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

Table of Contents

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

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

[1]Instituto Peruano de Orientación Psicológica Lima PE
[2]Instituto Nacional de Salud del Niño San Borja Lima PE
[3]Unidad de Telesalud Universidad Nacional Mayor de San Marcos Lima PE
[4]Department of Psychology, Health, and Technology University of Twente Enschede NL
[5]Department 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 v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

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

Authors:

David Villarreal-Zegarra[1,2]
Jackeline García-Serna[1]
Gleni Quispe-Callo[1]
Gabriel Lázaro-Cruz[1]
Gianfranco Centeno-Terrazas[1]
Ricardo Galvez-Arevalo[3]
Stefan Escobar-Agreda[5]
Alejandro Dominguez-Rodriguez[6]
C. Mahony Reategui-Rivera[1,4]
Joseph Finkelstein [4]

Affiliations

[1] Instituto Peruano de Orientación Psicológica, Lima, Peru.
[2] Escuela de Psicología, Universidad Continental, Lima, Peru
[3] Instituto Nacional de Salud del Niño San Borja, Lima, Peru.
[4] Department of Biomedical Informatics, University of Utah, Utah, United States of America.
[5] Unidad de Telesalud, Universidad Nacional Mayor de San Marcos, Lima, Peru.
[6] Department of Psychology, Health, and Technology, University of Twente, Enschede, The Netherlands.

Emails

David Villarreal-Zegarra	dvillareal@continental.edu.pe
Jackeline García-Serna	jackelinegarcia@ipops.pe
Gleni Quispe-Callo	gquispecal@unsa.edu.pe
Gabriel Lázaro-Cruz	jose.lazaro@upch.pe
Gianfranco Centeno-Terrazas	gianfranco.centeno.t@gmail.com
Ricardo Galvez-Arevalo	galvezarevaloricardo@gmail.com
Alejandro Dominguez-Rodriguez	a.dominguezrodriguez@utwente.nl
Stefan Escobar-Agreda	stefan1090ea@gmail.com
Mahony Reateguí-Rivera	mahony.reategui@utah.edu
Joseph Finkelstein	joseph.finkelstin@utah.edu

ORCID

David Villarreal-Zegarra	https://orcid.org/0000-0002-2222-4764
Jackeline García-Serna	https://orcid.org/0000-0001-9260-1505
Gleni Quispe-Callo	https://orcid.org/0000-0003-3994-9523
Ricardo Galvez-Arevalo	https://orcid.org/0000-0002-1006-1523
Gianfranco Centeno-Terrazas	https://orcid.org/0000-0002-0773-9866
Gabriel Lázaro-Cruz	https://orcid.org/0000-0002-5618-1002
Alejandro Dominguez-Rodriguez	https://orcid.org/0000-0003-3547-8824
Stefan Escobar-Agreda	https://orcid.org/0000-0002-8355-4310
Mahony Reateguí-Rivera	https://orcid.org/0000-0002-4030-8777
Joseph Finkelstein	https://orcid.org/0000-0002-8084-7441

Corresponsal author:

David Villarreal-Zegarra.
Email: dvillareal@continental.edu.pe; davidvillarreal@ipops.pe

Statements

Acknowledgments

We thank Piero Segobia and Carlos García-Navarrete for their collaboration in the initial stages of the manuscript.

Funding

Not applicable.

Data availability

Not applicable.

Ethical approval and consent to participate

Not applicable.

Consent to publication

Not applicable.

Competing 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.

Author Contributions

David Villarreal-Zegarra: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization.
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.
Stefan Escobar-Agreda: Investigation, Writing - Original Draft, Writing - Review & Editing
Alejandro Dominguez-Rodriguez: Investigation, Writing - Original Draft, Writing - Review & Editing, Supervision.
C. Mahony Reategui-Rivera: Conceptualization, Methodology, Investigation, Writing - Original Draft, Writing - Review & Editing.
Joseph Finkelstein: Investigation, Resources, Writing - Review & Editing, Supervision.

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.

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

BACKGROUND

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 annually due to loss of productivity [3]. Additionally, its increasing incidence and lack of health resources challenge the healthcare systems and workforce to adequately cover their rising demand [4].

In response, self-administered technology-based interventions have emerged as promising solutions for the management of 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].

Furthermore, 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 [7], NLP has the potential to make self-directed interventions more cost-effective and accessible and facilitate fidelity and engagement of patients through better interaction [8].

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 [9]. Conversely, AI-based NLP, encompassing machine learning and deep learning techniques, learns from extensive data to process language, showing remarkable success in various NLP tasks due to its scalability and ability to manage linguistic ambiguities [10].

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 has been demonstrated in specific psychotherapeutic settings [11].

Previously, other systematic reviews were conducted analyzing the impact of NLP on Mental Health, such as the one conducted by Le Glaz [12] and Zhang [13]. However, they were mostly focused on the general use of NLP. Also, a systematic review and research framework regarding the use of NLP in Mental Health interventions, however broadly without a concrete focus on Online Self-Applied Interventions [8].

Despite these advancements, analysis of their effectiveness and safety in managing mental health concerns such as depression and anxiety remains fragmented [14]. 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

Design and protocol register

Our study was a systematic review. We used the PRISMA reporting guidelines (see Multimedia Appendix 1), and the protocol was registered in PROSPERO (CRD42023472120).

Eligibility criteria

The PICO of the study is to determine if interventions based on NLP models can reduce depressive and anxious symptoms. 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. Anxious symptoms, in turn, are characterized by the anticipation of imagined events that are perceived as potential threats, causing emotional distress and physiological tension.

Population: Our population was focused on any age group (child, adolescent, adult, and older adult) with or without previous comorbidities. 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), as well as those who have been assessed with psychometric instruments to measure depressive or anxious symptoms, will be included.

Intervention: Our 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 whether they were designed to improve participants' mental health or whether they were designed for a purpose other than improving mental health. We considered only the self-administered interventions.

Comparator: We consider as comparators any passive comparator (i.e., waiting lists, non-intervention control groups, or placebos) and 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: Our study had two primary outcomes: the participants' depressive and anxiety symptom scores. Only scores on a psychometric questionnaire with evidence of validity and reliability for measuring depressive or anxiety symptoms (i.e., Patient Health Questionnaire [PHQ-9], Beck Depression Inventory [BDI], Hamilton Depression Rating Scale [HDRS], Generalised Anxiety Disorder Scale [GAD-7], Beck Anxiety Inventory [BAI], Hamilton Anxiety Rating Scale [HAM] or other) will be included as outcomes. Our study focused on the difference between pre- and post-intervention scores. However, the study does not report information on the effect size of the intervention. In that case, it will be calculated manually from the means, standard deviations, or effect sizes available in the articles.

Design: We included randomized clinical trials and quasi-experimental studies (without a control arm or randomization groups) to determine the effect of interventions based on NLP models to reduce depressive symptoms and anxiety. We excluded narrative reviews, systematic reviews, scoping reviews, or other designs. Only peer-reviewed publications (original articles or briefs) were included in this review; proceedings, posters, or other similar items were excluded. There were no exclusion criteria based on language, time interval of publication, or setting (i.e., clinical to community settings).

Information Sources and Search Strategy

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 search strategy included NLP, depression and anxiety terms, and health science descriptors (see Multimedia Appendix 2). Our search included any document available from January 01, 1900, to November 03, 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. The resulting retrieved text was reviewed independently by two pairs of authors (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.

Data collection process and Data items

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, which were reviewed by all raters with 85% agreement among them. Subsequently, minor changes were made to the final version of the extraction form to improve the clarity of the extraction form, which includes the following data: 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).

Study risk of bias assessment

We use the JBI Critical Appraisal Tools, which are designed 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 [15] was used, a checklist with a total of nine questions for assessing potential bias in this type of study. For randomized clinical trials, the JBI Critical Appraisal Tool for Risk Assessment of Bias in Randomised Controlled Trials [16] 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 or not 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 first resolved by discussion. If the disagreement persisted, a third reviewer was asked to arbitrate.

Synthesis methods

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 Hedge 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 [17].

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 [18]. 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 [19].

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 Large Language Models (LLM), Multimodal Large Language Models (MLLM), artificial intelligence-led (AI-led), 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.

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. Figure 1 shows the complete review process, and Multimedia Appendix 3 lists the articles excluded from the full-text review.

Characteristics of the included studies

Our review identified 19 randomized clinical trials [20–38] and two quasi-experimental studies without a control group [39,40]. 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%). Regarding the outcomes assessed, depressive symptoms were analyzed in 95.2% of the studies, and anxiety symptoms in 90.5%. We found 29 potential comparisons between interventions and controls because five studies reported three or more arms. The most common intervention was conversational agent alone (75.9%), and the most common control conditions were waiting list or no intervention (31%; n=9) and information, psychoeducation, or bibliotherapy (31%; n=9). The most commonly used scales to measure depressive and anxiety symptoms were the PHQ-9 (n=21; 72.4%) and the GAD-7 (n=18; 62.1%). Table 1 shows the characteristics of the studies, divided into randomized clinical trials and uncontrolled quasi-experimental studies.

Synthesis of results and meta-analysis

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, 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]; τ^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 4).

For the outcome of anxiety symptoms, 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]; τ^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 4).

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 than 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.

Risk of bias in studies

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 intervention delivery description. Detailed risk of bias analyses for each study are available in Multimedia Appendix 5 for depressive symptoms and Multimedia Appendix 6 for anxiety symptoms.

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 [41–43], 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 [44]. However, our review found that in the case of randomized clinical trials focusing on depressive and anxiety symptoms, validated instruments such as the PHQ and GAD 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 machine learning models showed significant results compared to other strategies, such as rule-based chatbots [45]. Unlike simpler NLP models, conversational agents offer better performance on various tasks [45]. However, more complex models also require high computational costs and large amounts of data for optimization [46,47], which may limit their adaptability to the different linguistic and cultural needs of different regions [48]. 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 [49,50]. 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 machine learning in mental health agrees with us on the potential of NLP-based interventions to improve population mental health [12]. 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 [51,52]. Given their accessibility through digital platforms, these interventions have the potential to reduce the burden of anxiety and depressive disorders at the population level [53,54], while also being cost-effective and a way to optimize mental health resources [55]. To ensure successful implementation within the public health system, the use of the Artificial Intelligence-Quality Implementation Framework (AI-QIF) could be beneficial [56]. 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 [57]. In addition, it is important to consider the digital determinants of health [58], 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 randomized clinical trials, 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 different 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 better understand the effect of these interventions. To overcome these limitations in future reviews, we recommend focusing research 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. However, there is still a need for randomized clinical trials with large sample sizes that can further strengthen the inferential power of future meta-analyses. Although our findings are encouraging, systematic reviews that examine the implementation process of these interventions in-depth, as well as qualitative studies that evaluate the usability and feasibility of these interventions, are needed to recommend their adoption in the public health system. Our study provides a valuable starting point for future research aimed at validating the efficacy of these interventions as components of standard mental health care within public health systems.

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. 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
8. 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
9. 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
10. 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
11. 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
12. Le Glaz A, Haralambous Y, Kim-Dufor D-H, Lenca P, Billot R, Ryan TC, Marsh J, DeVylder J, Walter M, Berrouiguet 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
13. 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
14. 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
15. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Chapter 3: Systematic reviews of effectiveness. Aromataris E Munn Z Ed JBI Man Evid Synth JBI; 2020.
16. 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. JBI Evid Synth 2023 Mar 1;21(3):494–506. PMID:36727247
17. 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
18. 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
19. 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
20. 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
21. 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
22. 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
23. 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
24. 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
25. 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
26. 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
27. 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
28. 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
29. 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
30. 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
31. 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
32. 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
33. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 6;4(2):e19.

PMID:28588005
34. 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
35. 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
36. 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
37. 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
38. 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
39. 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
40. 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
41. 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
42. 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
43. 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
44. 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
45. 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
46. 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
47. 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
48. 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

49. 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
50. 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
51. 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
52. 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
53. 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
54. Pham KT, Nabizadeh A, Selek S. Artificial Intelligence and Chatbots in Psychiatry. Psychiatr Q 2022 Mar;93(1):249–253. PMID:35212940
55. 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
56. 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
57. Thakkar A, Gupta A, De Sousa A. Artificial intelligence in positive mental health: a narrative review. Front Digit Health 2024;6:1280235. PMID:38562663
58. 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

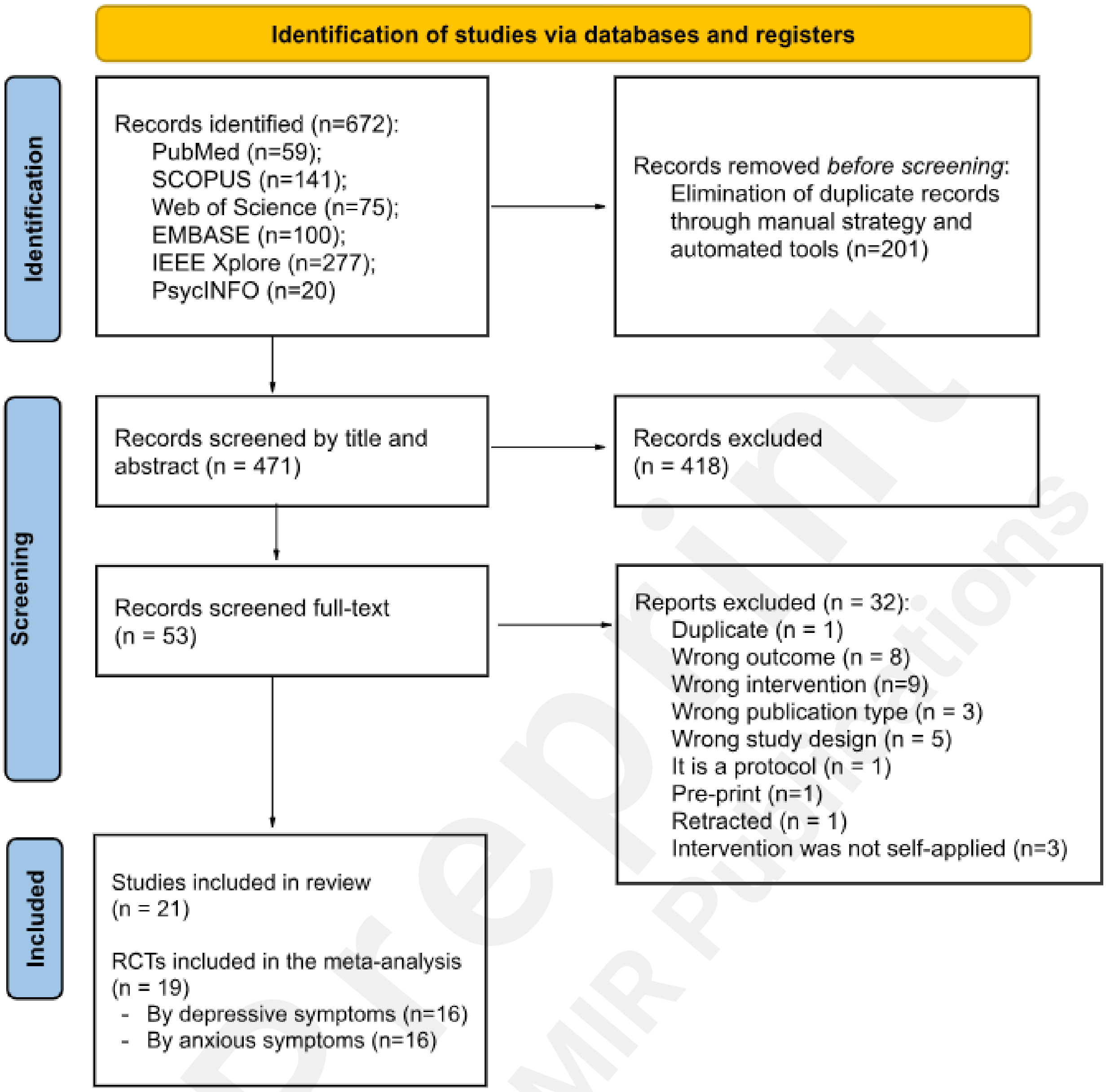

Figure 1. Flowchart.

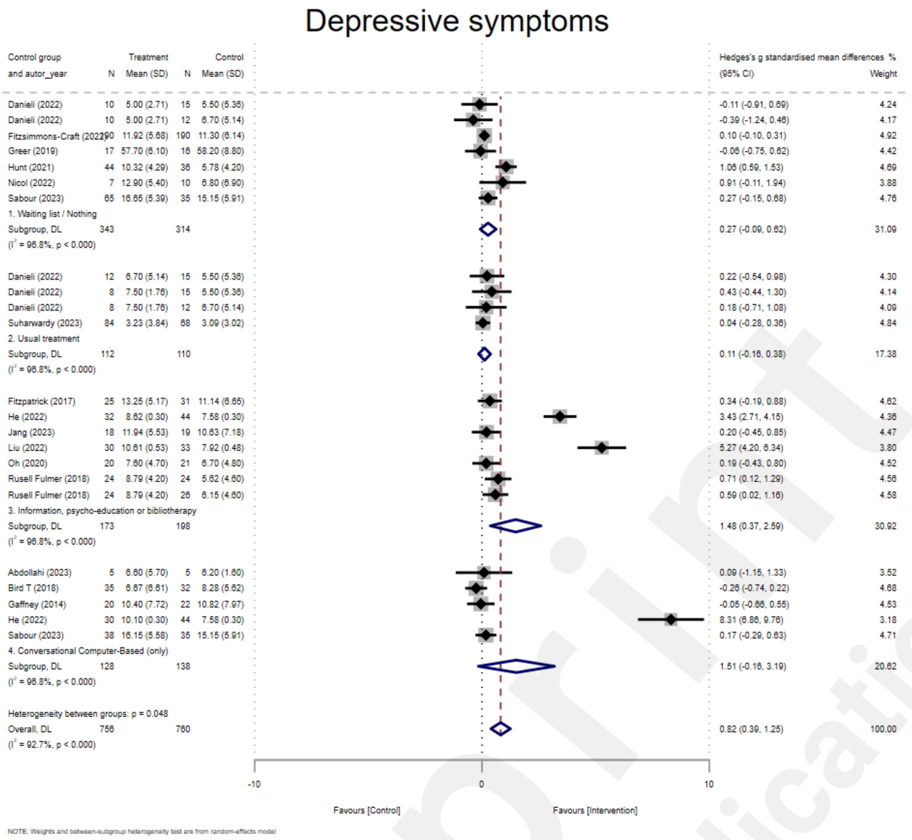

Figure 2. Forest plot for control conditions vs self-administered interventions based on natural language processing models to reduce depressive symptoms.

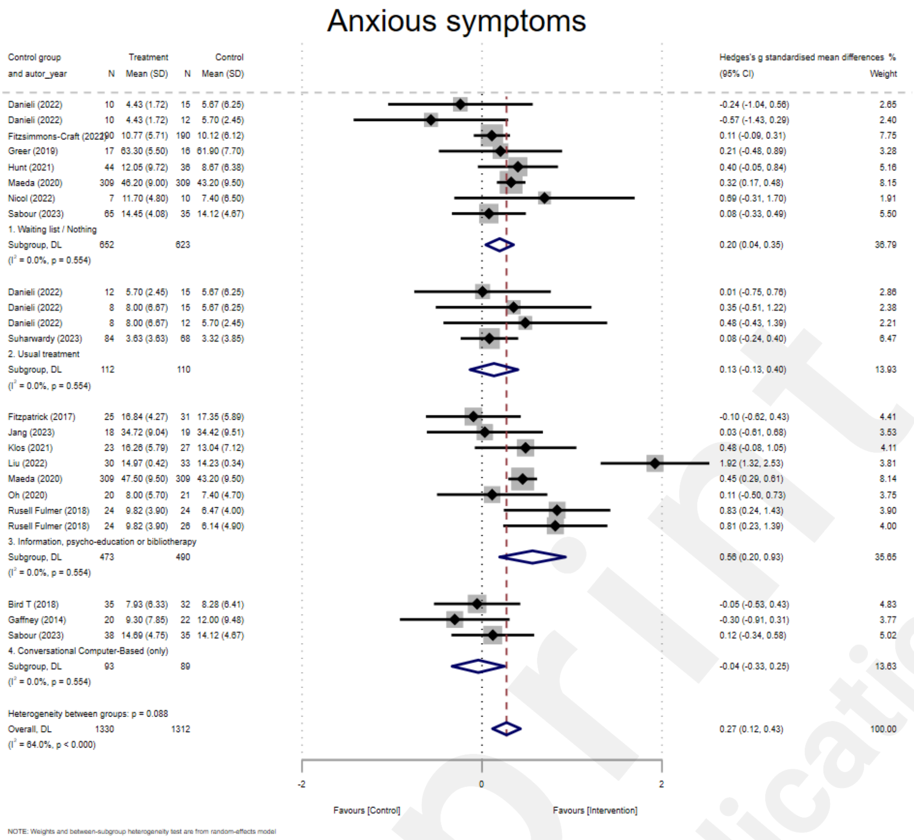

Figure 3. Forest plot for control conditions vs self-administered interventions based on natural language processing models to reduce anxious symptoms.

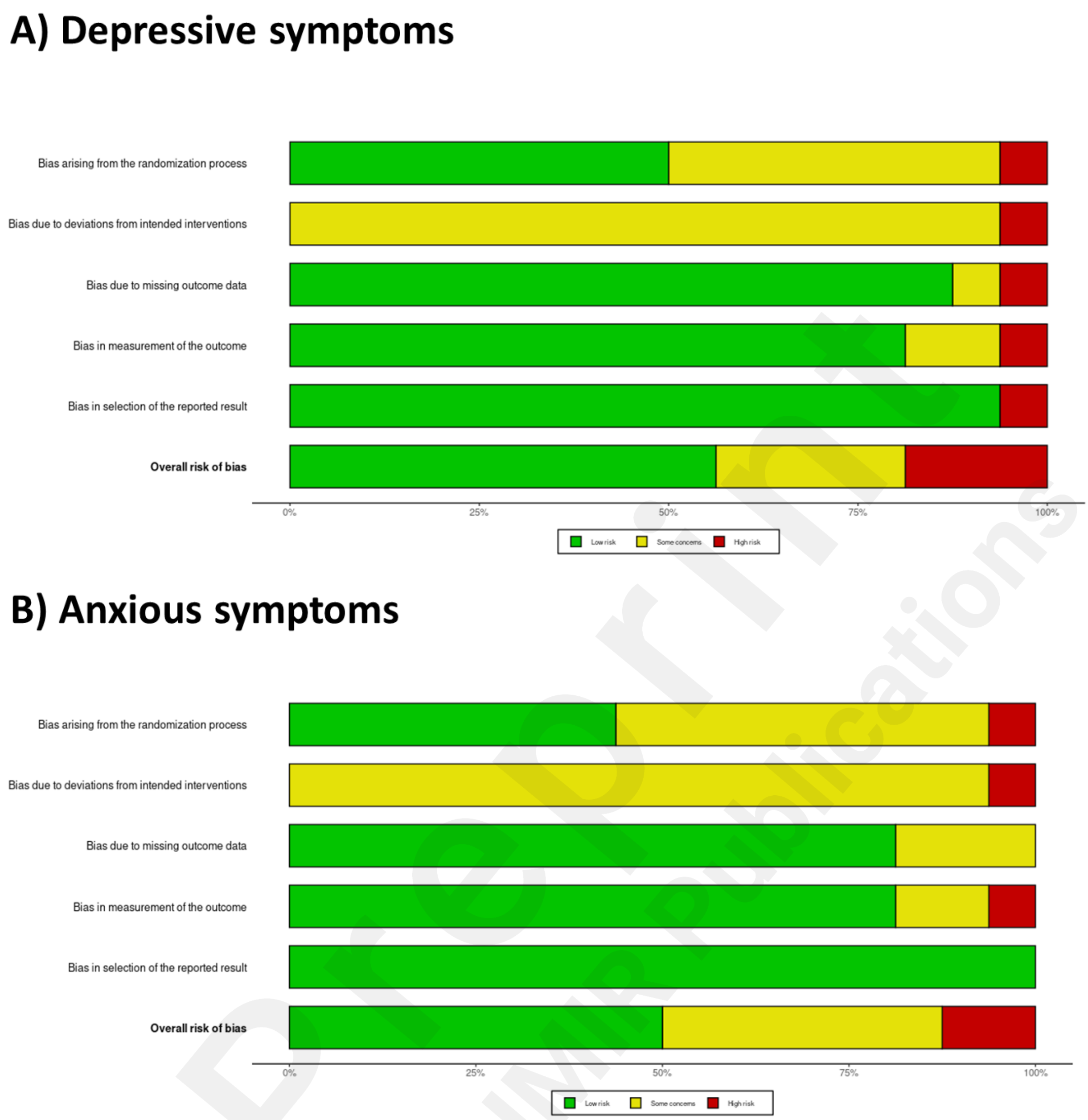

Figure 4. Risk of bias grouped for the outcome of depressive and anxious symptoms.

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

		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	26.3%	0	0.0%
	Parallel	14	73.7%	0	0.0%
Target population	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%
Depression was the main outcome	Main outcome	11	52.4%	2	9.5%
	Secondary outcome	7	33.3%	0	0.0%
	Not evaluated	1	4.8%	0	0.0%
Anxiety was the main outcome	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%
Interest conflict	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%
Study has 3 or more arms	No	14	66.7%	2	9.5%
	Yes	5	23.8%	0	0.0%
Control group*	Waiting list / Nothing	9	31.0%	0	0.0%
	Usual treatment	4	13.8%	0	0.0%
	Information, psycho-education or bibliotherapy	9	31.0%	0	0.0%
	Conversational Computer-Based	5	17.2%	0	0.0%
Type of NLP application*	Rule-based NLP	3	10.3%	0	0.0%
	AI-based NLP	20	69.0%	2	6.9%
	AI-based NLP combine with other tech intervention	1	3.4%	0	0.0%
	AI-based NLP combine with other no-tech intervention	3	10.3%	0	0.0%
Interventions focused on*	Depressive symptoms	8	27.6%	1	3.4%
	Anxious symptoms	7	24.1%	1	3.4%
	Other mental health problems	13	44.8%	2	6.9%
Interventions based on*	Cognitive behavior therapy	22	75.9%	2	6.9%
	Others	4	13.8%	0	0.0%
	Unclear	1	3.4%	0	0.0%
Scale used to measure depression*	Patient Health Questionnaire (PHQ-9, PHQ-8)	19	65.5%	2	6.9%
	Depression, Anxiety and Stress Scales (DASS-21)	2	6.9%	0	0.0%
	Other	3	10.3%	0	0.0%
Scale used to measure anxiety*	Generalized Anxiety Disorder (GAD-7)	16	55.2%	2	6.9%
	Depression, Anxiety and Stress Scales (DASS-21)	3	10.3%	0	0.0%
	Other	5	17.2%	5	17.2%

Note: *An intervention can have multiple control situations or interventions (n=29). CA = conversacional agent. AI = Artificial inteligence.

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

			n (measurements)	SMD (95% CI)	p	Heterogeneity (I^2)	Cochran's Q (p value)
Depressive symptoms	**By control group**	Waiting list / Nothing	6 (7)	0.267 [-0.085 to 0.620]	0.137	67.50%	0.005
		Usual treatment	2 (4)	0.111 [-0.155 to 0.378]	0.413	0.00%	0.847
		Information, psycho-education or bibliotherapy	6 (7)	**1.481 [0.368 to 2.594]**	**0.009**	95.20%	<0.001
		Conversational Computer-Based (only)	5 (5)	1.513 [-0.162 to 3.188]	0.077	96.80%	<0.001
	By intervention group	Rule-based NLP	1 (1)	-	-	-	-
		AI-based NLP	13 (18)	**1.059 [0.520 to 1.597]**	**<0.001**	94.20%	<0.001
		AI-based NLP combine with other tech intervention	1 (1)	0.086 [-1.154 to 1.327]	0.892	-	-
		AI-based NLP combine with other no-tech intervention	2 (3)	0.006 [-0.277 to 0.290]	0.964	0.00%	0.601
	By scale used	Patient Health Questionnaire (PHQ-9, PHQ-8)	11 (17)	**0.914 [0.417 to 1.410]**	**<0.001**	92.80%	<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**	24.20%	0.236
		Usual treatment	2 (4)	0.133 [-0.134 to 0.400]	0.329	0.00%	0.799
		Information, psycho-education or bibliotherapy	7 (8)	**0.561 [0.195 to 0.927]**	**0.003**	78.40%	<0.001
		Conversational Computer-Based (only)	3 (3)	-0.041 [-0.333 to 0.250]	0.782	0.00%	0.554
	By intervention group	Rule-based NLP	1 (3)	0.379 [0.268 to 0.489]	0.000	0.0%	0.369
		AI-based NLP	13 (17)	**0.302 [0.073 to 0.532]**	**0.010**	68.10%	<0.001
		AI-based NLP combine with other tech intervention	0 (0)	-	-	-	-
		AI-based NLP combine with other no-tech intervention	2 (3)	0.025 [-0.41 to 0.46]	0.910	30.70%	0.236
	By scale used	Generalized Anxiety Disorder (GAD-7)	9 (15)	**0.333 [0.074 to 0.592]**	**0.012**	71.70%	<0.001
		Depression, Anxiety and Stress Scales (DASS-21)	3 (3)	0.050 [-0.352 to 0.453]	0.806	47.30%	0.150

Note: n = Number of individual studies. Measurements = Number of measurements between comparators and intervention. CA = conversacional agent. AI = Artificial inteligence. SMD = Standardized Mean Difference. Bold values are significant. Our study only presents meta-analyses with at least three measurements.

Supplementary Files

Multimedia Appendixes

PRISMA 2020 Checklist.
URL: http://asset.jmir.pub/assets/1d2f5889f11b8034b205c32c80e81f3c.docx

Search strategy.
URL: http://asset.jmir.pub/assets/e938006a302b912b746c3644d2ba1bdd.docx

Excluded records.
URL: http://asset.jmir.pub/assets/c7ada47e14d9f4f66b418bb7d3d5e0ae.docx

Funnel plot by depressive and anxious symptoms.
URL: http://asset.jmir.pub/assets/8919b8e16f9c976499c08b0025b58993.docx

Risk of bias for individual studies for the outcome of depressive symptoms.
URL: http://asset.jmir.pub/assets/78e33ca084d09385a31f9f5dae1e78b1.docx

Risk of bias for individual studies for the outcome of anxious symptoms.
URL: http://asset.jmir.pub/assets/00700da5428868768af870a5c8a6c641.docx