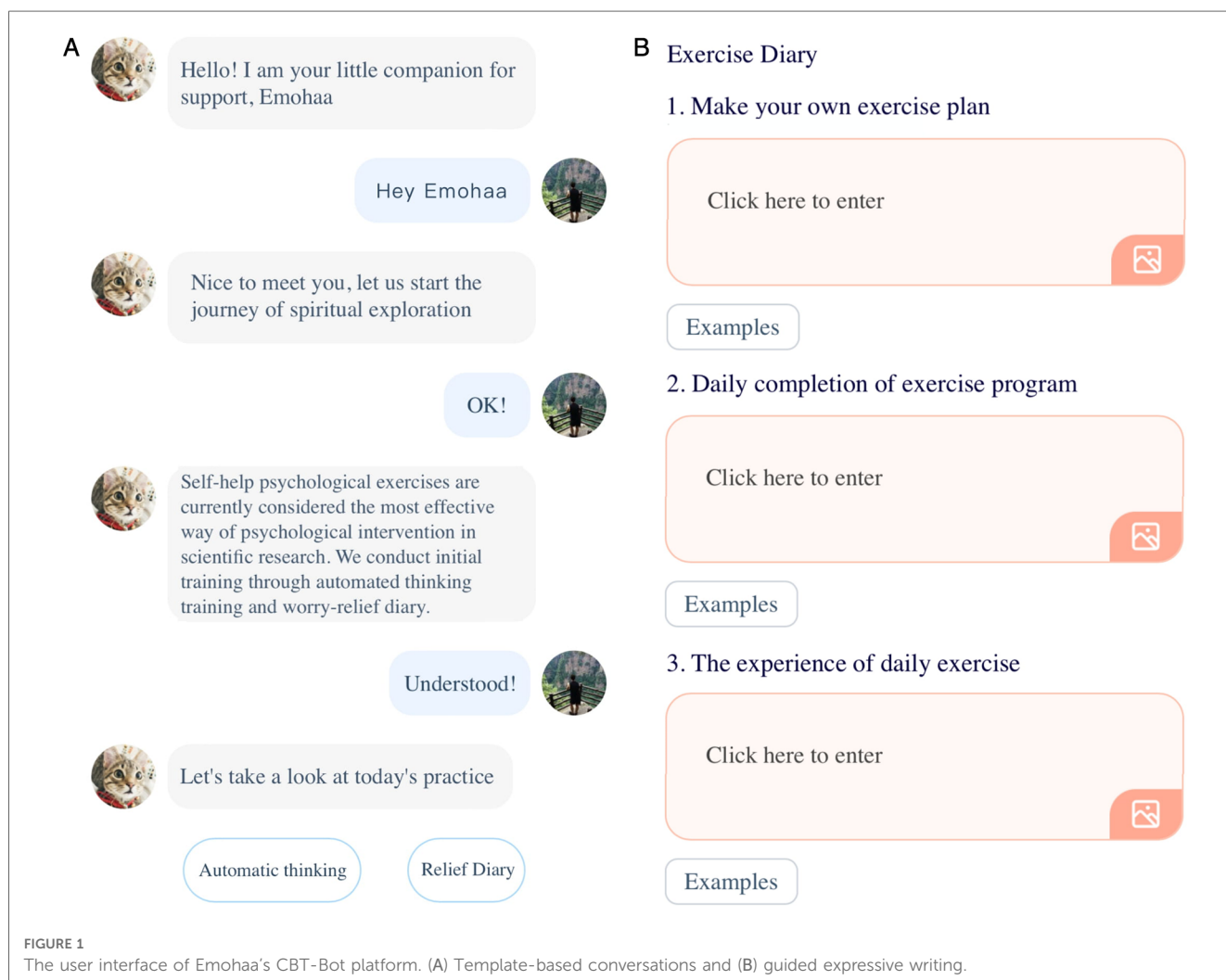
2.1.1. Cognitive behavioral therapy chatbot (CBT-Bot)

Creating a platform based on CBT principles postulates a direct and reciprocal interaction between thoughts, feelings, and behaviors that helps illuminate understanding of one’s overall emotional distress and situational responses while highlighting areas for intervention (38). As a tool for cognitive support, we integrated two different practices: automatic thoughts training and guided expressive writing. Individuals have automatic thoughts in response to a trigger, often outside of that one’s conscious awareness. These thoughts could often be irrational and harmful when associated with mental distress (39). As one of the core elements of CBT (38, 40), automatic thoughts training aims to identify and dismantle these thoughts (i.e., replace negative thoughts with rational perspectives), which could reduce mental distress and improve one’s mood (38, 41). In addition, previous studies have shown that writing about stressful or emotional events improves physical and psychological health in non-clinical and clinical populations (42, 43). Therefore, we adopted over 20 guided expressive writing exercises that cover a variety of topics and instruct users throughout each step of the exercise via interactive conversations.

On this platform, participants are initially given a set of conversational choices on this platform and are accordingly introduced to CBT and how to use this platform (Figure 1A). Accordingly, they are provided with two types of exercises: guided expressive writing and automatic thinking. An example of a guided writing exercise is shown in Figure 1B, where users are asked to fill out parts of their diary about a given topic in several steps. Automatic thinking exercises (Figure 2) present the user with a hypothetical scenario and require them to take the person’s perspective in that situation. Accordingly, they are asked a question regarding the correct approach to take in that person’s situation and report their confidence in their answer. Lastly, they are shown the correct answer about how to approach and gain a new perspective in such scenarios. To assess the efficacy of the exercises, we require users to report their mood after completing an exercise and describe their emotions using a set of pre-defined keywords. This platform is publicly available on WeChat, China’s most popular social media platform.

2.1.2. Emotional support chatbot (ES-Bot)

Several studies have shown that emotional support is beneficial for reducing mental and emotional distress (34, 44, 45). In



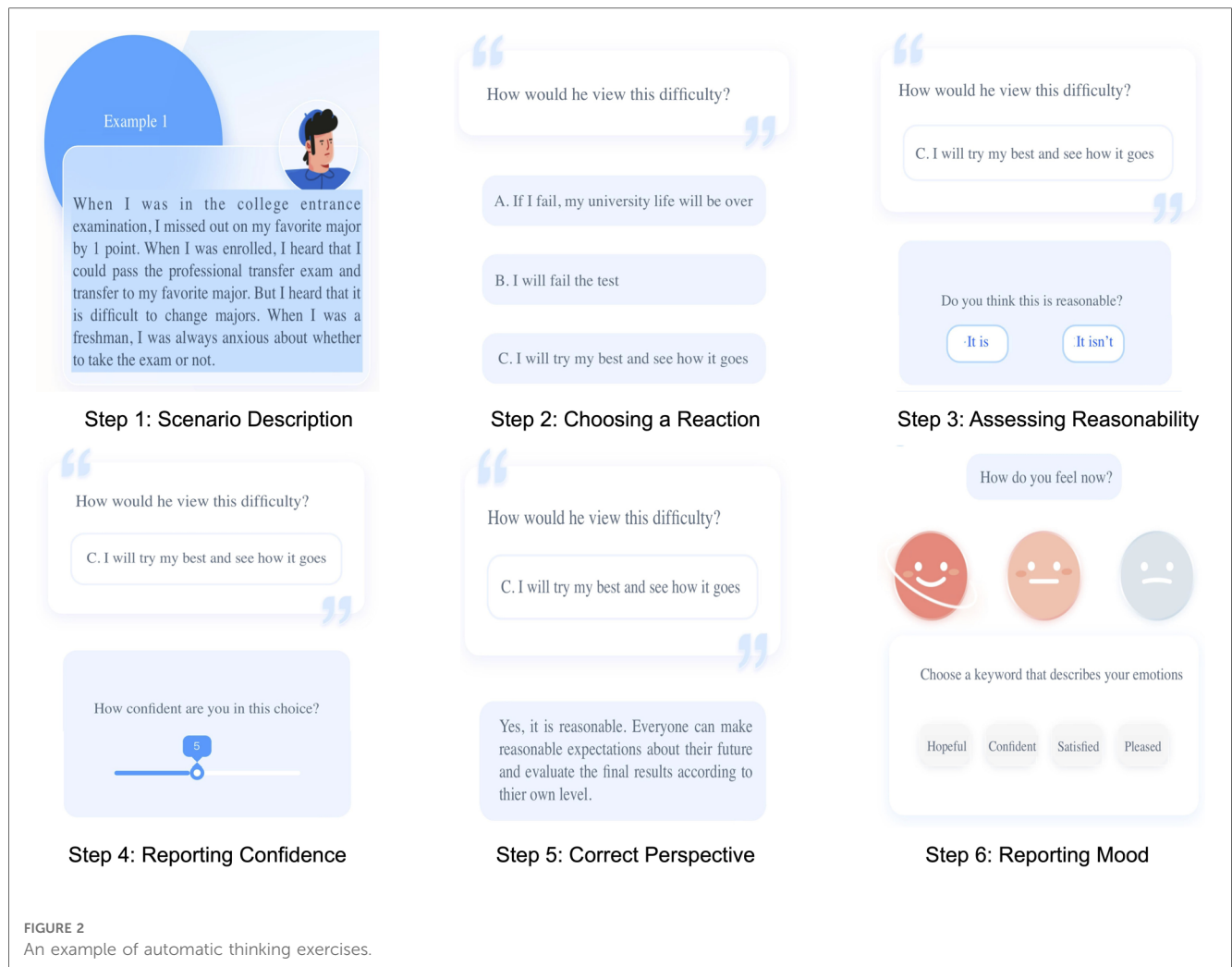


FIGURE 2 An example of automatic thinking exercises.

addition, allowing users to discuss their desired topics freely is crucial for creating anthropomorphic conversational agents. Therefore, we aimed to construct an agent that could openly converse with users about their emotional problems and generate responses based on their situation. To this end, we found Liu et al.'s (23) dataset of emotional support conversations (ESConv) suitable for this study. This dataset was constructed based on the Helping Skills Theory (22), in which trained human supporters leverage appropriate support strategies (e.g., self-disclosure, affirmation, and suggestions) to provide adequate emotional support. As the original dataset was curated in English, the conversations were manually translated into Chinese by trained professionals for the purpose of this study. Back-translation procedures (46) were followed to ensure the precision of the items. Specifically, we asked an English major who also had a psychology background to translate the English version into Chinese. Then another student translated back the Chinese version into English. Then we invited three professionals to compare, revise and decide on the final translation.

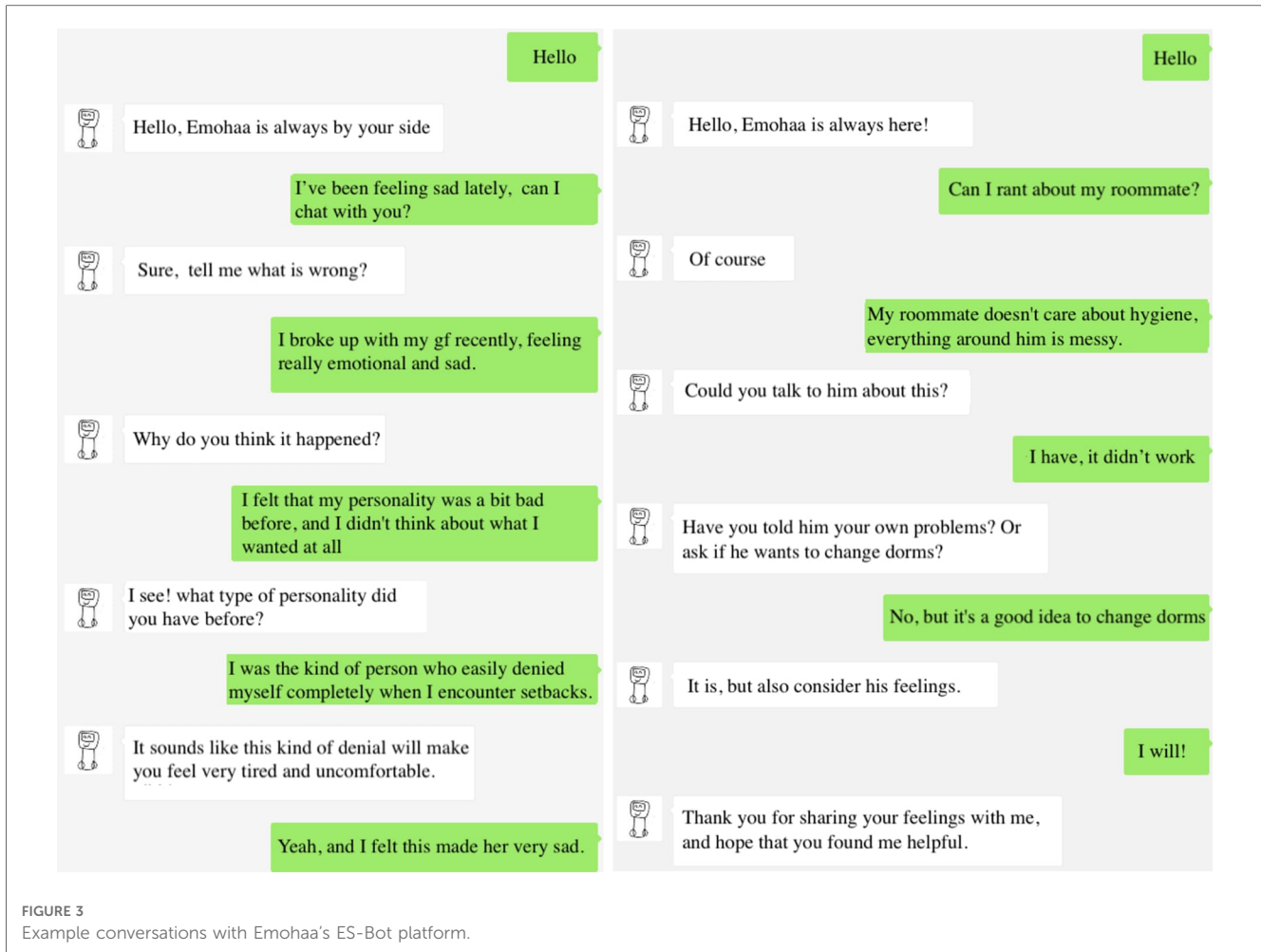
For building the ES-Bot platform, a large-scale Chinese dialogue model (47) was leveraged as the backbone to build a strategy-controlled emotional support dialogue model. Specifically, the model chooses an appropriate support strategy

given the conversation history. Accordingly, it generates responses that are coherent with the user's messages and conform to the chosen strategy. Given the free-flow design of this platform, as opposed to users choosing pre-defined options for the conversation, an additional model (48) was trained to classify whether users' messages demonstrated signs of suicidal thoughts to ensure users' safety. As individuals with the risk of suicide require immediate professional help, the platform recommends contact information of relevant authorities when corresponding signs are detected. Example conversations with this platform are demonstrated in Figure 3. Similarly, this platform is also publicly available on WeChat.

2.2. Measures

2.2.1. PHQ-9

Participants' depression was measured with the Patient Health Questionnaire (PHQ-9) (49), the most widely used measure in psychological depression trials (14, 50). PHQ-9 is a 9-item self-report questionnaire that measures the frequency and severity of depressive symptoms over the last two weeks. Participants were asked to score each item from 1 (not at all) to 4 (nearly every



day). In this study, the internal reliability of the scale (Cronbach's alpha) in the pre-test and the post-test were 0.78 and 0.85.

2.2.2. GAD-7

To measure participants' anxiety, we adopted the Generalized Anxiety Disorder (GAD-7) scale (51), a 7-item questionnaire assessing the frequency and severity of symptoms, thoughts, and related behaviors to anxiety within the last two weeks. Like PHQ-9, participants were required to score each item from 1 (not at all) to 4 (nearly every day). The Cronbach's alpha of this scale in the pre-test and the post-test were both 0.84.

2.2.3. PANAS

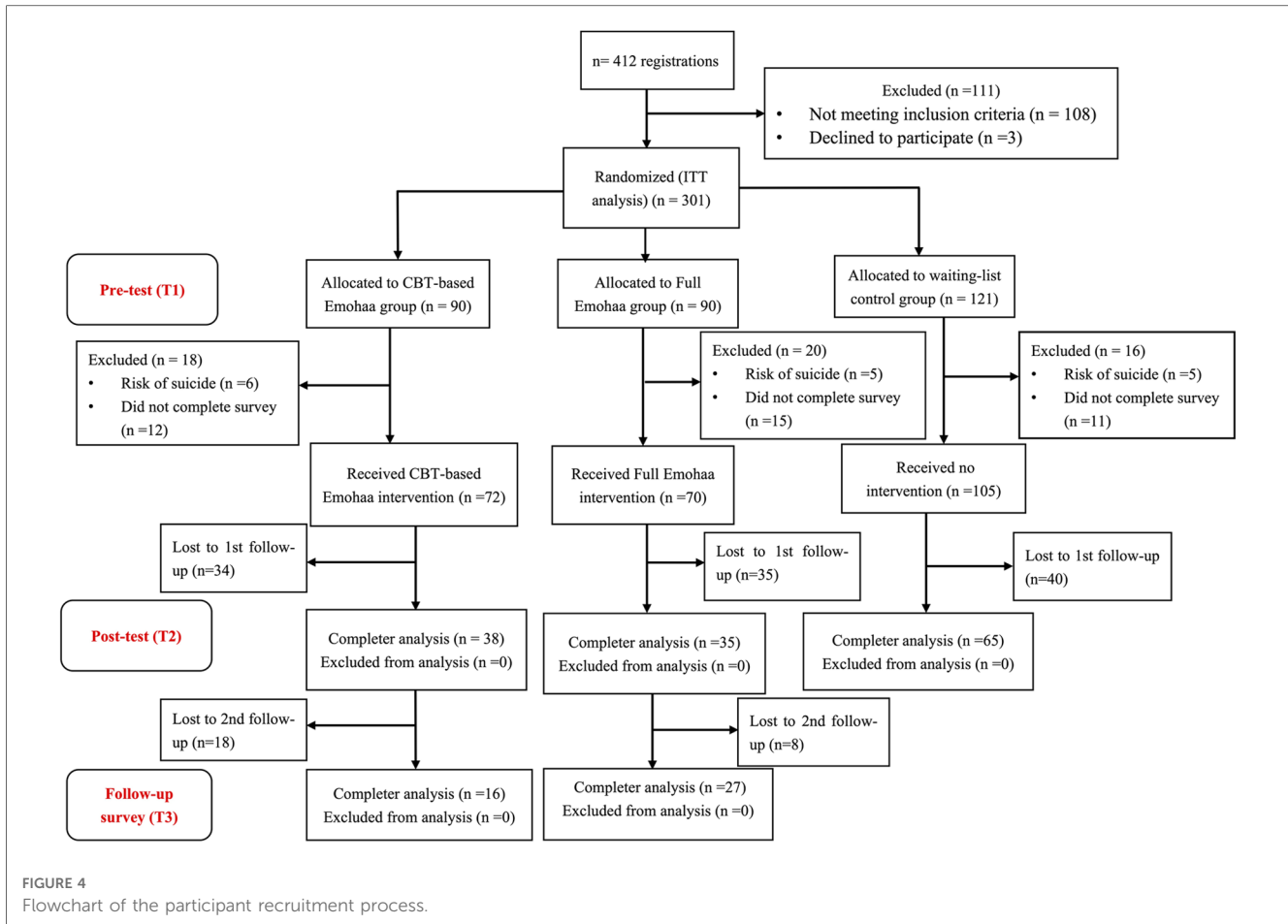
Participants' affect was measured by Watson et al.'s (52) 20-item Positive and Negative Affect Schedule (PANAS). In this questionnaire, half of the items represent positive affect (e.g., active, enthusiastic, and proud), and the remaining half corresponds to negative affect (e.g., upset, guilty, and irritable). All items are scored on a 5-Likert scale, and higher scores indicate higher levels of affect. The Cronbach's alpha of the positive affect dimension in the pre-test and the post-test were 0.88 and 0.82, and 0.85 and 0.82 for the negative affect dimension at the two-time points.

2.2.4. ISI

The 7-item Insomnia Severity Index (ISI; (53)) was used to measure participants' perceptions of their insomnia. This questionnaire assesses the severity of sleep-onset and maintenance difficulties, their interference with daily functioning, and the degree of distress caused by sleep problems. Participants were asked to score each item from 1 (none) to 4 (very severe). The Cronbach's alpha of the scale in the pre-test and the post-test were both 0.87.

2.3. Participants

Prior to the study, we used G*Power 3.1 to calculate the required number of participants. We set the large effect size f to be 0.40 while setting the power ($1 - \beta$ error probability) and α error probability to be 0.90 and 0.05, respectively. Thus, the required number of participants was calculated as 102. An online poster was made to recruit participants. We asked colleagues and friends to help release the recruitment information on their social media platforms, such as WeChat and Weibo. Participants who were interested in the study could contact our team based on the provided information in the advert. The following criteria were used to recruit participants through online posters: participants were required to be at least 18 years old, able to use



a smartphone, not currently in therapy as it would interfere with our study, and not suffering from physical issues such as physical illness or not taking medicine as they might influence their psychological state.

A total of 412 participants registered for the intervention, and 301 met all the above criteria. A research assistant, who was blinded to the purpose of the study, assigned a code based on the order that the participants contacted them. Accordingly, the participants were randomly assigned to three groups: Emohaa (CBT-Bot), Emohaa (Full), and the control group. Considering the relatively long waiting time and the potential number loss in the control group, we randomly allocated 30 more participants to the control group, adopting an approximate 3:3:4 allocation ratio for the 3 groups. The current study used a blank control group in which participants were asked to wait for a month before they would receive mindfulness intervention.

After signing the consent form, participants were instructed to take the pre-test (T1) questionnaires, including PHQ-9, GAD-7, PANAS, and ISI (Section 2.2), and their demographic information. Sixteen participants were excluded from the study and referred to relative authorities for professional help as they were at risk of suicide according to their scores on an item from PHQ-9 (i.e., “how often have you been bothered by the thoughts that you would be better off dead or thoughts of

hurting yourself in some way?”), and 38 participants were excluded because they did not complete the pre-test survey. Overall, 72 participants in the Emohaa (CBT-Bot) and 70 participants in the Emohaa (Full) completed the pre-test questionnaires. 105 participants in the control group completed pre-test questionnaires.

The entire intervention lasted for three consecutive weeks (i.e., 21 days). Then, one day after the end of the intervention, all the participants were asked to fill in the post-test (T2) questionnaire, which included the same items as T1. Additionally, one month after the end of the experiment, participants were invited to fill in a follow-up questionnaire (T3) intervention to track the lasting effect of the intervention. From the perspective of health ethics and practical reasons, other forms of intervention could have been provided to the control group after the intervention. Hence, there were no valid data at T3 for the control group. The above recruitment process is illustrated in Figure 4.

Of the randomized participants, 54.2% (134/247) went on to provide partial or complete data at T2. Independent *t*-tests analyses did not detect evidence of significant differences at T1 between those who dropped out of the study versus those who did not on age ($t = 1.51$; $p = 0.132$); gender ($t = 0.37$; $p = 0.709$); working tenure ($t = 0.92$; $p = 0.357$); PHQ-9 ($t = 0.95$; $p = 0.342$); GAD-7 ($t = 0.59$; $p = 0.558$); PANAS of positive ($t = 1.02$; $p = 0.145$) and negative ($t = 1.21$; $p = 0.227$)

affect scores; or on insomnia ($t = 1.17$; $p = 0.244$). Additionally, we employed MCAR analyses (54) to test whether the data are missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). The results showed that Chi-Square = 42.98 ($p < 0.001$), demonstrating the MCAR pattern.

2.4. Data collection and privacy

2.4.1. Experiment design

As mentioned, the ES-Bot platform allows users to send their desired text messages and employs a generative model for producing its responses, as opposed to the template-based conversations (i.e., providing users with limited conversational options and producing pre-defined answers) in the CBT-Bot platform. Due to the existing limitations of generative models, such as problems with response coherence and fluency, we believed a direct comparison between the effectiveness of the two platforms was inappropriate. Therefore, we required participants in the Emohaa (Full) group to use both platforms and aimed to study the complementary effect of the ES-Bot platform rather than analyzing its respective efficacy.

Accordingly, a research assistant informed the participants of our code of conduct and asked for consent through WeChat. Participants were assured that their participation was voluntary and anonymous. All participants were required to complete the mental distress questionnaires (Section 2.2) at T1 and T2. Excluding the control group, all participants were instructed to use the CBT-Bot platform daily, which required completing at least one automatic thinking exercise and writing a guided expressive diary. However, participants were encouraged to complete more exercises for better outcomes. In addition, participants from the Emohaa (Full) group were tasked to converse with the ES-Bot platform at least once daily. Each conversation session was required to last for 5–10 conversational turns. Participants were encouraged to continue chatting with the platform if they felt engaged in the conversation. Although there were no limitations on the conversational topics, participants were encouraged to talk about two main types of emotional experiences and problems: Event-based (e.g., breaking up with a partner, problems with work/study, and nuisance complaints); Emotion-based (i.e., topics that cause anger, sadness, anxiety).

2.4.2. Quality control

Participants' usage of the platform was manually checked every three days. Those who failed to conform to the guidelines were notified and required to complete the relative tasks to ensure high adherence. In addition, conversations with the ES-Bot platform were analyzed to monitor the chatbot's performance and the reliability of the conversations. For instance, during the first check of this intervention, it was found that 3/34 participants sent the same message multiple times to meet the requirements. These participants were

contacted and asked to repeat these conversations to ensure the experiment's integrity.

2.5. Privacy and ethics statement

Regarding the conversations with Emohaa, participants were instructed not to share any personal information (e.g., name, address, and date of birth) that could be used to identify them. The data collected during the experiment are anonymized, stored securely, and will be available for research purposes through a request to the corresponding author. The studies involving human participants were reviewed and approved by Beijing Normal University's Institutional Review Board (IRB Number: 202209150101). Written informed consent to participate in this study was provided by the participants.

3. Results

We used two strategies for analyzing our results. In the main context, we followed the Intention-To-Treat analysis (ITT) principle (55) by including all the participants who initially participated in the research. In the supplemental analyses, we used completer cases by excluding participants who dropped out during the intervention period.

3.1. User demographics

Demographic information of our studied sample ($n = 247$) is provided in Table 1. Overall, the majority of participants were female (107/134, 79.85%). The average age of the studied sample was 30.90 years old ($SD = 7.92$). Participants had worked for an average of 7.87 years ($SD = 8.45$) prior to the experiment. All of the participants were from Mainland China. As the baseline for participants' mental distress, on average, the samples showed moderate ranges of depression ($Mean = 16.43$, $SD = 5.01$), moderate anxiety ($Mean = 16.23$, $SD = 4.37$), and moderate insomnia ($Mean = 16.45$, $SD = 5.38$). In regards to PANAS, participants, on average, demonstrated moderate levels of positive ($Mean = 24.76$, $SD = 7.20$) and negative affect ($Mean = 22.34$, $SD = 6.35$).

We employed ANOVA and chi-squared test to see whether there were significant differences in baseline variables (age, gender, education, PHQ-9, GAD-7, PA, NA, Insomnia) among

TABLE 1 User demographics of our studied sample ($n = 247$).

		Emohaa CBT-Bot ($n = 72$)	Emohaa full ($n = 70$)	Control group ($n = 105$)
Gender, n (%)	Male	11 (15.28%)	18 (25.71%)	28 (26.67%)
	Female	61 (84.72%)	52 (74.29%)	77 (73.33%)
Age, Mean (SD)		31.07 (7.45)	29.00 (6.64)	32.05 (7.92)
Work experience, Mean (SD)		7.74 (7.48)	6.31 (6.60)	9.01 (9.95)

the three groups. The results showed that the three groups were not different in terms of baseline demographics of age ($F = 2.17, p = 0.117$) and gender ($\chi^2 = 3.56, p = 0.173$). Additionally, the baseline variables of PHQ-9 ($F = 2.45, p = 0.088$), GAD-7 ($F = 0.93, p = 0.396$), PA ($F = 0.83, p = 0.438$), NA ($F = 2.82, p = 0.061$), and Insomnia ($F = 2.76, p = 0.065$) showed no significant differences among the three groups.

3.2. Effects of Emohaa intervention

We used the last observation forward (LOCF) method to conduct ITT analyses. Previous systematic reviews have shown that LOCF is one of the most commonly used and relatively conservative strategies in ITT analysis (56, 57). To investigate whether the effects of interventions were different from each other and from that of the control group, we conducted a one-way repeated measures MANOVA with time (two levels: pre-test and post-test) and group type (three levels: Emohaa (CBT-Bot) vs. Emohaa (Full) vs. Control) as the independent variables and the five mental health indicators as the dependent variables. First, as presented in Table 2, there were significant Group \times Time interaction effects on depression, $F[2, 244] = 6.26, p = 0.002, \eta^2 = 0.050$, indicating a significant difference in participants' depression changes among the three groups, and

such difference had a relatively small effect size that is smaller than 0.06 (58). Specifically, as Figure 5A shows, the intervention effects on depression stemmed from decreases in both Emohaa (CBT-Bot) ($t = -2.25, p = 0.027$) and Emohaa (Full) group ($t = -2.09, p = 0.040$) from pre-test to post-test, but there was an increase of depression in the control group ($t = 2.04, p = 0.044$) from pre-test to post-test. Moreover, we did not find significant differences between the two types of interventions ($F[1, 140] = 0.76, p = 0.386$).

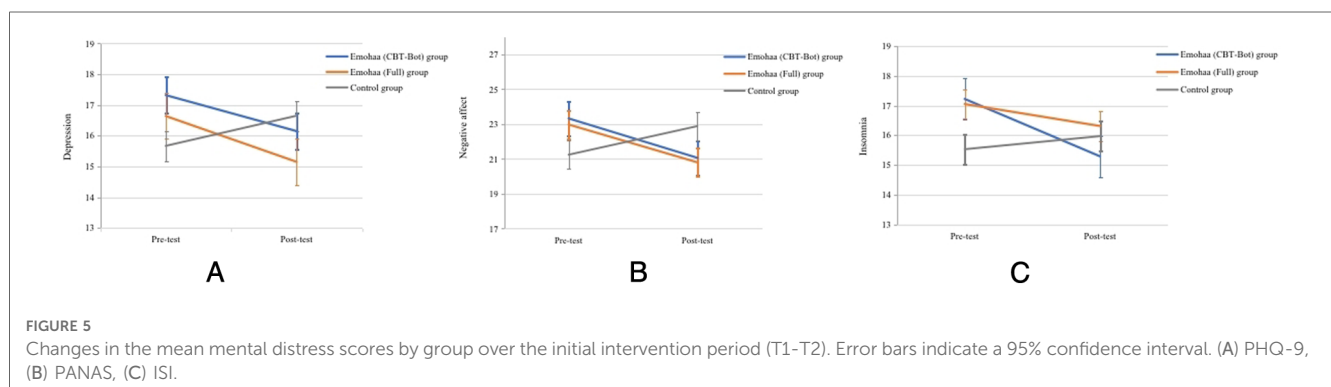
Similarly, we conducted the same MANOVA analyses to test whether participants' anxiety changed differently over the intervention period. The results revealed a main effect of time on anxiety, $F[1, 244] = 27.66, p < 0.001, \eta^2 = 0.102$, indicating participants' anxiety decreased over time. As Table 2 shows, the interaction effect of Group \times Time was not significant on anxiety $F[2, 224] = 0.60, p = 0.556, \eta^2 = 0.006$. No significant differences were found in the effects of the two types of interventions ($F[1, 140] = 0.38, p = 0.538$).

Additionally, the results showed that there was no main effect of time on positive affect ($F[1, 244] = 2.39, p = 0.123, \eta^2 = 0.010$) nor on negative affect ($F = 2.88, p = 0.091, \eta^2 = 0.012$), meaning that participants' positive affect increased and negative affect did not change over time. We did not find a significant interaction effect of Group \times Time on positive affect ($F[2, 244] = 1.58, p = 1.208, \eta^2 = 0.013$). The interaction effect of Group \times Time

TABLE 2 Analyses results of variance in mental health outcomes.

Variables	Time	Emohaa CBT-Bot ($n = 72$)		Emohaa full ($n = 70$)		Control group ($n = 105$)		F	p	η^2
		Mean	SD	Mean	SD	Mean	SD			
Depression	1	17.32	5.05	16.64	4.83	15.67	5.05	6.26	0.002	0.050
	2	16.15	5.58	15.15	5.91	16.65	5.39			
Anxiety	1	16.72	3.86	16.34	3.99	15.83	4.91	0.60	0.556	0.006
	2	14.69	4.75	14.12	4.67	14.45	4.08			
Positive affect	1	24.28	6.90	25.69	7.53	24.47	7.77	1.58	1.208	0.013
	2	24.30	6.65	23.68	6.52	24.21	8.05			
Negative affect	1	23.33	6.51	22.98	6.92	21.25	5.70	6.09	0.003	0.048
	2	21.07	6.55	20.81	6.67	22.89	7.73			
Insomnia	1	17.22	5.40	17.05	5.37	15.53	5.28	3.69	0.026	0.024
	2	15.29	5.63	16.31	5.89	15.97	5.92			

The F-test tests the Condition \times Time interaction effect to detect a significant difference among the conditions in the rate of change across time.



was significant on negative affect ($F[2, 244] = 6.09, p = 0.003, \eta^2 = 0.048$). Specifically, as shown in **Figure 5B**, participants' negative affect decreased significantly in both Emohaa (CBT-Bot) ($t = -2.20, p = 0.031$) and Emohaa (Full) ($t = -2.04, p = 0.045$) groups from pre-test to post-test, but their negative affect significantly increased in the control group ($t = 2.11, p = 0.037$). The post-hoc results showed that there were no significantly different effects of PA ($F[1, 140] = 0.01, p = 0.943$) and NA ($F[1, 140] = 0.01, p = 0.955$) between types of interventions.

Finally, the MANOVA results demonstrated a main effect of time on participants' insomnia ($F[1, 244] = 4.49, p = 0.035, \eta^2 = 0.018$), indicating that participants' insomnia decreased during the period of intervention. Besides, the results showed a significant interaction effect of Group \times Time on insomnia ($F[2, 244] = 3.69, p = 0.026, \eta^2 = 0.024$), but the differences had a relatively small effect size that is slightly larger than the small effect size of .01 (58). Specifically, as **Figure 5C** reveals, the effects stemmed from a significant insomnia decrease in the Emohaa (CBT-Bot) group ($t = -2.28, p = 0.026$), no significant change of insomnia in Emohaa (Full) group ($t = -2.03, p = 0.055$), and no difference in the control group between the pre-test and post-test ($t = 0.88, p = 0.379$). We did not find significantly different effects of the two types of interventions ($F[1, 140] = 0.02, p = 0.936$).

3.2.1. Supplemental analyses

We additionally conducted a completer analysis with 134 cases. Specifically, the results of one-way repeated measures MANOVAs showed that there were significant Time \times Group interaction effect on depression ($F[2, 131] = 19.11, p < 0.001, \eta^2 = 0.230$). Participants' depression decreased significantly in both Emohaa (CBT-Bot) ($t = -4.19, p < 0.001$) and Emohaa (Full) ($t = -4.05, p < 0.001$) groups from pre-test to post-test, but their depression increased in the control group ($t = 2.54, p = 0.013$).

Additionally, the results showed that there was a significant interaction effect of Time \times Group on anxiety ($F[2, 131] = 45.04, p < 0.001, \eta^2 = 0.260$). Participants' anxiety decreased significantly in both Emohaa (CBT-Bot) ($t = -5.69, p < 0.001$) and Emohaa (Full) ($t = -4.53, p < 0.001$) groups from pre-test to post-test, but participants' anxiety remained unchanged in the control group ($t = -0.39, p = 0.397$).

The results showed no interaction effect of Time \times Group on PA ($F[2, 131] = 1.85, p = 0.162, \eta^2 = 0.031$), but such effect existed on NA ($F[2, 131] = 12.11, p < 0.001, \eta^2 = 0.160$). Participants' NA decreased significantly in both Emohaa (CBT-Bot) ($t = -7.47, p < 0.001$) and Emohaa (Full) ($t = -2.14, p = 0.040$) groups from pre-test to post-test, but remained unchanged in the control group ($t = 0.26, p = 0.795$).

Finally, the results showed that there was a significant interaction effect of Time and Group on insomnia ($F[2, 131] = 3.52, p = 0.031, \eta^2 = 0.005$). The differences stemmed from a significant decrease in insomnia for the Emohaa (CBT-Bot) group ($t = -3.84, p < 0.001$). Participants' insomnia in the Emohaa (Full) group slightly decreased ($t = -2.01, p = 0.053$). Participants' insomnia in the control group did not change ($t = 0.19, p = 0.869$).

To conclude, the intervention effects robustly existed in depression, NA, and insomnia. However, the intervention effects on anxiety were not significant using the ITT analysis strategy. The different results on Anxiety may be that the main effect of time on Anxiety is strong, which means that anxiety declined in all three groups, and with a stricter analysis strategy of ITT, the differences among the three groups become less salient. The intervention effect did not exist on PA, regardless of which analysis strategy we adopted.

Furthermore, We collected participants' data on the mental health indicators (**Section 2.2**) three weeks after the post-test. Due to practical reasons that participants in the control group received other forms of interventions after the post-test, we only collected two intervention groups' data. 16 participants in the Emohaa (CBT-Bot) group and 27 in the Emohaa (Full) group returned the questionnaires.

To compare the effects between the two intervention groups, we conducted MANOVA with time (three levels: pre-test, post-test, and three weeks after post-test) and group type (two levels: CBT-Bot Emohaa vs. full Emohaa) as the independent variables and the five mental health indicators as the dependent variables. We also adopted ITT analysis in comparing the results of the two groups. Results showed that there were no significant interaction effects of Time \times Group on depression ($F[2, 139] = 0.16, p = 0.853, \eta^2 = 0.002$), anxiety ($F[2, 139] = 0.37, p = 0.693, \eta^2 = 0.003$), positive affect ($F[2, 139] = 2.03, p = 0.133, \eta^2 = 0.014$) or negative affect ($F[2, 139] = 1.04, p = 0.354, \eta^2 = 0.007$), indicating that the changes of participants' four mental health indicators did not vary from each other between the two groups. However, such interaction effect was significant on insomnia ($F[2, 139] = 3.18, p = 0.043, \eta^2 = 0.022$). The difference stemmed from that participants' insomnia symptoms returned to the pre-test level in the Emohaa (CBT-Bot) group. Still, participants' insomnia in the Emohaa (Full) group continued improving after the intervention.

3.3. Conversation analysis

During the experiment, participants had 7 conversation sessions with Emohaa on average ($SD = 6.62, Max = 18, Min = 4$). These sessions had a mean of 17 conversational turns ($SD = 10.68, Max = 87, Min = 5$). N-gram analysis was used to investigate the characteristics of participants' conversations with Emohaa. The most discussed keywords were found to be 感觉 (feeling; 33%), 工作 (work; 20%), 心情 (mood; 11%), 学习 (pressure; 10.5%), 朋友 (friends; 9.2%), and 孩子 (children; 7.7%). Percentages indicate the proportion of conversations that included the keyword. The main problems that participants wanted to talk about were 工作环境 (Work environment), 工作压力 (Work pressure), 浪费时间 (Wasting Time), 集中注意力 (Keeping focus), 牺牲休息时 (Sacrificing leisure time), and 转移注意力 (Diverted attention). In general, participants were mainly interested in 正念冥想 (Mindful meditation), 早点休息 (Resting early), 身体健康 (Being healthy), and 提供情绪价值 (Providing emotional value).

3.4. Acceptability and feasibility

After the end of the intervention, participants who had used Emohaa during the experiment were instructed to complete an additional survey to evaluate the agent's performance. Most participants (60/69, 86.9%) reported that they had never received psychological counseling before the experiment, and only two had taken psychotropic medication.

Initially, participants were asked to rate the CBT-Bot platform's ease of use, provided content, and interface quality on a 10-point Likert scale. Most participants reported moderate to high levels of satisfaction (scores ranging from 7 to 10) with the platform's functionality (56/69, 81.16%) and the designed exercises (47/69, 68.12%). In addition, more than half of the participants (43/69, 62.32%) were satisfied with the interface design. Overall, the majority (49/69, 71%) reported that they would recommend this platform to others.

Similarly, participants who had used Emohaa's ES-Bot platform were instructed to rate its performance. This platform was considered by most of the participants as an appropriate chatting partner (24/31, 77.42%) and channel for emotional venting (20/31, 64.5%), and more than half of the participants (18/31, 58.1%) reported that chatting with this platform made them feel heard. When asked about their expectations of the platform, the majority believed it to be a suitable companion for emotional companionship and support that can accurately interpret their emotions and provide emotional counseling (21/31, 67.74%). In addition, more than half of the participants were satisfied with the interface (20/31, 64.5%) and reported that they would recommend it to others (19/31, 61.3%). Independent *t*-tests showed that there was no significant difference between participants' satisfaction with the Emohaa CBT-Bot platform and ES-Bot platform ($t = 1.16$; $p = 0.250$).

Lastly, participants were asked to provide feedback on their experience with Emohaa. **Table 3** summarizes the most common themes in the collected responses. The most frequently raised concerns were technical issues, unclear instructions, and limited content and choices. The reported technical issues were mainly regarding the user interface (e.g., "Cannot click the next page" and "Accidentally closing the app removed all my progress"). Many participants were overwhelmed with the number of categories in the guided writing exercises and felt some topics were illogical and not applicable to real life. Participants also felt that the stories in

the automatic thinking exercises were excessive while there were not enough options to describe their mood and emotions after completing the exercise. Moreover, it was suggested that the dialogue options in the platform's template were inadequate.

Issues regarding the performance of the ES-Bot platform were also raised. Several participants reported that the conversations were rigid, and the system needed user guidance to continue the conversation. In some instances, the generated responses were reported as irrelevant or incoherent to the conversation. Participants also highlighted the platform's occasional inability to remember what had been said in the early stages of the conversation, initiate conversation topics, and understand various input types (i.e., audio, video, and image).

In addition, participants were also asked to provide suggestions on how to improve Emohaa. As concerns regarding lack of content and options were mentioned, it was suggested that additional scenarios, stories, instructions, and options be included in the CBT-Bot platform. The importance of regularly updating the platform and promptly fixing technical issues was also highlighted. Regarding the ES-Bot platform, nearly half of the participants (14/31, 45.1%) believed that improvements for making the generated responses less rigid were necessary. It was also suggested that support for different input types be added to create a more interactive and engaging experience. Many also believed that recommending mental health-related content during conversations and taking the initiative in conversations would benefit this platform.

4. Discussion

4.1. Main findings

The obtained results demonstrated Emohaa's efficacy as a short-term intervention for depression, negative affect, and insomnia. Based on the survey results, users experienced reduced levels of mental distress in the measured categories after using Emohaa. Compared to the control group, there was a significant decrease in depression among the participants who used Emohaa, as measured by the PHQ-9 questionnaire. Similarly, as measured by the PANAS and the ISI questionnaires, their negative affect and insomnia were also considerably reduced. Therefore, as shown by the experimental results, Emohaa can be seen as an effective tool for mental health support.

Regarding the difference in outcomes between the two groups that used Emohaa, no significant differences were found in the short term. Both interventions effectively relieved individuals' mental health symptoms. However, as provided by the supplemental analyses (**Section 3.2.1**), participants who used the ES-Bot platform showed comparatively fewer indicators of insomnia. This finding highlights a potential benefit of emotional venting in improving problems regarding sleep in the long term.

Based on the obtained feedback, most participants were satisfied with this agent and considered recommending it to others. In line with previous research (16, 17), the results of the conversation analysis indicated that participants were willing to

TABLE 3 Summary of participants' feedback on Emohaa.

Platform		Problems	<i>n</i>
CBT-Bot	Content	Repetitiveness	5
		Limited exercises and options	11
		Unclear instructions	12
	Design	Impractical tips and recommendations	6
		Technical (i.e., Bugs, glitches, and lags)	14
		Ambiguous exercise tracking and transition	5
ES-Bot	Functionality	Lack of initiative in conversations	2
		Unable to understand image/video/audio inputs	3
	Generation	Rigid conversations	10
		Unrelated and out-of-context responses	8

self-disclose their emotional problems, as shown by their most discussed keywords and topics. Moreover, most participants considered Emohaa's ES-Bot platform a chatting partner that can effectively listen to their problems and provide a channel for them to vent their emotions. Notably, the majority felt that this platform could understand their emotions, an essential feature of conversational agents for support and a crucial trait for establishing a therapeutic connection (26). Therefore, our findings suggest that Emohaa can also be seen as an acceptable and feasible tool for support.

In addition to highlighting Emohaa's effectiveness in mental health support, this study demonstrated the potential of generative conversational agents and combining emotional and cognitive support to reduce mental distress symptoms. Our findings suggest that allowing users to converse about their desired topics with the agent freely has a complementary effect when added to more common forms of machine-based support (i.e., template-based conversations and exercises for cognitive support through CBT).

4.2. Limitations and future work

This study had several limitations regarding its design and methodology. The study duration was limited; thus, only two assessments of participants' mental distress were made. Although a follow-up screening for participants that had used Emohaa during the experiment was conducted, no data regarding the control group's participants were gathered as they might have received other interventions after the initial two-week screening. Furthermore, the number of remaining participants in the follow-up survey is inadequate to draw a conclusion that the conversational agent (ES-Bot) is better than the CBT-Bot. Future studies would benefit from collecting more data from the three groups in the follow-up surveys to support the complementary effects of generative dialogue platforms for emotional support.

It is believed that the number of participants was sufficient to demonstrate the preliminary effects of employing conversational agents for mental health support in theory. However, the sample size and the experiment duration are inadequate for generalizing the obtained results of this study to the public. Future experiments will include a larger sample size and longer study duration to further ensure the generalizability of Emohaa's effectiveness in reducing mental distress. In addition, our adopted method of advertisement for this study could have introduced a bias in our recruitment process, in which individuals who were in some way connected to our helping colleagues and friends were more likely to participate in the study. This could have also affected the male-to-female ratio among the participants, leading to the over-representation of female participants in our sample.

As mentioned, Emohaa's several technical issues could substantially impact the users' perceived level of empathy and support (14), so they should be resolved promptly. A management system for addressing similar issues on time should also be implemented in future work. Moreover, several

participants raised issues regarding Emohaa's limited content (i.e., exercises and options) and unclear instructions. Similar to Liu et al. (26), a wider variety of psychological resources will be consulted in future work to expand the provided content in the CBT-Bot platform and revise the instructions to avoid user misinterpretations or confusion. Lastly, although our requirements regarding the daily usage of this platform could be applicable in a trial, such constraints are not practical in real-life applications. Hence, future work could further improve user engagement within machine-based interventions.

Regarding the ES-Bot platform, several reported instances suggested that Emohaa forgets the information in previous turns and that the generated responses are irrelevant to the context, which could impair user engagement and rapport. This is a well-known issue in current language models (59), and the main reason could be the limited number of words in the model's input (128 words for Emohaa). A feasible approach to address this issue is to add a module that could summarize the essential information of the previous turns in the conversation (60). In addition, previous work has demonstrated the benefits of adding persona (61–63) and commonsense knowledge (64, 65) for improving user experience with generative conversational agents. Future work could explore these additions to study their efficacy and corresponding improvements in mental health support.

4.3. Conclusions

The present study introduced Emohaa, a Chinese conversational agent for mental health support. Emohaa employs CBT principles to provide cognitive support through template-based guided conversations for expressive writing and automatic thinking exercises. In addition, it includes a platform for providing emotional support in which users can discuss their desired emotional problems. This study examined the effectiveness of Emohaa in reducing mental distress and investigated its feasibility and acceptability as a tool for mental health support in China. Our findings demonstrated that participants experienced fewer symptoms of mental distress after using Emohaa for the duration of the study. Hence, we believe this agent could serve as a valuable tool for reducing users' mental distress, namely depression, negative affect, and insomnia. In addition, we found that there might be a complementary effect on long-term insomnia when implementing the generative dialogue platform for emotional support. This finding highlights the potential of generative conversational agents for the future of mental health support. In the future, we hope our work can inspire other studies to expand upon our research, leverage generative models for providing support, and investigate their comparative efficacy.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by The studies involving human participants were reviewed and approved by Beijing Normal University's Institutional Review Board (IRB Number: 202209150101). The patients/participants provided their written informed consent to participate in this study.

Author contributions

SS, XX, and YZha contributed to the study design and data collection. SS and WZ wrote the manuscript with comparatively contributions from XX and YZha to several sections. SS and WZ performed data and statistical analysis, respectively. YZhe and JW created and supervised the conversational agents used in this work. All authors contributed to the article and approved the submitted version.

References

- [Dataset] WHO. World mental health report: transforming mental health for all (2022).
- Lakhan R, Agrawal A, Sharma M. Prevalence of depression, anxiety, stress during COVID-19 pandemic. *J Neurosci Rural Pract.* (2020) 11:519–25. doi: 10.1055/s-0040-1716442
- Taylor DJ, Gardner CE, Bramoweth AD, Williams JM, Roane BM, Grieser EA, et al. Insomnia, mental health in college students: behavioral sleep medicine. *Behav Sleep Med.* (2011) 9(2):107–116. doi: 10.1080/15402002.2011.557992
- Hanna M, Strober LB. Anxiety, depression in multiple sclerosis (MS): antecedents, consequences, differential impact on well-being and quality of life. *Mult Scler Relat Disord.* (2020) 44:102261. doi: 10.1016/j.msard.2020.102261
- Que J, Lu L, Shi L. Development and challenges of mental health in China. *General Psychiatry.* (2019) 32:e100053. doi: 10.1136/gpsych-2019-100053
- Ju Y, Zhang Y, Wang X, Li W, Ng RM, Li L. China's mental health support in response to COVID-19: progression, challenges and reflection. *Global Health.* (2020) 16:1–9. doi: 10.1186/s12992-020-00634-8
- Li W, Yang Y, Liu Z-H, Zhao Y-J, Zhang Q, Zhang L, et al. Progression of mental health services during the COVID-19 outbreak in China. *Int J Biol Sci.* (2020) 16:1732. doi: 10.7150/ijbs.45120
- Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. *Lancet Psychiatry.* (2019) 6(3):211–24. doi: 10.1016/S2215-0366(18)30511-X.
- Xiang Y-T, Ng CH, Yu X, Wang G. Rethinking progress, challenges of mental health care in China. *World Psychiatry.* (2018) 17:231. doi: 10.1002/wps.20500
- Fang M, Hu SX, Hall BJ. A mental health workforce crisis in China: a pre-existing treatment gap coping with the COVID-19 pandemic challenges. *Asian J Psychiatr.* (2020) 54:102265. doi: 10.1016/j.ajp.2020.102265
- Zhang X, Lewis S, Firth J, Chen X, Bucci S. Digital mental health in China: a systematic review. *Psychol Med.* (2021) 51:2552–70. doi: 10.1017/S0033291721003731
- Yu S, Kowitz SD, Fisher EB, Li G. Mental health in China: stigma, family obligations, the potential of peer support. *Community Ment Health J.* (2018) 54:757–64. doi: 10.1007/s10597-017-0182-z
- Yin H, Wardenaar KJ, Xu G, Tian H, Schoevers RA. Mental health stigma and mental health knowledge in Chinese population: a cross-sectional study. *BMC Psychiatry.* (2020) 20:1–10. doi: 10.1186/s12888-020-02705-x
- Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR Ment Health.* (2017) 4:e19. doi: 10.2196/mental.7785_
- Gratch J, Kang S-H, Wang N. Using social agents to explore theories of rapport and emotional resonance. In: *Social emotions in nature and artifact.* Oxford: Oxford University Press Oxford (2013). p. 181, 2568173.

Conflict of interest

XX, YZha, YZhe, and MH were employed by Beijing Lingxin Intelligent Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Ho A, Hancock J, Miner AS. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *J Commun.* (2018) 68:712–33. doi: 10.1093/joc/jqy026
- [Dataset] Liang K-H, Shi W, Oh Y, Zhang J, Yu Z. Discovering chatbot's self-disclosure's impact on user trust, affinity, and recommendation effectiveness (2021).
- Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. *J Med Internet Res.* (2021) 23:e17828. doi: 10.2196/17828
- Rimé B. Emotion elicits the social sharing of emotion: theory and empirical review. *Emot Rev.* (2009) 1:60–85. doi: 10.1177/1754073908097189
- Rimé B, Finkenauer C, Luminet O, Zech E, Philippot P. Social sharing of emotion: New evidence and new questions. *Eur Rev Soc Psychol.* (1998) 9:145–89. doi: 10.1080/14792779843000072
- Pauw LS, Sauter DA, van Kleef GA, Lucas GM, Gratch J, Fischer AH. The avatar will see you now: support from a virtual human provides socio-emotional benefits. *Comput Human Behav.* (2022) 136:107368. doi: 10.1016/j.chb.2022.107368
- Hill CE. *Helping skills: facilitating exploration, insight, and action.* Washington: American Psychological Association (2009).
- Liu S, Zheng C, Demasi O, Sabour S, Li Y, Yu Z, et al. Towards emotional support dialog systems. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers).* Online: Association for Computational Linguistics (2021). p. 3469–83. Available from: <https://doi.org/10.18653/v1/2021.acl-long.269>.
- Daley K, Hungerbuehler I, Cavanagh K, Claro HG, Swinton PA, Kapps M. Preliminary evaluation of the engagement and effectiveness of a mental health chatbot. *Front Digit Health.* (2020) 2:576361. doi: 10.3389/fdgth.2020.576361
- Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth.* (2018) 6:e12106. doi: 10.2196/12106
- Liu H, Peng H, Song X, Xu C, Zhang M. Using ai chatbots to provide self-help depression interventions for university students: a randomized trial of effectiveness. *Internet Interv.* (2022) 27:100495. doi: 10.1016/j.invent.2022.100495
- Harrigian K, Aguirre CA, Dredze M. On the state of social media data for mental health research (2020). Available from: <https://doi.org/10.48550/arXiv.2011.05233>.
- Valizadeh M, Parde N. The AI doctor is in: a survey of task-oriented dialogue systems for healthcare applications. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* Dublin, Ireland: MIT Press (2022). p. 6638–60. Available from: <https://doi.org/10.18653/v1/2022.acl-long.458>.
- Chen X, Zhang X, Zhu X, Wang G. Efficacy of an internet-based intervention for subclinical depression (moodbox) in China: study protocol for a randomized controlled trial. *Front Psychiatry.* (2021) 11:585960. doi: 10.3389/fpsy.2020.585920

30. Yeung A, Wang F, Feng F, Zhang J, Cooper A, Hong L, et al. Outcomes of an online computerized cognitive behavioral treatment program for treating chinese patients with depression: a pilot study. *Asian J Psychiatr.* (2018) 38:102–7. doi: 10.1016/j.ajp.2017.11.007
31. Lin L-Y, Wang K, Kishimoto T, Rodriguez M, Qian M, Yang Y, et al. An internet-based intervention for individuals with social anxiety and different levels of Taijin Kyofusho in China. *J Cross Cult Psychol.* (2020) 51:387–402. doi: 10.1177/0022022120920720
32. Peng W, Hu Y, Xing L, Xie Y, Sun Y, Li Y. Control globally, understand locally: a global-to-local hierarchical graph network for emotional support conversation [Preprint] (2022). Available at: <https://doi.org/10.48550/arXiv.2204.12749>.
33. Tu Q, Li Y, Cui J, Wang B, Wen J-R, Yan R. Misc: a mixed strategy-aware model integrating comet for emotional support conversation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland (2022). p. 308–19.
34. [Dataset] Zheng C, Sabour S, Wen J, Huang M. AugESC: large-scale data augmentation for emotional support conversation with pre-trained language models (2022). Available from: <https://doi.org/10.48550/arXiv.2202.13047>.
35. Xu X, Meng X, Wang Y. Poke: prior knowledge enhanced emotional support conversation with latent variable [Preprint] (2022). Available at: <https://doi.org/10.48550/arXiv.2210.12640>
36. Cheng Y, Liu W, Li W, Wang J, Zhao R, Liu B, et al. Improving multi-turn emotional support dialogue generation with lookahead strategy planning [Preprint] (2022). Available at: <https://doi.org/10.48550/arXiv.2210.04242>.
37. [Dataset] Cheng J, Sabour S, Sun H, Chen Z, Huang M. Pal: Persona-augmented emotional support conversation generation (2022). Available from: <https://doi.org/10.48550/arXiv.2212.09235>.
38. Beck JS, Beck AT. Cognitive behavior therapy: basics and beyond (2nd ed.). *J Autism Dev Disord.* (2011) 17:81–93. ISBN: 978-1609185046.
39. Hollon SD, Kendall PC. Cognitive self-statements in depression: Development of an automatic thoughts questionnaire. *Cognit Ther Res.* (1980) 4:383–95. doi: 10.1007/BF01178214
40. Lustman PJ, Griffith LS, Freedland KE, Kissel SS, Clouse RE. Cognitive behavior therapy for depression in type 2 diabetes mellitus: a randomized, controlled trial. *Ann Intern Med.* (1998) 129:613. doi: 10.7326/0003-4819-129-8-199810150-00005
41. Fukui I, Sakano Y. The relationship among irrational beliefs, automatic thoughts and depressive/anxious mood. *Human welfare studies* (2000).
42. Baikie KA, Wilhelm KA. Emotional and physical health benefits of expressive writing. *Adv Psychiatr Treat.* (2005) 11:338–46. doi: 10.1192/apt.11.5.338
43. Pennebaker JW, Chung CK. Expressive writing: connections to physical and mental health. In: *The Oxford handbook of health psychology*. Oxford University Press (2011). p. 417–37. Available from: <https://doi.org/10.1093/oxfordhb/9780195342819.013.0018>.
44. Burlinson BR. Emotional support skill. In: *Handbook of communication and social interaction skills* (2003). p. 551.
45. Heaney CA, Israel BA. Social networks and social support. *Health behavior and health education: theory, research, and practice*. Vol. 4 San Francisco: Jossey-Bass A Wiley Imprint (2008). p. 189–210.
46. Brislin RW. Back-translation for cross-cultural research. *J Cross Cult Psychol.* (1970) 1:185–216. doi: 10.1177/135910457000100301
47. [Dataset] Gu Y, Wen J, Sun H, Song Y, Ke P, Zheng C, et al. Eva2.0: investigating open-domain chinese dialogue systems with large-scale pre-training (2022). Available from: <https://doi.org/10.48550/arXiv.2203.09313>.
48. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. Roberta: A robustly optimized BERT pretraining approach. *CoRR* (2019). Available from: <https://arxiv.org/abs/1907.11692>.
49. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann.* (2002) 32:509–15. doi: 10.3928/0048-5713-20020901-06
50. von Glischinski M, von Brachel R, Thiele C, Hirschfeld G. Not sad enough for a depression trial? A systematic review of depression measures and cut points in clinical trial registrations. *J Affect Disord.* (2021) 292:36–44. doi: 10.1016/j.jad.2021.05.041
51. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder. *Arch Intern Med.* (2006) 166:1092. doi: 10.1001/archinte.166.10.1092
52. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the panas scales. *J Pers Soc Psychol.* (1988) 54:1063–70. doi: 10.1037/0022-3514.54.6.1063
53. Morin CM. *Insomnia: psychological assessment and management*. Washington: Guilford Press (1993).
54. Little R, Rubin D. *Statistical analysis with missing data*. Wiley Series in Probability and Mathematical Statistics. Wiley (2002).
55. McCoy E. Understanding the intention-to-treat principle in randomized controlled trials. *West J Emerg Med.* (2017) 18:1075–8. doi: 10.5811/westjem.2017.8.35985
56. Alshurafa M, Briel M, Akl EA, Haines T, Moayyedi P, Gentles SJ, et al. Inconsistent definitions for intention-to-treat in relation to missing outcome data: systematic review of the methods literature. *PLoS ONE.* (2012) 7:e49163. doi: 10.1371/journal.pone.0049163
57. Harrison CN, Schaap N, Vannucchi AM, Kiladjian J, Jourdan E, Silver RT, et al. Fedratinib in patients with myelofibrosis previously treated with ruxolitinib: An updated analysis of the jakarta2 study using stringent criteria for ruxolitinib failure. *Am J Hematol.* (2020) 95:594–603. doi: 10.1002/ajh.25777
58. Cohen J. *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates (1988).
59. Pelau C, Dabija D-C, Ene I. What makes an ai device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. *Comput Human Behav.* (2021) 122:106855. doi: 10.1016/j.chb.2021.106855
60. [Dataset] Xu J, Szlam A, Weston J. Beyond goldfish memory: long-term open-domain conversation (2021). Available from: <https://doi.org/10.48550/ARXIV.2107.07567>.
61. Zheng Y, Chen G, Huang M, Liu S, Zhu X. Personalized dialogue generation with diversified traits [Preprint] (2019). Available at: <https://doi.org/10.48550/arXiv.1901.09672>.
62. Wang W, Cai X, Huang CH, Wang H, Lu H, Liu X, et al. Emily: developing an emotion-affective open-domain chatbot with knowledge graph-based persona. *CoRR* (2021). Available from: <https://arxiv.org/abs/2109.08875>.
63. Wu CH, Zheng Y, Mao X, Huang M. Transferable persona-grounded dialogues via grounded minimal edits. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* Bangkok, Thailand: MIT Press (2021). p. 2368–82.
64. [Dataset] Li Q, Li P, Chen Z, Ren Z. Towards empathetic dialogue generation over multi-type knowledge (2020).
65. Sabour S, Zheng C, Huang M. CEM: commonsense-aware empathetic response generation. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. Palo Alto California: AAAI Press (2022). p. 11229–37. Available from: <https://doi.org/10.1609/aaai.v36i10.21373>.