

Self-administered interventions based on natural language processing models for reducing depressive and anxious symptoms: Systematic review and meta-analysis

David Villarreal-Zegarra, Jackeline García-Serna, Gleni Quispe-Callo, Gabriel Lázaro-Cruz, Gianfranco Centeno-Terrazas, Ricardo Galvez-Arevalo, Stefan Escobar-Agreda, Alejandro Dominguez-Rodriguez, C Mahony Reategui-Rivera, Joseph Finkelstein

Submitted to: JMIR Mental Health
on: April 15, 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 its 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 expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 31

 Figures 32

 Figure 1..... 33

 Figure 2..... 34

 Figure 3..... 35

 Figure 4..... 36

 Multimedia Appendixes 37

 Multimedia Appendix 1..... 38

 Multimedia Appendix 2..... 38

 Multimedia Appendix 3..... 38

 Multimedia Appendix 4..... 38

 Multimedia Appendix 5..... 38

 Multimedia Appendix 6..... 38

 Multimedia Appendix 7..... 38

Self-administered interventions based on natural language processing models for reducing depressive and anxious symptoms: Systematic review and meta-analysis

David Villarreal-Zegarra¹ MPH; Jackeline García-Serna¹ BSc; Gleni Quispe-Callo¹ BSc; Gabriel Lázaro-Cruz¹ Mg; Gianfranco Centeno-Terrazas¹ MPH; Ricardo Galvez-Arevalo² Mg; Stefan Escobar-Agreda³ MD; Alejandro Dominguez-Rodriguez⁴ PhD; C Mahony Reategui-Rivera⁵ MD; Joseph Finkelstein⁵ PhD, MD, FAMIA

¹Instituto Peruano de Orientación Psicológica Lima PE

²Instituto Nacional de Salud del Niño San Borja Lima PE

³Unidad de Telesalud Universidad Nacional Mayor de San Marcos Lima PE

⁴Department of Psychology, Health, and Technology University of Twente Enschede NL

⁵Department of Biomedical Informatics University of Utah Salt Lake City US

Corresponding Author:

David Villarreal-Zegarra MPH

Instituto Peruano de Orientación Psicológica

Jirón Hernán Velarde 8

Lima

PE

Abstract

Background: The introduction of Natural Language Processing (NLP) technologies has significantly enhanced the potential of self-directed interventions for treating anxiety and depression by improving human-computer interactions. Despite these advancements, particularly in AI and Large Language Models (LLMs), robust evidence validating their effectiveness remains sparse.

Objective: To determine whether interventions based on NLP models can reduce depressive and anxiety symptoms.

Methods: Our study was a systematic review, and the protocol was registered in PROSPERO (CRD42023472120). The databases we used for the systematic review are Web of Science, SCOPUS, MEDLINE (via PubMed), PsycINFO (via EBSCO), IEEE Xplore, EMBASE (via EBSCO), and Cochrane Library. The quality of the included studies was assessed using the JBI Critical Appraisal Tools.

Results: 21 articles were selected for review, and 16 were included in the meta-analysis for each outcome. The overall meta-analysis showed that self-administered interventions based on NLP models were significantly more effective in reducing depressive symptoms (SMD=0.819; 95%CI: 0.389-1.250; $p<0.001$) and anxiety symptoms (SMD=0.272; 95% CI: 0.116-0.428; $p=0.001$) compared with various control conditions. In subgroup analysis, AI-based NLP was shown to be effective in reducing depressive (SMD=1.059 [0.520 to 1.597]; $p<0.001$) and anxiety symptoms (SMD=0.302 [0.073 to 0.532]; $p=0.010$) compared with pooled control conditions. Also, NLP-based interventions overall outperform psychoeducation and bibliotherapy in reducing both depressive (SMD=1.481 [0.368 to 2.594]; $p=0.009$) and anxiety symptoms (SMD=0.561 [0.195 to 0.927]; $p=0.003$). In addition, these interventions are more effective than waitlist or no intervention in reducing anxious symptoms (SMD=0.196 [0.042 to 0.351]; $p=0.013$).

Conclusions: Our findings support the usefulness of self-applied NLP-based interventions in alleviating widely prevalent mental health problems such as depressive and anxious symptoms. Clinical Trial: Protocol was registered in PROSPERO (CRD42023472120)

(JMIR Preprints 15/04/2024:59560)

DOI: <https://doi.org/10.2196/preprints.59560>

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 visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/preprint/59560>



Original Manuscript

Self-administered interventions based on natural language processing models for reducing depressive and anxious symptoms: Systematic review and meta-analysis

Abstract

Background: The introduction of natural language processing (NLP) technologies has significantly enhanced the potential of self-directed interventions for treating anxiety and depression by improving human-computer interactions. Despite these advancements, particularly in complex models such as generative AI, robust evidence validating their effectiveness remains sparse.

Objective: To determine whether self-administered interventions based on NLP models can reduce depressive and anxious symptoms.

Methods: We conducted a systematic review and meta-analysis, adhering to PRISMA guidelines and the Cochrane Collaboration recommendations. We searched Web of Science, SCOPUS, MEDLINE, PsycINFO, IEEE Xplore, EMBASE, and Cochrane Library from inception to November 3, 2023. Eligibility criteria followed the PICO framework, including 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 RCTs and quasi-experimental studies but excluded narrative, systematic, and scoping reviews. Data extraction was performed independently by pairs of authors using a predefined form. The risk of bias was assessed, and the GRADE methodology was used to evaluate the certainty of evidence. Meta-analysis was conducted employing standardized mean differences (SMD) and random-effects models to account for heterogeneity.

Results: Twenty-one articles were selected for review, and 16 were included in the meta-analysis for each outcome. Most studies were recent (2020-2023), with interventions predominantly AI-based NLP (52.4%, 11/21) and most delivering some form of therapy, primarily Cognitive Behavioral Therapy (90.5%, 19/21). The overall meta-analysis showed that self-administered interventions based on NLP models were significantly more effective in reducing depressive symptoms (SMD=0.819; 95%CI: 0.389-1.250; $p<0.001$) and anxious symptoms (SMD=0.272; 95% CI: 0.116-0.428; $p=0.001$) compared with various control conditions. Subgroup analysis indicated that AI-based NLP was effective in reducing depressive symptoms (SMD=0.821; 95% CI: 0.207-1.436; $p<0.001$) compared with pooled control conditions. Rule-based NLP showed effectiveness in reducing both depressive (SMD=0.854; 95% CI: 0.172-1.537; $p=0.014$) and anxious symptoms (SMD=0.347; 95% CI: 0.116-0.578; $p=0.003$). The meta-regression showed no significant association between participants' mean age and treatment outcomes. Despite these positive findings, the overall certainty of evidence was very low, mainly due to high risk of bias, significant heterogeneity, and potential publication bias.

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

Keywords: Natural language processing, depression, anxiety, systematic review

Introduction

Depression and anxiety represent conditions with a substantial worldwide burden. In 2020, they 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 anxiety and depression result in a loss of 1 trillion dollars annually due to loss of productivity [3]. Additionally, its increasing incidence and lack of health resources challenge the healthcare systems and workforce to cover their rising demand adequately [4].

In response, self-administered technology-based interventions have emerged promising solutions for managing these conditions. These self-guided interventions enable users to progress through treatments independently, without external support [4]. These interventions have demonstrated the potential to reduce costs, save health providers' time, and improve satisfaction and access to care for patients with mental health conditions living in remote areas, those with disabilities, or those unable to afford traditional care, especially during crises and quarantine periods [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 interventions, 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, usually of 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 two main approaches: rule-based and artificial intelligence (AI)-based. Rule-based NLP utilizes 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 shows remarkable success in various NLP tasks due to its scalability and ability to manage linguistic ambiguities [17].

The advent of large language models (LLMs) and multimodal large language models (MLLMs) has further enhanced the capabilities of NLP-based health interventions. These advancements 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 [19] and Zhang [20], analyzed the impact of NLP on mental health. However, these reviews primarily focused on the

general applications of NLP in mental health. Additionally, another systematic review demonstrated promising results for NLP-based interventions in mental health, but these findings encompassed a broad range of mental health disorders and did not specifically address self-administered interventions [15].

Despite these advancements, 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 the available literature to determine the effect of self-administered NLP-based interventions on symptoms of depression and anxiety.

Methods

This study systematically searched the available literature in the principal health databases and synthesized the main quantitative results in a meta-analysis. Our study adheres to the reporting standards proposed by PRISMA (see Multimedia Appendix 1) and the recommendations for meta-analysis of the Cochrane Collaboration [22]. The protocol for this systematic review was registered in the PROSPERO repository (CRD42023472120).

Eligibility Criteria

Our study follows the PICO framework to evaluate whether interventions based on NLP models can effectively reduce depressive and anxious 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) anxious symptoms are characterized by the anticipation of imagined events that are perceived as potential threats, causing emotional distress and physiological tension.

We established the following eligibility criteria for our review:

Population: we included studies with participants in 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 (e.g., physician, psychologist, or psychiatrist) or assessed using validated psychometric instruments.

Intervention: the intervention must be based on NLP models such as the Large Language Model (LLM), Multimodal Large Language Model (MLLM), Artificial Intelligence-led (AI-led) (i.e., digital conversational agent, chatbots, interactive voice response), and other NLP models. We included interventions regardless of their primary design purpose, provided they were self-administered.

Comparator: We considered both passive and active comparators, including passive comparators (i.e., waiting lists, non-intervention control groups, or placebos) or active comparators (i.e., online 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).

Outcomes: We included studies measuring depressive and anxiety symptom scores using validated psychometric questionnaires (e.g., Patient Health Questionnaire [PHQ-9], Beck Depression Inventory [BDI], Hamilton Depression Rating Scale [HDRS], Generalized Anxiety

Disorder Scale [GAD-7], Beck Anxiety Inventory [BAI], Hamilton Anxiety Rating Scale [HAM], or similar instruments).

Design: we included randomized clinical trials (RCTs) and quasi-experimental studies (without a control arm or randomization groups) that assessed the effect of NLP-based interventions on depressive and anxious symptoms. We excluded narrative reviews, systematic reviews, scoping reviews, and other non-original 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 (i.e., clinical or community settings).

Information Sources and Search Strategy

The databases we used for the systematic review were Web of Science, SCOPUS, MEDLINE (via PubMed), PsycINFO (via EBSCO), IEEE Xplore, EMBASE, and Cochrane Library. The search strategy included NLP, depression and anxiety terms, and health science descriptors (see Multimedia Appendix 2). Our search included any document available from inception to November 3rd, 2023.

Selection Process

We downloaded all records identified by the search strategy in RIS format and compiled them into an Endnote file, which served as a repository for all retrieved records. However, this file could contain duplicate entries, so we used automated and manual methods to remove duplicate records. We exported the list of unique records from Endnote to Rayyan for the selection process. First, two pairs of authors (JGS with RGA, and GQC with GLC) 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 (JGS with RGA, and GQC with GLC). Any excluded studies were recorded along with the reasons for their exclusion (see Multimedia Appendix 3). If disagreements arose between the authors at either stage, they were resolved by discussion. A third reviewer (DVZ) 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 of the study. 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 would be 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, as the databases save the metadata in this language. The full-text review and results extraction were mainly performed in English and Spanish (the languages the reviewers speak). When finding studies in other languages, the reviewers would use the translator DeepL to translate the documents into English and proceed 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 on a predefined collection form in a Microsoft Excel sheet. Initially, a pilot data extraction process was conducted on five datasets 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 includes the following: a) General information (i.e., authors, year of publication, title, country, and language); b)

Characteristics of participants (i.e., age range, gender, number of participants, diagnosis) c) Characteristics of the interventions (i.e., type of NLP model, duration, frequency, and brief description of the intervention); d) Comparator (passive or active); e) Main outcomes (i.e., means, standard deviations, pre- and post-intervention measures, the effect size of control and intervention groups).

Risk of Bias Assessment and Certainty of Evidence

We use 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, the JBI Critical Appraisal Checklist for Quasi-experimental Studies [23] was used, a checklist with nine questions for assessing potential bias in this type of study. For RCTs, the JBI Critical Appraisal Tool for Risk Assessment of Bias in Randomized Controlled Trials [24] was used, which is also a 13-question checklist for the internal and statistical validity of the conclusions of this type of trial. Based on the answers given in both assessment tools, the 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 utilized the GRADE methodology to assess the certainty of evidence regarding the intervention's effects. This methodology evaluates the certainty of evidence based on several criteria: 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 reached a consensus on the combination of results from randomized and non-randomized 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 on symptoms of depression and anxiety, we adopted a comprehensive framework for data synthesis based on an adaptation of the categories from the NLP in the Mental Health framework proposed by Malgaroli et al. in the context of self-administered NLP interventions [15]. 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 latter category is presented through the findings of the meta-analysis and analysis of subgroups.

Meta-analysis

We performed analyses using STATA version 18 software. Meta-analysis would only be performed if at least three studies of the same design type (i.e., randomized or quasi-experimental clinical trials) assessing the same outcome were available. The analysis was differentiated by outcome and by study type. Standardized mean differences (SMD) were used for meta-analyses and summary statistics of the studies because the results of the included studies were measured with different scales, with a 95% confidence interval (CI). SMD is the mean difference between the intervention and control groups divided by the pooled Standard Deviation (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 sizes were used to evaluate the combined effect of the analyzed interventions using Hedge's g . This procedure is done on the basis that Hedge's g is a type of effect size for the SMD, which corrects for the possible risk of bias for small samples as opposed to Cohen's d [26].

Heterogeneity Analysis

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

Publication bias analysis

If there are more than ten studies in the meta-analysis, we conducted visual and quantitative tests to detect biases. Our visual exam used the funnel plot; the quantitative test was Egger's regression test. This test can capture the effects of small studies and other potential information biases that could exist [27]. We identified selection bias if we observed an asymmetric funnel plot distribution and a significant Egger's test result with a p -value < 0.05 . If there is asymmetry, the trim-and-fill method of Duval and Tweedie was implemented as a bias correction technique to estimate the number of missing studies for the meta-analysis [28].

Analysis of subgroups

If the meta-analysis data allows, we assessed intervention effects using the NLP models from the selected studies. Such models may include Rule-based NLP, AI-based NLP, or other natural language models. Also, we assessed the impact of interventions on subgroups, including gender, disease severity, prior therapies, concurrent anxiety and depression disorders, and age ranges.

We performed a random effects meta-regression using aggregate-level data. Our analysis specified the variables containing the standard error within each study using the "metareg" command and the "wsse" option within 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 obtained a meta-regression coefficient and 95% confidence interval.

Results

Study Selection

Initially, 672 records were identified in the different databases; after eliminating 201 duplicates, 471 records advanced to title and abstract review. Of these, 418 were discarded, leaving 53 records for full-text review. Subsequently, 32 records were excluded, resulting in 21 articles selected for review and 19 included in the meta-analysis by depressive and anxiety symptoms. Of the studies included in the meta-analysis, a group of 16 studies reported enough data to the meta-analysis for depressive symptoms, and other group of 16 studies reported enough data to the meta-analysis for anxiety symptoms. Figure 1 shows the complete review process, and Multimedia Appendix 3 and 4 lists the articles excluded and included from the full-text review, respectively.

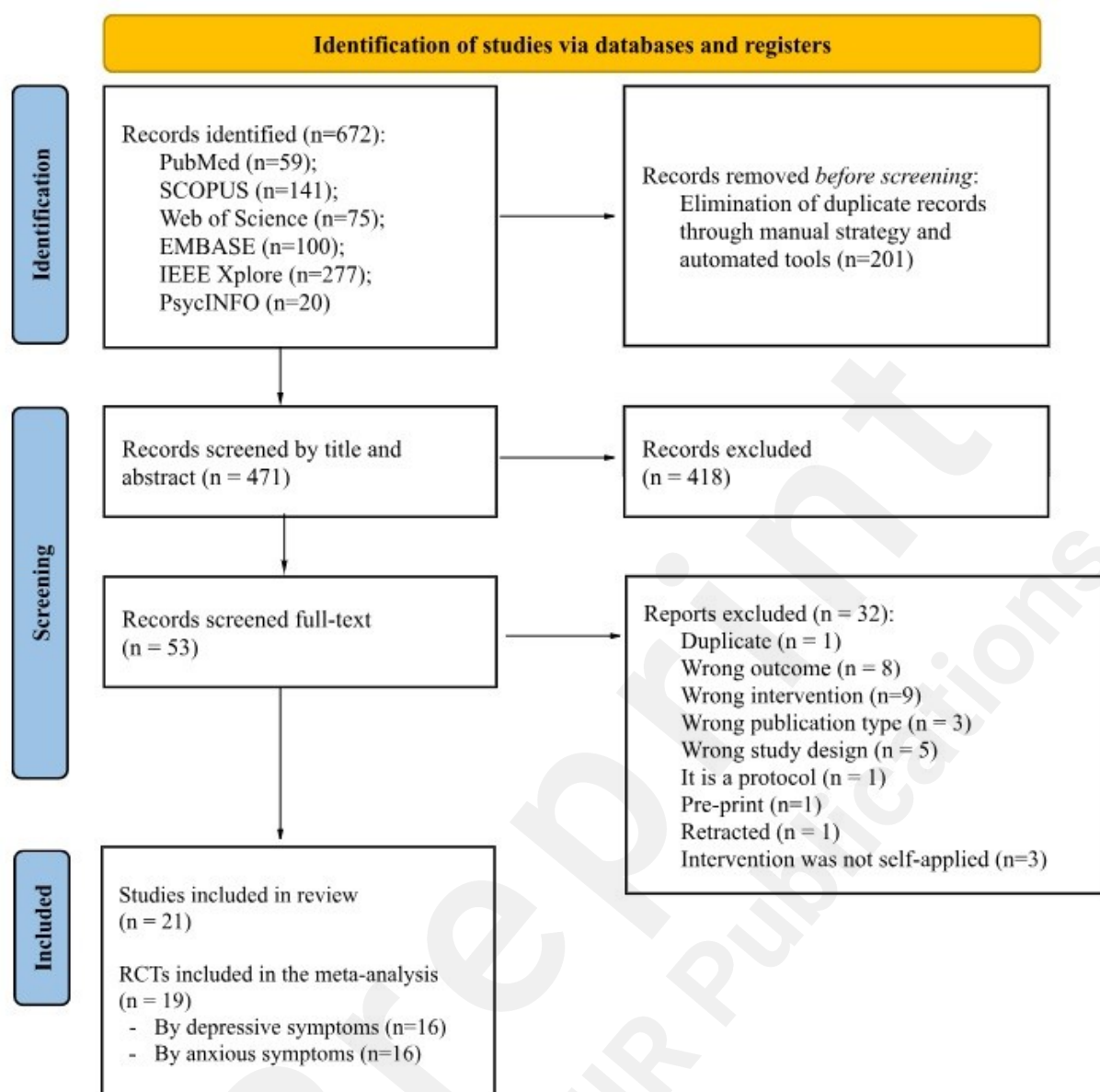


Figure 1. Flowchart of the selection process.

Characteristics of the Included Studies

Our review identified 19 RCTs [29–47] and two quasi-experimental studies without a control group [48,49]. Most of the studies (76.2%) were published between 2020 and 2023, and 81% were conducted in high-income countries, with the United States being the country with the most publications on this topic (47.6%). Regarding the characteristics of the population studied, the majority focused on adults (76.2%; 16/21). Regarding the outcomes assessed, depressive symptoms were analyzed in 95.2% (20/21) of the studies, and anxiety symptoms in 90.5% (19/21). We found 29 potential comparisons between interventions and controls because five studies reported three or more arms. The most common intervention was AI-based NLP (52.4%; 11/21), and the most common control conditions were waiting list or no intervention (38.1%; 8/21) and information, psychoeducation, or bibliotherapy (38.1%; 8/21). The most commonly used scales to measure depressive and anxiety symptoms were the Patient Health Questionnaire (PHQ-9, PHQ-8) (61.9%; 13/21) and the GAD-7 (47.6%; 10/21), 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).^a

	Randomized controlled trial (n=19)		Quasi-experimental without control (n=2)	
Year				
2014-2015	1	4.8%	0	0.0%
2016-2019	4	19.0%	0	0.0%
2020-2023	14	66.7%	2	9.5%
Income				
Upper-middle-income countries	4	19.0%	0	0.0%
High-income countries	15	71.4%	2	9.5%
Country				
Argentina	1	4.8%	0	0.0%
China	3	14.3%	0	0.0%
Italy	1	4.8%	1	4.8%
Japan	1	4.8%	0	0.0%
South Korea	2	9.5%	0	0.0%
United Kingdom	2	9.5%	0	0.0%
United States	9	42.9%	1	4.8%
Strategy (if is RCT)				
Cross-over	5	23.8%	0	0.0%
Parallel	14	66.7%	0	0.0%
Not applicable	0	0.0%	2	9.5%
Participants' age group				
Adolescent	2	9.5%	0	0.0%
Adult	14	66.7%	2	9.5%
Older adult	2	9.5%	0	0.0%
Pregnant women	1	4.8%	0	0.0%
Included in meta-analysis				
By depressive symptoms	16	76.2%	0	0.0%
By anxious symptoms	16	76.2%	0	0.0%
Depressive symptoms				
Main outcome	11	52.4%	2	9.5%
Secondary outcome	7	33.3%	0	0.0%
Not evaluated	1	4.8%	0	0.0%
Anxious symptoms				
Main outcome	11	52.4%	2	9.5%
Secondary outcome	6	28.6%	0	0.0%
Not evaluated	2	9.5%	0	0.0%
Funding				
Corporations	8	38.1%	1	4.8%
Government	4	19.0%	0	0.0%
Self-financed	2	9.5%	1	4.8%
No report	5	23.8%	0	0.0%
Conflict of interest				
Presents conflict of	4	19.0%	2	9.5%

No conflict of interest	12	57.1%	0	0.0%
No report	3	14.3%	0	0.0%
The study has 3 or more arms				
No	14	66.7%	2	9.5%
Yes	5	23.8%	0	0.0%
Control group^b				
Waiting list / Nothing	8	38.1%	0	0.0%
Usual treatment	2	9.5%	0	0.0%
Information, psychoeducation or bibliotherapy	8	38.1%	0	0.0%
Conversational Computer-Based	5	23.8%	0	0.0%
Not applicable	0	0.0%	2	9.5%
Type of NLP application^b				
Rule-based NLP	10	47.6%	1	4.8%
AI-based NLP	11	52.4%	1	4.8%
Interventions focused on^b				
Depressive symptoms	8	38.1%	1	4.8%
Anxious symptoms	7	33.3%	1	4.8%
Other mental health problems	13	61.9%	2	9.5%
Therapeutical approach^b				
Cognitive behavior therapy	15	71.4%	2	9.5%
Others	3	14.3%	0	0.0%
Unclear	1	4.8%	0	0.0%
Scale used to measure depression^b				
Patient Health Questionnaire (PHQ-9, PHQ-8)	13	61.9%	2	9.5%
Depression, Anxiety and Stress Scales (DASS-21)	2	9.5%	0	0.0%
Other	3	14.3%	0	0.0%
Not evaluated	1	4.8%	0	0.0%
Scale used to measure anxiety^b				
Generalized Anxiety Disorder (GAD-7)	10	47.6%	2	9.5%
Depression, Anxiety and Stress Scales (DASS-21)	3	14.3%	0	0.0%
Other	4	19.0%	0	0.0%
Not evaluated	2	9.5%	0	0.0%

^aAI = Artificial intelligence. NLP = Natural language processing.

^bBased on the number of studies that used this intervention, the totals do not add up to 100% because there are studies with three and four arms that evaluate more than one type of intervention at the same time.

NLP Technical Aspects

Among these studies, 47.6% (10/21) employed rule-based approaches, while 52.4% (11/21) utilized AI-based techniques. Within the AI-based category, 4 out of 11 studies (36.4%) implemented deep learning methods, 6 studies (54.5%) did not specify the exact AI technique used, and one study (9.1%) employed ML algorithms. Regarding the specific NLP techniques used, sentiment analysis was employed in two studies (18.2%), and natural language understanding was utilized in two studies (18.2%). Notably, seven studies, representing 63.6% of those that use AI-based NLP, did not specify the NLP techniques employed in their interventions. This distribution highlights a diverse application of NLP methods in addressing symptoms of depression and anxiety, with a significant portion of studies leveraging advanced AI techniques, albeit often without detailed specification.

The input modality for the NLP interventions was primarily text-based in 19 studies (90.5%), with one study (4.8%) using either text or voice, and another study (4.8%) using voice alone. For output modalities, text was predominantly used in 20 studies (95.2%), while only one study (4.8%) utilized voice. The language of the NLP input and output varied among the studies. English was used in seven studies (33.3%), Chinese in three studies (14.3%), and Japanese, Spanish, and Italian each in one study (4.8%). However, 38.1% (8/21) of the studies did not specify the language used for the NLP input and output.

Demographics and Sample Descriptions

The study participants' demographic characteristics were analyzed for NLP-based and AI-based studies. In total, 21 studies provided demographic information regarding the sample or testing dataset used for the intervention. Demographic data for rule-based NLP studies are reported only for the intervention samples. In contrast, AI-based NLP studies were expected to provide demographic information for the training data used to develop the AI 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 three studies mentioned the sources of their training data (Stanford Sentiment Treebank dataset, ad-hoc user utterances from a not declared source, and Emotion support conversation dataset), they did not describe the demographic characteristics of these datasets.

Testing Data or Intervention Sample Description

Across all studies, gender distribution varied significantly. Three studies (14.3%) had all women participants, 16 studies (76.2%) had more than 50% women participants, and 2 studies (9.5%) had more than 50% men participants. Regarding the age of the participants, 20 studies reported the mean age of their samples. Out of these, 45.0% (9 studies) involved participants older than 30 years, 50.0% (10 studies) included participants aged between 18 and 29 years, and one study (5.0%) included participants younger than 18 years. Participants' special conditions were also considered in the analysis. Four studies (19.0%) included participants with chronic diseases, seven studies (33.3%) focused on individuals with mental disorders, and another seven studies (33.3%) included university students. Four studies (19.0%) had participants with other conditions. Specifically, among those with mental disorders, 2 studies included participants with a screening for depression, and 2 studies focused on participants with a positive screening for substance use disorder. Among those with chronic diseases, there were diverse conditions such as diabetes mellitus, cancer, inflammatory bowel disease, and dementia, each represented by one study.

Focusing on the 11 AI-based NLP studies, the gender distribution for the actual intervention samples was as follows: 9 studies (81.8%) had more women participants, while 2 studies (18.2%) had more men participants. For age distribution, out of the 10 studies that reported mean ages, 5 studies (50.0%) had participants older than 30, and 5 studies (50.0%) included participants aged between 18 and 29. Regarding special conditions in the intervention samples of AI-based NLP studies, 2 studies (18.2%) included participants with chronic diseases, 2 studies (18.2%) focused on individuals with mental disorders, 6 studies (54.5%) included university students, and 2 studies (18.2%) had participants with other conditions. Specifically, 1 study included participants with panic disorder, and 1 study focused on participants with a

positive screening for depression. For chronic conditions, 1 study involved patients with dementia, and another 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 one study out of 21 (4.8%) reported having a component of diagnosis and screening for mental health problems, though it did not specify which disease, or the methods used for diagnosis.

In most studies, 90.5% (19/21) declared that they delivered some form of therapy through their NLP interventions. In contrast, 9.5% (2/21) did not include any therapeutic component. Among the studies that delivered therapy, which accounted for 19 in total, 16 studies (84.2%) implemented Cognitive Behavioral Therapy (CBT). Additionally, one study (5.3%) combined CBT with Dialectical Behavior Therapy (DBT), and two studies (10.5%) reported delivering therapy but did not specify the therapeutic approach used.

Meta-analysis Findings

Only 16 studies were included in the meta-analysis, excluding quasi-experimental studies without control (n=2) and those with insufficient data for meta-analysis (n=3). Therefore, only 76.2% of all identified trials were included in the meta-analysis. The rationale for excluding the two quasi-experimental studies is that a meta-analysis specific to this study design requires at least three studies. For the depressive symptoms (see Figure 2), the overall meta-analysis showed that self-administered interventions based on NLP models were significantly more effective in reducing depressive symptoms compared with various control conditions (waitlist/nothing, treatment as usual, psychoeducation, and other computer-based conversational interventions) (SMD=0.819; 95%CI: 0.389-1.250; $p<0.001$). In addition, high heterogeneity was observed in the overall meta-analysis ($I^2 = 92.7\%$ [78.3% to 96.4%]; $H^2 = 3.71$ [2.15 to 5.27]; $\tau^2 = 0.97$; $p<0.001$). Regarding publication bias, the funnel plot analysis showed evidence of bias (coefficient = 3.61 [0.45 to 6.78]; $p = 0.027$) (see Multimedia Appendix 5).

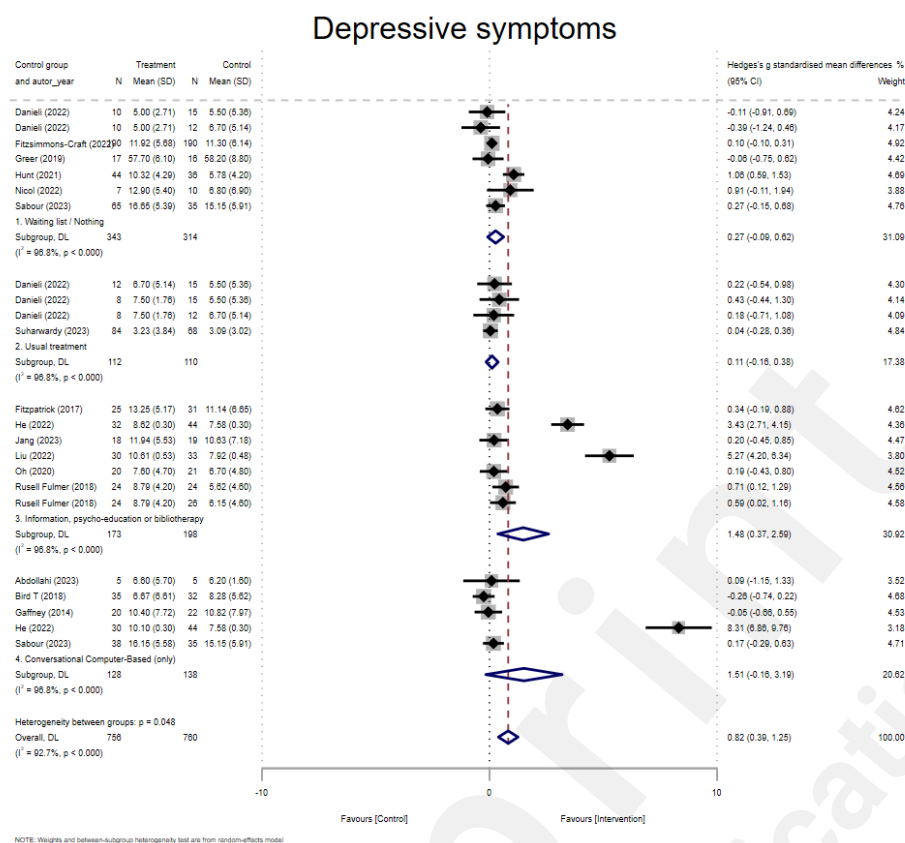


Figure 2. Forest plot for control conditions vs self-administered interventions based on natural language processing models to reduce depressive symptoms.

For the outcome of anxious symptoms (see Figure 3), the global meta-analysis showed that self-administered NLP model-based interventions were significantly more effective in reducing depressive symptoms compared with various control conditions (waitlist/nothing, treatment as usual, psychoeducation, and other computerized conversational interventions) (SMD=0.272; 95% CI: 0.116-0.428; p=0.001). In addition, high heterogeneity was observed in the overall meta-analysis ($I^2 = 64.0\%$ [0.5% to 81.6%]; $H^2 = 1.67$ [1.00 to 2.33]; $\tau^2 = 0.07$; p<0.001). Regarding publication bias, the funnel plot analysis showed no evidence of bias (coefficient = -0.22 [-1.55 to 1.11]; p = 0.734) (see Multimedia Appendix 5).

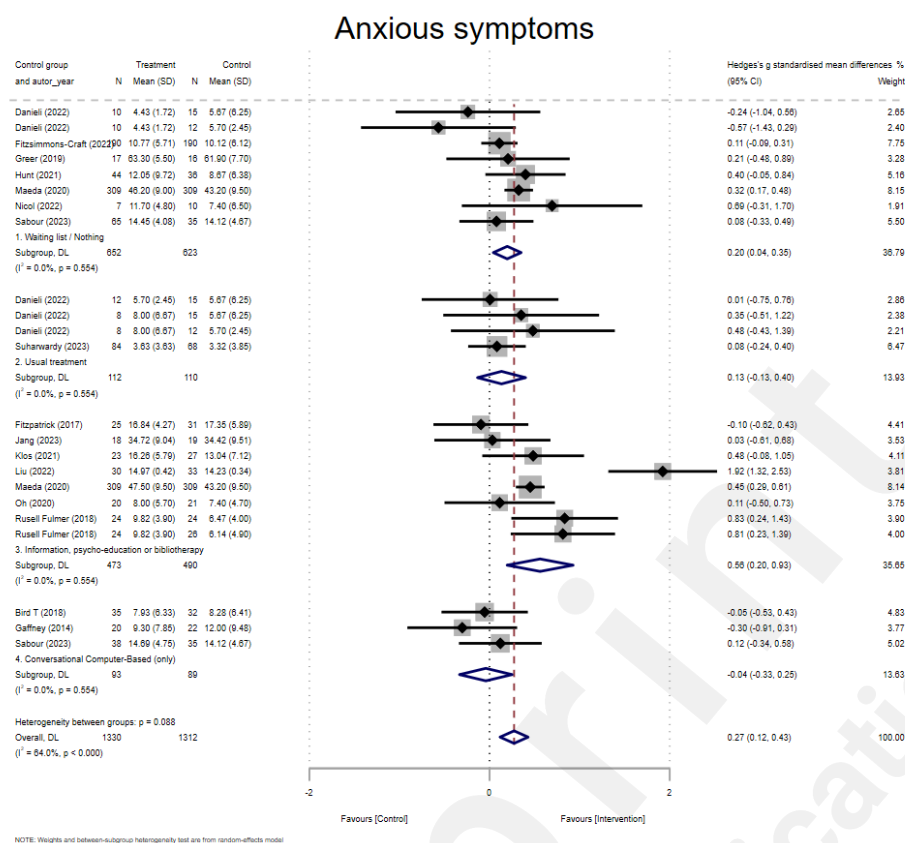


Figure 3. Forest plot for control conditions vs self-administered interventions based on natural language processing models to reduce anxious symptoms.

Subgroup Analyses

We also conducted a detailed analysis according to the type of comparator, intervention, and 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 (1.481 [0.368 to 2.594]; $p=0.009$). Similarly, AI-based NLP was more effective than the set of control conditions (1.059 [0.520 to 1.597]; $p<0.001$) 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 (0.914 [0.417 to 1.410]; $p<0.001$).

For the outcome of anxiety symptoms, self-administered interventions based on NLP models were more effective than wait-list/do-nothing (0.196 [0.042 to 0.351]; $p=0.013$) and then information, psychoeducation, or bibliotherapy (0.561 [0.195 to 0.927]; $p=0.003$). In addition, the use of AI-based NLP had a higher effect than the average of the control conditions (0.302 [0.073 to 0.532]; $p=0.010$) 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 (0.333 [0.074 to 0.592]; $p=0.012$). Full details of this subgroup analysis are shown in Table 2.

Table 2. Meta-analysis by subgroups for depressive and anxiety symptoms.^{a,b}

	Number of	SMD (95% CI)	p value	Heterogeneity	Cochran's
--	-----------	--------------	---------	---------------	-----------

	studies (Number of groups)			(I ²)	Q (p value)
DEPRESSIVE SYMPTOMS					
By control group					
Waiting list / Nothing	6 (7)	0.267 [-0.085 to 0.620]	0.137	67.5%	0.005
Usual treatment	2 (4)	0.111 [-0.155 to 0.378]	0.413	0.0%	0.847
Information, psychoeducation or bibliotherapy	6 (7)	1.481 [0.368 to 2.594]	0.009	95.2%	<0.001
Conversational Computer- Based (only)	5 (5)	1.513 [-0.162 to 3.188]	0.077	96.8%	<0.001
By intervention group					
Rule-based NLP	7 (7)	0.854 [0.172 to 1.537]	0.014	94.0%	<0.001
AI-based NLP	9 (16)	0.821 [0.207 to 1.436]	0.009	92.5%	<0.001
By scale used					
Patient Health Questionnaire (PHQ-9, PHQ-8)	11 (17)	0.914 [0.417 to 1.410]	<0.001	92.8%	<0.001
Depression, Anxiety and Stress Scales (DASS-21)	2 (2)	-	-	-	-
ANXIOUS SYMPTOMS					
By control group					
Waiting list / Nothing	7 (8)	0.196 [0.042 to 0.351]	0.013	24.2%	0.236
Usual treatment	2 (4)	0.133 [-0.134 to 0.400]	0.329	0.0%	0.799
Information, psychoeducation or bibliotherapy	7 (8)	0.561 [0.195 to 0.927]	0.003	78.4%	<0.001
Conversational Computer- Based (only)	3 (3)	-0.041 [-0.333 to 0.250]	0.782	0.0%	0.554
By intervention group					
Rule-based NLP	8 (9)	0.347 [0.116 to 0.578]	0.003	79.7%	<0.001
AI-based NLP	8 (14)	0.198 [-0.011 to 0.406]	0.063	34.4%	0.100
By scale used					
Generalized Anxiety Disorder (GAD-7)	9 (15)	0.333 [0.074 to 0.592]	0.012	71.7%	<0.001
Depression, Anxiety and Stress Scales (DASS-21)	3 (3)	0.050 [-0.352 to 0.453]	0.806	47.3%	0.150

^an = Number of individual studies. Measurements = Number of measurements between comparators and intervention. CA = conversational agent. AI = Artificial intelligence. SMD =

Standardized Mean Difference.

^bBold values are significant. Our study only presents meta-analyses with at least three measurements.

Given that factors such as age may influence the outcomes of depressive and anxious 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=0.182) or anxious symptoms (coefficient=-0.010; 95%CI=-0.030 to 0.010; p=0.294). Detailed results of the meta-regression are presented in Table 3.

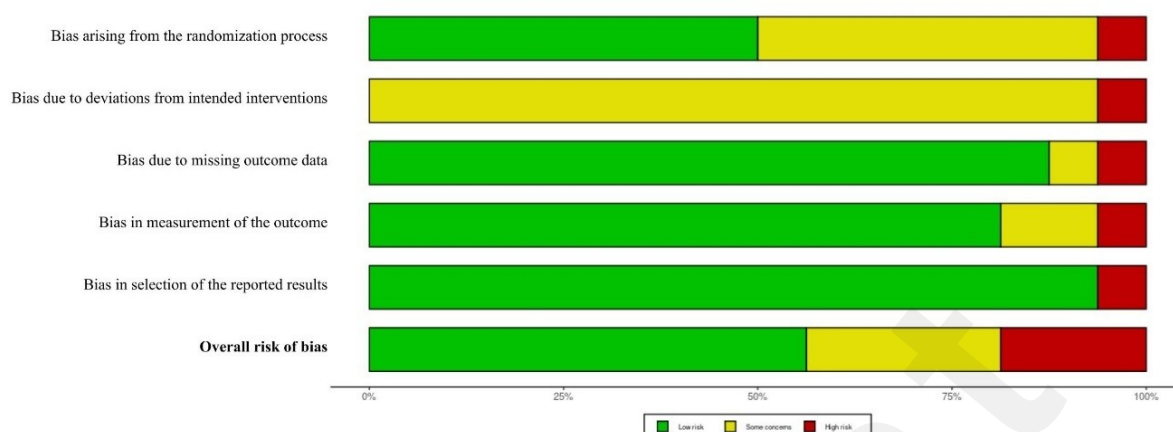
Table 3. Meta-regression analysis by overall meta-analysis of depressive and anxious symptoms.

Variable	Coeff.	95%CI	SE	t	p value
Depressive symptoms					
Age mean	-0.037	-0.092 to 0.019	0.026	-1.390	0.182
Intercept	2.108	-0.063 to 4.279	1.033	2.040	0.056
Anxious symptoms					
Age mean	-0.010	-0.030 to 0.010	0.009	-1.080	0.294
Intercept	0.553	-0.145 to 1.251	0.329	1.680	0.113

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 studies (n=9; 56.3%) had an overall low risk of bias, while only 18.7% (n=3) had an overall high risk of bias (see Figure 4A). Regarding the dimensions assessed, the lowest risk of bias was observed in reporting and analysis strategies (n=15; 93.8%), followed by participant loss or missing data (n=14; 87.5%). However, intervention delivery showed an unclear risk of bias due to limited reporting in the reviewed manuscripts. On the other hand, for the outcome of anxiety symptoms, half of the studies (n=8) had an overall low risk of bias, and only 12.5% (n=2) had an overall high risk of bias (see Figure 4B). At the level of each dimension assessed, all studies had a low risk of bias in reporting and analysis strategies, and 81.3% (n=13) had a low risk of bias in outcome measurement and retention throughout the study. However, most studies did not sufficiently detail the description of intervention delivery. Detailed risk of bias analyses for each study are available in Multimedia Appendix 6 for depressive symptoms and Multimedia Appendix 7 for anxious symptoms.

A) Depressive symptoms



B) Anxious symptoms

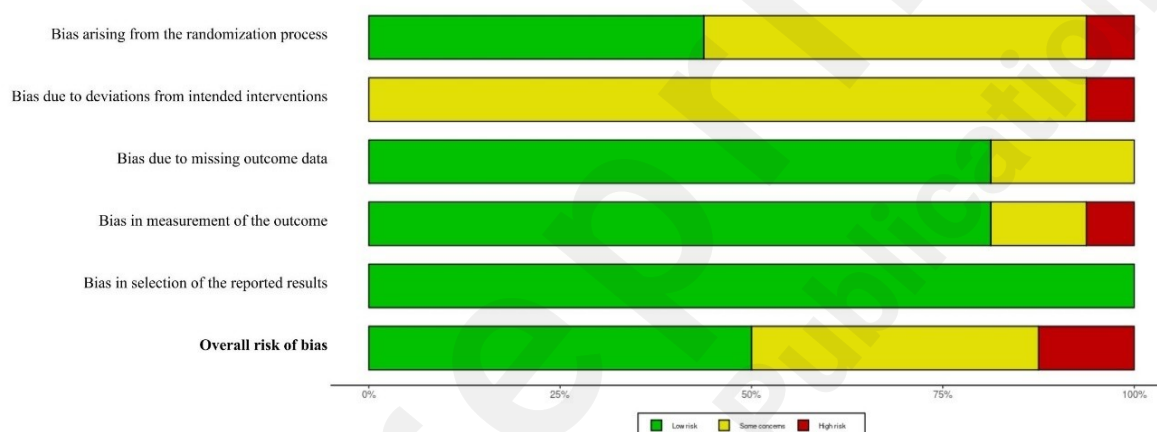


Figure 4. Risk of bias grouped for the outcome of depressive and anxious symptoms.

We found that, for the outcomes studied (depressive symptoms and anxious symptoms), the evidence was of very low certainty (See Table 4). This was mainly due to several factors. Firstly, there was a high risk of bias, with three studies presenting an overall high risk of bias for depressive symptoms and two studies for anxious symptoms. Secondly, there was significant inconsistency, as indicated by an overall I^2 higher than 60%. Additionally, indirectness was a major concern due to the high variability in the interventions, controls, and sample characteristics across the studies. Lastly, publication bias was strongly suspected, as the funnel plot revealed a marked asymmetry skewed to the right side. Despite these limitations, the findings provide a preliminary understanding of the potential effects of self-administered NLP-based interventions on depressive and anxious symptoms.

Table 4. Summary of findings and certainty of the evidence using the GRADE methodology

Outcome	Certainty assessment						Nº of patients		Effect	Certainty
	Nº of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Control	Intervention	Absolute (95%)	

									CI)	
Depressive symptoms	1516 (16 RCTs)	Very serious ^a	Very serious ^b	Very serious ^c	Not serious	Strongly suspected ^d	760	756	Hedges's SDM 0.82 lower (0.39 to 1.25)	⊕○○○ ○ Very low
Anxious Symptoms	2642 (16 RCTs)	Very serious ^e	Very serious ^b	Very serious ^c	Not serious	Strongly suspected ^d	1312	1330	Hedges's SDM 0.27 lower (0.12 to 0.43)	⊕○○○ ○ Very low

^aThree studies present an overall high risk of bias.

^bOverall I^2 higher than 60%.

^cThere is a high variability in the interventions, controls, and sample characteristics.

^dThe funnel plot reveals a marked asymmetry skewed to the right side.

^eTwo studies present an overall high risk of bias.

Discussion

Main findings

Our results indicate that self-administered interventions based on NLP models have a significant overall effect on reducing depressive and anxiety symptoms, in contrast to various control conditions. Our study used random effects models to estimate this overall effect, thus accounting for heterogeneity among the interventions analyzed. Thus, we consider the results to be robust. At the level of each intervention 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-only interventions were shown to be effective in reducing depressive and anxiety symptoms compared with pooled control conditions. In addition, NLP-based interventions overall outperform psychoeducation and bibliotherapy in reducing both depressive and anxiety symptoms. In addition, these interventions are more effective than waitlist or no intervention in reducing anxious symptoms.

These findings support the usefulness of self-applied NLP-based interventions in alleviating such common mental health problems as depressive and anxious 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 models at various stages of mental health care in both clinical and community settings [50–52], as 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 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 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 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 models compared to other low- and middle-income countries [58,59]. This situation represents a challenge and a potential source of inequity in access to and implementation of NLP-based interventions within public health systems.

Implications for clinical practice and public health

A previous systematic review study 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, as it 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 anxiety and depressive 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 (AI-QIF) 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, internet, and stable connectivity, which is a challenge 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 the 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 MDS. Third, the lack of clarity in the description of the studied groups may have introduced a risk of bias in assessing their effectiveness, as 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]. Sixth, the GRADE assessment shows that the evidence for self-administered NLP-based interventions on depressive and anxious 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, as studies with non-significant 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 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 to recommend the adoption of NLP-based self-administered interventions in public health systems effectively.

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.

Acknowledgments

This project was in part funded by the grant R33HL143317 from the National Institutes of Health. We thank Piero Segobia and Carlos García-Navarrete for their collaboration in the initial stages of the manuscript.

Conflict of Interests

The authors do not report any conflict of interest when conducting the study, analyzing the data, or writing the manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process

We used DeepL to translate specific sections of the manuscript and Grammarly to improve the wording of certain sections. All authors reviewed and approved the final version of the manuscript.

Data availability

Not applicable.

Author Contributions

David Villarreal-Zegarra: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization.

C. Mahony Reategui-Rivera: Conceptualization, Methodology, Investigation, Writing - Original Draft, Writing - Review & Editing.

Jackeline García-Serna: Methodology, Validation, Investigation, Data Curation.

Gleni Quispe-Callo: Validation, Data Curation.

Gabriel Lázaro-Cruz: Validation, Data Curation.

Gianfranco Centeno-Terrazas: Validation, Investigation, Writing - Original Draft, Writing - Review & Editing.

Ricardo Galvez-Arevalo: Validation, Data Curation.

Alejandro Dominguez-Rodriguez: Investigation, Writing - Original Draft, Writing - Review & Editing, Supervision.

Stefan Escobar-Agreda: Investigation, Writing - Original Draft, Writing - Review & Editing

Joseph Finkelstein: Investigation, Resources, Writing - Review & Editing, Supervision.

Abbreviations

NLP: Natural Language Processing

AI: Artificial Intelligence

SMD: Standardized Mean Differences

PHQ-9: Patient Health Questionnaire

CBT: Cognitive Behavioral Therapy

RCT: Randomized Clinical Trials

ML: Machine Learning

LLM: Large Language Model

PROSPERO: International prospective register of systematic reviews

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

JB: Joanna Briggs Institute

GAD: Generalized Anxiety Disorder Scale

GRADE: Grading of Recommendations, Assessment, Development, and Evaluation

HAM: Hamilton Anxiety Rating Scale

BDI: Beck Depression Inventory

HDRS: Hamilton Depression Rating Scale

BAI: Beck Anxiety Inventory

AI-QIF: Artificial Intelligence-Quality Implementation Framework

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 Lond Engl* 2021 Nov 6;398(10312):1700–1712. PMID:34634250
2. Arias D, Saxena S, Verguet S. Quantifying the global burden of mental disorders and their economic value. *EClinicalMedicine* 2022 Dec;54:101675. PMID:36193171
3. World Health Organization. Mental health at work. 2022. Available from: <https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work> [accessed Apr 5, 2024]
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* 2023 May;164:104292. PMID: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* 2021 Nov 26;8(11):e27404. PMID:34842556
6. Bauer AM, Iles-Shih M, Ghomi RH, Rue T, Grover T, Kincler N, Miller M, Katon WJ. Acceptability of mHealth augmentation of Collaborative Care: A mixed methods pilot study. *Gen Hosp Psychiatry* 2018;51:22–29. PMID:29272712
7. Philippe TJ, Sikder N, Jackson A, Koblanski ME, Liow E, Pilarinos A, Vasarhelyi K. Digital Health Interventions for Delivery of Mental Health Care: Systematic and Comprehensive Meta-Review. *JMIR Ment Health* 2022 May 12;9(5):e35159. PMID:35551058
8. Shelton CR, Kotsiou A, Hetzel-Riggin MD. Digital Mental Health Interventions: Impact and Considerations. IGI Global; 1AD. Available from: <https://www.igi-global.com/gateway/chapter/www.igi-global.com/gateway/chapter/281753> [accessed Jun 7, 2024] ISBN:9781799864530
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. PMID:37492117
10. Cuijpers P, Schuurmans J. Self-help interventions for anxiety disorders: an overview. *Curr Psychiatry Rep* 2007 Aug;9(4):284–290. PMID:17880859
11. World Health Organization. Psychological interventions implementation manual: integrating evidence-based psychological interventions into existing services. Geneva: World Health Organization; 2024. Available from: <https://www.who.int/publications/i/item/9789240087149>
12. World Health Organization. Consolidated telemedicine implementation guide. Geneva: World Health Organization; 2022. Available from: <https://www.who.int/publications-detail-redirect/9789240059184> [accessed Jun 7, 2024]
13. Dwyer DB, Falkai P, Koutsouleris N. Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annu Rev Clin Psychol* 2018 May 7;14:91–118. PMID:29401044
14. Caldarini G, Jaf S, McGarry K. A Literature Survey of Recent Advances in Chatbots. Information Multidisciplinary Digital Publishing Institute; 2022 Jan;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* 2023 Oct 6;13(1):309. PMID:37798296
 16. Friedman C, Hripcsak G. Natural language processing and its future in medicine. *Acad Med J Assoc Am Med Coll* 1999 Aug;74(8):890–895. PMID: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* 2018 Aug;13(3):55–75. doi: 10.1109/MCI.2018.2840738
 18. Nie J, Shao H, Fan Y, Shao Q, You H, Preindl M, Jiang X. LLM-based Conversational AI Therapist for Daily Functioning Screening and Psychotherapeutic Intervention via Everyday Smart Devices. *arXiv*; 2024. doi: 10.48550/arXiv.2403.10779
 19. Le Glaz A, Haralambous Y, Kim-Dufor D-H, Lenca P, Billot R, Ryan TC, Marsh J, DeVlyder J, Walter M, Berrouguet S, Lemey C. Machine Learning and Natural Language Processing in Mental Health: Systematic Review. *J Med Internet Res* 2021 May 4;23(5):e15708. PMID:33944788
 20. Zhang T, Schoene AM, Ji S, Ananiadou S. Natural language processing applied to mental illness detection: a narrative review. *NPJ Digit Med* 2022 Apr 8;5(1):46. PMID:35396451
 21. De Choudhury M, Pendse SR, Kumar N. Benefits and Harms of Large Language Models in Digital Mental Health. *arXiv*; 2023. doi: 10.48550/arXiv.2311.14693
 22. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page M, Welch V, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.4. Chichester: John Wiley & Sons; 2023. ISBN:978-0-470-69951-5
 23. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Chapter 3: Systematic reviews of effectiveness. Aromataris E Munn Z Ed *JBIMan Evid Synth* JBI; 2020.
 24. Barker TH, Stone JC, Sears K, Klugar M, Tufanaru C, Leonardi-Bee J, Aromataris E, Munn Z. The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. *JBIMan Evid Synth* 2023 Mar 1;21(3):494–506. PMID:36727247
 25. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011 Apr;64(4):383–394. PMID:21195583
 26. Ferguson CJ. An effect size primer: A guide for clinicians and researchers. *Prof Psychol Res Pract US: American Psychological Association*; 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* 1997 Sep 13;315(7109):629–634. PMID: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* 2000 Jun;56(2):455–463. PMID: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* 2023 Jul;14(3):2020–2032. doi: 10.1109/TAFFC.2022.3143803
 30. Sabour S, Zhang W, Xiao X, Zhang Y, Zheng Y, Wen J, Zhao J, Huang M. A chatbot for mental health support: exploring the impact of Emohaa on reducing mental distress in China. *Front Digit Health* 2023;5:1133987. PMID:37214342
 31. Suharwardy S, Ramachandran M, Leonard SA, Gunaseelan A, Lyell DJ, Darcy A, Robinson A, Judy A. Feasibility and impact of a mental health chatbot on postpartum mental health: a randomized controlled trial. *AJOG Glob Rep* 2023 Aug;3(3):100165. PMID:37560011

32. Danieli M, Ciulli T, Mousavi SM, Silvestri G, Barbato S, Di Natale L, Riccardi G. Assessing the Impact of Conversational Artificial Intelligence in the Treatment of Stress and Anxiety in Aging Adults: Randomized Controlled Trial. *JMIR Ment Health* 2022 Sep 23;9(9):e38067. PMID:36149730
33. Fitzsimmons-Craft EE, Chan WW, Smith AC, Firebaugh M-L, Fowler LA, Topooco N, DePietro B, Wilfley DE, Taylor CB, Jacobson NC. Effectiveness of a chatbot for eating disorders prevention: A randomized clinical trial. *Int J Eat Disord* 2022 Mar;55(3):343–353. PMID: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* 2022 Nov 22;6(11):e40242. PMID:36413390
35. He Y, Yang L, Zhu X, Wu B, Zhang S, Qian C, Tian T. 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
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* 2022 Mar;27:100495. PMID: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* 2021 May 20;9(5):e26152. PMID: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* 2021 Aug 12;5(8):e20678. PMID:34092548
39. Oh J, Jang S, Kim H, Kim J-J. Efficacy of mobile app-based interactive cognitive behavioral therapy using a chatbot for panic disorder. *Int J Med Inf* 2020 Aug;140:104171. PMID:32446158
40. Greer S, Ramo D, Chang Y-J, 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* 2019 Oct 31;7(10):e15018. PMID: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* 2018 Dec 13;5(4):e64. PMID: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* 2017 Jun 6;4(2):e19. PMID: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* 2018 Sep;46(5):570–582. PMID: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* 2014 Nov;42(6):731–746. PMID:23899405
45. Jang S, Kim J-J, Kim S-J, 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 Inf* 2021 Jun;150:104440. PMID:33799055

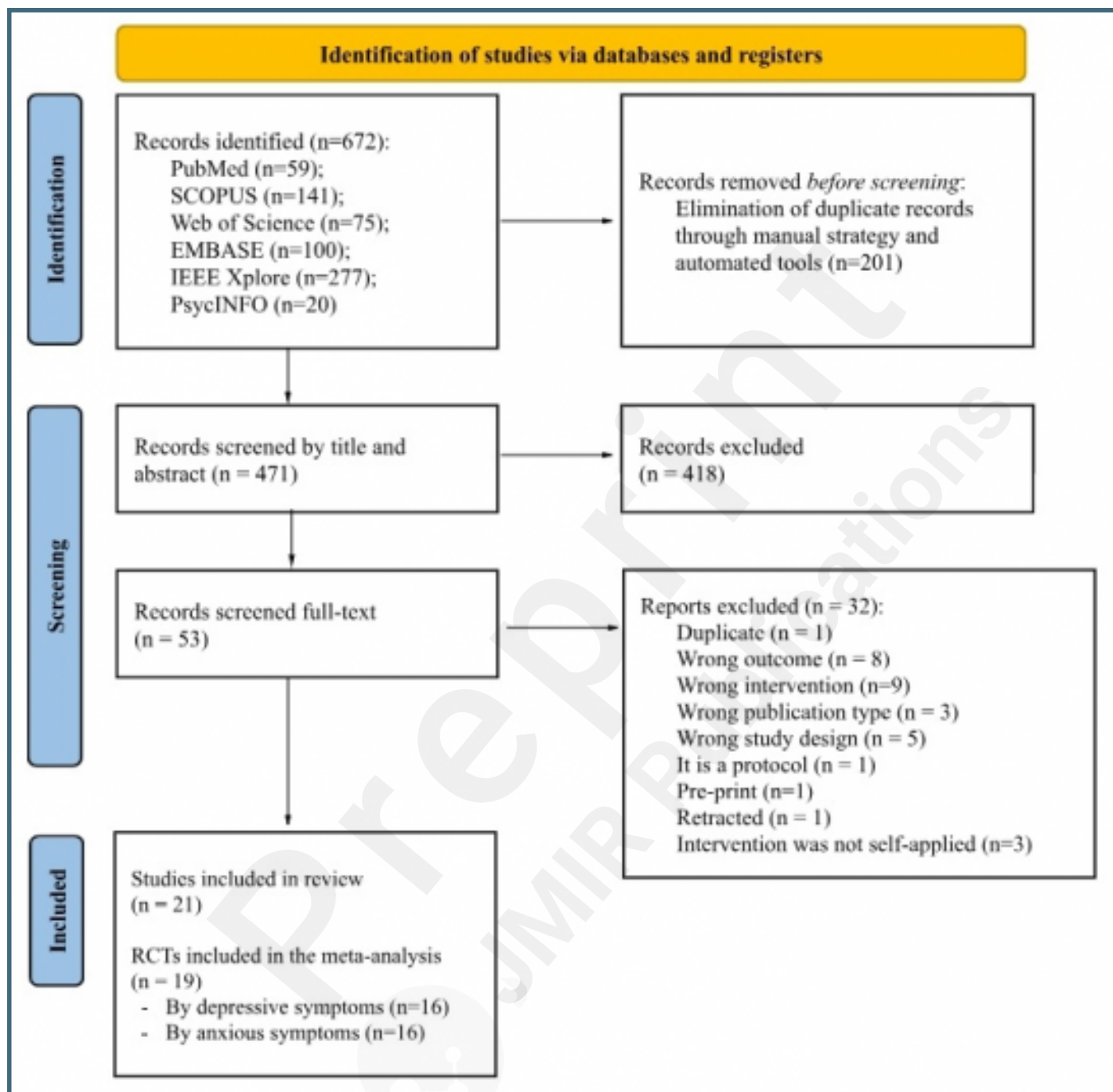
46. Maeda E, Miyata A, Boivin J, Nomura K, Kumazawa Y, Shirasawa H, Saito H, Terada Y. Promoting fertility awareness and preconception health using a chatbot: a randomized controlled trial. *Reprod Biomed Online* 2020 Dec;41(6):1133–1143. PMID:33039321
47. Prochaska JJ, Vogel EA, Chieng A, Baiocchi M, Maglalang DD, Pajarito S, Weingardt KR, Darcy A, Robinson A. A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. *Drug Alcohol Depend* 2021 Oct 1;227:108986. PMID: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* 2022 Jan 21;9(1):e32211. PMID:35060918
49. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, Robinson A. A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study. *J Med Internet Res* 2021 Mar 23;23(3):e24850. PMID:33755028
50. Lázaro E, Yepez J-C, Marín-Maicas P, López-Masés P, Gimeno T, de Paúl S, Moscardó V. Efficiency of natural language processing as a tool for analysing quality of life in patients with chronic diseases. A systematic review. *Comput Hum Behav Rep* 2024 May 1;14:100407. doi: 10.1016/j.chbr.2024.100407
51. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. *Nat Lang Eng* 2017 Sep;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* 2022 Apr 22;9(4):e35928. PMID: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* 2023 Apr 19;25:e44548. PMID:37074762
54. Treviso M, Lee J-U, Ji T, Aken B van, Cao Q, Ciosici MR, Hassid M, Heafield K, Hooker S, Raffel C, Martins PH, Martins AFT, Forde JZ, Milder P, Simpson E, Slonim N, Dodge J, Strubell E, Balasubramanian N, Derczynski L, Gurevych I, Schwartz R. Efficient Methods for Natural Language Processing: A Survey. *Trans Assoc Comput Linguist* 2023 Jul 12;11:826–860. doi: 10.1162/tacl_a_00577
55. Lin T, Wang Y, Liu X, Qiu X. A survey of transformers. *AI Open* 2022 Jan 1;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* 2023 Sep 1;4:100026. doi: 10.1016/j.nlp.2023.100026
57. Chin H, Song H, Baek G, Shin M, Jung C, Cha M, Choi J, Cha C. The Potential of Chatbots for Emotional Support and Promoting Mental Well-Being in Different Cultures: Mixed Methods Study. *J Med Internet Res* 2023 Oct 20;25:e51712. PMID:37862063
58. Oduoye MO, Fatima E, Muzammil MA, Dave T, Irfan H, Fariha FNU, Marbell A, Ubechu SC, Scott GY, Elebesunu EE. 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* 2024 Jan;7(1):e1794. PMID:38186931
59. Alami H, Rivard L, Lehoux P, Hoffman SJ, Cadeddu SBM, Savoldelli M, Samri MA, Ag Ahmed MA, Fleet R, Fortin J-P. Artificial intelligence in health care: laying the Foundation for Responsible, sustainable, and inclusive innovation in low- and middle-income countries. *Glob Health* 2020 Jun 24;16(1):52. PMID: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. PMID:34713142
61. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim H-C, Jeste DV. Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Curr Psychiatry Rep* 2019 Nov 7;21(11):116. PMID:31701320
62. Boucher EM, Harake NR, Ward HE, Stoeckl SE, Vargas J, Minkel J, Parks AC, Zilca R. Artificially intelligent chatbots in digital mental health interventions: a review. *Expert Rev Med Devices* 2021 Dec;18(sup1):37–49. PMID:34872429
63. Pham KT, Nabizadeh A, Selek S. Artificial Intelligence and Chatbots in Psychiatry. *Psychiatr Q* 2022 Mar;93(1):249–253. PMID:35212940
64. Dawoodbhoy FM, Delaney J, Cecula P, Yu J, Peacock I, Tan J, Cox B. AI in patient flow: applications of artificial intelligence to improve patient flow in NHS acute mental health inpatient units. *Heliyon* 2021 May;7(5):e06993. PMID:34036191
65. Nilsen P, Svedberg P, Neher M, Nair M, Larsson I, Petersson L, Nygren J. A Framework to Guide Implementation of AI in Health Care: Protocol for a Cocreation Research Project. *JMIR Res Protoc* 2023 Nov 8;12:e50216. PMID:37938896
66. Thakkar A, Gupta A, De Sousa A. Artificial intelligence in positive mental health: a narrative review. *Front Digit Health* 2024;6:1280235. PMID:38562663
67. Chidambaram S, Jain B, Jain U, Mwavu R, Baru R, Thomas B, Greaves F, Jayakumar S, Jain P, Rojo M, Battaglino MR, Meara JG, Sounderajah V, Celi LA, Darzi A. An introduction to digital determinants of health. *PLOS Digit Health* 2024 Jan;3(1):e0000346. PMID:38175828

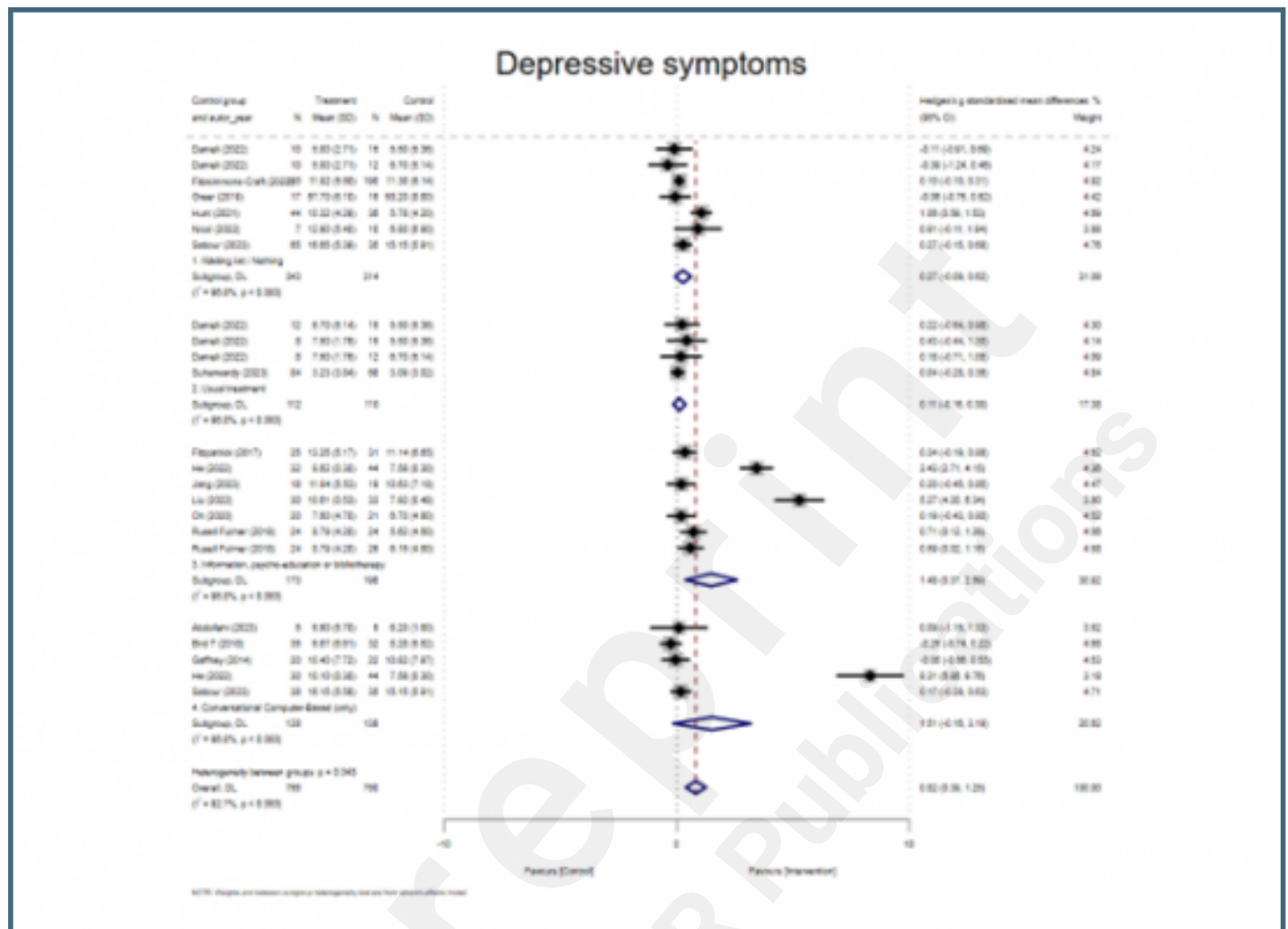
Supplementary Files

Figures

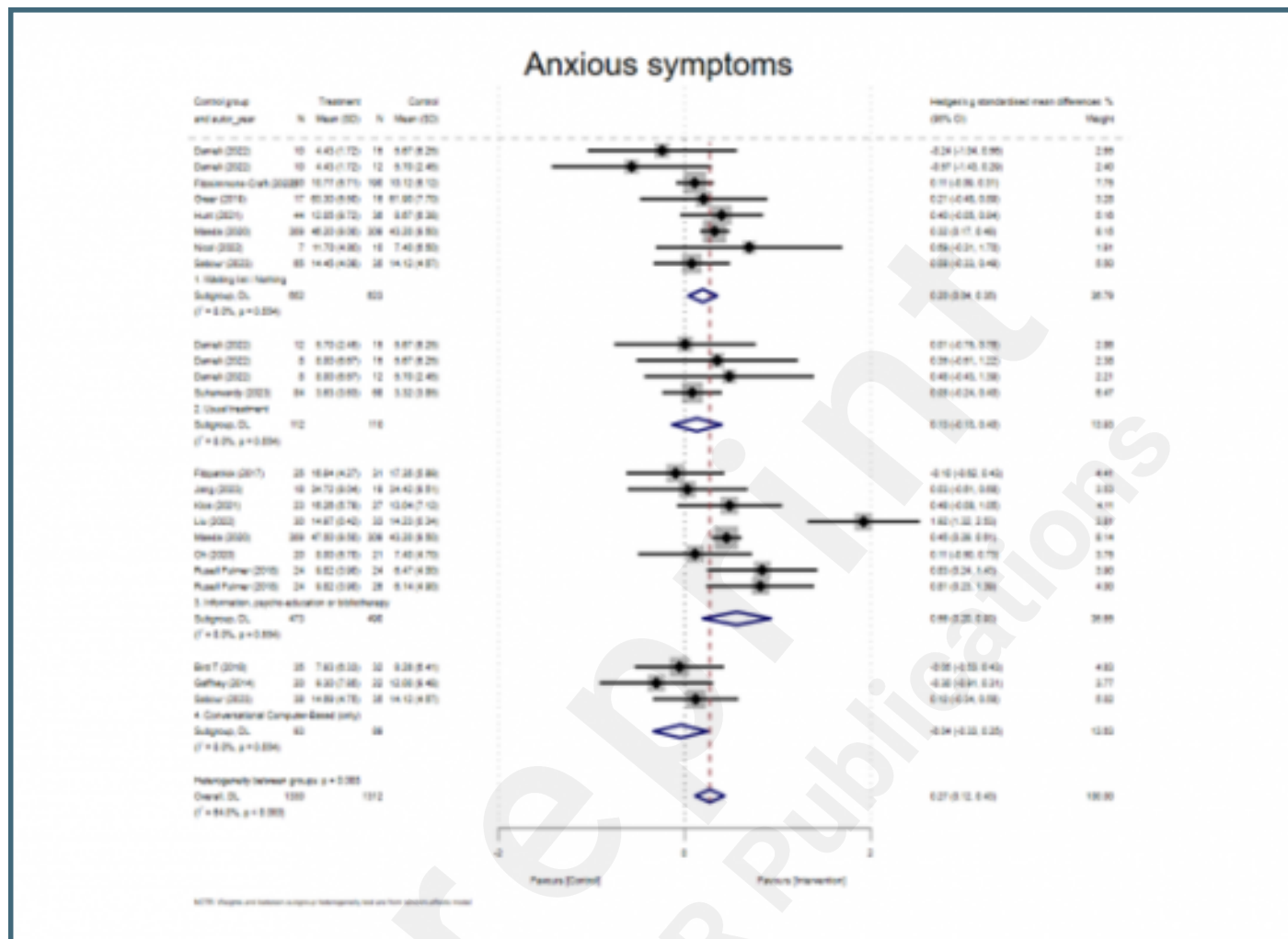
Flowchart of the selection process.



Forest plot for control conditions vs self-administered interventions based on natural language processing models to reduce depressive symptoms.

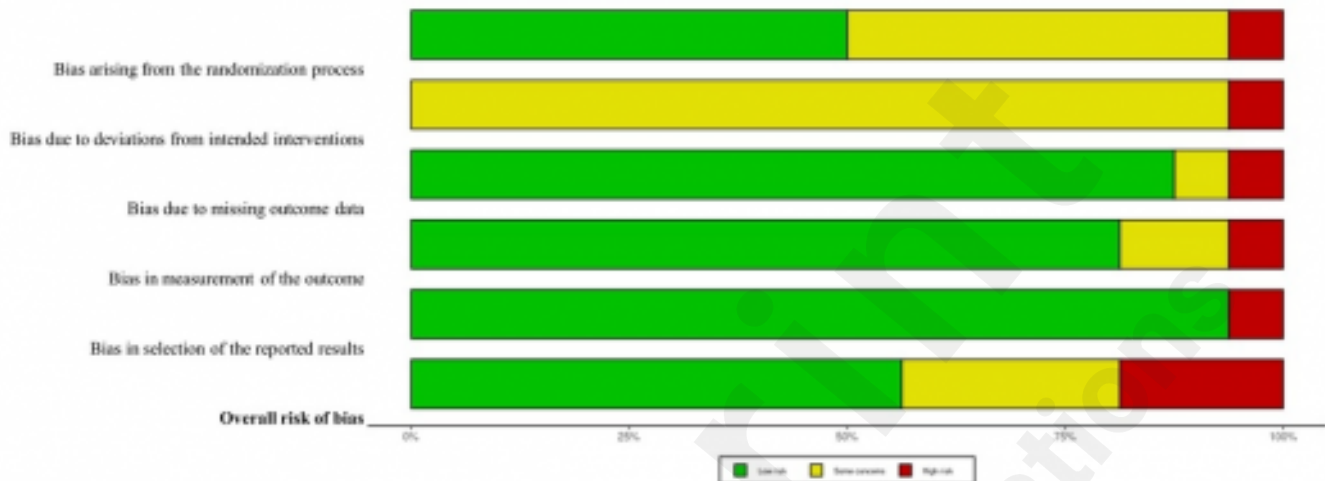


Forest plot for control conditions vs self-administered interventions based on natural language processing models to reduce anxious symptoms.

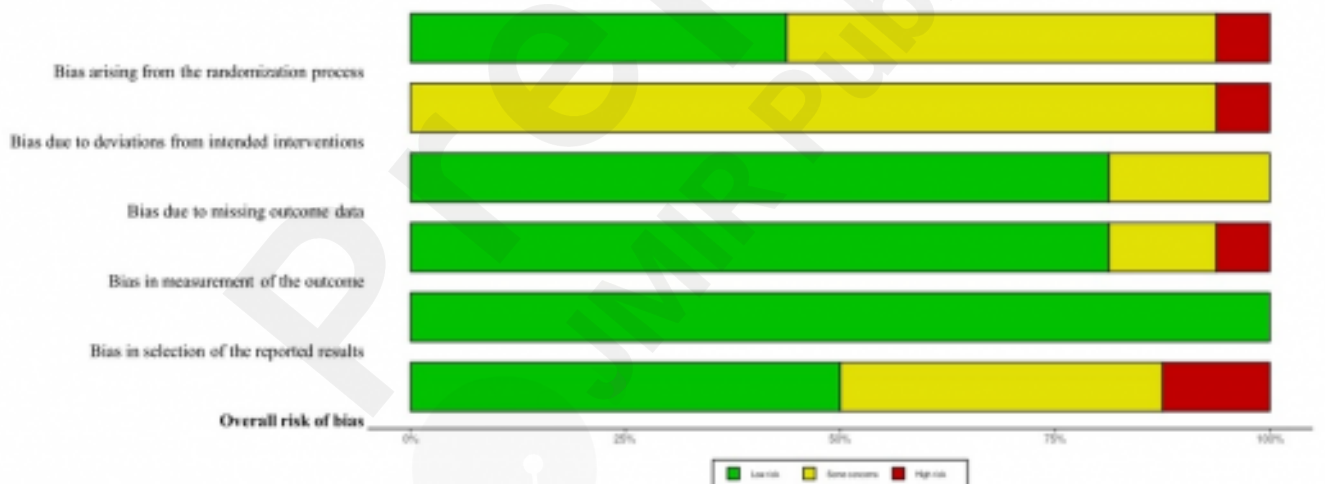


Risk of bias grouped for the outcome of depressive and anxious symptoms.

A) Depressive symptoms



A) Anxious symptoms



Multimedia Appendixes

PRISMA 2020 Checklist.

URL: <http://asset.jmir.pub/assets/33711eb795ca9a18f221f3c6db8077dd.docx>

Search strategy.

URL: <http://asset.jmir.pub/assets/7fc68e185f68f6ed9f5ec723fd4e1b69.docx>

Excluded records.

URL: <http://asset.jmir.pub/assets/2356b5bdbd00b9d85ba32d2cccf13a02.docx>

Included records.

URL: <http://asset.jmir.pub/assets/192b177817a17b05d47f1947fcfd72a7.docx>

Funnel plot by depressive and anxious symptoms.

URL: <http://asset.jmir.pub/assets/5ad83d6cc86f52658fc9c21bede23e4b.docx>

Risk of bias for individual studies for the outcome of depressive symptoms.

URL: <http://asset.jmir.pub/assets/f14395720f92f5db326efda68ea9ecaf.docx>

Risk of bias for individual studies for the outcome of anxious symptoms.

URL: <http://asset.jmir.pub/assets/4853f6fb3c5a1d3d9febd978ea8d8e1e.docx>