Review

Self-Administered Interventions Based on Natural Language Processing Models for Reducing Depressive and Anxiety Symptoms: Systematic Review and Meta-Analysis

David Villarreal-Zegarra[1,2*], MPH; C Mahony Reategui-Rivera[2*], MD; Jackeline García-Serna[1], BSc; Gleni Quispe-Callo[1], BSc; Gabriel Lázaro-Cruz[1], Mg; Gianfranco Centeno-Terrazas[1], MPH; Ricardo Galvez-Arevalo[3], Mg; Stefan Escobar-Agreda[4], MD; Alejandro Dominguez-Rodriguez[5], PhD; Joseph Finkelstein[2], MD, PhD

[1]Instituto Peruano de Orientación Psicológica, Lima, Peru

[2]Department of Biomedical Informatics, School of Medicine, University of Utah, Salt Lake City, UT, United States

[3]Instituto Nacional de Salud del Niño San Borja, Lima, Peru

[4]Telehealth Unit, Universidad Nacional Mayor de San Marcos, Lima, Peru

[5]Department of Psychology, Health, and Technology, University of Twente, Enschede, Netherlands

[*]these authors contributed equally

Corresponding Author:
C Mahony Reategui-Rivera, MD
Department of Biomedical Informatics
School of Medicine
University of Utah
421 Wakara Way
Salt Lake City, UT, 84108
United States
Phone: 1 (801) 581 4080
Email: mahony.reategui@utah.edu

Abstract

Background: The introduction of natural language processing (NLP) technologies has significantly enhanced the potential of self-administered interventions for treating anxiety and depression by improving human-computer interactions. Although these advances, particularly in complex models such as generative artificial intelligence (AI), are highly promising, robust evidence validating the effectiveness of the interventions remains sparse.

Objective: The aim of this study was to determine whether self-administered interventions based on NLP models can reduce depressive and anxiety symptoms.

Methods: We conducted a systematic review and meta-analysis. We searched Web of Science, Scopus, MEDLINE, PsycINFO, IEEE Xplore, Embase, and Cochrane Library from inception to November 3, 2023. We included studies with participants of any age diagnosed with depression or anxiety through professional consultation or validated psychometric instruments. Interventions had to be self-administered and based on NLP models, with passive or active comparators. Outcomes measured included depressive and anxiety symptom scores. We included randomized controlled trials and quasi-experimental studies but excluded narrative, systematic, and scoping reviews. Data extraction was performed independently by pairs of authors using a predefined form. Meta-analysis was conducted using standardized mean differences (SMDs) and random effects models to account for heterogeneity.

Results: In all, 21 articles were selected for review, of which 76% (16/21) were included in the meta-analysis for each outcome. Most of the studies (16/21, 76%) were recent (2020-2023), with interventions being mostly AI-based NLP models (11/21, 52%); most (19/21, 90%) delivered some form of therapy (primarily cognitive behavioral therapy: 16/19, 84%). The overall meta-analysis showed that self-administered interventions based on NLP models were significantly more effective in reducing both depressive (SMD 0.819, 95% CI 0.389-1.250; $P<.001$) and anxiety (SMD 0.272, 95% CI 0.116-0.428; $P=.001$) symptoms compared to various control conditions. Subgroup analysis indicated that AI-based NLP models were effective in reducing depressive symptoms (SMD 0.821, 95% CI 0.207-1.436; $P<.001$) compared to pooled control conditions. Rule-based NLP models showed effectiveness in reducing both depressive (SMD 0.854, 95% CI 0.172-1.537; $P=.01$) and anxiety (SMD 0.347, 95% CI 0.116-0.578; $P=.003$) symptoms. The meta-regression showed no significant association between participants' mean age and treatment outcomes (all

P>.05). Although the findings were positive, the overall certainty of evidence was very low, mainly due to a high risk of bias, heterogeneity, and potential publication bias.

Conclusions: Our findings support the effectiveness of self-administered NLP-based interventions in alleviating depressive and anxiety symptoms, highlighting their potential to increase accessibility to, and reduce costs in, mental health care. Although the results were encouraging, the certainty of evidence was low, underscoring the need for further high-quality randomized controlled trials and studies examining implementation and usability. These interventions could become valuable components of public health strategies to address mental health issues.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42023472120; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42023472120

Introduction

Background

Depression and anxiety pose a substantial worldwide burden. In 2020, depression and anxiety affected approximately 246 million and 374 million people, respectively [1]. Moreover, these conditions reduce individuals' quality of life and have significant economic repercussions [2]. The World Health Organization estimates that depression and anxiety result in a loss of US $1 trillion annually due to loss of productivity [3]. In addition, their increasing incidence and a lack of health resources challenge the health care systems and workforce to meet the growing demand for mental health care services adequately [4].

In response, self-administered technology-based interventions have emerged as promising solutions for managing these conditions. These self-guided interventions enable users to progress through treatments independently, without external support [4], and they have demonstrated the potential to reduce costs; save health providers' time; and improve satisfaction and access to care, especially during crises and quarantine periods, for patients with mental health conditions living in remote areas, those with disabilities, or those unable to afford traditional care [5]. However, despite the potential of self-directed interventions to manage mental health problems, many of these interventions face important challenges in user engagement and adherence [6].

Self-administered interventions that are effective vary by delivery format, including web-based platforms, mobile apps, and virtual or augmented reality [7,8]. These interventions can be integrated within a professional intervention package or be completely independent of any external support [9,10]. Furthermore, they can be based solely on the presentation of relevant therapeutic information, typically based on a behavioral cognitive approach [10-12], or rely on machine learning (ML) models to process the natural language of clients' responses [13].

Natural language processing (NLP) offers a promising avenue for enhancing the efficacy of self-administered interventions. Defined as a cross-disciplinary field focused on enabling computers to comprehend, process, and interact with human language [14], NLP has the potential to make self-directed interventions more cost-effective and accessible and facilitate fidelity and engagement of patients through better interaction [15].

Moreover, NLP can be categorized into 2 main approaches: rule based and artificial intelligence (AI) based. Rule-based NLP uses predefined linguistic rules to guide text interpretation, offering high explainability but limited flexibility in handling complex language nuances [16]. Conversely, AI-based NLP, encompassing ML and deep learning techniques, learns from extensive data to process language. It has shown remarkable success in various NLP tasks due to its scalability and ability to manage linguistic ambiguities [17].

The advent of large language models and multimodal large language models has further enhanced the capabilities of NLP-based health interventions. These advances are not limited to enhanced user interaction but extend to personalizing therapeutic modalities to the patient's unique requirements, as demonstrated in specific psychotherapeutic settings [18].

Previously, other systematic reviews, such as those conducted by Le Glaz et al [19] and Zhang et al [20], analyzed the impact of NLP on mental health. However, these reviews primarily focused on the general applications of NLP in mental health. In addition, another systematic review demonstrated promising results for NLP-based interventions in mental health, but the findings encompassed a broad range of mental health disorders and did not specifically address self-administered interventions [15].

Objectives

Although these advances are highly promising, analysis of their effectiveness and safety in managing mental health concerns such as depression and anxiety remains fragmented [21]. This study aims to systematically review available literature to determine the effect of self-administered NLP-based interventions on symptoms of depression and anxiety.

Methods

Design and Protocol Registration

This study systematically searched available literature in the principal health databases and synthesized the main quantitative results in a meta-analysis. Our study adheres to the PRISMA

XSL•FO
RenderX

(Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (refer to Multimedia Appendix 1 for the PRISMA 2020 checklist) and the Cochrane Collaboration recommendations for meta-analyses [22]. The protocol for this systematic review was registered in the PROSPERO repository (CRD42023472120).

Eligibility Criteria

Our study follows the PICOS (Population, Intervention, Comparison, Outcomes, and Study Design) framework to evaluate whether interventions based on NLP models can effectively reduce depressive and anxiety symptoms. We define these symptoms as follows: (1) depressive symptoms are defined as a mood disorder characterized by the persistent presence of a profound sense of sadness, loss of interest or pleasure in daily activities, and a general lack of energy; and (2) anxiety symptoms are characterized by the anticipation of imagined events that are perceived as potential threats, causing emotional distress and physiological tension.

The eligibility criteria for our review are presented in Textbox 1.

Textbox 1. Eligibility criteria for the review determined using the PICOS (Population, Intervention, Comparison, Outcomes, and Study Design) framework.

Review eligibility criteria

- Population: we included studies with participants of any age group (child, adolescent, adult, and older adult) with or without previous comorbidities. Eligible studies must report participants who have been diagnosed with depression or anxiety through an interview or consultation with a mental health professional (eg, physician, psychologist, or psychiatrist) or assessed using validated psychometric instruments.

- Intervention: the intervention must be based on natural language processing (NLP) models such as large language models, multimodal large language models, artificial intelligence–led systems (ie, digital conversational agent, chatbot, or interactive voice response), and other NLP models. We included interventions regardless of their primary design purpose, provided they were self-administered.

- Comparison: we considered both passive (ie, waiting lists, nonintervention control groups, or placebos) and active (ie, web-based or face-to-face psychological interventions, virtual reality, serious games, biofeedback for mental health problems, pharmacological therapies to treat symptoms of depression and anxiety, or animal-assisted therapies) comparators.

- Outcomes: we included studies measuring depressive and anxiety symptom scores using validated psychometric questionnaires (eg, Patient Health Questionnaire-9, Beck Depression Inventory, Hamilton Depression Rating Scale, Generalized Anxiety Disorder-7, Beck Anxiety Inventory, Hamilton Anxiety Rating Scale, or similar instruments).

- Study Design: we included randomized controlled trials and quasi-experimental studies (without a control arm or randomization groups) that assessed the effect of NLP-based interventions on depressive and anxiety symptoms. We excluded narrative reviews, systematic reviews, scoping reviews, and other nonoriginal research designs. Only peer-reviewed publications (original articles or briefs) were included; proceedings, posters, and other similar items were excluded. There were no exclusion criteria based on language, publication date, or setting (ie, clinical or community settings).

Information Sources and Search Strategy

The databases we used for the systematic review were Web of Science, Scopus, MEDLINE (by PubMed), PsycINFO (by EBSCO), IEEE Xplore, Embase, and Cochrane Library. The search strategy included terms related to NLP as well as depression and anxiety, along with health science descriptors (refer to Multimedia Appendix 2 for the search strategy). Our search included any document available from inception to November 3, 2023.

Selection Process

We downloaded all records identified by the search strategy in RIS format and compiled them into an EndNote (Clarivate) file, which served as a repository for all retrieved records. Next, we used automated and manual methods to remove duplicate records. We exported the list of unique records from EndNote to Rayyan (Rayyan Systems Inc) for the selection process. First, 2 pairs of authors (JG-S with RG-A and GQ-C with GL-C) independently assessed the abstracts and titles of the studies to ensure that they met the inclusion criteria. Two pairs of authors reviewed the resulting retrieved text independently (JG-S with RG-A and GQ-C with GL-C). Any excluded studies were recorded along with the reasons for their exclusion (refer to Multimedia Appendix 3 for a list of the excluded studies). If disagreements arose between the reviewers at either stage, they were resolved by discussion. A third reviewer (DV-Z) was consulted if disagreement persisted to decide whether the study met the inclusion criteria. Records were included or excluded depending on whether they met the inclusion criteria. At the title and abstract stage, if it was unclear whether a record met all the inclusion criteria, it could proceed to the full-text stage, where a more detailed review was carried out (a sensitive approach). However, at the full-text stage, all inclusion criteria had to be met for final acceptance.

The title and abstract review were performed in English because this is the language in which the databases save the metadata. The full-text review and results extraction were mainly performed in English and Spanish (the languages the reviewers speak). When studies in other languages were found, the reviewers used DeepL Translator (DeepL SE) to translate the documents into English before proceeding with the review and extraction. Therefore, our review had no language limitations. It is important to note that all papers evaluated in the full-text review and extraction were in English.

Data Collection

Two pairs of authors (JGS with RGA and GQC with GLC) independently collected the information from the included studies using a predefined collection form in a Microsoft Excel sheet. Initially, a pilot data extraction process was conducted

on 5 data sets reviewed by all raters with 85% agreement. Subsequently, minor changes were made to the final version of the extraction form to improve the clarity of the extracted data, which included the following: (1) general information (ie, authors, year of publication, title, country, and language); (2) participant characteristics (ie, age range, sex, number of participants, and diagnosis); (3) intervention characteristics (ie, type of NLP model, duration, frequency, and brief description of the intervention); (4) comparator (passive or active); and (5) main outcomes (ie, means, SDs, preintervention and postintervention measures, and the effect size for control and intervention groups).

Risk-of-Bias Assessment and Certainty of Evidence

We used the JBI critical appraisal tools to identify potential biases that may have occurred during the design, conduct, and analysis of the studies. For quasi-experimental studies, we used the JBI critical appraisal checklist for quasi-experimental studies [23], which is a checklist with 9 questions for assessing potential bias. For randomized controlled trials (RCTs), we used the JBI critical appraisal tool for the assessment of risk of bias in RCTs [24], which is a 13-question checklist evaluating the internal and statistical validity of the conclusions of RCTs. On the basis of the answers from both assessment tools, reviewers decide whether to include the reviewed study. Two reviewers used these tools independently to assess the risk of bias in the studies included in the meta-analysis. Any disagreement between the reviewers about whether to include or exclude a study was resolved by discussion. If the disagreement persisted, a third reviewer was asked to arbitrate.

We used the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) methodology to assess the certainty of evidence regarding the intervention's effects. This methodology evaluates the certainty of evidence based on several criteria, including risk of bias, inconsistency, indirectness, and imprecision [25]. Given that the GRADE approach is primarily focused on RCTs, and the GRADE working group has not yet reached a consensus on the combination of results from randomized and nonrandomized trials, we applied this evaluation exclusively to the RCTs included in our review.

Synthesis Methods

Narrative Synthesis

To address the multifaceted nature of the factors involved in self-administered NLP-based interventions for symptoms of depression and anxiety, we adopted a comprehensive framework for data synthesis based on an adaptation of the categories from the framework for NLP applications for mental health interventions proposed by Malgaroli et al [15] in the context of self-administered NLP interventions. This systematic approach thoroughly integrates all relevant factors, providing a coherent structure for our analysis. We categorized data from eligible studies into four primary domains: (1) demographic and sample descriptions, (2) NLP technical aspects, (3) clinical categories, and (4) intervention results. Due to the nature of our study, the last category is presented through the findings of the meta-analysis and analysis of subgroups.

Meta-Analysis

We performed analyses using Stata (version 18.0; StataCorp LLC). Meta-analysis was only performed if at least 3 studies of the same design type (ie, randomized or quasi-experimental controlled trials) assessing the same outcome were available. The analysis was differentiated by outcome and by study type. Standardized mean differences (SMDs) with 95% CIs were used for meta-analyses and summary statistics of the studies because the results of the included studies were measured using different scales. SMD is the mean difference between the intervention and control groups divided by the pooled SD.

The standard measure of effect size to be considered for the Hedges g analyses includes small (SMD 0.2), moderate (SMD 0.5), and large (SMD >0.8) effect sizes. These thresholds were used to evaluate the combined effect of the analyzed interventions using Hedges g. Hedges g, unlike Cohen d, corrects for possible risk of bias associated with small sample sizes, making it a more appropriate measure for our analyses [26].

Heterogeneity Analysis

The assessment of statistical heterogeneity involved the following tests: the Cochran Q test statistic to detect the presence of heterogeneity between studies, the I^2 Higgins and H^2 index statistics to measure the extent of variability between studies due to heterogeneity, and the between-study variance (τ^2) to assess the variance between the effects observed across the studies. If the overall assessment indicated high heterogeneity, random effects models were used to estimate the effect of the interventions in general.

Publication Bias Analysis

If there were >10 studies in the meta-analysis, we conducted both visual and quantitative tests to detect biases. Our visual examination used a funnel plot; the quantitative test used was the Egger regression test, which can capture the effects of small studies and other potential information biases [27]. We identified selection bias if we observed an asymmetric funnel plot distribution and a significant Egger regression test result ($P<.05$). In cases of asymmetry, the trim-and-fill method proposed by Duval and Tweedie [28] was implemented as a bias correction technique to estimate the number of missing studies for the meta-analysis.

Analysis of Subgroups

If the meta-analysis data allowed, we assessed intervention effects using the NLP-based models from the selected studies. Such models could include rule-based NLP, AI-based NLP, or other NLP. In addition, we assessed the impact of interventions on subgroups, including gender, disease severity, prior therapies, concurrent depression and anxiety disorders, and age ranges.

We performed a random effects meta-regression using aggregate-level data. Our analysis specified the variables containing the SE within each study using the *metareg* command and the *wsse* option in Stata. The meta-regression was a function of the mean age of the participants and was only applied to the overall meta-analysis. Our analysis yielded a meta-regression coefficient with 95% CI.

Results

Study Selection

Initially, 672 records were identified in the different databases; after eliminating 201 (29.9%) duplicates, 471 (70.1%) records advanced to title and abstract review. Of these 471 records, 418 (88.7%) were discarded, leaving 53 (11.3%) records for full-text review. Subsequently, 32 (60%) of the 53 records were excluded, resulting in 21 (40%) articles selected for review. Of these 21 articles, 19 (90%) were included in the meta-analysis on depressive and anxiety symptoms. Of the 19 studies included in the meta-analysis, 16 (84%) reported sufficient data for the meta-analysis of depressive symptoms, and another 16 (84%) reported sufficient data for the meta-analysis of anxiety symptoms. Figure 1 shows the complete review process, and Multimedia Appendices 3 and 4 [29-49] list the articles excluded and included, respectively.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the selection process. RCT: randomized clinical trial.
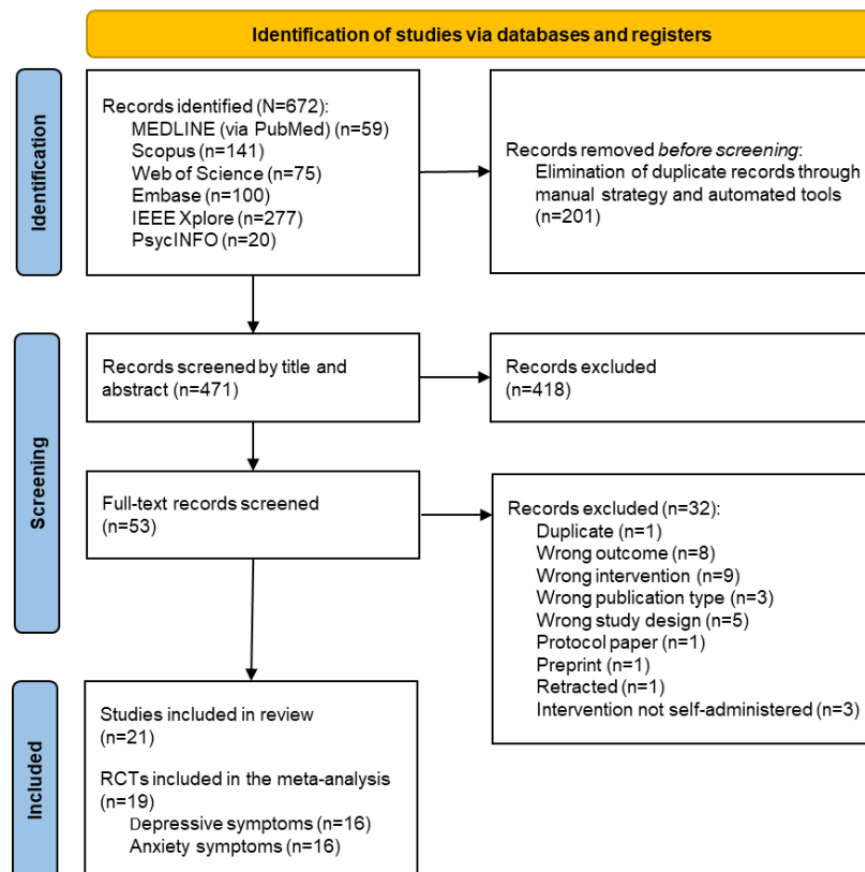
Characteristics of the Included Studies

Of the 21 studies identified, 19 (90%) were RCTs [29-47], and 2 (10%) were quasi-experimental studies without a control group [48,49]. Most of the studies (16/21, 76%) were published between 2020 and 2023, and 81% (17/21) were conducted in high-income countries. The United States was the country with the most publications among the selected studies (10/21, 48%). Regarding the characteristics of the populations studied, the majority (16/21, 76%) focused on adults. With regard to the outcomes assessed, depressive symptoms were analyzed in 95% (20/21) of the studies and anxiety symptoms in 90% (19/21).

We found 29 potential comparisons between interventions and controls because 5 (24%) of the 21 studies reported ≥3 arms. AI-based NLP applications were the most common intervention (11/21, 52%), while the most common control conditions were waiting list or no intervention (8/21, 38%) and information, psychoeducation, or bibliotherapy (8/21, 38%). The most commonly used scales to measure depressive and anxiety symptoms were the Patient Health Questionnaire (PHQ; PHQ-9 and PHQ-8; 13/21, 62%) and the Generalized Anxiety Disorder-7 (GAD-7; 10/21, 48%), respectively. Table 1 shows the characteristics of the studies, divided into RCTs and uncontrolled quasi-experimental studies.

Table 1. Characteristics of the included studies (n=21).

Characteristics	Randomized controlled trials (n=19), n (%)	Uncontrolled quasi-experimental studies (n=2), n (%)
Publication year		
2014-2015	1 (5)	0 (0)
2016-2019	4 (21)	0 (0)
2020-2023	14 (74)	2 (100)
Country income level		
Upper-middle income	4 (21)	0 (0)
High income	15 (79)	2 (100)
Country		
Argentina	1 (5)	0 (0)
China	3 (16)	0 (0)
Italy	1 (5)	1 (50)
Japan	1 (5)	0 (0)
South Korea	2 (11)	0 (0)
United Kingdom	2 (11)	0 (0)
United States	9 (47)	1 (50)
Study design		
Crossover	5 (26)	0 (0)
Parallel	14 (74)	0 (0)
Not applicable	0 (0)	2 (100)
Participants' life stage		
Adolescent	2 (11)	0 (0)
Adult	14 (74)	2 (100)
Older adult	2 (11)	0 (0)
Pregnant	1 (5)	0 (0)
Included in meta-analysis		
Depressive symptoms	16 (84)	0 (0)
Anxiety symptoms	16 (84)	0 (0)
Not applicable	0 (0)	2 (100)
Depressive symptoms		
Main outcome	11 (58)	2 (100)
Secondary outcome	7 (37)	0 (0)
Not evaluated	1 (5)	0 (0)
Anxiety symptoms		
Main outcome	11 (58)	2 (100)
Secondary outcome	6 (32)	0 (0)
Not evaluated	2 (11)	0 (0)
Funding		
Corporations	8 (42)	1 (50)
Government	4 (21)	0 (0)
Self-financed	2 (11)	1 (50)
Not reported	5 (26)	0 (0)
Conflicts of interest		

Characteristics	Randomized controlled trials (n=19), n (%)	Uncontrolled quasi-experimental studies (n=2), n (%)
Yes	4 (21)	2 (100)
No	12 (63)	0 (0)
Not reported	3 (16)	0 (0)
Study has ≥3 arms		
No	14 (74)	2 (100)
Yes	5 (26)	0 (0)
Control group[a]		
Waiting list or no intervention	8 (42)	0 (0)
Usual treatment	2 (11)	0 (0)
Information, psychoeducation, or bibliotherapy	8 (42)	0 (0)
Conversational computer-based intervention	5 (26)	0 (0)
Not applicable	0 (0)	2 (100)
Type of NLP[b] application[a]		
Rule based	10 (53)	1 (50)
AI[c] based	11 (58)	1 (50)
Focus of intervention[a]		
Depressive symptoms	8 (42)	1 (50)
Anxiety symptoms	7 (37)	1 (50)
Other mental health problems	13 (68)	2 (100)
Therapeutical approach[a]		
Cognitive behavioral therapy	15 (79)	2 (100)
Other	3 (16)	0 (0)
Unclear	1 (5)	0 (0)
Scale used to measure depression[a]		
PHQ[d]-9 and PHQ-8	13 (68)	2 (100)
DASS-21[e]	2 (11)	0 (0)
Other	3 (16)	0 (0)
Not evaluated	1 (5)	0 (0)
Scale used to measure anxiety[a]		
GAD-7[f]	10 (53)	2 (100)
DASS-21	3 (16)	0 (0)
Other	4 (21)	0 (0)
Not evaluated	2 (11)	0 (0)

[a]The totals do not add up to 100% because there are studies with 3 and 4 arms that evaluated >1 type of intervention at the same time.

[b]NLP: natural language processing.

[c]AI: artificial intelligence.

[d]PHQ: Patient Health Questionnaire.

[e]DASS-21: Depression, Anxiety, and Stress Scale-21.

[f]GAD-7: Generalized Anxiety Disorder-7.

NLP Technical Aspects

Of the 21 included studies, 10 (48) used rule-based approaches, while 11 (52%) used AI-based techniques. Within the AI-based category, of the 11 studies, 4 (36%) implemented deep learning methods, 6 (55%) did not specify the AI technique used, and 1 (9%) used ML algorithms. Regarding the specific NLP techniques used, sentiment analysis was used in 18% (2/11) of the studies, and natural language understanding was used in 18% (2/11). Notably, 7 (64%) of the 11 studies did not specify the NLP techniques used in their interventions. This distribution highlights a diverse application of NLP methods in addressing symptoms of depression and anxiety, with more than half of the studies (11/21, 52%) leveraging advanced AI techniques, albeit often without detailed specification (7/11, 64%).

The input modality for the NLP interventions was primarily text based in 19 (90%) of the 21 studies, while 1 (5%) study used either text or voice, and 1 (5%) study used voice alone. Regarding output modalities, text was predominantly used in 20 (95%) of the 21 studies, while only 1 (5%) study used voice. The language of the NLP input and output varied among the studies. Of the 21 studies, 7 (33%) used English, and 3 (14%) used Chinese, while Japanese, Spanish, and Italian were used in 1 (5%) study each. However, 38% (8/21) of the studies did not specify the language used for the NLP input and output.

Demographics and Sample Descriptions

Overview

The study participants' demographic characteristics were analyzed for rule-based NLP studies and AI-based NLP studies. All 21 studies provided demographic information regarding the sample or testing data set used for the intervention. Demographic data for rule-based NLP studies are reported only for the intervention samples. By contrast, AI-based NLP studies were expected to provide demographic information for the training data used to develop the AI-based models and the participants involved in the intervention or experiment.

Training Sample Description

None of the AI-based NLP studies provided detailed demographic information regarding the training data. While 3 (27%) of the 11 AI-based NLP studies mentioned the source of their training data (Stanford Sentiment Treebank data set, ad hoc user utterances from an unspecified source, and Emotion Support Conversation data set), they did not describe the demographic characteristics of these data sets.

Testing Data or Intervention Sample Description

Across all studies, gender distribution varied significantly. Of the 21 studies, in 3 (14%), only women participated; in 16 (76%), >50% of the participants were women; and in 2 (10%), >50% of the participants were men. Regarding the age of the participants, 20 (95%) of the 21 studies reported the mean age of their samples. Of these 20 studies, 9 (45%) involved participants aged >30 years, 10 (50%) included participants aged between 18 and 29 years, and 1 (5%) included participants aged <18 years. Participants' special conditions were also considered in the analysis. Of the 21 studies, 4 (19%) included participants with chronic diseases, 7 (33%) focused on

individuals with mental disorders, and 7 (33%) included university students, while 4 (19%) involved participants with other conditions. Specifically, among the 7 studies that focused on mental disorders, 2 (29%) included participants with a positive screening for depression, and 2 (29%) focused on participants with a positive screening for substance use disorder. Among the 4 studies that included participants with chronic diseases, there were diverse conditions, such as diabetes mellitus (n=1, 25%), cancer (n=1, 25%), inflammatory bowel disease (n=1, 25%), and dementia (n=1, 25%).

Focusing on the 11 AI-based NLP studies, the gender distribution of the intervention samples was as follows: in 9 (82%) studies, the majority of the participants were women; and in 2 (18%) studies, the majority of the participants were men. Regarding age distribution, of the 10 studies that reported mean ages, 5 (50%) involved participants aged >30 years, and 5 (50%) included participants aged between 18 and 29 years. With regard to special conditions in the intervention samples, of the 11 studies, 2 (18%) included participants with chronic diseases, 2 (18%) focused on individuals with mental disorders, 6 (55%) included university students, and 2 (18%) involved participants with other conditions (participants with panic disorder: n=1, 50%; and participants with a positive screening for depression: n=1, 50%). For chronic conditions, of the 2 studies, 1 (50%) involved patients with dementia, and 1 (50%) included patients with diabetes mellitus.

Clinical Categories

The included studies were evaluated for their focus on clinical presentation and the delivery of therapeutic interventions. Only 1 (5%) of the 21 studies reported having a component of diagnosis and screening for mental health problems, although it did not specify the disease or the methods used for diagnosis.

Most of the studies (19/21, 90%) declared that they delivered some form of therapy through their NLP interventions. By contrast, 2 (10%) of the 21 studies did not include any therapeutic component. Among the 19 studies that delivered therapy, 16 (84%) implemented cognitive behavioral therapy, 1 (5%) combined cognitive behavioral therapy with dialectical behavioral therapy, and 2 (11%) reported delivering therapy but did not specify the therapeutic approach used.

Meta-Analysis Findings

Main Meta-Analysis

Only 16 (76%) of the 21 studies were included in the meta-analysis, excluding the uncontrolled quasi-experimental studies (n=2, 10%) and the RCTs with insufficient data for meta-analysis (n=3, 14%). The rationale for excluding the 2 quasi-experimental studies was that a meta-analysis specific to this study design required at least 3 studies of the same design type assessing the same outcome. For the depressive symptoms (Figure 2 [29-37,39-45]), the overall meta-analysis showed that self-administered interventions based on NLP models were significantly more effective in reducing depressive symptoms compared to various control conditions (waiting list or no intervention, treatment as usual, psychoeducation, and other computer-based conversational interventions; SMD 0.819, 95% CI 0.389-1.250; $P<.001$). In addition, high heterogeneity was

observed in the overall meta-analysis (I^2=92.7%, 95% CI 78.3%-96.4%; H^2=3.71, 95% CI 2.15-5.27; τ^2=0.97; $P<.001$). Regarding publication bias, the funnel plot analysis showed evidence of bias (Egger test coefficient=3.61, 95% CI 0.45-6.78; P=.03; Multimedia Appendix 5).

For the outcome of anxiety symptoms (Figure 3 [30-34,36-46]), the global meta-analysis showed that self-administered NLP model–based interventions were significantly more effective in reducing depressive symptoms compared to various control conditions (waitlist or no intervention, treatment as usual, psychoeducation, and other conversational computer-based interventions; SMD 0.272; 95% CI 0.116-0.428; P=.001). In addition, high heterogeneity was observed in the overall meta-analysis (I^2=64%, 95% CI 0.5%-81.6%; H^2=1.67, 95% CI 1.00-2.33; τ^2=0.07; $P<.001$). Regarding publication bias, the funnel plot analysis showed no evidence of bias (Egger test coefficient=–0.22, 95% CI –1.55 to 1.11; P=.73; Multimedia Appendix 5).

Figure 2. Forest plot for control conditions versus self-administered interventions based on natural language processing models to reduce depressive symptoms.
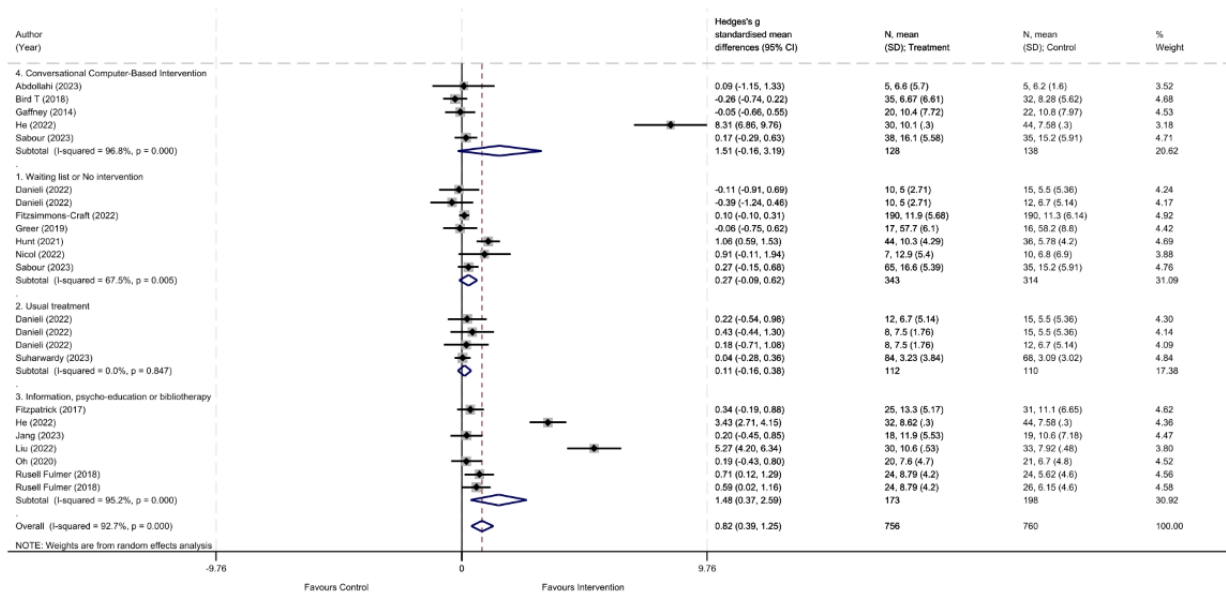
Figure 3. Forest plot for control conditions versus self-administered interventions based on natural language processing models to reduce anxiety symptoms.
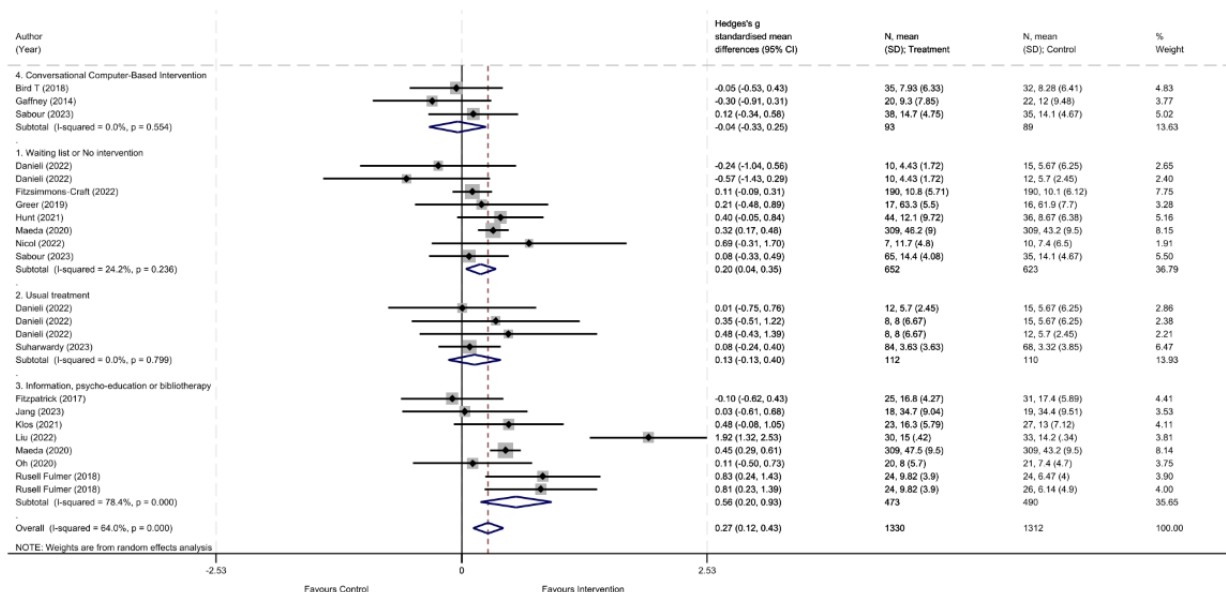
Subgroup Analyses

We also conducted a detailed analysis according to the type of comparator, intervention, and the scale used, evaluating the results for depressive symptoms and anxiety symptoms separately. For depressive symptoms, self-administered interventions based on NLP models were found to be more effective than information, psychoeducation, or bibliotherapy (SMD 1.481, 95% CI 0.368-2.594; *P*=.009). Similarly, AI-based NLP models were more effective than the set of control conditions (SMD 1.059, 95% CI 0.520-1.597; *P*<.001) for reducing depressive symptoms. Regarding the scale used, studies using the PHQ-9 or PHQ-8 showed that self-administered interventions based on NLP outperformed the set of control conditions (SMD 0.914, 95% CI 0.417-1.410; *P*<.001).

For the outcome of anxiety symptoms, self-administered interventions based on NLP models were more effective than waitlist or no intervention (SMD 0.196, 95% CI 0.042-0.351; *P*=.01) and information, psychoeducation, or bibliotherapy (SMD 0.561, 95% CI 0.195-0.927; *P*=.003). In addition, the use of AI-based NLP models had a higher effect than the average of the control conditions (SMD 0.302, 95% CI 0.073-0.532; *P*=.01) in reducing anxiety symptoms. Regarding the scale used, studies using the GAD-7 showed that self-administered interventions based on NLP had a higher effect than the average of the control conditions in reducing anxiety symptoms (SMD 0.333, 95% CI 0.074-0.592; *P*=.01). Full details of this subgroup analysis are presented in Table 2.

Given that factors such as age may influence the outcomes of depressive and anxiety symptoms, we performed a meta-regression to assess whether the mean age of participants affected the overall meta-analysis results. Our analysis revealed that the mean age was not significantly associated with the point estimates for either depressive symptoms (coefficient=–0.037, 95% CI –0.092 to 0.019; *P*=.18) or anxiety symptoms (coefficient=–0.010, 95% CI –0.030 to 0.010; *P*=.29). Detailed results of the meta-regression are presented in Table 3.

Table 2. Meta-analysis by subgroup for depressive and anxiety symptoms.

Symptoms and subgroups	Studies, n (%); groups, n	SMD[a] (95% CI)	P value	Heterogeneity (I^2; %)	Cochran Q test (P value)
Depressive symptoms (n=16)					
By control group					
Waiting list or no intervention	6 (38); 7	0.267 (–0.085 to 0.620)	.14	67.5	.005
Usual treatment	2 (12); 4	0.111 (–0.155 to 0.378)	.41	0	.85
Information, psychoeducation, or bibliotherapy	6 (38); 7	*1.481 (0.368 to 2.594)* [b]	*.009*	95.2	<.001
Conversational computer-based intervention	5 (31); 5	1.513 (–0.162 to 3.188)	.08	96.8	<.001
By intervention group					
Rule-based NLP[c] model	7 (44); 7	*0.854 (0.172 to 1.537)*	*.01*	94	<.001
AI[d]-based NLP model	9 (56); 16	*0.821 (0.207 to 1.436)*	*.009*	92.5	<.001
By scale used					
PHQ[e]-9 and PHQ-8	11 (69); 17	*0.914 (0.417 to 1.410)*	*<.001*	92.8	<.001
DASS-21[f]	2 (12); 2	__g	—	—	—
Anxiety symptoms (n=16)					
By control group					
Waiting list or no intervention	7 (44); 8	*0.196 (0.042 to 0.351)*	*.01*	24.2	.24
Usual treatment	2 (12); 4	0.133 (–0.134 to 0.400)	.33	0	.80
Information, psychoeducation, or bibliotherapy	7 (44); 8	*0.561 (0.195 to 0.927)*	*.003*	78.4	<.001
Conversational computer-based intervention	3 (19); 3	–0.041 (–0.333 to 0.250)	.78	0	.55
By intervention group					
Rule-based NLP model	8 (50); 9	*0.347 (0.116 to 0.578)*	*.003*	79.7	<.001
AI-based NLP model	8 (50); 14	0.198 (–0.011 to 0.406)	.06	34.4	.10
By scale used					
GAD-7[h]	9 (56); 15	*0.333 (0.074 to 0.592)*	*.01*	71.7	<.001
DASS-21	3 (19); 3	0.050 (–0.352 to 0.453)	.81	47.3	.15

[a]SMD: standardized mean difference.

[b]Italicized values are significant. Only meta-analyses with at least 3 measurements are presented in this study.

[c]NLP: natural language processing.

[d]AI: artificial intelligence.

[e]PHQ: Patient Health Questionnaire.

[f]DASS-21: Depression, Anxiety, and Stress Scale-21.

[g]There are not enough trials to do a meta-analysis.

[h]GAD-7: Generalized Anxiety Disorder-7.

Table 3. Meta-regression analysis by overall meta-analysis of depressive and anxiety symptoms.

Variable	Coefficient (SE; 95% CI)	t (df)	P value
Depressive symptoms			
Age, mean	–0.037 (0.026; –0.092 to 0.019)	–1.390 (18)	.18
Intercept	2.108 (1.033; –0.063 to 4.279)	2.040 (18)	.06
Anxiety symptoms			
Age, mean	–0.010 (0.009; –0.030 to 0.010)	–1.080 (16)	.29
Intercept	0.553 (0.329; –0.145 to 1.251)	1.680 (16)	.11

Risk of Bias and Certainty of Evidence

In the overall analysis of the risk of bias for the outcome of depressive symptoms, the majority of the studies (9/16, 56%) had an overall low risk of bias, while only 19% (3/16) had an overall high risk of bias (Figure 4A). Regarding the dimensions assessed, the lowest risk of bias was observed in reporting and analysis strategies (15/16, 94%), followed by participant loss or missing data (14/16, 88%). However, intervention delivery showed an unclear risk of bias due to limited reporting in the reviewed manuscripts. By contrast, for the outcome of anxiety symptoms, half of the studies (8/16, 50%) had an overall low risk of bias, while only 12% (2/16) had an overall high risk of bias (Figure 4B). At the level of each dimension assessed, all studies had a low risk of bias in reporting and analysis strategies, and 81% (13/16) had a low risk of bias in outcome measurement and retention throughout the study. Detailed risk-of-bias analyses for each study are available in Multimedia Appendix

6 for depressive symptoms and Multimedia Appendix 7 for anxiety symptoms.

We found that, for the outcomes studied (depressive symptoms and anxiety symptoms), the evidence was of very low certainty (Table 4). This was mainly due to several factors. First, there was a high risk of bias, with 3 (19%) of the 16 studies presenting an overall high risk of bias for depressive symptoms and 2 (12%) of the 16 studies presenting an overall high risk of bias for anxiety symptoms. Second, there was significant inconsistency, as indicated by an overall I^2 value of >60%. In addition, indirectness was a major concern due to the high variability in the interventions, controls, and sample characteristics across the studies. Finally, publication bias was strongly suspected due to the marked right-side asymmetry revealed by the funnel plot. Notwithstanding these limitations, the findings provide a preliminary understanding of the potential effects of self-administered NLP-based interventions on depressive and anxiety symptoms.

Figure 4. Risk of bias grouped for the outcomes of (A) depressive symptoms and (B) anxiety symptoms.
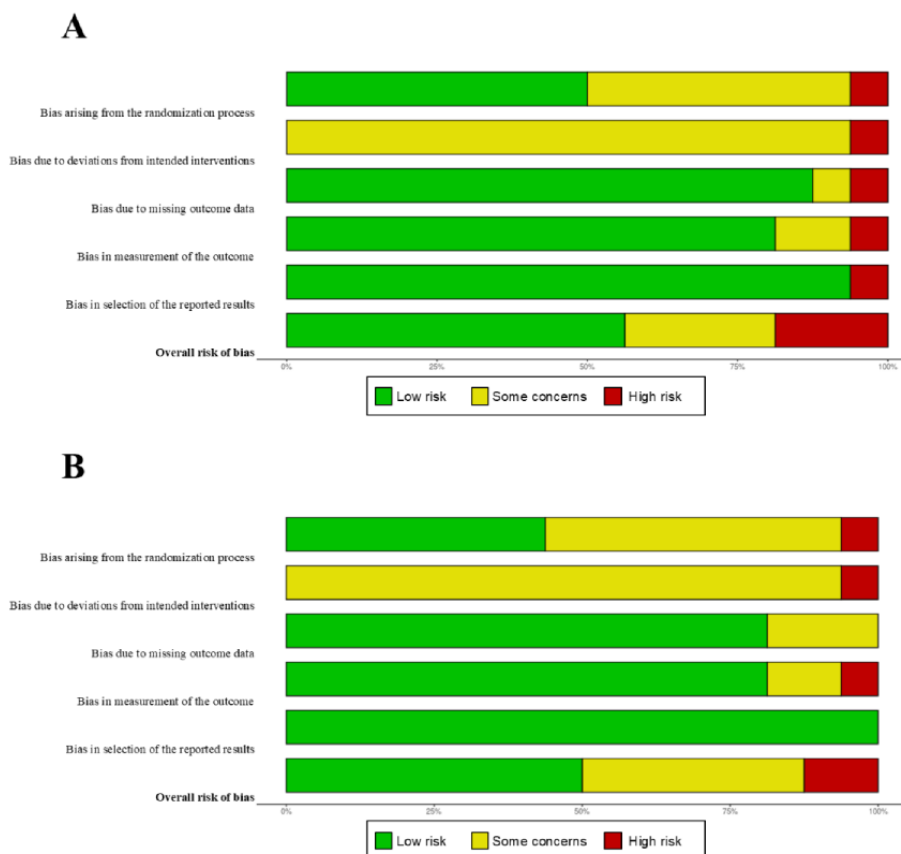
Table 4. Summary of findings and certainty of evidence using the Grading of Recommendations Assessment, Development, and Evaluation methodology.

Outcome	Assessment of certainty of evidence						Effect: Hedges g, SMD[a] (95% CI)	Certainty of evidence
	Studies (RCTs[b]), n (participants, n)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias		
Depressive symptoms	16 (1516; control: 760, intervention: 756)	Very serious[c]	Very serious[d]	Very serious[e]	Not serious	Strongly suspected[f]	0.82, lower (0.39-1.25)	⊕OOO[g]
Anxiety symptoms	16 (2642; control: 1312, intervention: 1330)	Very serious[h]	Very serious[d]	Very serious[e]	Not serious	Strongly suspected[f]	0.27, lower (0.12-0.43)	⊕OOO

[a]SMD: standardized mean difference.

[b]RCT: randomized controlled trial.

[c]Of the 16 studies, 3 (19%) present an overall high risk of bias.

[d]Overall I^2 value >60%.

[e]There is a high variability in the interventions, controls, and sample characteristics.

[f]The funnel plot reveals a marked right-side asymmetry.

[g]Very low (each filled circle [⊕] signifies a higher level of certainty, while each empty circle [O] indicates a lower level of certainty).

[h]Of the 16 studies, 2 (12%) present an overall high risk of bias.

Discussion

Principal Findings

Our results indicate that self-administered interventions based on NLP models have a significant overall effect on reducing depressive and anxiety symptoms compared to various control conditions. Our study used random effects models to estimate this overall effect, thus accounting for heterogeneity among the interventions analyzed. Therefore, we consider the results to be robust. At the level of each intervention group and control group, we observed variability in their effectiveness in reducing symptoms of depression and anxiety, which could be due to the limited number of studies available for meta-analysis. In particular, conversational computer-based interventions were shown to be effective in reducing depressive and anxiety symptoms compared to pooled control conditions. In addition, NLP-based interventions overall outperformed psychoeducation and bibliotherapy in reducing both depressive and anxiety symptoms. Furthermore, these interventions were more effective than waitlist or no intervention in reducing anxiety symptoms.

These findings support the usefulness of self-administered NLP-based interventions in alleviating such common mental health problems as depressive and anxiety symptoms. Thus, they have the potential to be implemented in primary care settings, where they could represent a valuable public health strategy to improve the mental health of the population.

Comparison With Other Studies

Our findings are consistent with previous research that has examined the application of NLP-based models at various stages of mental health care in both clinical and community settings [50-52], indicating that NLP-based interventions may effectively alleviate symptoms of emotional disorders. The robustness of our research is strengthened by the fact that most of the studies included in the meta-analysis of depressive (9/16, 56%) and anxiety symptoms (8/16, 50%) have a low risk of bias, indicating that our findings are derived from rigorous and reliable research.

A previous scoping review highlighted the heterogeneity of the tools used to assess the effects of dialogue interventions on mental health [53]. However, our review found that in the case of RCTs focusing on depressive and anxiety symptoms, validated instruments such as the PHQ-9 and GAD-7 were used, reducing the risk of bias and making the results more robust. Nevertheless, we highlight the lack of studies using experiential sampling or real-time measures to assess depressive and anxiety symptoms, which could provide a more accurate assessment of the impact of these self-administered NLP-based interventions.

The subgroup analysis showed variability in the effectiveness of the interventions in reducing depressive and anxiety symptoms, which may be due to the limited number of studies analyzed. Another possible explanation lies in the variety of NLP-based models used and their level of sophistication. Interventions using conversational agents based on advanced deep ML models showed significant results compared to other strategies, such as rule-based chatbots [54]. Unlike simpler NLP-based models, conversational agents offer better performance on various tasks [54]. However, more complex models also require high computational costs and large amounts of data for optimization [55,56], which may limit their adaptability to the different linguistic and cultural needs of different regions [57]. It is important to note that high-income countries have led research in this field and have advanced technological resources for developing these AI-based models compared to low- and middle-income countries [58,59]. This situation represents a challenge and a potential source of inequity in access to, and the implementation of, NLP-based interventions within public health systems.

Implications for Clinical Practice and Public Health

A previous systematic review on the general use of NLP and ML in mental health also identified the potential of NLP-based interventions to improve population mental health [19]. However, our study differs in that it focuses only on self-applied interventions to reduce depressive and anxiety symptoms, thus contributing to a specific aspect of NLP-based interventions. Our study provides a valuable starting point for future research to confirm the effectiveness of NLP-based interventions in the real world and their ability to be implemented within the public health system. There is a need to evaluate the implementation and promotion of these interventions as part of mental health strategies because this could be an effective strategy to reduce depressive and anxiety symptoms in health service users [60,61]. Given their accessibility through digital platforms, these interventions have the potential to reduce the burden of depressive and anxiety disorders at the population level [62,63] while also being cost-effective and a way to optimize mental health resources [64]. To ensure successful implementation within the public health system, using the Artificial Intelligence–Quality Implementation Framework could be beneficial [65]. However, it is crucial to develop protocols that ensure confidentiality and respect for the ethics and privacy of patient data at all stages of implementation and use [66]. In addition, it is important to consider the digital determinants of health [67], such as access to appropriate devices, the internet, and stable connectivity, because these factors pose challenges for implementation in low- and middle-income countries.

Strengths and Limitations

The main strength of our study is that we conducted an exhaustive review of available literature on the subject and that the main meta-analysis was based on RCTs, which is the most robust design for determining the effect of an intervention. However, our study has several limitations. First, the methodological variability of the included studies led to high heterogeneity in both outcomes, which could affect the interpretation of our findings despite using random effects models for their management. Second, the various measurement tools used in the studies could introduce measurement bias. However, we believe that our study minimized this risk by including only studies that used validated instruments and an effect size that controls for heterogeneity among measures such as the SMD. Third, the lack of clarity in the description of the studied groups may have introduced a risk of bias in assessing their effectiveness because there is no clear taxonomy for grouping NLP-based interventions. Fourth, the global meta-analysis for depressive symptoms identified the potential existence of publication bias, which could overestimate results in favor of trials with positive effects. Therefore, we encourage researchers to report their studies, even if they have negative results, to understand the effect of these interventions better. Fifth, variability in the standards for diagnosing and treating depression and anxiety, as well as in the criteria for determining recovery among the included studies, may have affected the interpretation of the efficacy of the interventions and the generalizability of the findings to different populations. This heterogeneity highlights the importance of considering the context in which NLP-based interventions are applied and the need to adapt them to the characteristics of different populations [11]. Finally, the GRADE assessment shows that the evidence for self-administered NLP-based interventions on depressive and anxiety symptoms is of very low certainty. This suggests caution in interpreting these potential benefits. High risk of bias, significant inconsistency (high I^2 values), and high indirectness complicate the findings. Suspected publication bias further skews the results because studies with nonsignificant or negative outcomes may be underreported. To overcome these limitations in future reviews, we recommend focusing on specific interventions and encouraging researchers to share their primary data to strengthen the quality and reliability of meta-analytic analyses.

Conclusions

Our systematic review and meta-analysis support the use of self-administered interventions based on NLP models to reduce depressive and anxiety symptoms. These findings enhance the theoretical understanding of how advanced NLP tools can effectively deliver psychological therapy, improving cognitive and emotional self-regulation in individuals. By demonstrating the efficacy of various NLP-based interventions, our study advances the theoretical framework by elucidating the mechanisms through which these technologies can replicate and potentially enhance traditional therapeutic processes.

The integration of NLP with different therapeutic modalities offers a novel approach to mental health treatment, expanding the accessibility and scalability of evidence-based interventions. However, the certainty of evidence for the effectiveness of these interventions remains very low, primarily due to a high risk of bias, significant inconsistency, and indirectness in the included studies. Therefore, there is a crucial need for RCTs with larger sample sizes and rigorous methodologies to strengthen the inferential power of future meta-analyses.

Moreover, while our findings are encouraging, there is a need for systematic reviews that examine the implementation processes of these interventions in depth, as well as qualitative studies that evaluate their usability and feasibility. Such research will be essential for effectively recommending the adoption of NLP-based self-administered interventions in public health systems.

Our study provides a valuable starting point for future research to validate the efficacy and practical implementation of these interventions as components of standard mental health care. Ensuring their integration into public health strategies could enhance the mental health outcomes of diverse populations, particularly those who may have limited access to traditional therapeutic resources.

Authors' Contributions

DVZ was responsible for conceptualization, methodology, validation, formal analysis, investigation, data curation, writing the original draft, and visualization. CMRR was responsible for conceptualization, methodology, investigation, writing the original draft, and reviewing and editing the manuscript. JGS was responsible for methodology, validation, investigation, and data curation. GQC, GLC, and RGA were responsible for validation and data curation. GCT was responsible for validation, investigation, writing the original draft, and reviewing and editing the manuscript. ADR was responsible for investigation, writing the original draft, reviewing and editing the manuscript, and supervision. SEA was responsible for investigation, writing the original draft, and reviewing and editing the manuscript. Joseph Finkelstein was responsible for investigation, resources, reviewing and editing the manuscript, and supervision. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[DOCX File , 32 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Search strategy.
[DOCX File , 22 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Excluded records.
[DOCX File , 25 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Included records.
[DOCX File , 21 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Funnel plot by depressive and anxiety symptoms.
[DOCX File , 295 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Risk of bias for individual studies for the outcome of depressive symptoms.
[DOCX File , 154 KB-Multimedia Appendix 6]

Multimedia Appendix 7

Risk of bias for individual studies for the outcome of anxiety symptoms.
[DOCX File , 160 KB-Multimedia Appendix 7]

References

1. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet. Nov 06, 2021;398(10312):1700-1712. [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
2. Arias D, Saxena S, Verguet S. Quantifying the global burden of mental disorders and their economic value. EClinicalMedicine. Dec 2022;54:101675. [FREE Full text] [doi: 10.1016/j.eclinm.2022.101675] [Medline: 36193171]

3.   Mental health at work. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work [accessed 2024-04-05]

4.   Edge D, Watkins ER, Limond J, Mugadza J. The efficacy of self-guided internet and mobile-based interventions for preventing anxiety and depression - a systematic review and meta-analysis. Behav Res Ther. May 2023;164:104292. [FREE Full text] [doi: 10.1016/j.brat.2023.104292] [Medline: 37003138]

5.   Saad A, Bruno D, Camara B, D'Agostino J, Bolea-Alamanac B. Self-directed technology-based therapeutic methods for adult patients receiving mental health services: systematic review. JMIR Ment Health. Nov 26, 2021;8(11):e27404. [FREE Full text] [doi: 10.2196/27404] [Medline: 34842556]

6.   Bauer AM, Iles-Shih M, Ghomi RH, Rue T, Grover T, Kincler N, et al. Acceptability of mHealth augmentation of collaborative care: a mixed methods pilot study. Gen Hosp Psychiatry. 2018;51:22-29. [FREE Full text] [doi: 10.1016/j.genhosppsych.2017.11.010] [Medline: 29272712]

7.   Philippe TJ, Sikder N, Jackson A, Koblanski ME, Liow E, Pilarinos A, et al. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. JMIR Ment Health. May 12, 2022;9(5):e35159. [FREE Full text] [doi: 10.2196/35159] [Medline: 35551058]

8.   Shelton CR, Kotsiou A, Hetzel-Riggin MD. Digital mental health interventions: impact and considerations. In: Lum H, editor. Human Factors Issues and the Impact of Technology on Society. Hershey, PA. IGI Global; 2021.

9.   Pineda BS, Mejia R, Qin Y, Martinez J, Delgadillo LG, Muñoz RF. Updated taxonomy of digital mental health interventions: a conceptual framework. Mhealth. 2023;9:28. [FREE Full text] [doi: 10.21037/mhealth-23-6] [Medline: 37492117]

10.   Cuijpers P, Schuurmans J. Self-help interventions for anxiety disorders: an overview. Curr Psychiatry Rep. Aug 2007;9(4):284-290. [FREE Full text] [doi: 10.1007/s11920-007-0034-6] [Medline: 17880859]

11.   Psychological interventions implementation manual: integrating evidence-based psychological interventions into existing services. World Health Organization. URL: https://www.who.int/publications/i/item/9789240087149 [accessed 2024-04-29]

12.   Consolidated telemedicine implementation guide. World Health Organization. URL: https://www.who.int/publications-detail-redirect/9789240059184 [accessed 2024-06-07]

13.   Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol. May 07, 2018;14:91-118. [doi: 10.1146/annurev-clinpsy-032816-045037] [Medline: 29401044]

14.   Caldarini G, Jaf S, McGarry K. A literature survey of recent advances in chatbots. Information. Jan 15, 2022;13(1):41. [doi: 10.3390/info13010041]

15.   Malgaroli M, Hull TD, Zech JM, Althoff T. Natural language processing for mental health interventions: a systematic review and research framework. Transl Psychiatry. Oct 06, 2023;13(1):309. [FREE Full text] [doi: 10.1038/s41398-023-02592-2] [Medline: 37798296]

16.   Friedman C, Hripcsak G. Natural language processing and its future in medicine. Acad Med. Aug 1999;74(8):890-895. [doi: 10.1097/00001888-199908000-00012] [Medline: 10495728]

17.   Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing [review article]. IEEE Comput Intell Mag. Aug 2018;13(3):55-75. [doi: 10.1109/MCI.2018.2840738]

18.   Nie J, Shao H, Fan Y, Shao Q, You H, Preindl M, et al. LLM-based conversational AI therapist for daily functioning screening and psychotherapeutic intervention via everyday smart devices. arXiv. Preprint posted online on March 16, 2024. [FREE Full text] [doi: 10.1145/3560905.3568074]

19.   Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res. May 04, 2021;23(5):e15708. [FREE Full text] [doi: 10.2196/15708] [Medline: 33944788]

20.   Zhang T, Schoene AM, Ji S, Ananiadou S. Natural language processing applied to mental illness detection: a narrative review. NPJ Digit Med. Apr 08, 2022;5(1):46. [FREE Full text] [doi: 10.1038/s41746-022-00589-7] [Medline: 35396451]

21.   de Choudhury M, Pendse SR, Kumar N. Benefits and harms of large language models in digital mental health. arXiv. Preprint posted online on November 7, 2023. [FREE Full text] [doi: 10.31234/osf.io/y8ax9]

22.   Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions. London, UK. The Cochrane Collaboration; 2023.

23.   Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Systematic reviews of effectiveness. In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. JBI Manual for Evidence Synthesis. Sydney, Australia. Joanna Briggs Institute; 2020.

24.   Barker TH, Stone JC, Sears K, Klugar M, Tufanaru C, Leonardi-Bee J, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. JBI Evid Synth. Mar 01, 2023;21(3):494-506. [doi: 10.11124/JBIES-22-00430] [Medline: 36727247]

25.   Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. Apr 2011;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]

26.   Ferguson CJ. An effect size primer: a guide for clinicians and researchers. Prof Psychol Res Pract. Oct 2009;40(5):532-538. [doi: 10.1037/a0015808]

27.   Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. Sep 13, 1997;315(7109):629-634. [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]

28.   Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. Jun 2000;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]

29.   Abdollahi H, Mahoor MH, Zandie R, Siewierski J, Qualls SH. Artificial emotional intelligence in socially assistive robots for older adults: a pilot study. IEEE Trans Affect Comput. Jul 1, 2023;14(3):2020-2032. [FREE Full text] [doi: 10.1109/taffc.2022.3143803] [Medline: 37840968]

30.   Sabour S, Zhang W, Xiao X, Zhang Y, Zheng Y, Wen J, et al. A chatbot for mental health support: exploring the impact of Emohaa on reducing mental distress in China. Front Digit Health. 2023;5:1133987. [FREE Full text] [doi: 10.3389/fdgth.2023.1133987] [Medline: 37214342]

31.   Suharwardy S, Ramachandran M, Leonard SA, Gunaseelan A, Lyell DJ, Darcy A, et al. Feasibility and impact of a mental health chatbot on postpartum mental health: a randomized controlled trial. AJOG Glob Rep. Aug 2023;3(3):100165. [FREE Full text] [doi: 10.1016/j.xagr.2023.100165] [Medline: 37560011]

32.   Danieli M, Ciulli T, Mousavi SM, Silvestri G, Barbato S, Di Natale L, et al. Assessing the impact of conversational artificial intelligence in the treatment of stress and anxiety in aging adults: randomized controlled trial. JMIR Ment Health. Sep 23, 2022;9(9):e38067. [FREE Full text] [doi: 10.2196/38067] [Medline: 36149730]

33.   Fitzsimmons-Craft EE, Chan WW, Smith AC, Firebaugh ML, Fowler LA, Topooco N, et al. Effectiveness of a chatbot for eating disorders prevention: a randomized clinical trial. Int J Eat Disord. Mar 2022;55(3):343-353. [doi: 10.1002/eat.23662] [Medline: 35274362]

34.   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. Nov 22, 2022;6(11):e40242. [FREE Full text] [doi: 10.2196/40242] [Medline: 36413390]

35.   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. Nov 21, 2022;24(11):e40719. [FREE Full text] [doi: 10.2196/40719] [Medline: 36355633]

36.   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. Mar 2022;27:100495. [FREE Full text] [doi: 10.1016/j.invent.2022.100495] [Medline: 35059305]

37.   Hunt M, Miguez S, Dukas B, Onwude O, White S. Efficacy of Zemedy, a mobile digital therapeutic for the self-management of irritable bowel syndrome: crossover randomized controlled trial. JMIR Mhealth Uhealth. May 20, 2021;9(5):e26152. [FREE Full text] [doi: 10.2196/26152] [Medline: 33872182]

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. Aug 12, 2021;5(8):e20678. [FREE Full text] [doi: 10.2196/20678] [Medline: 34092548]

39.   Oh J, Jang S, Kim H, Kim JJ. Efficacy of mobile app-based interactive cognitive behavioral therapy using a chatbot for panic disorder. Int J Med Inform. Aug 2020;140:104171. [doi: 10.1016/j.ijmedinf.2020.104171] [Medline: 32446158]

40.   Greer S, Ramo D, Chang YJ, Fu M, Moskowitz J, Haritatos J. Use of the chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR Mhealth Uhealth. Oct 31, 2019;7(10):e15018. [FREE Full text] [doi: 10.2196/15018] [Medline: 31674920]

41.   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. Dec 13, 2018;5(4):e64. [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]

42.   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. Jun 06, 2017;4(2):e19. [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]

43.   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. Sep 2018;46(5):570-582. [FREE Full text] [doi: 10.1017/S1352465817000820] [Medline: 29366432]

44.   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. Nov 2014;42(6):731-746. [doi: 10.1017/S135246581300060X] [Medline: 23899405]

45.   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. Jun 2021;150:104440. [doi: 10.1016/j.ijmedinf.2021.104440] [Medline: 33799055]

46.   Maeda E, Miyata A, Boivin J, Nomura K, Kumazawa Y, Shirasawa H, et al. Promoting fertility awareness and preconception health using a chatbot: a randomized controlled trial. Reprod Biomed Online. Dec 2020;41(6):1133-1143. [doi: 10.1016/j.rbmo.2020.09.006] [Medline: 33039321]

47. Prochaska JJ, Vogel EA, Chieng A, Baiocchi M, Maglalang DD, Pajarito S, et al. A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. Drug Alcohol Depend. Oct 01, 2021;227:108986. [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108986] [Medline: 34507061]

48. Bassi G, Giuliano C, Perinelli A, Forti S, Gabrielli S, Salcuni S. A virtual coach (Motibot) for supporting healthy coping strategies among adults with diabetes: proof-of-concept study. JMIR Hum Factors. Jan 21, 2022;9(1):e32211. [FREE Full text] [doi: 10.2196/32211] [Medline: 35060918]

49. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, et al. A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. J Med Internet Res. Mar 23, 2021;23(3):e24850. [FREE Full text] [doi: 10.2196/24850] [Medline: 33755028]

50. Lázaro E, Yepez JC, Marín-Maicas P, López-Masés P, Gimeno T, de Paúl S, et al. Efficiency of natural language processing as a tool for analysing quality of life in patients with chronic diseases. A systematic review. Comput Human Behav Rep. May 2024;14:100407. [doi: 10.1016/j.chbr.2024.100407]

51. Calvo RA, Milner DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. Nat Lang Eng. Jan 30, 2017;23(5):649-685. [doi: 10.1017/S1351324916000383]

52. Harvey D, Lobban F, Rayson P, Warner A, Jones S. Natural language processing methods and bipolar disorder: scoping review. JMIR Ment Health. Apr 22, 2022;9(4):e35928. [FREE Full text] [doi: 10.2196/35928] [Medline: 35451984]

53. Jabir AI, Martinengo L, Lin X, Torous J, Subramaniam M, Tudor Car L. Evaluating conversational agents for mental health: scoping review of outcomes and outcome measurement instruments. J Med Internet Res. Apr 19, 2023;25:e44548. [FREE Full text] [doi: 10.2196/44548] [Medline: 37074762]

54. Treviso M, Lee JU, Ji T, Aken BV, Cao Q, Ciosici M, et al. Efficient methods for natural language processing: a survey. Trans Assoc Comput Linguist. 2023;11:826-860. [FREE Full text] [doi: 10.1162/tacl_a_00577]

55. Lin T, Wang Y, Liu X, Qiu X. A survey of transformers. AI Open. 2022;3:111-132. [doi: 10.1016/j.aiopen.2022.10.001]

56. Khan W, Daud A, Khan K, Muhammad S, Haq R. Exploring the frontiers of deep learning and natural language processing: a comprehensive overview of key challenges and emerging trends. Nat Lang Process J. Sep 2023;4:100026. [doi: 10.1016/j.nlp.2023.100026]

57. Chin H, Song H, Baek G, Shin M, Jung C, Cha M, et al. The potential of chatbots for emotional support and promoting mental well-being in different cultures: mixed methods study. J Med Internet Res. Oct 20, 2023;25:e51712. [FREE Full text] [doi: 10.2196/51712] [Medline: 37862063]

58. Oduoye MO, Fatima E, Muzammil MA, Dave T, Irfan H, Fariha FN, et al. Impacts of the advancement in artificial intelligence on laboratory medicine in low- and middle-income countries: challenges and recommendations-A literature review. Health Sci Rep. Jan 2024;7(1):e1794. [FREE Full text] [doi: 10.1002/hsr2.1794] [Medline: 38186931]

59. Alami H, Rivard L, Lehoux P, Hoffman SJ, Cadeddu SB, Savoldelli M, et al. Artificial intelligence in health care: laying the foundation for responsible, sustainable, and inclusive innovation in low- and middle-income countries. Global Health. Jun 24, 2020;16(1):52. [FREE Full text] [doi: 10.1186/s12992-020-00584-1] [Medline: 32580741]

60. Babel A, Taneja R, Mondello Malvestiti F, Monaco A, Donde S. Artificial intelligence solutions to increase medication adherence in patients with non-communicable diseases. Front Digit Health. 2021;3:669869. [FREE Full text] [doi: 10.3389/fdgth.2021.669869] [Medline: 34713142]

61. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, et al. Artificial intelligence for mental health and mental illnesses: an overview. Curr Psychiatry Rep. Nov 07, 2019;21(11):116. [FREE Full text] [doi: 10.1007/s11920-019-1094-0] [Medline: 31701320]

62. Boucher EM, Harake NR, Ward HE, Stoeckl SE, Vargas J, Minkel J, et al. Artificially intelligent chatbots in digital mental health interventions: a review. Expert Rev Med Devices. Dec 2021;18(sup1):37-49. [doi: 10.1080/17434440.2021.2013200] [Medline: 34872429]

63. Pham KT, Nabizadeh A, Selek S. Artificial intelligence and chatbots in psychiatry. Psychiatr Q. Mar 2022;93(1):249-253. [FREE Full text] [doi: 10.1007/s11126-022-09973-8] [Medline: 35212940]

64. Dawoodbhoy FM, Delaney J, Cecula P, Yu J, Peacock I, Tan J, et al. AI in patient flow: applications of artificial intelligence to improve patient flow in NHS acute mental health inpatient units. Heliyon. May 2021;7(5):e06993. [FREE Full text] [doi: 10.1016/j.heliyon.2021.e06993] [Medline: 34036191]

65. Nilsen P, Svedberg P, Neher M, Nair M, Larsson I, Petersson L, et al. A framework to guide implementation of AI in health care: protocol for a cocreation research project. JMIR Res Protoc. Nov 08, 2023;12:e50216. [FREE Full text] [doi: 10.2196/50216] [Medline: 37938896]

66. Thakkar A, Gupta A, de Sousa A. Artificial intelligence in positive mental health: a narrative review. Front Digit Health. 2024;6:1280235. [FREE Full text] [doi: 10.3389/fdgth.2024.1280235] [Medline: 38562663]

67. Chidambaram S, Jain B, Jain U, Mwavu R, Baru R, Thomas B, et al. An introduction to digital determinants of health. PLOS Digit Health. Jan 2024;3(1):e0000346. [FREE Full text] [doi: 10.1371/journal.pdig.0000346] [Medline: 38175828]

68. Type to translate. DeepL. URL: https://www.deepl.com/en/translator [accessed 2024-04-29]

69. Responsible AI that ensures your writing and reputation shine. Grammarly. URL: https://www.grammarly.com/ [accessed 2024-06-29]

XSL•FO

RenderX

Abbreviations

AI: artificial intelligence
GAD-7: Generalized Anxiety Disorder-7
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
ML: machine learning
NLP: natural language processing
PHQ: Patient Health Questionnaire
PICOS: Population, Intervention, Comparison, Outcomes, and Study Design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference

XSL•FO

RenderX