Effectiveness of AI-driven Conversational Agents in Improving Mental Health Among Young People: A Systematic Review and Meta-analysis

Yaming Hang, Wenzhi Wu, Yi Feng, Xiaohang Song, Xiyao Xiao, Zhihong Qiao

Submitted to: Journal of Medical Internet Research
on: December 04, 2024

Disclaimer: **© The authors. All rights reserved.** This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Effectiveness of AI-driven Conversational Agents in Improving Mental Health Among Young People: A Systematic Review and Meta-analysis

Yaming Hang[1*] PhD; Wenzhi Wu[1*] PhD; Yi Feng[2] PhD; Xiaohang Song[1]; Xiyao Xiao[3]; Zhihong Qiao[1] Prof Dr

[1]Faculty of Psychology Beijing Normal University Beijing CN
[2]Lingxin AI beijing CN
[*]these authors contributed equally

Corresponding Author:
Yi Feng PhD

Abstract

Background: The increasing prevalence of mental health issues among adolescents and young adults, coupled with barriers to accessing traditional therapy, has led to growing interest in artificial-intelligence-driven (AI-driven) conversational agents (CAs) as a novel digital mental health intervention. Despite accumulating evidence suggesting the effectiveness of AI-driven CAs for mental health, there is still limited evidence on their effectiveness for different mental health conditions in adolescents and young adults.

Objective: This study aims to examine the effectiveness of AI-driven Conversational Agents for mental health among young people.

Methods: Five main databases (PubMed, PsycINFO, EMBASE, Cochrane Library, and Web of Science) were searched systematically, resulting in fifteen articles (including 16 randomized controlled trials) involving 1,974 participants. The quality of these studies, possible publication bias and moderators were then examined.

Results: The results indicated a moderate-to-large (Hedges' g = 0.60) effect of AI-driven CAs on reducing depressive symptoms, particularly in subclinical populations. However, their effectiveness in addressing other health issues, such as anxiety and stress, was not significant (p > 0.05).

Conclusions: The findings highlight the potential of AI-driven CAs for early intervention in depression among this population, and underscore the need for further improvements to enhance their efficacy across a broader range of mental health outcomes.

(JMIR Preprints 04/12/2024:69639)
DOI: https://doi.org/10.2196/preprints.69639

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Review

Yaming Hang[1], Wenzhi Wu[1], Yi Feng[2]*, Xiaohang Song[1], Xiyao Xiao[3], Zhihong Qiao[1]*
[1] Beijing Key Laboratory of Applied Experimental Psychology, Faculty of Psychology, Beijing Normal University, China
[2] Mental Health Center, Central University of Finance and Economics, Beijing, China
[3] Lingxin AI, Beijing, China

[1]Yaming Hang and Wenzhi Wu contributed equally to this work.
*Correspondence: Yi Feng (fengyi@cufe.edu.cn) or Zhihong Qiao (qiaozhihong@bnu.edu.cn)

Effectiveness of AI-driven Conversational Agents in Improving Mental Health Among Young People: A Systematic Review and Meta-analysis

Abstract

Background: The increasing prevalence of mental health issues among adolescents and young adults, coupled with barriers to accessing traditional therapy, has led to growing interest in artificial-intelligence-driven (AI-driven) conversational agents (CAs) as a novel digital mental health intervention. Despite accumulating evidence suggesting the effectiveness of AI-driven CAs for mental health, there is still limited evidence on their effectiveness for different mental health conditions in adolescents and young adults.
Objective: This study aims to examine the effectiveness of AI-driven Conversational Agents for mental health among young people.
Methods: Five main databases (PubMed, PsycINFO, EMBASE, Cochrane Library, and Web of Science) were searched systematically, resulting in fifteen articles (including 16 randomized controlled trials) involving 1,974 participants. The quality of these studies, possible publication bias and moderators were then examined.
Results: The results indicated a moderate-to-large (Hedges' g = 0.60) effect of AI-driven CAs on reducing depressive symptoms, particularly in subclinical populations. However, their effectiveness in addressing other health issues, such as anxiety and stress, was not significant ($p > 0.05$).
Conclusions: The findings highlight the potential of AI-driven CAs for early intervention in depression among this population, and underscore the need for further improvements to enhance their efficacy across a broader range of mental health outcomes.

Keywords: artificial intelligence; conversational agents; meta-analysis; mental health intervention; young people

Introduction

Mental health issues among adolescents and young adults are increasingly becoming a public health concern, affecting between 10% to 20% of the global youth population [1]. The early-onset mental health disorders are particularly alarming, with 50% of cases emerging before the age of 14 and 75% by the age of 25 [2]. Despite the significant impact of mental health disorders on young populations, these conditions remain underdiagnosed and undertreated [3]. The impact of these untreated conditions is profound, as persistent mental health problems often extend into adulthood, leading to impairments in educational achievement, psychosocial functioning, and overall quality of life [4, 5]. The COVID-19 pandemic has exacerbated these challenges, resulting in a marked increase in rates of depression, anxiety, and stress among young people [6].

In parallel with the rise in mental health issues, this generation of young people is growing up in a digital world. Over 90% of individuals aged 15-24 are “online”, and even in low-income countries, mobile access is widespread [7, 8]. Adolescents and young adults are also the earliest adopters and heaviest users of new technologies [9]. This level of digital engagement provides a unique opportunity to leverage digital mental health interventions, which can bridge the treatment gap by offering scalable, accessible, and cost-effective solutions [10]. Compared to traditional face-to-face therapy, web- and mobile-based interventions provide anonymity, reduce stigma, and offer greater flexibility and autonomy [11]. As a result, digital mental health interventions have gained increasing attention. For example, in the third global survey on eHealth, WHO reported that 58% of the surveyed countries have integrated digital health strategies as part of their healthcare frameworks [12]. However, early forms of digital mental health interventions, such as Internet-Based Cognitive Behavioral Therapy (ICBT), encounter several challenges, including limited interactivity and relatively high dropout rates [13]. Moreover, these interventions tend to be generalized, often lacking the personalization needed to meet the unique needs of individual users.

With the development of large language model in the field of artificial intelligence (AI), a promising avenue for digital mental health interventions is the use of AI-driven Conversational Agents (CAs), which use artificial intelligence to simulate human behavior and offer a task-oriented framework with evolving dialogue able to engage users in conversation [14]. These agents can provide psychoeducation and deliver treatment options [15], such as Cognitive Behavioral Therapy (CBT). AI-driven CAs offer personalized and interactive mental health support, engaging users in therapeutic dialogues that simulate human conversations [16]. Compared with the rule-based systems that rely on predefined responses, AI-driven CAs adapt and personalize their interactions based on user inputs, which may enhance therapeutic engagement and outcomes [17]. Although these systems are increasingly utilized among adults, their effectiveness for adolescents and young adults remains underexplored. Given the increasing mental health burden and the unique digital engagement patterns of younger individuals, understanding the potential of AI-driven CAs to support mental health among this group is crucial.

Despite the growing interest in AI-driven CAs for mental health, there is still limited evidence on their effectiveness for various mental health conditions in adolescents and young adults. Previous reviews have often combined rule-based and AI-driven CAs or included both young people and older adults [18, 19], which may lead to significant heterogeneity. To address these gaps, this meta-analysis aims to evaluate the effectiveness of AI-driven CAs in reducing mental health symptoms, particularly depression and anxiety, among adolescents and young adults aged 12-25. Additionally, this study explores the moderators that may influence treatment outcomes, such as characteristics of study population and AI-driven CAs, to better understand the factors that enhance the intervention effectiveness of these digital tools among this population.

Methods

Literature Search

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [20]. To locate studies assessing the effectiveness of AI-driven CAs for mental health problems in adolescents and young adults, two independent researchers (LYN and SXH) conducted a comprehensive search across five databases: PubMed, PsycINFO, EMBASE, Cochrane Library, and Web of Science. The search spanned from the inception of each database up to August 6th, 2024. The following search terms were used: (robot OR social bot OR dialogue system

OR conversational agent OR conversational bot OR conversational system OR conversational interface OR chatbot OR chat bot OR chatterbot OR chatter bot OR chat-bot OR smartbot OR smart bot OR smart-bot OR virtual coach OR virtual agent OR embodied agent OR relational agent OR avatar OR virtual character OR animated character OR virtual human OR virtual assistant OR digital assistant OR counseling agent) AND (mental illness OR mental disorder* OR affective disorder OR psychotic disorder OR post-traumatic stress disorder OR PTSD OR distress OR depress OR anxiety OR bipolar OR schizophrenia OR psychosis OR mental health OR mental wellness OR wellbeing OR well-being OR SWB OR happiness OR happy OR positive affect OR negative affect OR positive emotion OR negative emotion OR mood OR life satisfaction OR healthy relationship OR resilience OR self-efficacy). Detailed Search strategies were also provided in Multimedia Appendix 1. No filters were applied to ensure the inclusion of all relevant studies. Additionally, reference lists of included studies and previous reviews were manually searched to identify any further eligible studies. A detailed search strategy is provided in the Supplementary Material.

Inclusion and Exclusion Criteria

Studies were selected based on these criteria: (1) Population: studies using AI-driven CAs for managing mental health issues were included if the average age of participants was between 12 and 25 years. This age range followed previous meta-analyses conducted in adolescents and young adults [21]. No restrictions were imposed on the eligible participant populations regarding diagnoses of common mental disorders or any other clinical or demographic characteristics. Participants could be from clinical (formally diagnosed mental health conditions), subclinical (self-reported or screened mental health symptoms), or unselected/nonclinical populations; (2) Intervention: we included interventions delivered by AI-driven CAs. These CAs utilized artificial intelligence technologies such as natural language processing or machine learning to engage in human-like conversations, distinct from rule-based systems. This definition followed a previous study [19]; (3) Comparator: eligible studies included any control conditions, such as waitlist or active control groups (e.g., treatment as usual, therapist-led interventions); (4) Outcomes: studies were included if they reported at least one mental health outcome and provided sufficient data for effect size calculation; (5) Study design: only randomized controlled trials (RCTs) were included. Studies on rule-based CAs, review articles, conference abstracts, and non-English publications were excluded. Screening was performed independently by two researchers (WWZ and HYM), and full texts of potential studies were obtained for detailed eligibility assessment.

Data Extraction and Quality Assessment

For each included study, the following data were extracted: authorship, year of publication, participant characteristics (e.g., sample size, gender distribution, mean age), CA specifications (e.g., name, platform, response generation approach, interaction mode), intervention details (e.g., length, control group type), and measures. Methodological quality was assessed using the Cochrane Risk of Bias tool [22], considering factors such as random sequence generation, allocation concealment, blinding of participants and assessors, handling of incomplete outcome data, and selective reporting. Two authors (WWZ and SXH) independently conducted the data extraction and quality assessments. Disagreements were resolved through discussion, with the involvement of a third author (HYM) when necessary. The results of the risk of bias assessments are visually presented in a summary graph.

Meta-analytic Procedure

For each study, the means, standard deviations, and sample sizes at post-test were extracted to compute effect sizes (ESs). When multiple studies reported data for the same outcome, pooled ESs were calculated. Given the small sample sizes in some studies, Hedges' *g* was used to adjust for bias [23]. All ESs were coded so that positive Hedges' *g* values indicated superior outcomes for the treatment group relative to the control group. Where intention-to-treat and completer analyses were both available, data from the intention-to-treat analyses were used. Follow-up data were not analyzed due to insufficient reporting across studies. Effect size calculations and overall estimates were performed using Comprehensive Meta-Analysis software (Version 3.0) and Stata SE 15.1. For studies that did not report means and standard deviations, alternative statistics (e.g., Cohen's *d*, *t* values, *F* values) were used. Multi-arm trials were treated in accordance with Cochrane guidelines by combining means and standard deviations to create a single pairwise comparison [24]. Given the expected heterogeneity across studies, a random-effects model was applied to estimate the mean ESs [25]. Heterogeneity was examined using the *Q* statistic and I^2 index [23]. Outliers were identified as studies whose 95% confidence intervals lay outside the 95% confidence interval of the overall estimate [26]. A "leave-one-out" sensitivity analysis was conducted to assess the robustness of the results. Subgroup and meta-regression analyses were performed with outliers excluded.

To assess publication bias, three methods were used: (1) visual inspection of funnel plots, (2) Duval and Tweedie's trim-and-fill procedure [27] to adjust ES estimates for publication bias, and (3) Egger's test for funnel plot asymmetry [28]. Moderator analyses were performed for both categorical and continuous moderators where heterogeneity was significant ($p < 0.10$ or $I^2 > 25\%$). Subgroup analyses utilized a mixed-effects model, while meta-regression used unrestricted maximum likelihood estimation. Moderators included age, gender, intervention length, interaction mode, delivery platform, response generation approach, sample type, and control group type, selected based on previous meta-analytic studies [19, 29]. All visualization was conducted by *R* version 4.3.1 and Review Manager 5.3.

Results

Search Results

The initial database search identified 14,907 potentially relevant articles, supplemented by two additional studies found through manual reference list checks. After removing 7,412 duplicates, 7,497 unique articles remained for screening. Titles and abstracts were reviewed, resulting in the exclusion of 7,100 records. Subsequently, 397 full-text articles were assessed for eligibility. The PRISMA flow diagram outlining this process is provided (Figure 1). Ultimately, 15 articles with 16 RCTs were included in the systematic review and meta-analysis.

-----------------*Please insert Figure 1 here*-----------------

Study Characteristics

The characteristics of the 16 included RCTs are presented below (Table S1 in Multimedia Appendix 1 [30-44]). Sample size ranged from 42 to 415 participants, with a total pooled sample size of 1,974 across all studies. The meta-analysis incorporated studies conducted with clinical (*n* = 1), subclinical (*n* = 7), and nonclinical (*n* = 8) populations. Twelve studies used retrieval-based CAs, three employed generative CAs, and one study utilized both retrieval-based and generative CAs. Regarding interaction modes, thirteen studies employed text-based CAs, and three used multimodal CAs.

Risk of Bias in Included Studies

The overall quality of the included studies was suboptimal. Only one study [36] satisfied all six quality criteria. Three studies met four criteria, while five studies met three, and seven studies met fewer than three criteria. Notably, a significant number of studies lacked sufficient information to assess certain criteria: seven out of 16 studies did not report on blinding of participants or personnel, ten did not mention blinding of outcome assessors, and nine studies were not registered and lacked information on selective reporting. A summary of the risk of bias for each criterion is displayed below (Figure 2).

-----------------*Please insert Figure 2 here*-----------------

Depression

Overall Effect

AI-driven CAs demonstrated a medium effect on depression symptoms at post-test, with a Hedges' *g* of 0.50 (95% CI [0.20, 0.80], *N* = 11, *z* = 3.24, *p* < 0.01; Figure 3). Significant heterogeneity was observed among the studies [*Q* (10) = 39.66, *p* < 0.001, I^2 = 74.8%]. One study [30] was identified as an outlier with 95% confidence interval outside the 95% confidence interval of the pooled studies. After this study was removed, sensitivity analyses indicated that the ES increased to medium-to-large (Hedges' *g* = 0.60, 95% CI [0.36, 0.84], *N* = 10, *z* = 4.95, *p* < 0.001) and the heterogeneity reduced but remained significant [*Q* (9) = 17.86, *p* < 0.05, I^2 = 49.6%]. This outlier study was excluded from further analyses. Complementary sensitivity analyses confirmed that no single study significantly influenced the results.

-----------------*Please insert Figure 3 here*-----------------

Publication Bias

Duval and Tweedie's trim-and-fill method found no evidence of publication bias (Figure S1 in Multimedia Appendix 1). Similarly, Egger's test revealed no significant bias (b_0 = -2.03, SE = 1.73, *t* = 1.78, one-tailed *p* = 0.14).

Moderators

Subgroup analyses revealed that sample type significantly moderated the ESs at post-test (Q_b = 8.46, *p* < 0.05; Table 1). Only studies conducted in subclinical samples exhibited significant ESs (*g* = 0.73), and the ESs were larger than those conducted in nonclinical samples (*g* = 0.04). Additionally, heterogeneities within subgroups were reduced to non-significance (*p* > 0.10). Meta-regression analyses indicated that none of the variables (i.e., mean age, publication year, sex, and quality criteria) were significant moderators (*p* > 0.05; Table 2).

Table 1. Moderation analysis with categorical variables for depression symptoms.

Moderators		N	g	95% CI	Qw	p	Qb	p
Interaction mode	Text-based	8	0.52	[0.21,0.83]	15.58	0.03	1.93	0.17
	Multimodal	2	0.82	[0.54,1.09]	0	0.99		
Response generation approach	Retrieval based	7	0.52	[0.28,0.76]	7.36	0.29	0.09	0.77

	Generative	2	0.69	[-0.37,1.75]	7.97	0.005		
Sample type	Clinical sample	1	0.91	[-0.11,1.94]			8.46	0.02
	Nonclinical sample	2	0.04	[-0.38,0.46]	0.18	0.67		
	Subclinical sample	7	0.73	[0.51,0.95]	8.43	0.21		
Control group	Active control	6	0.52	[0.23,0.81]	14.98	0.01	0.61	0.74
	Information only	3	0.38	[0.17,0.58]	1.69	0.43		
	Waitlist or assessment only	1	0.91	[-0.11,1.94]				
Delivery platform	Instant messenger platform	5	0.59	[0.15,1.03]	11.15	0.03	0.06	0.97
	Standalone app	4	0.62	[0.29,0.96]	6.52	0.09		
	Serious gaming platform	1	0.52	[-0.24,1.28]				
Intervention length	0-4 weeks	7	0.55	[0.23,0.86]	16.79	0.01	0.78	0.38
	> 4 weeks	3	0.76	[0.40,1.12]	0.54	0.76		

Table 2. Moderation analysis with continuous variables for depression symptoms

Moderators	*N*	*β*	*SE*	*Z*	*p*
Mean age	10	-0.06	0.06	-1.03	0.30
Year	10	0.04	0.04	0.97	0.33
%—total Female	10	-0.01	0.01	-0.76	0.45
Quality criteria	10	0.07	0.07	1.04	0.30

Generalized Anxiety

Overall Effect

The results showed that AI-driven CAs had a non-significant impact on generalized anxiety symptoms compared to control conditions (g = 0.42, 95% CI [-0.04, 0.87], N = 10, z = 1.78, p = 0.08; Figure S2 in Multimedia Appendix 1). Large and significant heterogeneity was observed (Q (9) = 71.45, p < 0.001, I^2 = 87.4%). One study [42] was identified as an outlier with 95% confidence interval outside the 95% confidence interval of the pooled studies. After this study was removed, the ES reduced and remained non-significant (Hedges' g = 0.17, 95% CI [-0.07, 0.42], N = 9, z = 1.38, p = 0.17) and the heterogeneity remained significant [Q (5) = 16.41, p < 0.05, I^2 = 51.3%]. This outlier study was excluded from further analyses. Complementary sensitivity analyses confirmed the robustness of the findings.

Publication Bias

Duval and Tweedie's trim-and-fill method found no evidence of publication bias (Figure S1 in Multimedia Appendix 1). Similarly, Egger's test revealed no significant bias (b_0 = -2.03, SE = 1.73, t = 1.78, one-tailed p = 0.14). Duval and Tweedie's trim-and-fill analysis indicated that two studies were missing on the left side of the mean effect size. After imputing the missing studies under a random-effects model, the adjusted effect size remained non-significant (g = 0.06, 95% CI [-0.21, 0.32], Q (11) = 27.20). However, Egger's test failed to detect significant publication bias (b_0 = 1.97, SE = 1.68, t = 1.18, one-tailed p = 0.14).

Moderators

Given that the adjusted effect size was non-significant (p > 0.05), moderator analyses were not conducted.

Stress

Overall Effect

AI-driven CAs had a non-significant impact on stress at post-test compared to control groups (g = 0.002, 95% CI [-0.19, 0.20], N = 4, z = 0.02, p = 0.98; Figure S3 in Multimedia Appendix 1). No significant heterogeneity was detected among the studies (Q (3) = 2.39, p = 0.50, I^2 = 0.0%). Sensitivity analyses showed that the results were not driven by any single study, and no outliers were identified.

Publication Bias

Neither Duval and Tweedie’s trim-and-fill method nor Egger’s test found any evidence of publication bias (b_0 = -0.68, SE = 1.75, t = 0.39, one-tailed p = 0.37; Figure S1 in Multimedia Appendix 1).

Moderators

Since the overall effect size and heterogeneity were not significant (p > 0.05), moderator analyses were not performed.

Positive Affect

Overall Effect

The effect of AI-driven CAs on positive affect at post-test was non-significant (g = 0.01, 95% CI [-0.24, 0.27], N = 7, z = 0.11, p = 0.92; Figure S4 in Multimedia Appendix 1). There was significant heterogeneity among the studies (Q (6) = 16.28, p = 0.012, I^2 = 63.1%). Sensitivity analyses confirmed that the results were not driven by any individual study, and no outliers were detected.

Publication Bias

Duval and Tweedie’s trim-and-fill method and Egger’s test (b_0 = 2.12, SE = 2.21, t = 0.96, one-tailed p = 0.19; Figure S1 in Multimedia Appendix 1) both indicated no publication bias.

Moderators

Given the non-significant effect size (p > 0.05), no moderator analyses were conducted.

Negative Affect

Overall Effect

Similar to the effect on positive affect, AI-driven CAs demonstrated a non-significant effect on negative affect compared to control groups at post-test (g = 0.36, 95% CI [-0.04, 0.76], N = 7, z = 1.77, p = 0.08; Figure S5 in Multimedia Appendix 1). Heterogeneity was large and significant among the studies (Q (6) = 38.11, p < 0.001, I^2 = 84.3%). One study [42] was identified as an outlier with 95% confidence interval outside the 95% confidence interval of the pooled studies. After this study was removed, the ES reduced and remained non-significant (Hedges’ g = 0.11, 95% CI [-0.06, 0.28], N = 10, z = 4.95, p < 0.001) and the heterogeneity reduced to non-significance [Q (5) = 5.64, p = 0.34, I^2 = 11.3%]. This outlier study was excluded from further analyses. Complementary sensitivity analyses confirmed that no study disproportionately influenced the results.

Publication Bias

Duval and Tweedie’s trim-and-fill analysis indicated that one study was missing on the left side of the mean effect size. After imputing the missing study under a random-effects model, the adjusted effect size remained non-significant (*g* = 0.07, 95% CI [-0.13, 0.27], *Q* (6) = 8.96; Figure S1 in Multimedia Appendix 1). However, Egger’s test suggested no evidence of publication bias (b_0 = 1.53, SE = 1.40, *t* = 1.09, one-tailed *p* = 0.17).

Moderators

Given the non-significant overall effect size (*p* > 0.05), no moderator analyses were performed.

Mental Well-being

Overall Effect

The effect of AI-driven CAs on mental well-being at post-test was non-significant (*g* = 0.04, 95% CI [-0.21, 0.29], *N* = 4, *z* = 0.31, *p* = 0.76; Figure S6 in Multimedia Appendix 1). There was significant heterogeneity among the studies (*Q* (3) = 4.43, *p* = 0.22, I^2 = 32.3%). Sensitivity analyses confirmed that the results were not driven by any individual study, and no outliers were detected.

Publication Bias

Duval and Tweedie’s trim-and-fill method and Egger’s test (b_0 = 3.46, SE = 2.01, *t* = 1.72, one-tailed *p* = 0.11; Figure S1 in Multimedia Appendix 1) both indicated no publication bias.

Moderators

Given the non-significant effect size (*p* > 0.05), no moderator analyses were conducted.

Discussion

This meta-analysis was the first comprehensive evaluation for the effectiveness of AI-driven CAs mental health intervention among young people. Sixteen studies with a total of 1,974 participants were evaluated. Findings underscored the potential of AI-driven CAs to significantly alleviate depressive symptoms, particularly in subclinical populations. However, their effects on other mental health outcomes, such as anxiety, stress, and negative affect, were less robust, revealing important insights into both the promise and limitations of AI-driven interventions in this demographic.

Principal Results and Comparison with Prior Work

The results of this meta-analysis revealed that AI-driven CAs demonstrated a moderate-to-large intervention effect on depressive symptoms. This finding aligns with previous research that has demonstrated the efficacy of AI-driven CAs in reducing depression among all age groups [19], which also revealed that AI-driven CAs had a moderate-to-large effect on depression. This suggests that AI-driven CAs, especially when enhanced with natural language processing and machine learning, can be particularly effective in mitigating depressive symptoms. To note, these results are more favorable compared to earlier meta-analyses that included rule-based systems, which reported smaller effect sizes (*g* ranging from 0.26 to 0.29) for depression [29, 45]. One possible explanation is that as AI-driven CAs can offer greater flexibility and adaptability in delivering therapeutic interventions [46, 47], thus they are generally more effective in managing depressive symptoms than their rule-based counterparts. It is also possible that the larger ES found in this study reflects that AI-based CAs are more beneficial to young people, for whom digital interventions may be more acceptable and engaging due to their familiarity with digital platforms [48]. This aligns with a previous review which indicated that younger age was associated with larger effect of CAs on depressive symptoms

[29].

In contrast to the substantial effects observed for depression, the effects of AI-driven CAs on anxiety, stress, positive affect, negative affect, and mental well-being in this age group were all non-significant. This aligns with previous meta-analyses that have found AI-based CA interventions to be less effective for anxiety, positive affect, negative affect, and psychological well-being compared to depression [18, 19]. The non-significant findings for anxiety and stress in this meta-analysis may be explained by the limited inclusion of behavioral strategies, such as exposure therapy, in current AI-driven CAs. As anxiety and stress often require more intensive behavioral interventions [49], future iterations of AI-driven CAs may benefit from integrating these techniques to improve outcomes for anxiety-related symptoms. In addition, the small and non-significant effects on outcomes related to well-being may suggest that AI-based CAs were not yet able to enhance well-being in young people. It is possible that as most AI-driven CAs were based on CBT [45], they are less effective in cultivating positive psychological assets.

Subgroup analyses revealed the significant role of clinical vs. subclinical type in moderating the efficacy of AI-driven CAs on depression intervention. Specifically, AI-driven CAs were particularly effective in subclinical populations. This finding aligns with a previous meta-analysis [19], suggesting that subclinical populations are more likely to benefit from AI-based CAs. This is consistent with the broader literature on psychological interventions, which has shown that these interventions are often more effective in promoting mental well-being for people with mental or physical health conditions compared to the general population [50]. Subclinical depression is clinically significant not only because it can cause considerable impairment requiring intervention, but also due to the heightened risk of progression to major depressive disorder, which can potentially be prevented with early treatment [51]. The notable intervention effectiveness of AI-driven CAs among young people with subclinical depression provides an important insight that these digital tools may serve as valuable early interventions, to help mitigate the risk of developing more severe mental health conditions.

The non-significant moderating effects of interaction mode, response generation approach, and delivery platform revealed that CA technical features may not influence the effectiveness of AI-driven CAs in reducing depression among young people. It is possible that as young people are familiar with digital platforms, they can interact effectively with AI-driven CAs with different technical features. Age and sex also did not moderate post-test effect. This may reflect that AI-driven CAs could be effective for both males and females, and for young people in different age groups. In addition, intervention length did not moderate the ESs on depression, which may reflect that both short-course and long-course treatments delivered by AI-driven CAs could be effective in alleviating depressive symptoms. Finally, control group type, publication year and study quality did not moderate post-test ESs, which further support the robustness of the effectiveness of AI-driven CAs for depression among young people.

Limitations

Despite the promising results, several limitations should be acknowledged. First, the limited number of studies examining the long-term effects of AI-driven CAs prevented a thorough evaluation of the sustainability of treatment outcomes. As digital interventions continue to gain prominence, it is crucial for future studies to include follow-up assessments to better understand the durability of therapeutic effects. Second, the inclusion of only English-language studies may have introduced selection bias, limiting the generalizability of our findings. Finally, we included CAs based on various therapeutic orientations, which may lead to considerable heterogeneity in results.

Implications

Despite the promising results, several limitations should be acknowledged. First, the limited number of studies examining the long-term effects of AI-driven CAs prevented a thorough evaluation of the sustainability of treatment outcomes. As digital interventions continue to gain prominence, it is crucial for future studies to include follow-up assessments to better understand the durability of therapeutic effects. Second, the inclusion of only English-language studies may have introduced selection bias, limiting the generalizability of our findings. Finally, we included CAs based on various therapeutic orientations, which may lead to considerable heterogeneity in results.

Conclusions

With continued advancements in AI technologies, these digital tools have the potential to play a pivotal role in bridging the mental health treatment gap for young people. This meta-analysis provides robust evidence for the effectiveness of AI-driven CAs in reducing depressive symptoms among young people, particularly in subclinical populations, indicating great promise as a novel tool in scalable and accessible interventions. Their effectiveness for anxiety, stress, and outcomes related to well-being is not robust, highlighting the need for further development. Future research should focus on refining the therapeutic capabilities of AI-driven CAs and exploring long-term mental-health outcomes.

Acknowledgements

This study was funded by the National Social Science Foundation of China under Grant No. 20&ZD153 and Higher Education Teaching Reform research and practice project of Henan Agricultural University (2024XJGLX052). The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript. We thank Helin Zou for visualization for the results.

Conflicts of Interest

None declared.

Abbreviations

AI: artificial intelligence
CA: conversational agent
CBT: cognitive behavioral therapy
ES: effect size
ICBT: Internet-Based Cognitive Behavioral Therapy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
WHO: World Health Organization

Multimedia Appendix 1

The supplemental content of Search strategies, study characteristics, funnel plots, and forest plots.

References

1. Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, et al. Child

and adolescent mental health worldwide: evidence for action. Lancet 2011 Oct 22;378(9801):1515-25. PMID: 22008427. doi: 10.1016/s0140-6736(11)60827-1.

2. Fusar-Poli P. Integrated Mental Health Services for the Developmental Period (0 to 25 Years): A Critical Review of the Evidence. Front Psychiatry 2019;10:355. PMID: 31231250. doi: 10.3389/fpsyt.2019.00355.
3. Islam MI, Yunus FM, Isha SN, Kabir E, Khanam R, Martiniuk A. The gap between perceived mental health needs and actual service utilization in Australian adolescents. Sci Rep 2022 Mar 31;12(1):5430. PMID: 35361817. doi: 10.1038/s41598-022-09352-0.
4. Copeland WE, Wolke D, Shanahan L, Costello EJ. Adult Functional Outcomes of Common Childhood Psychiatric Problems: A Prospective, Longitudinal Study. JAMA Psychiatry 2015 Sep;72(9):892-9. PMID: 26176785. doi: 10.1001/jamapsychiatry.2015.0730.
5. Goodman A, Joyce R, Smith JP. The long shadow cast by childhood physical and mental problems on adult life. Proc Natl Acad Sci U S A 2011 Apr 12;108(15):6032-7. PMID: 21444801. doi: 10.1073/pnas.1016970108.
6. Ravens-Sieberer U, Kaman A, Erhart M, Devine J, Schlack R, Otto C. Impact of the COVID-19 pandemic on quality of life and mental health in children and adolescents in Germany. Eur Child Adolesc Psychiatry 2022 Jun;31(6):879-89. PMID: 33492480. doi: 10.1007/s00787-021-01726-5.
7. Keeley B, Little C. The State of the Worlds Children 2017: Children in a Digital World: ERIC; 2017. ISBN: 9280649302.
8. Park BK, Calamaro C. A Systematic Review of Social Networking Sites: Innovative Platforms for Health Research Targeting Adolescents and Young Adults. J Nurs Scholarsh 2013 Sep 01;45(3):256-64. doi: 10.1111/jnu.12032.
9. Schmidt ME, Anderson DR. Children and television: Fifty years of research: Lawrence Erlbaum Associates; 2007. ISBN: 9781410618047.
10. Lattie EG, Stiles-Shields C, Graham AK. An overview of and recommendations for more accessible digital mental health services. Nat Rev Psychol 2022 Feb 01;1(2):87-100. doi: 10.1038/s44159-021-00003-1.
11. Ebert DD, Van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and Mobile-Based Psychological Interventions: Applications, Efficacy, and Potential for Improving Mental Health. Eur Psychol 2018 May 01;23(2):167-87. doi: 10.1027/1016-9040/a000318.
12. World Health O. Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. Geneva: World Health Organization; 2016. ISBN: 9789241511780.
13. Schmidt ID, Forand NR, Strunk DR. Predictors of Dropout in Internet-Based Cognitive Behavioral Therapy for Depression. Cognit Ther Res 2019 Jun;43(3):620-30. PMID: 32879540. doi: 10.1007/s10608-018-9979-5.
14. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. CAN J PSYCHIAT 2019 Jul;64(7):456-64. PMID: 30897957. doi: 10.1177/0706743719828977.
15. Dingler T, Kwasnicka D, Wei J, Gong E, Oldenburg B. The Use and Promise of Conversational Agents in Digital Health. Yearb. Med. Inform 2021 Aug;30(1):191-9. PMID: 34479391. doi: 10.1055/s-0041-1726510.
16. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-58. PMID: 30010941. doi: 10.1093/jamia/ocy072.
17. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The Next Generation: Chatbots in Clinical Psychology and Psychotherapy to Foster Mental Health – A Scoping Review. Verhaltenstherapie 2019;32(Suppl. 1):64-76. doi: 10.1159/000501812.

18. Abd-alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: A scoping review. Int J Med Inform 2019 Dec 01;132:103978. doi: 10.1016/j.ijmedinf.2019.103978.
19. Li H, Zhang R, Lee YC, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. npj Digital Med 2023 Dec 19;6(1):236. PMID: 38114588. doi: 10.1038/s41746-023-00979-5.
20. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. PLoS Med 2009;6(7):e1000100. doi: 10.1371/journal.pmed.1000100.
21. Bourke M, Patten RK, Dash S, Pascoe M, Craike M, Firth J, et al. The Effect of Interventions That Target Multiple Modifiable Health Behaviors on Symptoms of Anxiety and Depression in Young People: A Meta-Analysis of Randomized Controlled Trials. J Adolesc Health 2022 Feb 01;70(2):208-19. doi: 10.1016/j.jadohealth.2021.08.005.
22. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928-d. doi: 10.1136/bmj.d5928.
23. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to meta-analysis: John Wiley & Sons; 2009. ISBN: 9780470057247.
24. Higgins JPT, Deeks JJ, Altman DG. Special Topics in Statistics. Cochrane Handbook for Systematic Reviews of Interventions 2008. p. 481-529. ISBN: 9780470712184.
25. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011 Feb 10;342:d549. PMID: 21310794. doi: 10.1136/bmj.d549.
26. Cuijpers P. Meta-analyses in mental health research: A practical guide: Vrije Universiteit Amsterdam; 2016. ISBN: 9789082530506.
27. Duval S, Tweedie R. Trim and Fill: A Simple Funnel-Plot–Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. Biometrics 2000 Jun 01;56(2):455-63. doi: 10.1111/j.0006-341X.2000.00455.x.
28. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315(7109):629. doi: 10.1136/bmj.315.7109.629.
29. He Y, Yang L, Qian C, Li T, Su Z, Zhang Q, et al. Conversational Agent Interventions for Mental Health Problems: Systematic Review and Meta-analysis of Randomized Controlled Trials. J Med Internet Res 2023 Apr 28;25:e43862. doi: 10.2196/43862.
30. Bird T, Mansell W, Wright J, Gaffney H, Tai S. Manage Your Life Online: A Web-Based Randomized Controlled Trial Evaluating the Effectiveness of a Problem-Solving Intervention in a Student Sample. Behav Cogn Psychother 2018 Sep;46(5):570-82. PMID: 29366432. doi: 10.1017/s1352465817000820.
31. Drouin M, Sprecher S, Nicola R, Perkins T. Is chatting with a sophisticated chatbot as good as chatting online or FTF with a stranger? J Computers in Human Behavior 2022;128:107100. doi: 10.1016/j.chb.2021.107100
32. Ehrlich C, Hennelly SE, Wilde N, Lennon O, Beck A, Messenger H, et al. Evaluation of an artificial intelligence enhanced application for student wellbeing: pilot randomised trial of the mind tutor Int J Appl Posit Psychol 2024;9(1):435-54.
33. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 6;4(2):e19. PMID: 28588005. doi: 10.2196/mental.7785.
34. Gaffney H, Mansell W, Edwards R, Wright J. Manage Your Life Online (MYLO): a pilot trial of a conversational computer-based intervention for problem solving in a student sample. Behav Cogn Psychother 2014 Nov;42(6):731-46. PMID: 23899405. doi:

10.1017/s135246581300060x.

35. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Ment Health 2018 Dec 13;5(4):e64. PMID: 30545815. doi: 10.2196/mental.9782.
36. 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 Nov 21;24(11):e40719. PMID: 36355633. doi: 10.2196/40719.
37. Jang S, Kim JJ, Kim SJ, Hong J, Kim S, Kim E. 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 2021 Jun;150:104440. PMID: 33799055. doi: 10.1016/j.ijmedinf.2021.104440.
38. 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 Aug 12;5(8):e20678. PMID: 34092548. doi: 10.2196/20678.
39. 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 Mar;27:100495. PMID: 35059305. doi: 10.1016/j.invent.2022.100495.
40. Liu I, Liu F, Xiao Y, Huang Y, Wu S, Ni S. Investigating the key success factors of chatbot-based positive psychology intervention with retrieval-and generative pre-trained transformer (GPT)-based chatbots. INT J HUM-COMPUT INT 2024:1-12.
41. Pinto MD, Hickman RL, Jr., Clochesy J, Buchner M. Avatar-based depression self-management technology: promising approach to improve depressive symptoms among young adults. Appl Nurs Res 2013 Feb;26(1):45-8. PMID: 23265918. doi: 10.1016/j.apnr.2012.08.003.
42. Romanovskyi O, Pidbutska N, Knysh A, editors. Elomia Chatbot: The Effectiveness of Artificial Intelligence in the Fight for Mental Health. International Conference on Computational Linguistics and Intelligent Systems; 2021.
43. Terblanche N, Molyn J, De Haan E, Nilsson VO. Coaching at Scale: Investigating the Efficacy of Artificial Intelligence Coaching. INT J EVID BASED COA 2022;20(2). DOI:10.24384/5cgf-ab69
44. 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 Nov 22;6(11):e40242. PMID: 36413390. doi: 10.2196/40242.
45. Zhong W, Luo J, Zhang H. The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: A systematic review and meta-analysis. J Affect Disord 2024 2024/07/01/;356:459-69. doi: 10.1016/j.jad.2024.04.057.
46. Kurniawan MH, Handiyani H, Nuraini T, Hariyati RTS, Sutrisno S. A systematic review of artificial intelligence-powered (AI-powered) chatbot intervention for managing chronic illness. Ann Med 2024 Dec 31;56(1):2302980. doi: 10.1080/07853890.2024.2302980.
47. López-Cózar R, Callejas Z, Espejo G, Griol D. Enhancement of Conversational Agents By Means of Multimodal Interaction. In: Perez-Marin D, Pascual-Nieto I, editors. Conversational Agents and Natural Language Interaction: Techniques and Effective Practices. Hershey, PA, USA: IGI Global; 2011. p. 223-52. doi: 10.4018/978-1-60960-617-6.ch010
48. Christensen H, Hickie IB. Using e-health applications to deliver new mental health services.

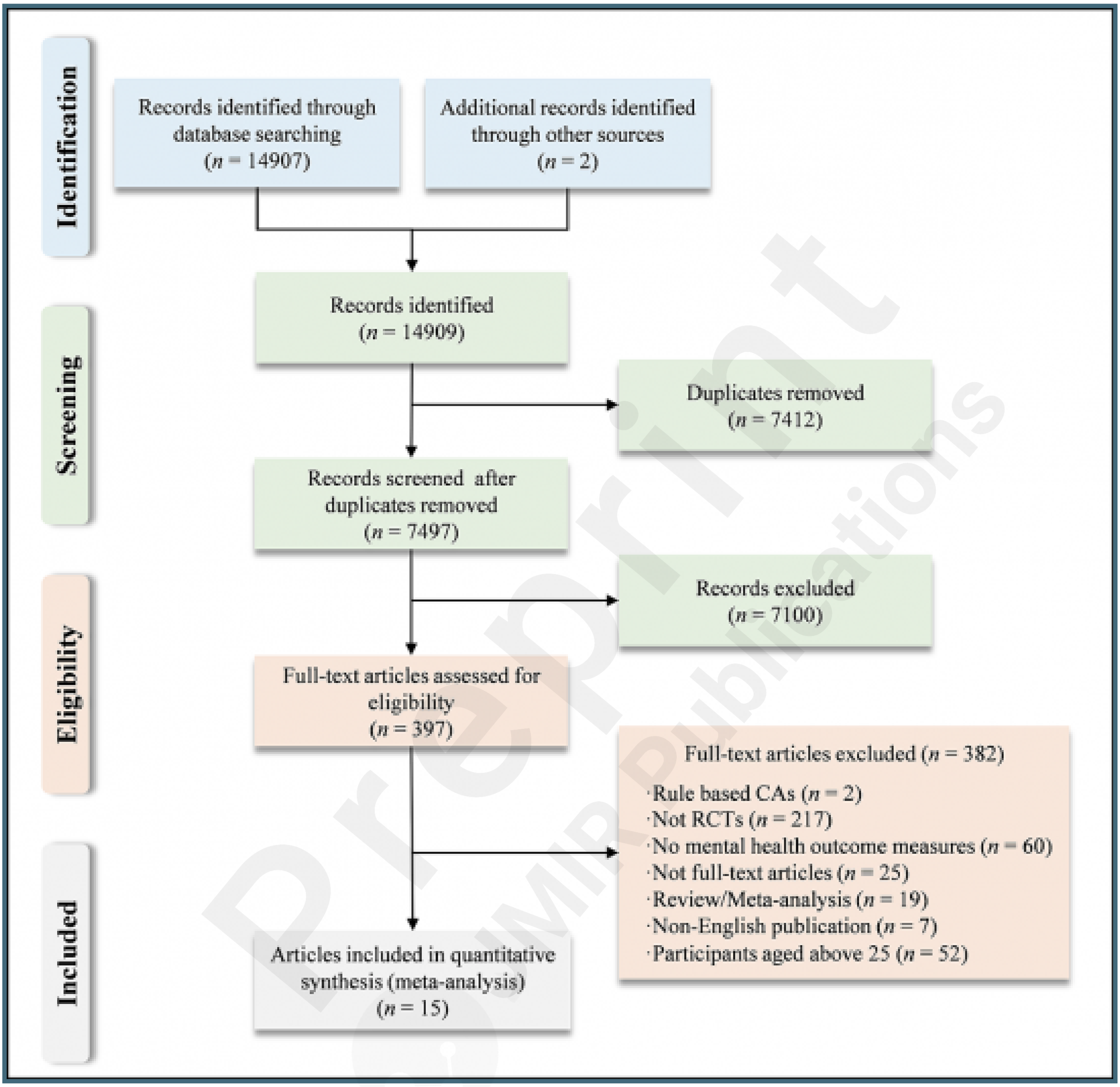
Med J Aust 2010 Jun 7;192(S11):S53-6. PMID: 20528711. doi: 10.5694/j.1326-5377.2010.tb03695.x.

49. Carpenter JK, Andrews LA, Witcraft SM, Powers MB, Smits JAJ, Hofmann SG. Cognitive behavioral therapy for anxiety and related disorders: A meta-analysis of randomized placebo-controlled trials. Depress Anxiety 2018 Jun 01;35(6):502-14. doi: 10.1002/da.22728.
50. van Agteren J, Iasiello M, Lo L, Bartholomaeus J, Kopsaftis Z, Carey M, et al. A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. Nature Human Behaviour. 2021 May 01;5(5):631-52. doi: 10.1038/s41562-021-01093-w.
51. Cuijpers P, Koole SL, van Dijke A, Roca M, Li J, Reynolds CF. Psychotherapy for subclinical depression: meta-analysis. Br J Psychiatry 2014;205(4):268-74. doi: 10.1192/bjp.bp.113.138784.

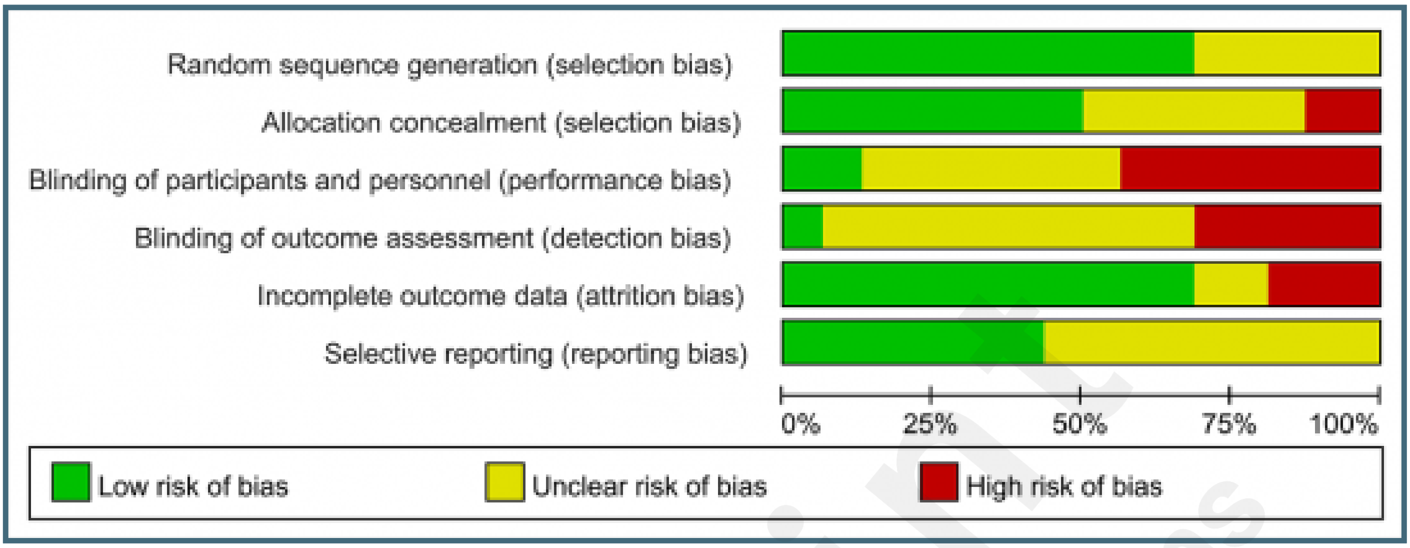
Supplementary Files

Figures

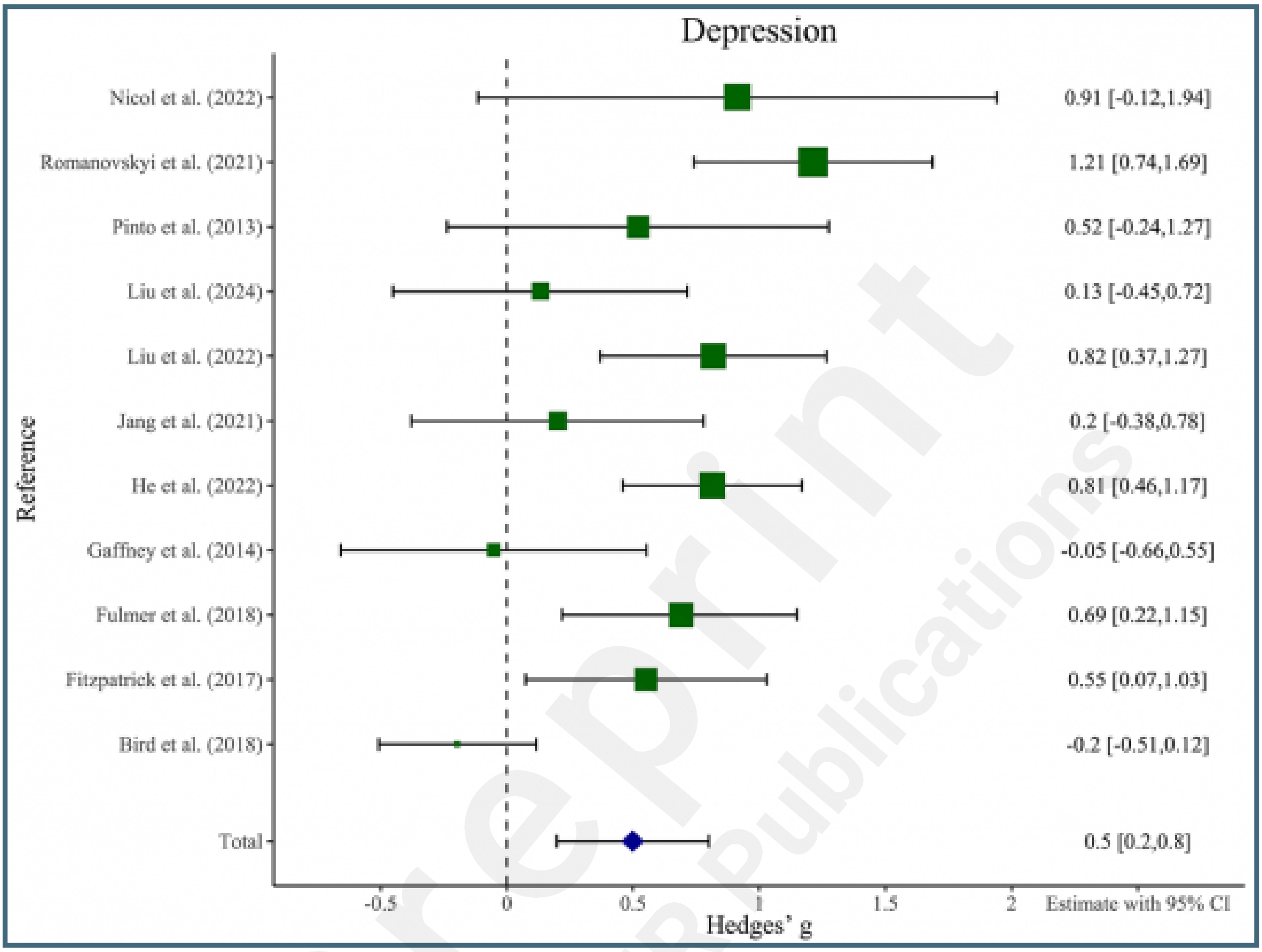
Flowchart of the study inclusion procedure.


Identification
Screening
Eligibility
Included
Records identified through database searching (n = 14907)
Additional records identified through other sources (n = 2)
Records identified (n = 14909)
Duplicates removed (n = 7412)
Records screened after duplicates removed (n = 7497)
Records excluded (n = 7100)
Full-text articles assessed for eligibility (n = 397)
Full-text articles excluded (n = 382)
·Rule based CAs (n = 2)
·Not RCTs (n = 217)
·No mental health outcome measures (n = 60)
·Not full-text articles (n = 25)
·Review/Meta-analysis (n = 19)
·Non-English publication (n = 7)
·Participants aged above 25 (n = 52)
Articles included in quantitative synthesis (meta-analysis) (n = 15)

Risk of bias graph.



Forest plot of effect sizes for AI-driven CAs in depression.



Multimedia Appendixes

The supplemental content of Search strategies, study characteristics, funnel plots, and forest plots.
URL: http://asset.jmir.pub/assets/2ac60a0af96f49445b0ab2617bfc1307.docx