

Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness[☆]

Hao Liu^a, Huaming Peng^a, Xingyu Song^b, Chenzi Xu^c, Meng Zhang^{c,*}

^a School of Design, South China University of Technology, B11 Building, University Town Campus, South China University of Technology, Guangzhou Higher Education Mega Center, Panyu District, 510006 Guangzhou, China

^b School of Psychology, Central China Normal University, The 8th Floor, Nanhu Complex Building, Central China Normal University, No. 152 Luoyu Road, Wuhan 430079, China

^c Social Work Research Center, South China University of Technology, 5th Building, Wushan Campus, South China University of Technology, 381 Wushan Road, Tianhe District, Guangzhou 510640, China

ARTICLE INFO

Keywords:

Public health informatics
AI Artificial Intelligence
mHealth

ABSTRACT

Background: Depression impacts the lives of a large number of university students. Mobile-based therapy chatbots are increasingly being used to help young adults who suffer from depression. However, previous trials have short follow-up periods. Evidence of effectiveness in pragmatic conditions are still in lack.

Objective: This study aimed to compare chatbot therapy to bibliotherapy, which is a widely accepted and proven-useful self-help psychological intervention. The main objective of this study is to add to the evidence of effectiveness for chatbot therapy as a convenient, affordable, interactive self-help intervention for depression.

Methods: An unblinded randomized controlled trial with 83 university students was conducted. The participants were randomly assigned to either a chatbot test group ($n = 41$) to receive a newly developed chatbot-delivered intervention, or a bibliotherapy control group ($n = 42$) to receive a minimal level of bibliotherapy. A set of questionnaires was implemented as measurements of clinical variables at baseline and every 4 weeks for a period of 16 weeks, which included the Patient Health Questionnaire-9 (PHQ-9), the Generalized Anxiety Disorder scale (GAD-7), the Positive and Negative Affect Scale (PANAS). The Client Satisfaction Questionnaire-8 (CSQ-8) and the Working Alliance Inventory-Short Revised (WAI-SR) were used to measure satisfaction and therapeutic alliance after the intervention. Participants' self-reported adherence and feedback on the therapy chatbot were also collected.

Results: Participants were all university students (undergraduate students ($n = 31$), postgraduate students ($n = 52$)). They were between 19 and 28 years old (mean = 23.08, standard deviation (SD) = 1.76) and 55.42% (46/83) female. 24.07% (20/83) participants were lost to follow-up. No significant group difference was found at baseline. In the intention-to-treat analysis, individuals in the chatbot test group showed a significant reduction in the PHQ-9 scores ($F = 22.89$; $P < 0.01$) and the GAD-7 scores ($F = 5.37$; $P = 0.02$). Follow-up analysis of completers suggested that the reduction of anxiety was significant only in the first 4 weeks. The WAI-SR scores in the chatbot group were higher compared to the bibliotherapy group ($t = 7.29$; $P < 0.01$). User feedback showed that process factors were more influential than the content factors.

Conclusions: The chatbot-delivered self-help depression intervention was proven to be superior to the minimal level of bibliotherapy in terms of reduction on depression, anxiety, and therapeutic alliance achieved with participants.

Abbreviations: IPI, Internet-based Psychological Interventions; AI, Artificial Intelligence; PHQ-9, the Patient Health Questionnaires-9; CBT, Cognitive Behavioral Therapy; GAD-7, the Generalized Anxiety Disorder Scale-7 (GAD-7); PANAS, the Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988); CSQ-8, the Client Satisfaction Questionnaires-8; WAI-SR, the Working Alliance Inventory-Short Revised; ANOVA, Analysis of Variance; ANCONA, Analysis of Covariance; ITT, Intent-to-Treat; SD, Standard Deviation; DST, Dialogue Status Tracking; DPO, Dialogue Policy Optimization.

[☆] All authors have read and approved to submit of our manuscript to Internet Interventions. There is no conflict of interest of any authors in relation to the submission. This paper has not been submitted elsewhere for consideration of publication.

* Corresponding author.

E-mail addresses: sdliuhao@mail.scut.edu.cn (H. Liu), hmpeng@scut.edu.cn (H. Peng), mengzh@scut.edu.cn (M. Zhang).

<https://doi.org/10.1016/j.invent.2022.100495>

Received 17 September 2021; Received in revised form 14 December 2021; Accepted 4 January 2022

Available online 6 January 2022

2214-7829/© 2022 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Mental health is an essential component of human well-being. Depression is one of the most common mental illnesses around the world. As of 2015 (World Health Organization, 2017), approximately 322 million people were suffering from depression worldwide, accounting for 4.4% of the global population. Evidence suggests that depression is extremely dangerous because it is linked to suicide attempts (Hawton et al., 2013). Depression is prevalent among university students (Guo et al., 2020; Sheldon et al., 2021), where it leads to poor academic performance (Andrews and Wilding, 2004; Hysenbegasi et al., 2005), unhealthy interpersonal relationships (Lee et al., 2021), and low quality of life (Zhong et al., 2019). Despite the fact that most universities offer free therapy for students, many students refuse to seek help when they are suffering from mental health issues due to the reason of low perceived need (Andrade et al., 2014), attitude barriers (Andrade et al., 2014; Neathery et al., 2020), and the lack of mental health education (Neathery et al., 2020). Scholars have proposed numerous solutions to the aforementioned problems, including IPI.¹ IPI (Andersson, 2016; Wang et al., 2018) has been shown to be effective because clients can have private, self-helping interventions whenever and wherever they want. IPI can help to reduce the stigma associated with seeing a therapist, but it struggles to establish a therapeutic alliance with the clients (Donkin et al., 2013) and have a relatively poor adherence rate, which was reported in (Kaiser et al., 2021) to have a positive impact on user acceptance and adherence.

The great leap of AI² technology over the last decade has resulted in a growing body of literature recognizing the significance of AI in the domain of IPI. Recent evidence indicated that AI chatbots are a potential effective mediator for fully automatic IPI (Gratzer and Goldbloom, 2020; Vaidyam et al., 2019) because they can better mirror human therapists. An early study (Bickmore et al., 2005) created an embodied anthropomorphic chatbot intended to increase exercise, demonstrating that a therapeutic alliance can be established between a client and an embodied chatbot. Previous randomized controlled trials (Burton et al., 2016; Shinichiro et al., 2018) validated the efficacy of psychological treatments delivered by embodied agents. However, most chatbots for customer use are not embodied. Instead, a review about synchronous text-based dialogue systems in mental health (Hoermann et al., 2017) recognized the significance of text-based therapy chatbots. In recent years, a substantial body of literature has accumulated on the topic of delivering therapy through chatbots on mobile devices. Therapy chatbots have been shown to have a considerably higher adherence rate (Vaidyam et al., 2019) than the previous IPI. In addition, a study (Lee et al., 2020) documented an increase in the level of client self-disclosure after using a chatbot for therapy. Researchers have so far endeavored to deploy mobile-based therapy chatbots for college students (Dekker et al., 2020), teenagers (Huang et al., 2015), the elderly (Ryu et al., 2020), patients (Greer et al., 2019), and so on. As documented in a review (Abd-Alrazaq et al., 2019), the most hotly discussed topics of chatbot-delivered therapy are depression and autism. Although a growing number of controlled studies are in favor of the use of chatbots in depression interventions, the current study about this topic is still at a preliminary level and requires further research (Abd-Alrazaq et al., 2020; Vaidyam et al., 2019). To the best of our knowledge, previous randomized controlled trials of chatbot-delivered self-help depression intervention were all conducted in a relatively short period of time (mostly 4 weeks) and failed to reveal the pragmatic effectiveness of the depression therapy chatbot because the disease course of depression can usually last for three to four months.

The objectives of this study were to examine the superiority of a newly developed chatbot-delivered self-help depression intervention to

a minimal level of bibliotherapy regarding (1) efficacy on depression symptoms reduction, (2) adherence, and (3) therapeutic alliance. We had three hypotheses: compared to the bibliotherapy, the chatbot intervention would (1) reduce self-reported depression symptoms more significantly over the 16-week follow-up period, (2) maintain a higher adherence rate in the 16-week follow-up period and (3) achieve a better therapeutic alliance with the participants.

In this study, a therapy chatbot called XiaoNan was developed. A 16-week randomized controlled trial of effectiveness was conducted with 83 Chinese university students. This study evaluated the potential of chatbot-delivered therapy as an effective, convenient, and scalable solution for self-help depression treatments.

2. Methods

2.1. Participants and sample size

Participants were recruited using an online poster from three different universities in China, respectively in the city of Harbin, Wuhan, and Guangzhou. The acceptable level of significance was set at $P = 0.05$, with an expected power of the study $1 - \beta = 80\%$. The expected efficacy obtained from previous studies was a reduction of depression as measured by PHQ-9³ (Kroenke et al., 2001) score of 3.6, and a standard deviation = 5. Accordingly, the expected sample size was calculated to be $n = 31$ in each group. Since there was an attrition rate of 15% in previous studies, the expected number of participants recruited to each group was 37.

The PHQ-9, which is a commonly used measure in psychological depression trials (von Glitschinski et al., 2021), was used to select participants. Individuals who met the following criteria were included in the trial: aged 18 years or older, being full-time university students, able to communicate in Chinese without difficulty, have skillful use of smartphones, have a PHQ-9 score of nine or higher, not currently undergoing any form of mental health intervention. The eligibility criterion of PHQ-9 score equal to or greater than nine was decided according to the average inclusion PHQ-9 score in previous depression trials (von Glitschinski et al., 2021). All participants had confirmed their consent to participation and were provided with the phone number of the local free Mental Assistance Hotline in consideration of safety reasons. Procedures of this trial were approved by the Institutional Ethics Committee of South China University of Technology.

2.2. Interventions

2.2.1. Chatbot-delivered intervention

A therapy chatbot named XiaoNan was developed for the trial. The chatbot was deployed through the “WeChat Official Accounts Platform”. Users can use the smartphone app “WeChat” to access the chatbot, which is available on Windows, macOS, Android, and iOS.

XiaoNan is a pipeline-based chatbot powered by the open-source conversational AI “RASA”. The text contents of the chatbot were based on CBT⁴ principles, which were reviewed and approved by professional therapists. The overall workflow of XiaoNan is illustrated in Fig. 1(A), and the structure of XiaoNan is presented in Fig. 1(B). Input messages can be either text or voice messages. When a voice message is received, it's first transferred to text via a natural speech recognition service that is provided by “IFLYTEK Open Platform”. Next, the text is sent to XiaoNan for natural language understanding. The language understanding module included three machine learning models: natural language processing, intention classification, and emotion recognition. The natural language processing model is responsible for collecting necessary information from the users to support the conversation by handling

¹ Internet-based Psychological Intervention.

² Artificial Intelligence.

³ The Patient Health Questionnaire.

⁴ Cognitive Behavior Therapy.

“entities”, “slots”, and “forms”. The intention classification and emotion recognition models were able to label the input text with pre-defined intention and emotion tags. Finally, the output text message will be decided by the dialogue management module using a template referred to as “domain” and generated in the natural language generation module based on a response database. The “domain” included a pre-defined set of “entities”, “slots”, “forms”, “rules” and “stories”. “Entities”, “slots”, and “forms” were used to contain dynamic information needed in the responses. The “rules” and “stories” were based on machine learning models provided by “RASA”. The “rules” specify commands that are enforced, for the intention “goodbye”, for example, a “rule” will force the chatbot to generate a respond to end the conversation. “Stories”, on the other hand, are a general guide to responding, with logical connections between the elements, which were created according to the principles of CBT. All machine learning models were trained to meet a minimal classification accuracy requirement of 90%. Fig. 2 depicts examples of interacting with XiaoNan. The chatbot was trained to alleviate depression. It will estimate the degree of depression in the input and then illustrate CBT to the user using a pre-written template. The chatbot can help the user to separate their emotions, thoughts, reactions, and behavior and to establish new automatic thoughts, as if in a CBT session. Users may also use XiaoNan to keep track of their daily emotions or have

random conversations, the chatbot will respond empathetically.

2.2.2. Bibliotherapy intervention

Bibliotherapy is one of the most significant forms of self-help therapy. It refers to the treatment method that uses literature to alleviate a patient's problems. In bibliotherapy, patients receive psychological intervention by reading literature following the advice of professionals. Evidence from previous studies shows that bibliotherapy is effective for the treatment of moderate depression. The long-term effects of bibliotherapy were also verified in randomized clinical trials as documented in a systematic review (Gualano et al., 2017). It's reported that adult patients who received cognitive bibliotherapy can maintain a good level of improvement in a 3-month follow-up period (Floyd et al., 2004; Jamison and Scogin, 1995). And the effects of bibliotherapy are especially noticeable in the change of automatic thoughts (Moldovan et al., 2013).

2.2.3. Participants administration

The participants were randomly allocated to either a chatbot test group or a bibliotherapy control group. Participants who were assigned to the chatbot test group were provided with a link to the therapy chatbot XiaoNan. They were asked to use the chatbot for 16 weeks. Meanwhile, participants in the bibliotherapy control group received a

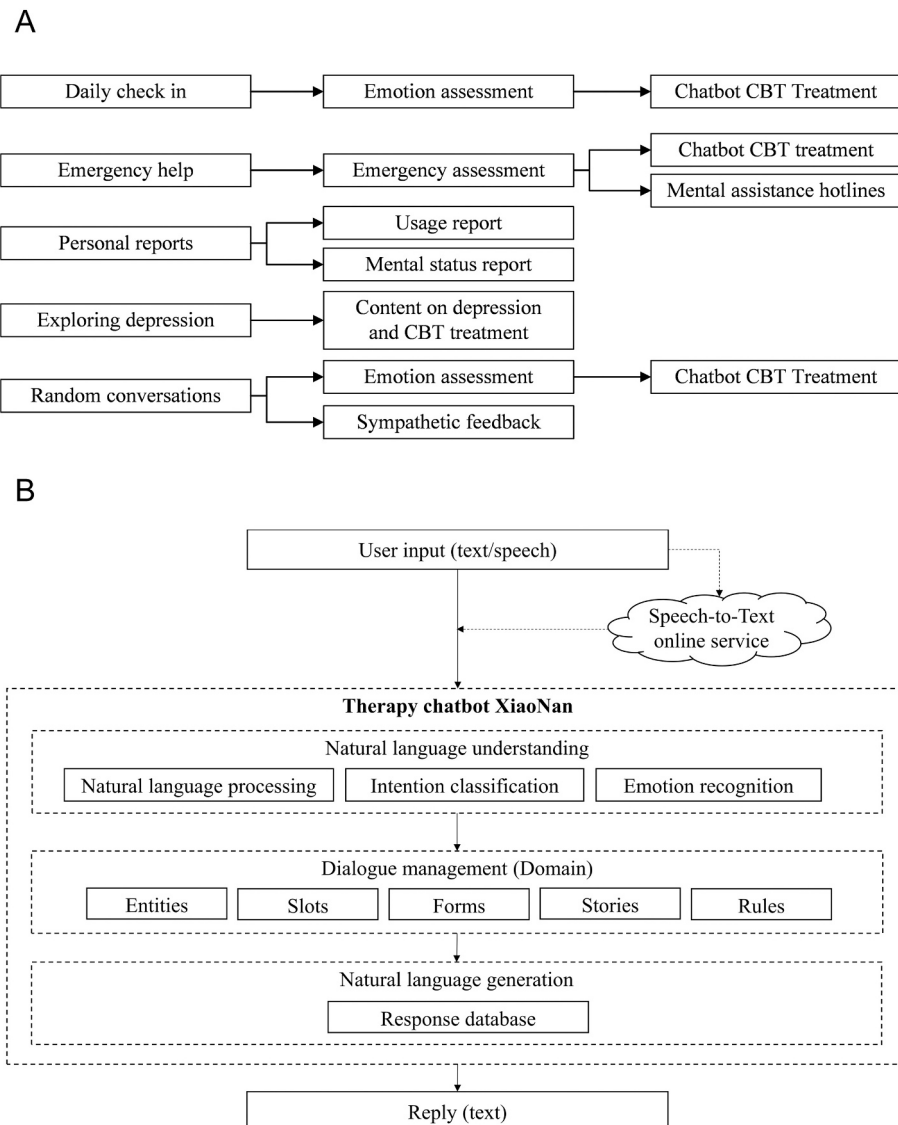


Fig. 1. (A) The workflow of the chatbot XiaoNan. (B) The structure of the chatbot XiaoNan.

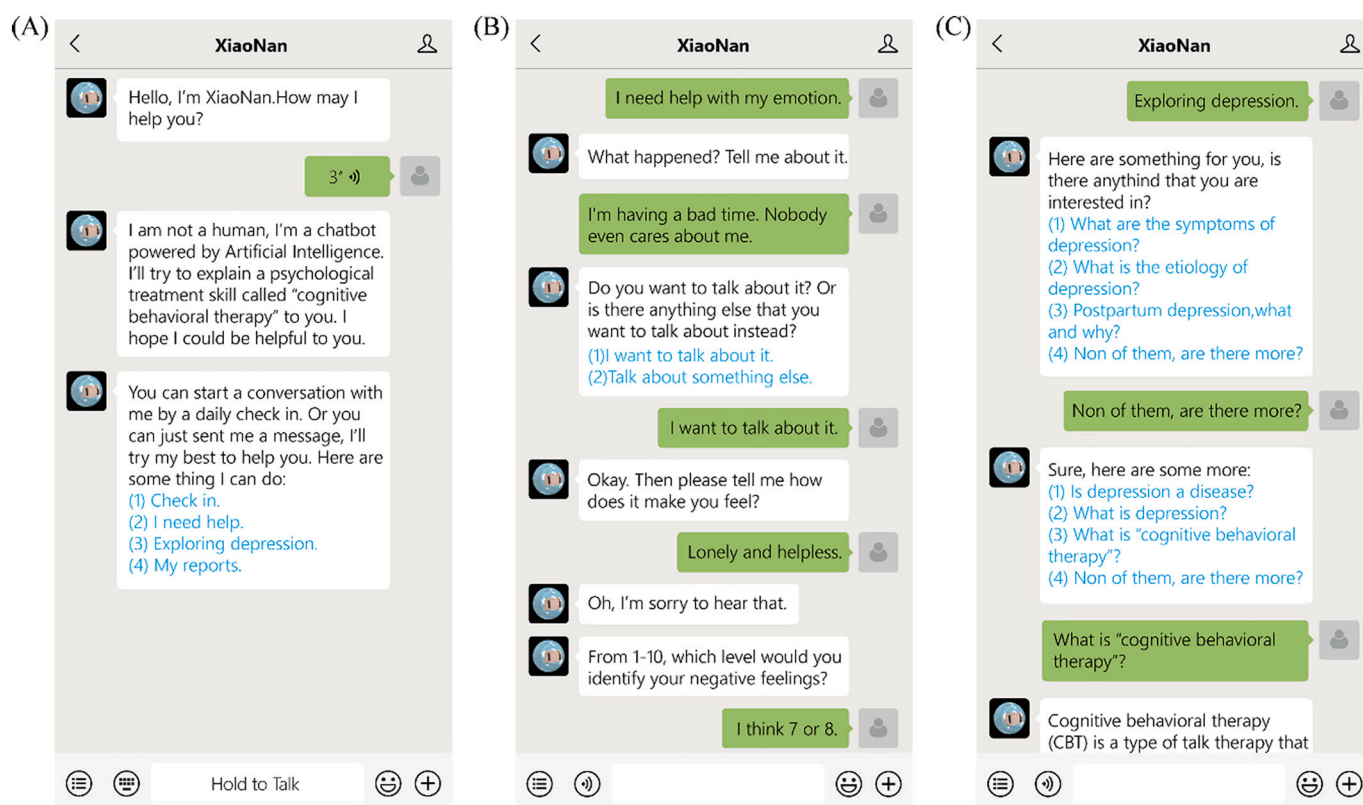


Fig. 2. Examples of using the chatbot. (A) Both text and voice messages are supported. There will be instructions when using the chatbot for the first time. Users can select the options in the choice list by clicking the text or replying with relevant number or contents. (B) An example of CBT treatment. The chatbot will try to recognize, evaluate, and deal with negative emotions from the input text. (C) “Exploring depression” provides a question answering system on the topic of depression disorder.

minimal level of bibliotherapy intervention: they were asked to read the book *Change Your Thinking-Positive and practical ways to overcome stress negative emotions and self-defeating behavior using CBT* (Edelman, 2007), a book intended to help people with common mental health issues including depression using CBT principles, in a self-help manner during the 16 weeks.

Participants were instructed not to seek any other treatment other than the intervention they were offered during the trial. For consideration of safety, professionals will intervene by telephone (Mental Assistance Hotline) when the participants report that they need emergency psychological assistance. The professional will remain in contact with the patient until the patient is ensured to be safe. However, in such cases, the data of the involved patient will be excluded from the statistical analysis.

2.3. Measures

The primary outcome was the reduction of depression, anxiety, positive and negative affect symptoms. Whereas user satisfaction, therapeutic alliance, and adherence were considered as the secondary outcome. Key measures used in the trial was the PHQ-9, GAD-7⁵ (Spitzer et al., 2006), PANAS⁶ (Watson et al., 1988). Other measures included CSQ-8⁷ (Kelly et al., 2018), WAI-SR⁸ (Munder et al., 2010), and self-reported adherence as measured by a 4-point Likert scale (not used = 1; several days = 2; over half the days = 3; nearly every day = 4).

⁵ The Generalized Anxiety Disorder Scale-7.

⁶ The Positive and Negative Affect Schedule.

⁷ The Client Satisfaction Questionnaire-8.

⁸ The Working Alliance Inventory-Short Revised.

Two open-ended questions were asked at the end of the intervention in the chatbot group to gain feedback on the chatbot. The questions were about their experience of using the chatbot. The first question was “What was the best thing about using XiaoNan?”, the other was “What was the worst thing about using XiaoNan?”

2.4. Randomization

Blinding was not applicable because chatbot interventions and bibliotherapy interventions are delivered in different ways, which was apparent to participants. The participants were given a random number $0 < n \leq 1$, which was generated with the random number generator in *Statistical Package for the Social Science* (SPSS) v.26 (IBM Corp., Armonk, NY). Those who received a number $0 < n \leq 0.5$ were assigned to the chatbot test group while others were allocated to the control group.

2.5. Statistical methods

SPSS v.26 was used to evaluate the results, with the significance level set at 0.05. ANOVA⁹ and chi-square analysis was used on baseline variables (age, gender, education, baseline clinical variables) to see whether there were any major discrepancies between groups. For primary outcomes, an univariable ANCOVA¹⁰ was used in ITT¹¹ analysis to evaluate the univariate impact of group membership on the results after corrected for baseline clinical variables. Missing data, which was assumed to be missing at random, was handled using the multiple

⁹ Analysis of Variance.

¹⁰ Analysis of Covariance.

¹¹ Intention to Treat.

imputation procedure in SPSS. Cohen's *d* effect sizes were calculated. Independent *t*-tests were conducted on the CSQ-8 scores, the WAI-SR scores, and self-reported adherence. Lastly, answers to the two open-ended questions were collected and a word frequency analysis was conducted.

3. Results

3.1. Participant flow

The recruitment started on February 11th and finished on April 15th, 2021. A total number of 187 responses were received, among which 83 participants (46 females) were considered eligible for the trial. All of the participants were university students (undergraduate students (*n* = 31) or postgraduate students (*n* = 52)). They were between 19 and 28 years old (mean = 23.08, SD¹² = 1.76). All of them were native Chinese speakers. The trial started on April 17th, 2021, and lasted for 16 weeks, during which period participants were required to provide follow-up data every four weeks (baseline = T₁, 4 weeks = T₂, 8 weeks = T₃, 12 weeks = T₄, 16 weeks = T₅). At T₅, client satisfaction and working alliance were evaluated with CSQ-8 and WAI-SR.

Twenty participants were lost to follow-up. The overall flow of the participants is shown in Fig. 3. There was an overall attrition rate of 24.10% (20/83). No evidence of significant differences at baseline were detected by independent *t*-tests and Chi-square analyses between those who completed the study versus those who dropped out on group membership ($\chi^2 = 0.93$; *P* = 0.34); gender ($\chi^2 = 0.22$; *P* = 0.64); age (*t* = 0.53; *P* = 0.60); education (*t* = 0.44; *P* = 0.66); PHQ-9 (*t* = 0.28; *P* = 0.78); GAD-7 (*t* = 1.70; *P* = 0.09); PANAS positive (*t* = 0.08; *P* = 0.93) and negative (*t* = 0.55; *P* = 0.58) affect scores.

3.2. Baseline data

As illustrated in Table 1, ANOVA and Chi-square analyses detected no significant difference between groups at baseline (T₁) in terms of age, gender, education, and clinical variables as measured by scales.

3.3. ITT analysis

The primary outcome variables measured at T₅ were used for the ITT analysis. As is presented in Table 2, univariate ANCOVA revealed that chatbot-enhanced self-help intervention showed a significant advantage to bibliotherapy in terms of the reduction of depression with a high effect size (*d* = 0.83) and anxiety with a low effect size (*d* = 0.30) as measured by the PHQ-9 (*F* = 22.89; *P* < 0.01) and the GAD-7 (*F* = 5.37; *P* = 0.02) in the 16-week period. The trends of the significantly changed variables (PHQ-9 score and GAD-7 score) are illustrated in Fig. 4. The reduction effect of depression (*P* < 0.01) and anxiety (*P* = 0.046) was robust after Bonferroni correction for multiple comparisons. No significant between-group difference was observed on positive and negative affect.

3.4. Ancillary analyses

Multivariable ANOVA was implemented among those who completed the trial to detect if any main effect existed. Results showed that group membership (*F* = 37.77; *P* < 0.01) and time (*F* = 8.40; *P* < 0.01) both had significant influences on the PHQ-9 scores; Group membership had a main effect on the GAD-7 scores (*F* = 15.89; *P* < 0.01). Following univariable ANOVAs between different-time variables showed that a main effect of time in the first 4 weeks (T₁ and T₂) (*F* = 0.50; *P* = 0.48), indicating that the group membership difference was made in the first 4 weeks. No other main effect was found.

Independent *t*-tests were conducted on completers at T₅, showing that the therapeutic alliance as measured by the WAI-SR in the chatbot tests group was significantly better than in the bibliotherapy control group (*t* = 7.29; *P* < 0.01) with a high effect size *d* = 1.85. Whereas no evidence of differences was found on the CSQ-8 scores (*t* = 0.88; *P* = 0.38).

The self-reported adherence rates of the chatbot test group and bibliotherapy group were 1.96 (SD = 0.70) and 1.97(0.62), respectively. Independent *t*-test detected no significant difference between the adherence rates of the two groups (*t* = 0.68; *P* = 0.50). However, as is shown in Fig. 5. During the period, the adherence rate of chatbot group participants witnessed a downward trend, whereas that of the bibliotherapy maintained a slight increase in the first 8 weeks. No adverse events were observed in the trial.

Table 3 illustrates the keywords of the answers to “What was the best thing about using XiaoNan?” and “What was the worst thing about using XiaoNan?”. A word frequency analysis showed two major themes: process and content. In addition, it is worth noting that one participant said that XiaoNan had increased her willingness to seek help from professional therapists. The feedback showed that process factors are slightly more important than content factors.

4. Discussion

The results showed that chatbots are an effective medium for self-help depression treatment in pragmatic conditions. The therapy chatbot reduced depression as measured by the PHQ-9 significantly in a 16 weeks period and reduced anxiety as measured by the GAD-7 slightly in the first 4 weeks. The effectiveness of the chatbot intervention was superior to that of the bibliotherapy. The client satisfaction and self-reported adherence were similar in the two groups. However, a decreasing trend was found in the chatbot adherence rate, this might have been resulted from the technical and content defects of the chatbot. The WAI-SR scores of the chatbot users were higher in the chatbot users, showing that the use of conversational AI can help build a therapeutic alliance. The results were consistent with some of the previous studies (Fitzpatrick et al., 2017; Fulmer et al., 2018), in which there were shorter intervention periods of 2 and 4 weeks. A previous study (Lee et al., 2020) has reported an increase in self-disclosure, this phenomenon was also observed in user feedback where expressions of higher self-disclosure were found. In addition, feedback on XiaoNan showed that process factors are more influential than the content.

Several limitations can be found in this study. Firstly, the content of the chatbot was limited as we aimed to make sure all the contents were approved by professionals. An expansion of topics and contents could reduce the repetitiveness of the conversation and thus bring potential improvements. Future work could analyze the content of repetitive conversations generated by the chatbot and add more clinically robust content to the chatbot accordingly. From the technical perspective, the future version of the chatbot could include DST¹³ and DPO¹⁴ modules to reduce repetitive and improve personalization. During a conversation, the DST module is responsible for handling the start, pause, and end of the conversation. It will also match the input text with previous context to find connections and keep a short memory to support the conversation. The DPO module will analyze the successfulness of each intervention after a conversation (Meinlschmidt et al., 2020) and keep a long-term memory to tag and file all conversations. Key information identified during conversations such as patient complaints and severe symptoms will be recorded in the long-term memory and discussed with the user in further sessions. In addition, the conversations will be compared regularly to keep track of the symptoms and requirements of the patient so that a personalized treatment plan could be decided for further

¹² Standard Deviation.

¹³ Dialogue Status Tracking.

¹⁴ Dialogue Policy Optimization.

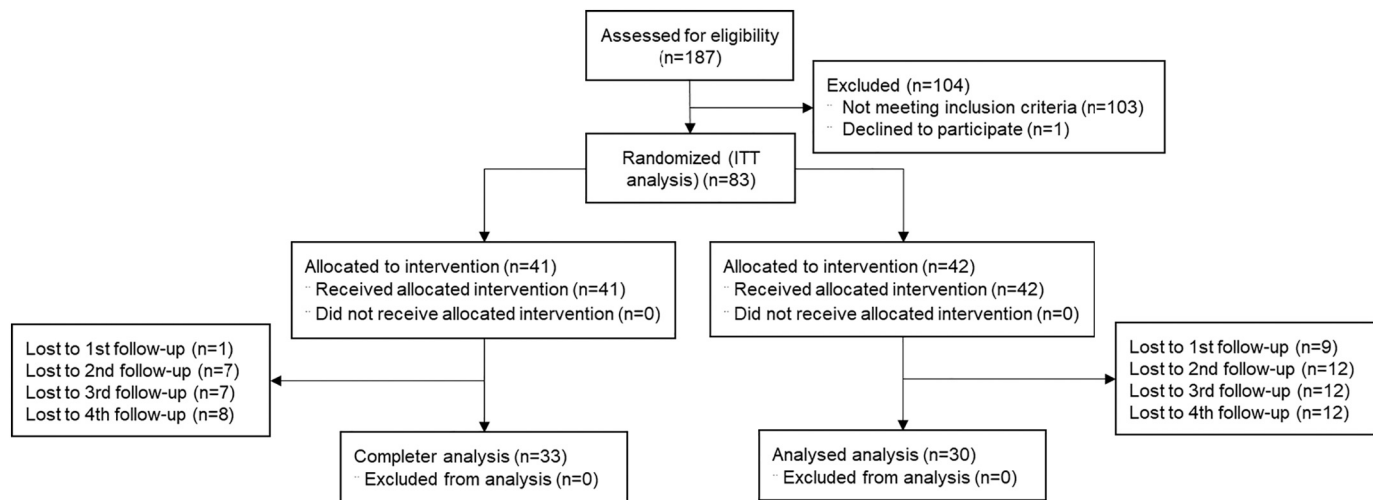


Fig. 3. The flow of participants.

Table 1
Demographics of participants and variables at baseline (T₁).

	Chatbot test group ^a	Bibliotherapy control group ^a	χ^2/t	P
Age (years)	23.41 (1.77)	22.76 (1.70)	1.69	0.09
Gender				
Male	17 (41.46)	20 (47.62)	0.32	0.57
Female	24 (58.54)	22 (52.38)		
Education (years)	17.29 (1.78)	16.62 (1.65)	1.77	0.08
Scale, mean (SD)				
Depression (PHQ-9)	13.17 (3.32)	13.59 (4.44)	0.49	0.63
Anxiety (GAD-7)	15.59 (3.70)	16.69 (3.77)	1.33	0.19
Positive affect	28.17 (8.49)	27.24 (9.39)	0.47	0.64
Negative affect	27.07 (9.60)	28.10 (8.92)	0.50	0.62

^a The numbers are mean (standard deviation) or n (%).

treatments. For instance, if insomnia is detected in a conversation, the DPO module will keep track of this symptom. In future conversations, the DPO module will compare the description of insomnia to previous records and provide necessary guidance for the user. Next, the therapy target of XiaoNan in this trial was restricted to depression. However, the symptoms of clients are often a mixture of two or more issues. Given the fact that chatbot-delivered therapy for other targets, such as panic disorder (Oh et al., 2020), substance addiction (Barnett et al., 2020), and attention deficit (Jang et al., 2021), have already been under evaluation. A combination of two or more topics might be beneficial to the effectiveness. Moreover, this study has recruited a participant sample that lacks variety. In future work, this randomized controlled trial should be repeated within different demographics to increase the generalizability of the results. Lastly, a longitude study as was implemented in (Bendig

et al., 2021) could be implemented to test the effectiveness over a longer period of time.

Funding

This work did not receive any funding.

Ethics approval

The study design and protocol were approved by the Institutional Ethics Committee of South China University of Technology. The research was carried out in conjunction with the Declaration of Helsinki. Written informed consent was obtained from all participants.

Availability of data and material

The datasets generated for this study are available on reasonable request to the corresponding author.

CRediT authorship contribution statement

Hao Liu: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft, Validation. **Huaming Peng:** Supervision, Validation. **Xingyu Song:** Methodology, Investigation, Validation. **Chenzi Xu:** Methodology, Investigation, Validation. **Meng Zhang:** Supervision, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial

Table 2
ITT analysis at T₅.

	Chatbot test group		Bibliotherapy control group		F	P	d ^c
	T ₅ ^a	95%CI ^b	T ₅ ^a	95%CI ^b			
PHQ-9	7.92 (0.48)	6.98–8.86	10.61 (0.53)	9.54–11.68	22.89	<0.01**	0.83
GAD-7	14.23 (0.34)	13.56–14.89	14.97 (0.42)	14.11–15.84	5.38	0.02*	0.30
Positive affect	28.29 (0.27)	27.76–28.83	28.65 (0.37)	27.87–29.43	2.77	0.10	0.17
Negative affect	27.80 (0.77)	26.28–29.31	27.27 (1.12)	24.80–19.75	3.53	0.64	0.08

** The result is significant at the 0.01 level.

* The result is significant at the 0.05 level.

^a The numbers are pooled mean (standard error).

^b 95% Confidence Interval.

^c Cohen d shown for between-subjects effects using means and standard errors at T₅.

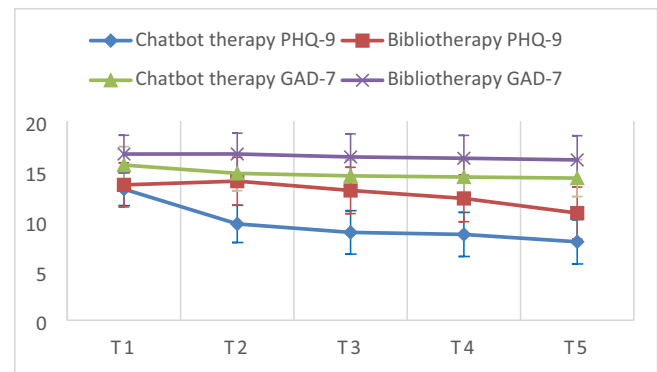


Fig. 4. Clinical variables during the period.
^aNumbers are mean (SD).

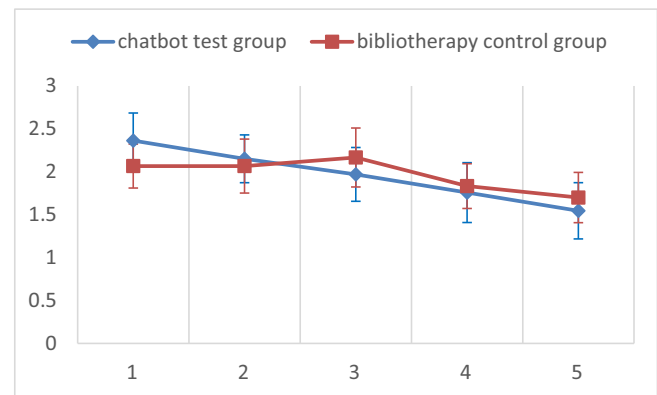


Fig. 5. Self-reported adherence rate.
^aNumbers are mean (SD).

Table 3
Answers to the question “What was the best and worst thing about using XiaoNan?”.^a

The best thing about using XiaoNan.	Process (25/33) ^b	Easy to access (11/33) Empathy/Friendly (8/33) Interesting (7/33) Educational (5/33)
	Content (15/33)	Exploring depression (9/33) Interactive CBT (5/33) Choice list (3/33)
		Impersonal (8/33) Unnatural (7/33) Rigid patterns (7/33) Misunderstanding (5/33)
The worst thing about using XiaoNan.	Process (25/33)	Repetitive contents (10/33) Too general (8/33) Irrelevant contents (4/33) Too simple (2/33)
	Content (21/33)	

^a Some answers have multiple themes and were counted multiple times.
^b Numbers are (counted number/total number of participants).

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank all participants for attending the trial.

References

Abd-alrazaq, A.A., Alajlani, M., Alalwan, A.A., Bewick, B.M., Gardner, P., Househ, M., 2019. An overview of the features of chatbots in mental health: a scoping review. *Int. J. Med. Inform.* 132, 103978 <https://doi.org/10.1016/j.ijmedinf.2019.103978>.

Abd-Alrazaq, A.A., Rababeh, A., Alajlani, M., Bewick, B.M., Househ, M., 2020. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. *J. Med. Internet Res.* 22, 1–17. <https://doi.org/10.2196/16021>.

Andersson, G., 2016. Internet-delivered psychological treatments. *Annu. Rev. Clin. Psychol.* <https://doi.org/10.1146/annurev-clinpsy-021815-093006>.

Andrade, L.H., Alonso, J., Mneimneh, Z., Wells, J.E., Al-Hamzawi, A., Borges, G., Bromet, E., Bruffaerts, R., De Girolamo, G., De Graaf, R., Xavier, M., Kessler, R.C., 2014. Barriers to mental health treatment: results from the WHO world mental health surveys. *Psychol. Med.* 44, 1303–1317. <https://doi.org/10.1017/S0033297113001943>.

Andrews, B., Wilding, J.M., 2004. The relation of depression and anxiety to life-stress and achievement in students. *Br. J. Psychol.* 95, 509–521. <https://doi.org/10.1348/0007126042369802>.

Barnett, A., Savic, M., Pienaar, K., Carter, A., Warren, N., Sandral, E., Manning, V., Lubman, D.I., 2020. Enacting ‘more-than-human’ care: clients’ and counsellors’ views on the multiple affordances of chatbots in alcohol and other drug counselling. *Int. J. Drug Policy* 102910. <https://doi.org/10.1016/j.drugpo.2020.102910>.

Bendig, E., Erb, B., Meißner, D., Bauereiß, N., Baumeister, H., 2021. Feasibility of a software agent providing a brief intervention for self-help to uplift psychological wellbeing (“SISU”). A single-group pretest-posttest trial investigating the potential of SISU to act as therapeutic agent. *Internet Interv.* 24, 100377 <https://doi.org/10.1016/j.invent.2021.100377>.

Bickmore, T., Gruber, A., Picard, R., 2005. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Educ. Couns.* 59, 21–30. <https://doi.org/10.1016/j.pec.2004.09.008>.

Burton, C., Tatar, A.S., McKinstry, B., Matheson, C., Matu, S., Moldovan, R., Macnab, M., Farrow, E., David, D., Pagliari, C., Blanco, A.S., Wolters, M., Help4Mood Consortium, 2016. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. *J. Telemed. Telecare* 22, 348–355. <https://doi.org/10.1177/1357633X15609793>.

Dekker, I., De Jong, E.M., Schippers, M.C., De Bruijn-Smolters, M., Alexiou, A., Giesbers, B., 2020. Optimizing students’ mental health and academic performance: AI-enhanced life crafting. *Front. Psychol.* 11 <https://doi.org/10.3389/fpsyg.2020.01063>.

Donkin, L., Hickie, I.B., Christensen, H., Naismith, S.L., Neal, B., Cockayne, N.L., Glozier, N., 2013. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. *J. Med. Internet Res.* 15, e231 <https://doi.org/10.2196/jmir.2771>.

Edelman, S., 2007. *Change Your Thinking-Positive and practical ways to overcome stress negative emotions and self-defeating behavior using CBT*. In: Da Capo Lifelong Books, Reprint ed.

Fitzpatrick, K.K., Darcy, A., Vierhile, M., 2017. 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. Heal.* 4, e19 <https://doi.org/10.2196/mental.7785>.

Floyd, M., Scogin, F., McKendree-Smith, N.L., Floyd, D.L., Rokke, P.D., 2004. Cognitive therapy for depression: a comparison of individual psychotherapy and bibliotherapy for depressed older adults. *Behav. Modif.* 28, 297–318. <https://doi.org/10.1177/0145445503259284>.

Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., Rauws, M., 2018. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR Ment Heal.* 5, e64 <https://doi.org/10.2196/mental.9782>.

Gratz, D., Goldbloom, D., 2020. Therapy and E-therapy—Preparing future psychiatrists in the era of apps and chatbots. *Acad. Psychiatry* 44, 231–234. <https://doi.org/10.1007/s40596-019-01170-3>.

Greer, S., Ramo, D., Chang, Y.-J., Fu, M., Moskowitz, J., Haritatos, J., 2019. Use of the chatbot “Vivibot” to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. *JMIR Mhealth Uhealth* 7, e15018. <https://doi.org/10.2196/15018>.

Gualano, M.R., Bert, F., Martorana, M., Voglino, G., Andriolo, V., Thomas, R., Gramaglia, C., Zeppegno, P., Siliquini, R., 2017. The long-term effects of bibliotherapy in depression treatment: systematic review of randomized clinical trials. *Clin. Psychol. Rev.* 58, 49–58. <https://doi.org/10.1016/j.cpr.2017.09.006>.

Guo, L.P., Li, Z.H., Chen, T.L., Liu, G.H., Fan, H.Y., Yang, K.H., 2020. The prevalence of mental distress and association with social changes among postgraduate students in China: a cross-temporal meta-analysis. *Public Health*. <https://doi.org/10.1016/j.puhe.2020.06.014>.

Hawton, K., Casañas, I., Comabella, C., Haw, C., Saunders, K., 2013. Risk factors for suicide in individuals with depression: a systematic review. *J. Affect. Disord.* 147, 17–28. <https://doi.org/10.1016/J.JAD.2013.01.004>.

Hoermann, S., McCabe, K.L., Milne, D.N., Calvo, R.A., 2017. Application of synchronous text-based dialogue Systems in Mental Health Interventions: systematic review. *J. Med. Internet Res.* 19, e267 <https://doi.org/10.2196/jmir.7023>.

Huang, J., Li, Q., Xue, Y., Cheng, T., Xu, S., Jia, J., Feng, L., 2015. Teenchat: A chatterbot system for sensing and releasing adolescents’ stress. In: *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 9085, pp. 133–145. https://doi.org/10.1007/978-3-319-19156-0_14.

- Hysenbegasi, A., Hass, S.L., Rowland, C.R., 2005. The impact of depression on the academic productivity of university students. *J. Ment. Health Policy Econ.* 8, 145–151.
- Jamison, C., Scogin, F., 1995. The outcome of cognitive bibliotherapy with depressed adults. *J. Consult. Clin. Psychol.* 63, 644–650. <https://doi.org/10.1037/0022-006x.63.4.644>.
- Jang, S., Kim, J.J., Kim, S.J., Hong, J., Kim, S., Kim, E., 2021. Mobile app-based chatbot to deliver cognitive behavioral therapy and psychoeducation for adults with attention deficit: a development and feasibility/usability study. *Int. J. Med. Inform.* 150, 104440 <https://doi.org/10.1016/j.ijmedinf.2021.104440>.
- Kaiser, J., Hanschmidt, F., Kersting, A., 2021. The association between therapeutic alliance and outcome in internet-based psychological interventions: a meta-analysis. *Comput. Human Behav.* 114, 106512 <https://doi.org/10.1016/j.chb.2020.106512>.
- Kelly, P.J., Kyngdon, F., Ingram, I., Deane, F.P., Baker, A.L., Osborne, B.A., 2018. The client satisfaction Questionnaire-8: psychometric properties in a cross-sectional survey of people attending residential substance abuse treatment. *Drug Alcohol Rev.* 37, 79–86. <https://doi.org/10.1111/dar.12522>.
- Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16, 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>.
- Lee, T.S.H., Wu, Y.J., Chao, E., Chang, C.W., Hwang, K.S., Wu, W.C., 2021. Resilience as a mediator of interpersonal relationships and depressive symptoms amongst 10th to 12th grade students. *J. Affect. Disord.* 278, 107–113. <https://doi.org/10.1016/j.jad.2020.09.033>.
- Lee, Y.C., Yamashita, N., Huang, Y., 2020. Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. In: *Proceedings of the ACM on Human-Computer Interaction*, pp. 1–27. <https://doi.org/10.1145/3392836>.
- Meinlschmidt, G., Tegethoff, M., Belardi, A., Stalujanis, E., Oh, M., Jung, E.K., Kim, H.C., Yoo, S.S., Lee, J.H., 2020. Personalized prediction of smartphone-based psychotherapeutic micro-intervention success using machine learning. *J. Affect. Disord.* 264, 430–437. <https://doi.org/10.1016/j.jad.2019.11.071>.
- Moldovan, R., Cobeau, O., David, D., 2013. Cognitive bibliotherapy for mild depressive symptomatology: randomized clinical trial of efficacy and mechanisms of change. *Clin. Psychol. Psychother.* 20, 482–493. <https://doi.org/10.1002/cpp.1814>.
- Munder, T., Wilmers, F., Leonhart, R., Linster, H.W., Barth, J., 2010. Working Alliance inventory-short revised (WAI-SR): psychometric properties in outpatients and inpatients. *Clin. Psychol. Psychother.* 17, 231–239. <https://doi.org/10.1002/cpp.658>.
- Neathery, M., Taylor, E.J., He, Z., 2020. Perceived barriers to providing spiritual care among psychiatric mental health nurses. *Arch. Psychiatr. Nurs.* 34, 572–579. <https://doi.org/10.1016/j.apnu.2020.10.004>.
- Oh, J., Jang, S., Kim, H., Kim, J.J., 2020. Efficacy of mobile app-based interactive cognitive behavioral therapy using a chatbot for panic disorder. *Int. J. Med. Inform.* 140, 104171 <https://doi.org/10.1016/j.ijmedinf.2020.104171>.
- Ryu, H., Kim, S., Kim, D., Han, S., Lee, K., Kang, Y., 2020. Simple and steady interactions win the healthy mentality: designing a chatbot Service for the Elderly. In: *Proceedings of the ACM on Human-Computer Interaction*. <https://doi.org/10.1145/3415223>.
- Sheldon, E., Simmonds-Buckley, M., Bone, C., Mascarenhas, T., Chan, N., Wincott, M., Gleeson, H., Sow, K., Hind, D., Barkham, M., 2021. Prevalence and risk factors for mental health problems in university undergraduate students: a systematic review with meta-analysis. *J. Affect. Disord.* 287, 282–292. <https://doi.org/10.1016/j.jad.2021.03.054>.
- Shinichiro, S., Daisuke, S., Shimoyama, H., 2018. An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: feasibility and acceptability pilot trial. *JMIR Ment. Heal.* 5, e10454 <https://doi.org/10.2196/10454>.
- Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Intern. Med.* 166, 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>.
- Vaidyam, A.N., Wisniewski, H., Halamka, J.D., Kashavan, M.S., Torous, J.B., 2019. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *Can. J. Psychiatr.* 64, 456–464. <https://doi.org/10.1177/0706743719828977>.
- von Glischinski, M., von Brachel, R., Thiele, C., Hirschfeld, G., 2021. Not sad enough for a depression trial? A systematic review of depression measures and cut points in clinical trial registrations. *J. Affect. Disord.* 292, 36–44. <https://doi.org/10.1016/j.jad.2021.05.041>.
- Wang, K., Varma, D.S., Prosperi, M., 2018. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. *J. Psychiatr. Res.* <https://doi.org/10.1016/j.jpsychires.2018.10.006>.
- Watson, D., Clark, L.A., Tellegen, A., 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. <https://doi.org/10.1037//0022-3514.54.6.1063>.
- World Health Organization, 2017. *Depression and Other Common Mental Disorders Global Health Estimates*. Geneva.
- Zhong, X., Liu, Y., Pu, J., Tian, L., Gui, S., Song, X., Xu, S., Zhou, X., Wang, H., Zhou, W., Chen, J., Qi, X., Xie, P., 2019. Depressive symptoms and quality of life among chinese medical postgraduates: a national cross-sectional study. *Psychol. Heal. Med.* 24, 1015–1027. <https://doi.org/10.1080/13548506.2019.1626453>.