Advances in Artificial Intelligence and Machine Learning; Review 4 (3) 2665-2686 Received 17-07-2024; Accepted 21-09-2024; Published 28-09-2024

Can Chatbots Alleviate Depression? Results of a Systematic Review

Marcela Tabares Tabares marcela.tabares@ucaldas.edu.co
Médica- Mg Salud pública- PhD Ciencias del comportamiento
Investigadora Telesalud- Universidad de Caldas
Colombia.

Edgar León Landa EdgarLeon390@gmail.com
Universidad Internacional de la Rioja (UNIR),
España
Spain.

Consuelo Vélez Álvarez consuelo.velez@ucaldas.edu.co
Grupo Promoción de la Salud y Prevención de la Enfermedad,
Universidad de Caldas,
Colombia.

Paula Carmona Gallego p.a.carmona.g@gmail.com
Grupo Telesalud,
Universidad de Caldas,
Colombia.

Joshua Bernal Salcedo rmosli@kau.edu.sa
Fundación para el Fomento de la Investigación Sanitario y
Biomédica de la Comunidad Valenciana
Unidad Mixta de Imagen Biomédica FISABIO-CIPF Valencia,
España.

Corresponding Author: Marcela Tabares Tabares

Abstract

Purpose: This article systematically reviews the impact of chatbots on the treatment of depression, evaluating studies published between January 2019 and April 2024. Depression, exacerbated by factors such as the COVID-19 pandemic, requires accessible and effective treatments. Chatbots, using artificial intelligence and natural language processing, emerge as accessible alternatives, offering interventions based on Cognitive-Behavioral Therapy (CBT). **Methods**: Using databases such as Cochrane Library, PubMed, Scopus, and ScienceDirect, 321 articles were identified, of which 12 met the inclusion criteria. These studies evaluated changes in depression symptoms using validated instruments in individuals who interacted with chatbots. **Results**: The results indicate that chatbots can significantly reduce depression symptoms, although their effectiveness varies depending on the design and implementation of the intervention. A lack of gender balance and variations in sample sizes and intervention durations, ranging from one to 16 weeks, were found. Methodological limita-

Citation: Marcela Tabares Tabares, et al. Can Chatbots Alleviate Depression? Results of a Systematic Review. Advances in Artificial Intelligence and Machine Learning. 2024;4(3):155.

tions include selection bias and lack of clear information on random allocation and blinding. **Conclusions**: The bibliometric analysis highlights the interrelation of key terms such as ”chatbot,” ”depression,” and ”cognitive-behavioral therapy,” underscoring the importance of advanced technologies in developing these tools. Despite their potential, chatbots should not be considered a substitute for professional treatment in severe cases. This study suggests that chatbots are promising for psychological support, but more research is needed to optimize their clinical effectiveness and user acceptance.

Keywords: Depression, Chatbot, Cognitive behavioral therapy, Artificial intelligence, Natural language processing.

1. INTRODUCTION

Depression is a common and debilitating mental disorder affecting millions of people worldwide. According to the World Health Organization (WHO), over 280 million people of all ages suffer from depression globally. This disorder is a leading cause of disability and significantly contributes to the global burden of disease [1]. Depression is characterized by deep sadness, loss of interest or pleasure, feelings of guilt or low self-esteem, sleep or appetite disturbances, fatigue, and poor concentration. Factors such as genetics, neurotransmitter changes, adverse life events, and physical health problems can influence its development [2].

The COVID-19 pandemic exacerbated depression cases due to social isolation, economic uncertainty, and fear of contagion. A WHO report [3], highlights the increase in the prevalence of depression and other mental disorders in the post-pandemic phase, calling for a coordinated global response to address this growing mental health crisis [3].

Treatment for depression includes psychotherapeutic therapies, such as Cognitive-Behavioral Therapy (CBT), and, in some cases, antidepressant medications. Early intervention is key to preventing the progression of depression to more severe states. Mental health education and early access to effective treatments are essential to reduce the burden of depression [4].

Numerous barriers can prevent patients from receiving timely intervention. The stigma associated with mental health is one of the most significant barriers. Many people with depression fear being judged by their communities, which can deter them from admitting their symptoms or seeking help, worsening the severity of depression in the long term [5]. Other barriers include inequality in the availability of mental health services, especially in rural or low-income areas, and the cost of therapy, which can be a limiting factor in countries without a robust public health system [6].

It is crucial to develop and promote more accessible and less costly treatment strategies. One example is digital therapies or interventions, such as chatbots, which can offer psychological support at a lower cost and with greater accessibility [7]. Thanks to technological advances and artificial intelligence algorithms, chatbots are increasingly used in the mental health field, providing consistent and evidence-based interventions 24/7. These systems can offer CBT, one of the most effective interventions for depression, and have been shown to significantly reduce depressive symptoms compared to control groups that did not use chatbots [8].

Chatbots can overcome the stigma and accessibility barriers that often prevent people from seeking treatment. Their discretion and immediate availability make them an accessible and non-intimidating first step towards seeking help [9]. However, their effectiveness may be reduced by the lack of human empathy and the inability to adequately handle acute mental health crises. While useful for managing mild symptoms and self-monitoring, they should not be used as a substitute for professional care in severe depression cases.

Chatbots offer a promising alternative for providing immediate, continuous, and low-cost psychological support, especially in contexts with limited mental health resources. However, uncertainties remain regarding their clinical effectiveness, the quality of interactions, data security, and user acceptance [10]. The aim of this systematic review is to analyze the scientific evidence of the effect of chatbot interventions on symptoms of depression. This review seeks to synthesize existing evidence, identify knowledge gaps, and guide future research and technological developments, promoting the effective use of chatbots in the mental health field, specifically in depression cases. This work could underpin public health policies and provide clear guidelines for integrating these technologies into healthcare systems, thereby maximizing their positive impact on depression treatment.

2. METHOD

In this review, the PRISMA methodology (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) was employed to ensure a rigorous and transparent process. PRISMA involves the comprehensive identification of relevant studies through searches across various databases, the careful selection of studies based on clearly defined inclusion and exclusion criteria, the systematic extraction of key data, and the assessment of the quality and risk of bias in the included studies. In this case, a meta-analysis was not conducted due to the significant heterogeneity among the studies. Finally, PRISMA guided the clear and complete reporting of all methods and results, following a standardized checklist, which ensures reproducibility and minimizes bias in the research (PRISMA, 2020) [11]. The following sections detail each stage of this methodology.

2.1 Protocol and Registration

The protocol was registered on 11 April 2024 in PROSPERO (International Prospective Register of Systematic Reviews) with ID CRD42024534759.

PROSPERO is an international database that allows researchers to register protocols for systematic reviews in health before beginning data analysis, ensuring transparency, preventing publication bias, and avoiding duplication of studies. This public registry facilitates access to information on ongoing and completed reviews, and promotes the quality and credibility of systematic reviews by adhering to international standards (PROSPERO, 2024) [12].

2.2 Eligibility Criteria

The inclusion criteria were studies that directly evaluated changes in depression symptoms through validated instruments in individuals who engaged in one or more conversations with a chatbot designed to alleviate these symptoms, published between January 2019 and April 2024. This period was selected due to the increased use of such artificial intelligence tools. Exclusion criteria were qualitative research, studies that did not report scores from instruments measuring depression symptoms before and after the intervention, and studies not conducted after conversations with the chatbot. No exclusions were made based on age, sex, or depression level.

The PICO question that directed the review was:

- Participants: people exhibiting symptoms of depression.
- Intervention: conversation with a chatbot as a treatment modality.
- Comparator: conventional in-person or virtual therapy, educational or informational interventions, or no intervention.
- Outcomes: change in depression symptoms measured with validated instruments such as the PHQ-9 (Patient Health Questionnaire-9).

2.3 Sources of Information

The information was searched in the Cochrane Library, PubMed, Scopus and ScienceDirect databases. The date of the last search was 11 April 2024.

2.4 Search

The keywords used in the search were ”chatbot AND depression”, in title, abstract or keywords.

2.5 Selection of Studies

The database searches and the review of inclusion and exclusion criteria were conducted by the principal investigator and an independent reviewer (P.C.G). Subsequently, a quality assessment of the articles included in the review was performed using the risk of bias assessment tool proposed by the Cochrane Collaboration, applied independently by two researchers to each article. This tool considers the following components: selection bias, performance bias, detection bias, attrition bias, and reporting bias; each area was rated according to three points (1=high, 2=unclear, 3=low) [13].

2.6 Data extraction

Data extraction was performed by one author and then independently verified by a second author. There were no disagreements. The following data were extracted from each study: duration of intervention, sample size, age, gender, type of intervention/conversation held with the chatbot, instrument used to measure depressive symptoms, variations in depressive symptoms, participant perception/acceptance and attrition rate.

3. RESULTS

The search criteria yielded 321 results. After removing duplicates and applying the inclusion and exclusion criteria, 12 studies were selected and included in the present analysis (FIGURE 1).

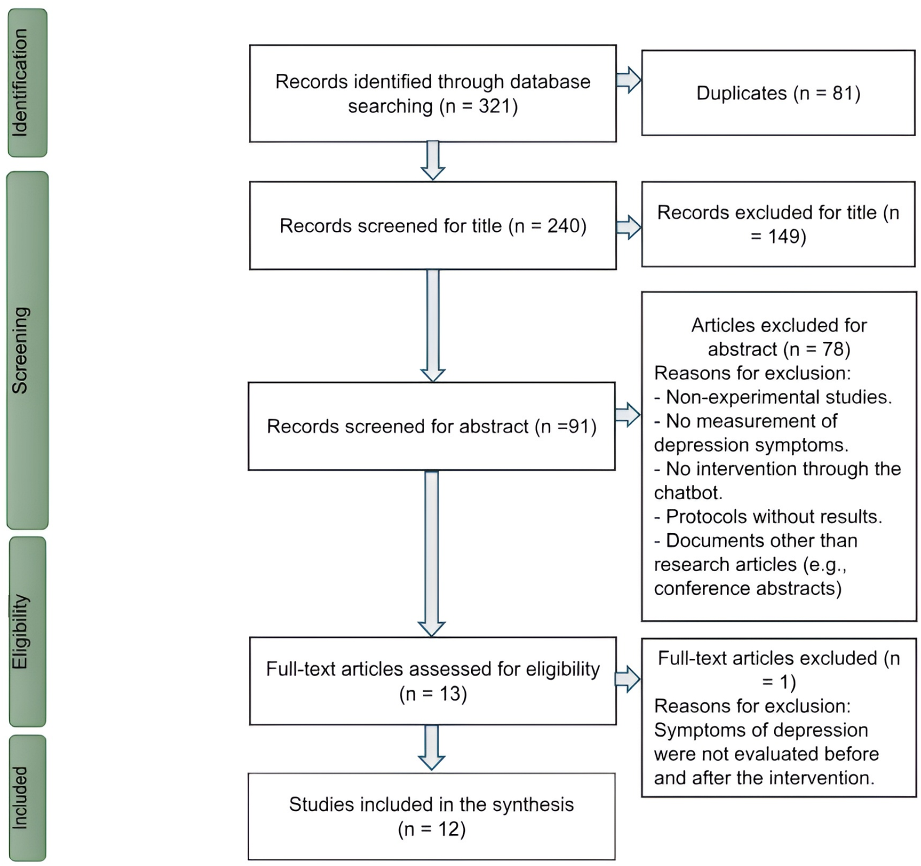

Figure 1: Systematic review process

3.1 Bibliometric Analysis

In the bibliometric analysis of the literature on the use of chatbots to reduce depression symptoms, co-occurrence maps of terms were generated using VOSviewer software. These maps reveal the network of terms most frequently associated in the reviewed articles, highlighting predominant themes and their interrelations.

FIGURE 2 presents a co-occurrence map showing a complex network of concepts, highlighting the interconnection between terms such as "chatbot," "depression," "mental health," and "cognitive-behavioral therapy." Other important terms include "artificial intelligence," "natural language processing," "sentiment analysis," and "machine learning," underscoring the crucial role of advanced technologies in the development and functioning of chatbots [14, 15].

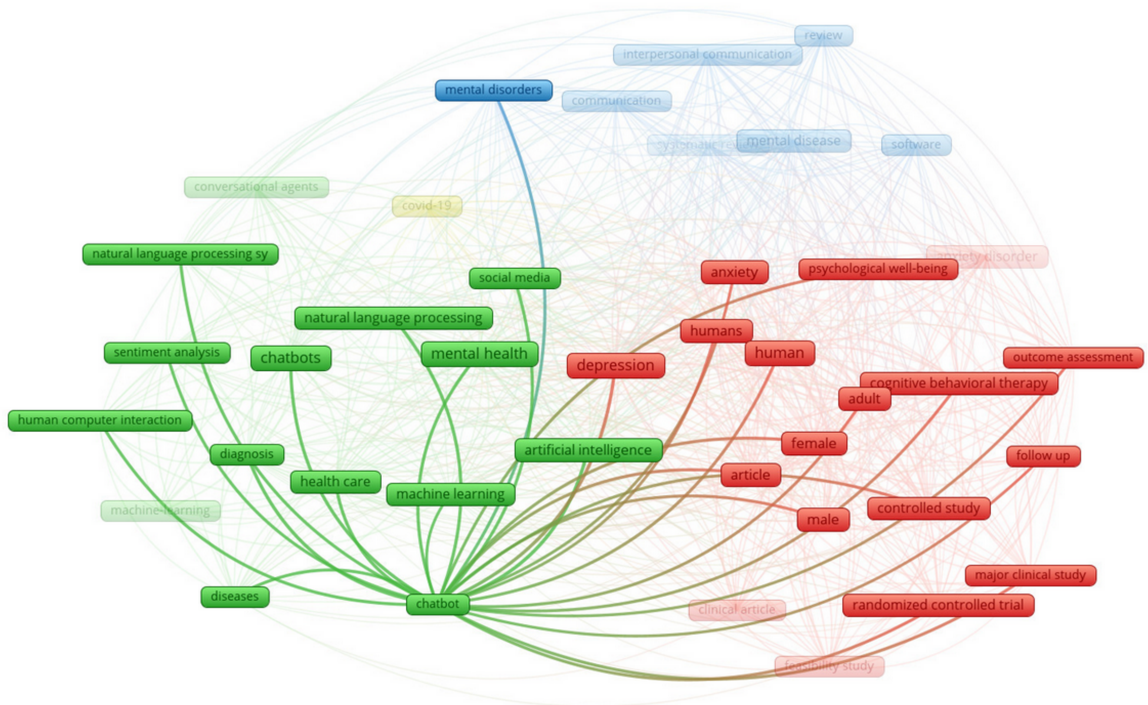

Figure 2: Co-Occurrence Analysis of Terms in Scopus. Own elaboration with VOSviewer

FIGURE 3 shows a heat map that highlights terms related to research on the use of chatbots in alleviating depression. The central terms of the map were "depression," "chatbot," "mental health," and "artificial intelligence." The proximity of "cognitive-behavioral therapy" to the core suggests that many studies have investigated the effectiveness of chatbots in providing interventions based on this therapeutic approach. Additionally, the terms "artificial intelligence" and "natural language processing" are highlighted, indicating the importance of these technologies in developing effective chatbots [16]. The presence of terms like "randomized controlled trial" and "controlled study" emphasizes researchers' commitment to rigorous methodological approaches.

Figure 3: Keyword heat map. Own elaboration with VOSviewer

Finally, the bibliometric analysis showed that countries such as the United States, China, Australia, Germany, and the United Kingdom lead research on the use of chatbots in mental health, while Latin American countries have a lower incidence in this research area.

3.2 Study Quality Analysis

Regarding selection bias, several studies, such as those by Daley et al. [17], and Mauriello et al. [18], did not provide clear information on random sequence generation, introducing a lack of clarity in this aspect. Additionally, studies like those by Ryu et al. [19], and Anmella et al. [20], presented a high risk of bias due to the lack of information on allocation concealment. Concerning performance bias, most studies did not specifically mention the implementation of blinding for participants and personnel, resulting in a high risk of performance bias in many cases.

In terms of detection bias, studies such as those by Romanovskyi et al. [21], and Nicol et al. [22], did not detail whether blinding was implemented in the evaluation of self-reported outcomes, introducing risks of detection bias. However, for objective measures, several studies, such as those by Mauriello et al. [18], and Liu et al. [23], presented low risk due to the standardized nature of the measures used. In terms of attrition bias, some studies, such as Ryu et al. [19], presented a high risk due to the lack of information on dropout rates and the management of incomplete data. Others, such as He et al. [8], showed low risk in this aspect.

Finally, regarding reporting bias, most studies presented a low risk of selective reporting bias, as all relevant results were reported. FIGURE 4 presents the risk of bias assessment in the 12 studies reviewed, highlighting three levels: low risk, unclear risk, and high risk. Most studies show an unclear risk in key areas such as allocation concealment and participant blinding, indicating potential methodological limitations. However, objective measures tend to have a low risk of bias, suggesting greater reliability in those results. This analysis underscores the need for methodological improvements in studies to ensure the validity and reproducibility of future systematic reviews.

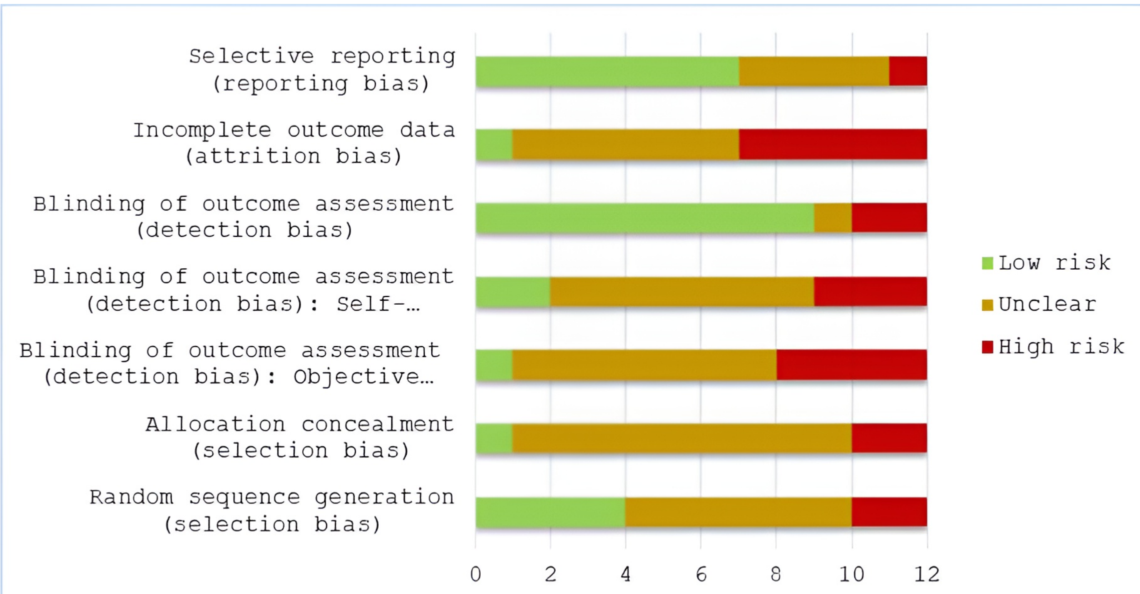

Figure 4: Risk of Bias Assessment Across 12 Reviewed Studies. Own elaboration.

3.3 Analysis of the Methodology and Results of the Studies

The following section describes the data found in the studies included in this review, concerning participants, types of interventions with chatbots and their duration, effectiveness of chatbots, and perception and acceptance by participants. More detailed information can be found in the table included as supplementary material to this article.

3.3.1 Participants

The reviewed studies presented variations in sample size, gender distribution, and the average age of participants. The study with the largest sample size was conducted by Daley et al. [17], involving 3,629 participants, followed by Romanovskyi et al. [21], with 412 participants. Two studies had sample sizes between 100 and 200: Klos et al. [24], with 181 participants and He et al. [8], with

149 participants. The remaining studies had sample sizes under 100: Qi [25], with 86 participants, Liu et al. [23], with 83 participants, and Karkosz et al. [26], with 80 participants. Four studies had sample sizes between 20 and 50: Mauriello et al. [18], with 47 participants, Anmella et al. [20], with 34 participants, Bendig et al. [27], with 30 participants, and Ryu et al. [19] with 25 participants. Finally, only one had a sample size of less than 20, the one conducted by Nicol et al. [22], involving 17 participants.

Regarding gender distribution, most studies reported a higher participation of women. Only two studies, by Romanovskyi et al. [21], and He et al. [8], had a higher participation of men. None of the reviewed studies achieved gender balance in participation.

Concerning age, five studies reported the mean age and Standard Deviation (SD) of the participants, ranging between 15.35 and 35.3 years. Nicol et al. [22], documented a mean age of 35.3 years (SD = 10.12), Liu et al. [23], noted a mean age of 25.18 years (SD = 4.69), Bendig et al. [27], reported a mean age of 23.17 years (SD = 3.85), and Anmella et al. [20], also reported a mean age of 23.08 years, though without clarity on the SD. He et al. [8], reported a mean age of 18.78 years (SD = 0.89), while Klos et al. [24], indicated a mean age of 15.35 years in the experimental group (SD = 5.75). On the other hand, six studies did not clearly mention these data, making comparison difficult. These studies include Daley et al. [17], Ryu et al. [19], Mauriello et al. [18], Romanovskyi et al. [21], Karkosz et al. [26], and Qi [28].

3.3.2 Types of interventions with chatbots and their duration

The reviewed studies included various chatbot interventions, each with different durations (see TABLE 1). Daley et al. [17], implemented the Viki chatbot for 4 weeks, providing support for managing stress, mood, and anxiety. Ryu et al. [19], used a chatbot via KakaoTalk for older adults over 2 weeks, offering simple and consistent interactions to reduce anxiety and depression. Mauriello et al. [18], employed a suite of chatbots to manage daily stress over 1 week. Bendig et al. [27], conducted a 2-day intervention with writing tasks and exercises based on Acceptance and Commitment Therapy. Klos et al. [24] used the Tess chatbot for 8 weeks, offering comprehensive mental health support. Romanovskyi et al. [21], applied the Elomia chatbot for 4 weeks, using cognitive-behavioral and conversational techniques to improve mental health.

Nicol et al. [22], used a chatbot for 12 weeks for adolescents with depression and anxiety, based on Cognitive-Behavioral Therapy (CBT) principles. Liu et al. [23], implemented the XiaoNan chatbot for 16 weeks, providing information and coping strategies. He et al. [8], employed the XiaoE chatbot for 1 week for emotional support and task reminders using CBT techniques. Anmella et al. [20], guided users with a chatbot for 4 weeks, using gamification, incentives, and module personalization among other strategies. Karkosz et al. [26], used the Fido chatbot for 2 weeks, based on CBT techniques. Finally, Qi [28], used the Woebot chatbot for 8 weeks to provide emotional support and coping tools also based on CBT.

These varied durations and intervention designs illustrate the diverse approaches taken in utilizing chatbots to address depression symptoms, reflecting different levels of engagement and therapeutic methods tailored to specific populations and needs.

Table 1: Characteristics of the studies

Authors	**Study design**	**Duration of the intervention**	**Sample size**	**Women (W) and men (M)**	**Mean age and standard deviation (SD)**	**Attrition rate**
Daley et al. [17]	Prospective observational study without control group	4 weeks	3629	2758 W, 871 M	No clarity	No clarity
Ryu et al. [19]	Uncontrolled intervention study	2 weeks	25	17 W, 7 M	No clarity	6 out of 25 participants
Mauriello et al., [18]	Uncontrolled intervention study	1 week	47	34 W, 13 M	No clarity	16 out of 47 participants
Bendig et al. [27]	Pre-test and post-test trial	2 days	30	24 W, 6 M	23,17 (SD =3.85)	0
Klos et al. [24]	Pilot Randomised Controlled Trial (RCT)	8 weeks	181	158 W, 23 M	Intervention group 15.35 (SD=5.75). Control group 15.59 (SD=5.30)	61% in the experimental group and 59% in the control group.
Romanovskyi et al. [21]	Pilot Randomised Controlled Trial (RCT)	4 weeks	412	202 W, 210 M	No clarity	Not specified
Nicol et al. [22]	Pilot Randomised Controlled Trial (RCT)	12 weeks	17	15 W, 2 M	No clarity	0
Liu et al. [23]	Non-blinded Randomised Controlled Trial (RCT)	16 weeks	83	46 W, 37 M	23.08 (No clarity in the SD)	24.10%
He et al. [8]	Single-blind, three-arm Randomised Controlled Trial (RCT)	1 week	149	55 W, 94 M	18.78 (SD=0.89)	XiaoE: 1 participantXiaoai: 8 participantse-book: 10 participants
Anmella et al. [20]	Feasibility and effectiveness study.	4 weeks	34	26 W, 8 M	35,3 (SD=10.12)	76%
Karkosz et al. [26]	Randomised Controlled Trial (RCT)	2 weeks	80	58 W, 22 M	25,18 (SD=4.69)	No clarity
Qi [28]	Quasi-experimental trial	8 weeks	86	46 W, 40 M	No clarity	No clarity

3.3.3 Effectiveness of chatbots

In terms of overall effectiveness, most interventions demonstrated significant reductions in depression symptoms. Specifically, the Viki chatbot [17], showed a reduction in PHQ-9 scores from 15.9 (SD = 6.5) to 10.4 (SD = 6.5). Similarly, the chatbot used via KakaoTalk [19], led to a reduction in CES-D scores from 19.429 (SD = 9.9) to 16.667 (SD = 6.7). Popbots [18], also demonstrated effectiveness, reducing PHQ-4 scores from 3.0 to 2.0. Fido [26], saw CESD-R scores drop from 28.33 (SD = 13.84) to 20.33 (SD = 11.94), and Woebot [28], reported a reduction in DASS-21 depression scores from 43.7 (SD = 4.6) to 36.2 (SD = 2.8). XiaoE [8], led to a significant reduction of 2.44 points on the PHQ-9 in the intervention group, while the chatbot used by Nicol et al. [22], resulted in a 3.3-point reduction in PHQ-9 scores for the intervention group compared to a 2-point reduction in the control group.

In contrast, the chatbots SISU [27], and Tess [24], did not show significant changes in depression symptoms, with the PHQ-9 scores remaining unchanged in the SISU trial and no significant changes reported in the Tess trial. Vickybot [20], also showed no significant changes in PHQ-9 scores. Results for Elomia [21], indicated a general improvement from ”moderate depression” to ”mild depression” on the PHQ-9 scale, but exact scores were not provided. Similarly, XiaoNan [23], mentioned a significant reduction in PHQ-9 scores without providing specific before-and-after figures, making exact comparison difficult.

3.3.4 Perception and acceptance by participants

The participants’ perception and acceptance of chatbot interventions varied across studies, presenting both positive and negative opinions, as well as different dropout rates. Daley et al. [17], reported a positive perception of the Viki chatbot, with users experiencing significant reductions in symptoms of anxiety, depression, and stress, although the dropout rate was not specified. Similarly, Ryu et al. [19], highlighted that several participants, especially older women, found interaction with the chatbot easier and more comfortable than initially expected, with a dropout rate of 6 out of 25 participants. On the other hand, Mauriello et al. [18], reported mostly positive feedback regarding the effectiveness of the Popbots chatbots, with most participants considering them ”slightly effective” to ”very effective,” but with a dropout rate of 16 out of 47 participants.

In contrast, Bendig et al. [27], noted that although there were reports of negative effects, most participants completed all sessions and showed a positive attitude towards the SISU chatbot intervention, with no dropout reported. Klos et al. [24], found that positive feedback towards the Tess chatbot was associated with a higher number of message exchanges, although the dropout rate was high: 61% in the experimental group and 59% in the control group. Romanovskyi et al. [21], reported a positive perception of the Elomia chatbot, with users expressing increased self-confidence and highlighting communication with the chatbot as beneficial for their emotional well-being and self-awareness, although the dropout rate was not specified. Nicol et al. [22], found that 80% of adolescents considered the chatbot intervention acceptable, and 70% agreed that ”using the app helped treat depression in the best possible way,” with no dropout recorded.

Liu et al. [23], noted that factors related to the interaction process with the XiaoNan chatbot were more important than the content of the conversations, reporting a dropout rate of 24.10%. He et al. [8], found greater acceptability towards the XiaoE chatbot compared to control groups, with a dropout rate of 1 participant in the XiaoE intervention group, 2 in the control group using Xiaoai, and 10 in the control group using the e-book. Anmella et al. [20], reported high subjective perception of usability, satisfaction, and acceptability with the Vickybot chatbot, but observed low participation and engagement in terms of objective usage metrics, with a dropout rate of 76%. Karkosz et al. [26], noted that some participants highlighted the Fido chatbot’s ability to recognize user intent, although there were reports of failures. However, using Fido encouraged users to consider starting traditional therapy, with no clarity on the dropout rate. Finally, Qi [28], did not specify the participants’ perception or acceptance regarding the Woebot chatbot intervention, nor did they describe the dropout rate.

4. DISCUSSION

The bibliometric analysis revealed a complex network of associated terms in the literature on the use of chatbots for depression treatment. The prominence of terms such as ”chatbot,” ”depression,” ”mental health,” and ”Cognitive-Behavioral Therapy” (CBT) suggests these are central themes in current research. This indicates a strong interrelation between chatbot use and mental health treatment approaches, especially CBT, which is widely recognized and effective for treating depression [26, 28]. However, according to Lin et al. [29], some articles claiming to implement CBT do not include its basic concepts. For example, none of the interventions in their study included ”Action Planning,” an essential component in behavioral activation. This aspect was not analyzed in the present review, so future studies should verify if CBT-based interventions adhere to its fundamentals.

The relevance of advanced technologies like artificial intelligence and natural language processing, highlighted by terms such as ”artificial intelligence” and ”natural language processing,” underscores the crucial role of these technologies in chatbot effectiveness. Terms like ”sentiment analysis” and ”machine learning” also reflect the sophistication of the technological tools used. Moreover, the emphasis on empirical validation of chatbots, evidenced by terms such as ”randomized controlled trial” and ”controlled study,” suggests a commitment from the scientific community to the efficacy and safety of these methods in clinical contexts, essential for their acceptance and adoption [8].

The presence of demographic terms such as ”adult,” ”male,” and ”female” indicates that studies have considered various population groups, although specific research for each group is not detailed. This underscores the need for future research to delve into how different population groups respond to chatbot use for mental health. Additionally, the interest in digital and mobile platforms, reflected in terms like ”digital health” and ”mhealth,” highlights the potential of chatbots as accessible tools for managing depression, especially given the increased demand for remote psychological support exacerbated by the COVID-19 pandemic [8].

The geographical distribution of publications shows a higher concentration of research in developed countries such as the United States, China, Australia, Germany, and the United Kingdom, while Latin American countries have a lower frequency. This suggests the need to promote research in underrepresented regions to better understand the impact and effectiveness of chatbots in diverse cultural and economic contexts. Latin American countries represent less than 5% of global research and development (R&D) investment, despite having over 20% of the world's population; in contrast, Asia, led by China, significantly invests more in R&D (Research & Development World, 2020). This disparity underscores the need to increase support for technology and mental health research in Latin America to ensure more equitable representation in global scientific literature [30].

The results suggest variation in sample sizes among the reviewed studies, ranging from 17 to 3629 participants. This variability can affect result generalization, as studies with larger samples, like Daley et al. [17], can offer robust and representative outcomes. In contrast, studies with smaller samples, such as Nicol et al. [22], with 17 participants, can provide detailed but limited reports in terms of general application. Furthermore, the predominance of women in most studies suggests a need to balance gender representation for a more comprehensive perspective on the impact of chatbots across diverse demographic groups. The wide range of average ages also indicates that chatbots are being used in various stages of life, though lack of detailed data in some studies limits the ability to make accurate comparisons.

The diversity in chatbot interventions and their durations reflects the flexibility of these tools to adapt to different contexts and needs. Longer interventions, such as those by Nicol et al. [22], and Liu et al. [23], lasting 12 and 16 weeks respectively, may provide more opportunity to observe significant changes in depression symptoms. On the other hand, shorter interventions, like Bendig et al. [27], lasting only 2 days, demonstrate that even brief periods can be beneficial.

The use of diverse techniques, from CBT to gamification, underscores the versatility of chatbots in depression treatment. However, the effectiveness of these interventions may depend on both duration and the appropriateness of techniques used to meet users' specific needs. This aligns with findings by Ulrich et al. [31], who emphasize that personalizing interventions and adjusting them to individual user preferences and circumstances is crucial to maximize therapeutic benefits. Several studies included in this review, such as Nicol et al. [22], and Anmella et al. [20], mention having personalized their interventions, resulting in increased user satisfaction and engagement.

Intervention duration could significantly influence user engagement. Users tend to maintain engagement with shorter interventions, whereas longer interventions may experience higher dropout rates. This is reflected in Bendig et al.'s study [27], which reported a 0% dropout rate in a 2-day intervention, compared to higher dropout rates in studies with longer durations. Therefore, finding an appropriate balance between intervention duration and technique used is crucial to optimize both user participation and clinical outcomes.

Most reviewed studies indicate that chatbots are effective in reducing depression symptoms, albeit with variations in efficacy levels. Chatbots like Viki, used by Daley et al. [17], and KakaoTalk's chatbot, used by Ryu et al. [19], showed significant reductions in symptoms. In contrast, chatbots like SISU [27], and Tess [24], did not show significant changes, suggesting effectiveness may depend on specific design and implementation factors for each chatbot. The lack of detailed efficacy data in some studies also indicates a need for more standardized and comprehensive reporting to facilitate precise and robust comparisons among different interventions. This variability highlights the importance of rigorously developing and evaluating different designs and approaches to optimize the therapeutic impact of chatbots in depression treatment.

Perception and acceptance of chatbots vary across studies. Some, like Nicol et al. [22], reported high acceptability and satisfaction among participants, while others, like Anmella et al. [20], observed low participation despite positive subjective perceptions of usability and satisfaction. Additionally, studies like Bendig et al. [27], which did not show significant symptom changes (PHQ-9 before and after 5.00), still reported positive attitudes toward the intervention. This variability may stem from differences in user interface, quality of interactions, and participant expectations. High dropout rates in some studies, such as Anmella et al. [20], with 76%, underscore the importance of designing chatbots that are not only effective but also attractive and user-friendly. Dropout rates and participant feedback provide valuable insights to improve adherence and effectiveness of future chatbot interventions [20].

It is essential to consider some limitations of this systematic review. Firstly, methodological limitations were identified in the reviewed studies, such as lack of blinding, as well as variable quality and potential biases. These deficiencies may influence interpretation of results and should be addressed in future studies. Moreover, variability in study designs, sample sizes, intervention durations, and evaluation methods complicates direct comparison and synthesis of results. This heterogeneity limits the ability to draw robust and comparative conclusions, emphasizing the need for a more standardized approach in future research.

Some studies did not provide complete demographic data, limiting the ability to analyze the impact of chatbots on different population groups. Additionally, adapting chatbots to diverse cultural and linguistic contexts is necessary to ensure their efficacy in various populations, as current research primarily focuses on developed countries, potentially limiting applicability of findings to other regions.

Most studies have focused on short-term interventions, so it is essential to conduct research evaluating the long-term impact of chatbot use in depression treatment to better understand its sustained efficacy and potential for integration into long-term mental health programs. Additionally, potential benefits of combining chatbot use with other forms of treatment, such as in-person therapy or group interventions, should be considered, which could provide a more holistic and effective approach to depression treatment.

5. CONCLUSIONS

The results of this systematic review demonstrate the effectiveness of chatbots in reducing symptoms of depression, highlighting their potential as a complementary tool in mental health care. The implementation of chatbots offers an accessible and scalable alternative to provide therapeutic support, especially in regions with a shortage of mental health professionals.

To maximize their impact, public health policies should consider personalizing these technological interventions. Adapting interventions to the specific needs of users enhances satisfaction, engagement, and therapeutic outcomes. The integration of emerging technologies such as artificial intelligence and deep learning can further improve user interaction and understanding, enabling greater personalization and more natural interaction. Fostering research and development of these technologies is crucial to ensure their effective and safe implementation.

Furthermore, guidelines should include adapting chatbots to different cultural and linguistic contexts to ensure their efficacy in diverse populations. Current research primarily focuses on developed countries, which may limit the applicability of findings to other regions. Investing in research and development of chatbots in underrepresented countries can help bridge this gap and ensure more equitable representation in global scientific literature.

Finally, long-term evaluation of chatbot use for depression treatment is crucial to understand its sustained efficacy and potential for integration into mental health programs. Public health policies should support long-term studies and the combination of chatbots with other forms of treatment, such as in-person therapy or group interventions, to provide a more holistic and effective approach to depression treatment.

References

[1] https://www.who.int/es/news-room/fact-sheets/detail/depression

[2] Dudek KA, Dion-Albert L, Kaufmann FN, Tuck E, Lebel M, Menard C. Neurobiology of Resilience in Depression: Immune and Vascular Insights From Human and Animal Studies. Eur J Neurosci. 2021;53(1):183-221.

[3] https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022.1

[4] Cuijpers P, Quero S, Dowrick C, Arroll B. Psychological Treatment of Depression in Primary Care: Recent Developments. Curr Psychiatry Rep. 2019;21:129.

[5] Li F, Jia H, Liu J, Liu S, Lu X, Yip PS. Self-Stigma and Its Relationship With Rates of Depression Diagnosis. Crisis. 2022;43:190-196.

[6] von Gaudecker JR, Agbapuonwu N, Kyololo O, Sathyaseelan M, Oruche U. Barriers and Facilitators to Treatment Seeking Behaviors for Depression, Epilepsy, and Schizophrenia in Low- And Middle-Income Countries: A Systematic Review. Arch Psychiatr Nurs. 2022;41:11-19.

[7] Martinengo L, Lum E, Car J. Evaluation of Chatbot-Delivered Interventions for Self-Management of Depression: Content Analysis. J Affect Disord. 2022;319:598-607.

[8] He Y, Yang L, Zhu X, Wu B, Zhang S, Qian C et al. Mental Health Chatbot for Young Adults With Depressive Symptoms During the COVID-19 Pandemic: Single-Blind, Three-Arm Randomized Controlled Trial. J Med Internet Res. 2022;24:e40719.

[9] Haque MD, Rubya S. An Overview of Chatbot-Based Mobile Mental Health Apps: Insights From App Description and User Reviews. JMIR mHealth uHealth. 2023;11:e44838.

[10] Ahmed A, Aziz S, Khalifa M, Shah U, Hassan A, et al. Thematic Analysis on User Reviews for Depression and Anxiety Chatbot Apps: Machine Learning Approach. JMIR Form Res. 2022;6:e27654.

[11] Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD ET AL. PRISMA 2020 Explanation and Elaboration: Updated Guidance and Exemplars for Reporting Systematic Reviews. BMJ. 2021;372:n160.

[12] https://www.crd.york.ac.uk/prospero/

[13] Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I et al. Rob 2: A Revised Tool for Assessing Risk of Bias in Randomized Trials. Br Med J (Clin Res Ed). 2019;366:l4898.

[14] Dikeç, G., Oban, V., & Barış Usta M. (2023). Qualitative and artificial intelligence-based sentiment analysis of Turkish tweets related to schizophrenia. Şizofreniye Yönelik Türkçe Twitter İletilerin Nitel ve Yapay Zekâ Temelli Duygu Çözümlemesi. Turk psikiyatri dergisi =. Turk J Psychiatry;34:145-153.

[15] Shanmugavadivel K, Sathishkumar VE, Raja S, Lingaiah TB, Neelakandan S, Subramanian M. Deep Learning Based Sentiment Analysis and Offensive Language Identification on Multilingual Code-Mixed Data. Sci Rep. 2022;12:21557.

[16] Grové C. Co-developing a Mental Health and Wellbeing Chatbot With and for Young People. Front Psychiatry. 2020;11:606041.

[17] 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.

[18] Mauriello ML, Tantivasadakarn N, Mora-Mendoza MA, Lincoln ET, Hon G, Nowruzi P et al. A Suite of Mobile Conversational Agents for Daily Stress Management (Popbots): Mixed Methods Exploratory Study. JMIR Form Res. 2021;5:e25294.

[19] Ryu H, Kim S, Kim D, Han S, Lee K, Kang Y. Simple and Steady Interactions Win the Healthy Mentality: Designing a Chatbot Service for the Elderly. Proc ACM Hum Comput Interact. 2020;4:1-25.

[20] Anmella G, Sanabra M, Primé-Tous M, Segú X, Cavero M, et al. Vickybot, a Chatbot for Anxiety-Depressive Symptoms and Work-Related Burnout in Primary Care and Health Care Professionals: Development, Feasibility, and Potential Effectiveness Studies. J Med Internet Res. 2023;25:e43293.

[21] Romanovskyi O, Pidbutska N, Knysh A. Elomia Chatbot: The Effectiveness of Artificial Intelligence in the Fight for Mental Health. In: Rt-PA 5th International Conference on Computational Linguistics and Intelligent Systems. 2021:1215-1224.

[22] Nicol G, Wang R, Graham S, Dodd S, Garbutt J. Chatbot-Delivered Cognitive Behavioral Therapy in Adolescents With Depression and Anxiety During the COVID-19 Pandemic: Feasibility and Acceptability Study. JMIR Form Res. 2022;6:e40242.

[23] 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.

[24] Klos MC, Escoredo M, Joerin A, Lemos VN, Rauws M, Bunge EL. Artificial Intelligence-Based Chatbot for Anxiety and Depression in University Students: Pilot Randomized Controlled Trial. JMIR Form Res. 2021;5:e20678.

[25] https://www.rdworldonline.com/global-rd-investments-unabated-in-spending-growth/

[26] Karkosz S, Szymański R, Sanna K, Michałowski J. Effectiveness of a Web-Based and Mobile Therapy Chatbot on Anxiety and Depressive Symptoms in Subclinical Young Adults: Randomized Controlled Trial. JMIR Form Res. 2024;8:e47960.

[27] Bendig E, Erb B, Meißner D, Bauereiß N, Baumeister H. 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. 2021;24:100377.

[28] Qi Y. Pilot Quasi-Experimental Research on the Effectiveness of the Woebot AI Chatbot for Reducing Mild Depression Symptoms Among Athletes. Int J Hum Comput Interact. 2024:1-8.

[29] Lin X, Martinengo L, Jabir AI, Ho AH, Car J, Atun R et al. Scope, Characteristics, Behavior Change Techniques, and Quality of Conversational Agents for Mental Health and Well-Being: Systematic Assessment of Apps. J Med Internet Res. 2023;25:e45984.

[30] https://uis.unesco.org/sites/default/files/documents/fs59-global-investments-rd-2020-en.pdf

[31] Ulrich S, Lienhard N, Künzli H, Kowatsch T. A Chatbot-Delivered Stress Management Coaching for Students (Misha App): Pilot Randomized Controlled Trial. JMIR mHealth uHealth. 2024;12:e54945.

Supplementary material: Summary of study results

Authors	Brief description of the intervention and monitoring (if applicable)	Instrument and symptoms before/after	Perception/acceptance of participants
Daley et al. [17],	The Viki chatbot intervention consisted of conversations designed to help users reflect on their experiences and learn techniques for managing stress, mood and anxiety. It used tools such as mood tracking, emojis, GIFs and gamification elements to mimic human interactions and encourage engagement. Developed by clinical psychologists and health professionals, the conversations were accessible and adaptive for the user. The chatbot initiated conversations, allowing responses through predefined options or free text at certain times. Control group: Not applicable.	Patient Health Questionnaire-9 (PHQ-9)Before: 15.9 (SD=6.5). After: 10.4 (SD=6.5). Significant decrease.	The study found a positive perception and acceptance of the chatbot. The intervention engaged users and significantly reduced symptoms of anxiety, depression and stress. In addition, greater engagement with the chatbot, as measured by the number of responses sent, was associated with greater reductions in symptoms of anxiety and depression.
Ryu et al. [19],	A mental health care chatbot for older adults was developed through KakaoTalk. Designed to reduce anxiety and depression in this group, it provided simple, consistent interactions. The focus on simple, repetitive conversations proved effective in increasing talk time, despite the brevity of the messages compared to human interactions. Control group: not applicable	Centre for Epidemiological Studies Depression Scale (CES-D)Before: 19.429 (SD=9,9). After: 16.667 (SD=6,7). Significant decrease.	Several participants found the interaction easier and more comfortable than expected. Initially, many users, especially older women, showed uncertainty, but then actively participated, especially when provided with response options via button navigation.

Mauriello et al., [18],	Participants interacted with a suite of chatbots designed to manage everyday stress, offering a variety of conversations and micro-interventions. It included a problem-solving bot and a laughter bot. Participants could choose which chatbot to interact with according to their preferences and needs. The micro-interventions covered techniques such as positive reframing, reflective reminders, practical problem solving and light conversation to generate laughter. These interactions were brief, 2-3 minutes, addressing everyday concerns and providing emotional support. Control group: Not applicable.	Patient Health Questionnaire-4 (PHQ-4)Before 3.0. After: 2.0. Significant decrease.	Feedback on the chatbots was mostly positive. Most found them ”slightly effective” to ”very effective” and described them as ”friendly and engaging”. They valued the variety of options and the ability to choose chatbots according to their needs.
Bendig et al. [27],	The chatbot intervention consisted of two sessions. Each session included writing tasks about negative autobiographical events and an exercise based on Acceptance and Commitment Therapy. Participants interacted with the software agent for two consecutive days, approximately 30 minutes each day, completing all sessions in that period. Control group: not applicable	Patient Health Questionnaire-9 (PHQ-9)Before: 5.00 (SD=3.54). After: 5.00 (SD=3.75). There was no significant change.	Although there were reports of negative effects attributed to the intervention, most participants completed all sessions and showed a positive attitude towards the intervention.

Klos et al. [24],	Tess, an artificial intelligence-based chatbot, was used to provide short text conversations as comprehensive mental health support. Tess sent reminders, psychoeducational content and emotionally supportive responses, using words and emojis to provide a friendly experience. She responded with prescribed statements to replicate empathetic responses appropriate to the emotions or concerns expressed by users. Tess' conversations were based on models of cognitive behavioural therapy, emotion-focused therapy, solution-focused brief therapy and motivational interviewing. Control group: Received an electronic psychoeducation e-book focusing on affective symptoms, providing evidence-based information and resources to help identify and seek treatment for depressive symptoms.	Patient Health Questionnaire-9 (PHQ-9)There was no significant change.	Positive feedback from participants was found to be associated with a higher number of message exchanges. The number of messages exchanged with Tess was related to acceptance and positive perception of the intervention. It is mentioned that participants accepted the chatbot favourably and were willing to actively interact with it during the study.
Romanovskyi et al. [21],	The Elomia intervention was based on a chatbot that offered emotional and therapeutic support to users with symptoms of depression, anxiety and negative affect. It used cognitive behavioural and talk therapy techniques to improve mental health. During conversations with Elomia, users were able to express their thoughts and feelings, allowing them to unburden themselves and reflect on their problems. In addition, Elomia offered techniques to reduce stress, improve self-esteem and promote a positive attitude. Control group: used a "Self-Help Guide for Depression" which contains a set of cognitive behavioural techniques to help deal with non-adaptive thoughts and reduce anxiety and depression.	Patient Health Questionnaire-9 (PHQ-9)Before: " Moderate Depression ". After: " Mild Depression ". The figure is not specified.	Participants had a positive perception towards Elomia Chatbot, reporting feeling more self-confident, seeing problems differently and experiencing emotional stability. Communication with Elomia helped them to understand their problems in a new way and to reflect on their way of living, valuing the experience as beneficial for their emotional well-being and self-awareness.

Nicol et al. [22],	The chatbot used natural language processing and machine learning techniques to provide a self-guided and personalised intervention to the adolescents. Conversations were tailored to each adolescent's situation to develop emotional regulation skills and address specific problems, focusing on mood monitoring and personalised goal-oriented conversations. The intervention was based on principles of cognitive behavioural therapy, interpersonal therapy for adolescents and elements of dialectical behavioural therapy, with the aim of improving the mental health of adolescents with depression and anxiety during the COVID-19 pandemic. Control group: Received usual care from their primary care physician for depression and anxiety.	Patient Health Questionnaire-9 (PHQ-9)Intervention group: decrease of 3.3 points. Control group: decrease of 2 points.	Adolescents found the app acceptable, with 80% agreeing that they liked using it and 70% believing it is viable for treating depression. Usability scored an average of 21.4 (on a scale of 5 to 25). Parents gave average scores of 16.6 for acceptability and 17 for feasibility (on a scale of 1 to 20).
Liu et al. [23],	The study implemented an intervention through a chatbot called XiaoNan. Participants assigned to the test group had access to XiaoNan and were asked to use it for 16 weeks. The chatbot provided information, coping strategies and tracked participants' progress throughout the study. Control group: received a minimal level of bibliotherapy as an intervention.	Patient Health Questionnaire-9 (PHQ-9)A "significant reduction in PHQ-9 scores" is mentioned (scores are not specified).	The results indicated that the process of interacting with the chatbot was more important than the content of the conversations. One participant noted that XiaoNan increased her willingness to seek professional help. Overall, participants had a positive perception towards the chatbot in terms of satisfaction and therapeutic alliance.

He et al. [8],	The intervention used XiaoE, a chatbot that focused on providing emotional support, counselling and daily task reminders using Cognitive Behavioural Therapy (CBT) techniques. The control group used Xiaoai, a chatbot that offered a general, non-specific approach to treating depressive symptoms in young adults during the COVID-19 pandemic. Control group 1 (e-book): e-book and daily article on depression. Control group 2 (Xiaoai): communication with a chatbot designed for informal conversations and not specifically for mental health services.	Patient Health Questionnaire-9 (PHQ-9)Intervention group: Significant decrease of 2.44. E-book control group: no significant differences. Xiaoai control group: small worsening of depressive symptoms.	The usability of the intervention was similar in all groups. However, the XiaoE group showed higher acceptability compared to the control groups, excelling in subscales such as satisfaction with the content, emotional awareness, learning new knowledge and relevance to daily life.
Anmella et al. [20],	The chatbot guided users through different activities and personalised reminders to complete their treatment plans. In addition, gamification strategies, user incentives, personalisation of psychological modules, reminders, and improvements to the chatbot's interpretability and responsiveness were included using Natural Language Processing (NLP) techniques. The intervention was adapted to offer users different usage flows and focused on flexibility and personalisation to improve engagement and effectiveness of the intervention. Control group: not applicable.	Patient Health Questionnaire-9 (PHQ-9)There was no significant change.	Participants showed high subjective perceptions of usability, satisfaction and acceptability with the chatbot. However, objective metrics of use, such as completion, adherence and engagement, indicated low participation. This suggests that, although they positively rated the usability and experience with the chatbot, their actual engagement with the intervention was limited.

Karkosz et al. [26],	The intervention involved interacting with a therapeutic chatbot called Fido, which used the Polish language. The chatbot was based on Cognitive Behavioural Therapy (CBT) techniques and addressed symptoms of depression and anxiety. Topics covered included general psychoeducation, identification and modification of cognitive distortions through Socratic questioning, identification of suicidal ideation, and redirection to suicide prevention hotlines. Control group: used therapeutic resources derived from a Cognitive Behavioural Therapy (CBT) manual provided through an online platform.	Centre for Epidemiologic Studies Depression Scale (CESD-R) Before: 28.33 (SD=13.84). After: 20.33 (SD=11.94). Significant decrease.	Some common responses on the negative aspects included failures to recognise the user's intention, leading the chatbot to indicate that it did not understand the conversation. In terms of positive effects, it was highlighted that the use of Fido encouraged users to decide to start traditional therapy.
Qi [28],	The intervention consisted of the experimental group interacting daily with the Woebot chatbot for 20 minutes each day. The Woebot chatbot was based on Cognitive Behavioural Therapy (CBT) and provided a conversational platform for participants. During these interactions, participants could communicate with the chatbot about their emotions, thoughts and experiences. The chatbot offered emotional support, coping tools and strategies to address symptoms of mild depression. Control group: It is not specifically mentioned whether they received any prompts or interventions.	Lovibond & Lovibond Self-Assessment Scale for Depression, Anxiety and Stress (DASS-21) Intervention grade: Before: 43.7 (SD=4.6). After: 36.2 (SD=2.8). Control group: Before: 43.5 (SD=3.8). After: 43.0 (SD=3.3).	Participants' perception or acceptance of the Woebot chatbot intervention is not specifically described.