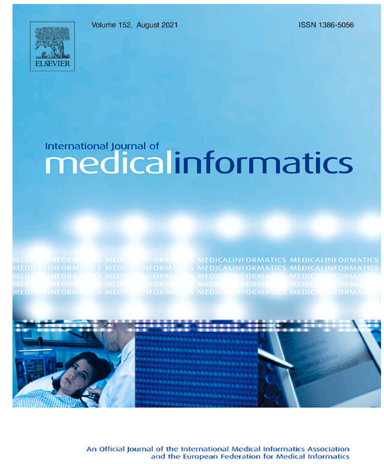
Journal Pre-proofs



Review article

Conversational Agents for depression screening: a systematic review

Iván Otero-González, Moisés R. Pacheco-Lorenzo, Manuel J. Fernández-Iglesias, Luis E. Anido-Rifón

PII: S1386-5056(23)00290-3
DOI: https://doi.org/10.1016/j.ijmedinf.2023.105272
Reference: IJB 105272

To appear in: *International Journal of Medical Informatics*

Received Date: 26 June 2023
Revised Date: 19 October 2023
Accepted Date: 23 October 2023

Please cite this article as: I. Otero-González, M.R. Pacheco-Lorenzo, M.J. Fernández-Iglesias, L.E. Anido-Rifón, Conversational Agents for depression screening: a systematic review, *International Journal of Medical Informatics* (2023), doi: https://doi.org/10.1016/j.ijmedinf.2023.105272

Conversational Agents for depression screening: a systematic review

Iván Otero-González[1], Moisés R. Pacheco-Lorenzo[1], Manuel J. Fernández-Iglesias[1], Luis E. Anido-Rifón[1].

[1] atlanTTic, University of Vigo, Vigo, Spain

Correspondence:

Iván Otero-González, atlanTTic, University of Vigo, Vigo, Spain.

Email: ivan.otero@det.uvigo.es

Abstract

Objective: This work explores the advances in conversational agents aimed at the detection of mental health disorders, and specifically the screening of depression. The focus is put on those based on voice interaction, but other approaches are also tackled, such as text-based interaction or embodied avatars.

Methods: PRISMA was selected as the systematic methodology for the analysis of existing literature, which was retrieved from Scopus, PubMed, IEEE Xplore, APA PsycINFO, Cochrane, and Web of Science. Relevant research addresses the detection of depression using conversational agents, and the selection criteria utilized include their effectiveness, usability, personalization, and psychometric properties.

Results: Of the 993 references initially retrieved, 36 were finally included in our work. The analysis of these studies allowed us to identify 30 conversational agents that claim to detect depression, specifically or in combination with other disorders such as anxiety or stress disorders. As a general approach, screening was implemented in the conversational agents taking as a reference standardized or psychometrically validated clinical tests, which were also utilized as a golden standard for their validation. The implementation of questionnaires such as Patient Health Questionnaire or the Beck Depression Inventory, which are used in 65% of the articles analyzed, stand out.

Conclusions: The usefulness of intelligent conversational agents allows screening to be administered to different types of profiles, such as patients (33% of relevant proposals) and caregivers (11%), although in many cases a target profile is not clearly of (66% of solutions analyzed). This study found 30 standalone conversational agents, but some proposals were explored that combine several approaches for a more enriching data acquisition. The interaction implemented in most relevant conversational agents is text-based, although the evolution is clearly towards voice integration, which in turns enhances their psychometric characteristics, as voice interaction is perceived as more natural and less invasive.

Keywords: depression, screening, conversational agents, Natural Language Processing, artificial intelligence, machine learning

1 - Introduction

Data provided by the World Health Organization in 2020 indicate that depression is a common illness worldwide, affecting an estimated 3.8% of the population; 5% in the case of adults and 5.7% among adults over 60 years of age. Worldwide, approximately 280 million people have depression [1] and in some developed countries, such as Spain, prevalence reaches 6.7% among 25- to 64-year-olds and 12.8% in the case of older adults over 65 years old. The expected prognosis is that by 2030 depression will be the leading cause of disability [2], [3]. This situation, if not conveniently addressed, could lead to suicidal behavior and the inevitable increase in mortality rates [4].

Recent literature identifies some relevant challenges in the context of depression screening [5]–[10], such as symptoms varying not only among individuals, but also at the time tests are performed; being typically associated with other disorders like anxiety or loneliness; the associated stigma and lack of social acceptance; the lack of personalized assistance, or the variability in clinical practice. These challenges drive research towards the integration of technology in the screening process (e.g., automated voice or video analysis), new approaches to personalized medicine, comprehensive mental health services and interdisciplinary collaboration.

This study aims to identify existing research for conversational agent[1] (CA)-mediated screening in the field of clinical depression. Therefore, the research question addressed is:

What conversational agents exist to detect depressive disorder and what are their characteristics?

To answer this question, an analysis of the state-of-the-art among the scientific community is carried out, which eventually will give us the possibility to explore and discover relevant lines of research and innovation within the field of neuropsychiatric disorders, focused on depression, using intelligent CAs.

1.1 - Characterization of a CA

Conversational agents are automated applications able to maintain a dialog with humans through an interactive interface, which can be text-based or voice-based [11].

To understand the conversational interface (CI) of a CA, Fig.1 illustrates the simple scheme of the processes that intervene internally during a machine - human dialogue, which in turn explains the internal workings of a CA with voice as an interactive component.

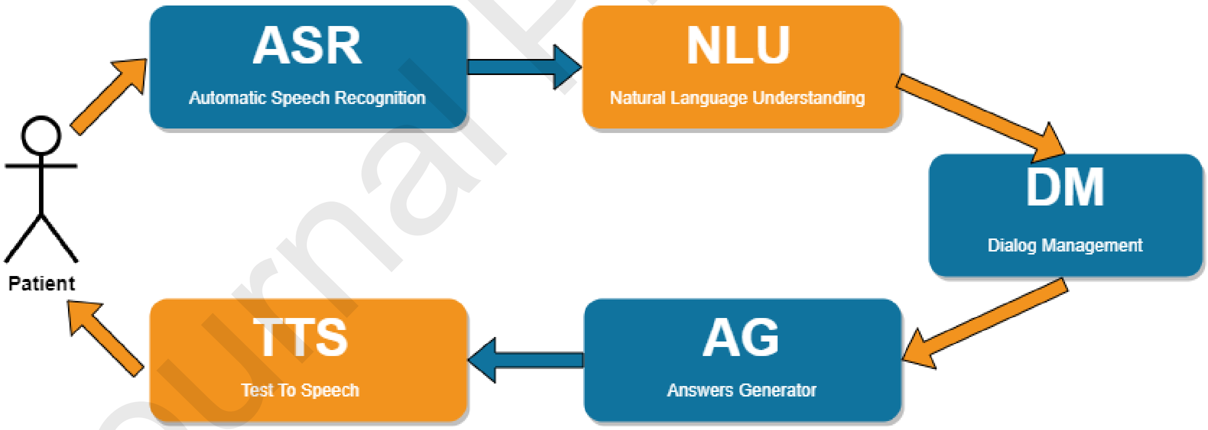


Figure 1: Data and process flows within a conversational agent.

1.2 - Motivation of a CA

The high demands in psychological attention in modern societies, the shortage of practitioners, and the dramatic impact of the COVID-19 pandemics, provided strong

[1] This and following abbreviations were added at the section 11.

motivation for further progress in the psychological healthcare field. According to many authors, a critical issue is therefore the screening of mental conditions, and specifically depressive disorders [12], [13], which would facilitate diagnosis, increase the perceived quality of care, and clearly differentiate among the various existing mental disorders and their urges. The benefits that CAs bring to the field of mental health are highlighted below.

- The perceived quality and ease of interaction among patients with mental disorders is greater with a CA than with a physician [14].
- A significant reduction of depressive symptoms during treatment is achieved in people with major depressive disorder [15].
- Efficiency in screening for depression, in terms of time and cost, compared to traditional clinical tests [16]–[18]. In addition, CAs are a promising approach in the case of patients who would not seek care because of stigma [14].
- Appropriateness of voice as a communication tool [19]. Voice interaction naturally overcomes the digital divide by means of a guided dialogue [20], which can also be perceived as an ally, a relationship, which benefits treatment outcomes [21]. Besides, voice interaction leverages naturalness [21].
- The configuration of CAs mimicking empathic interaction allows patients to feel socially connected, which can result in greater effectiveness [22].
- Contribution to partially overcome the shortage of healthcare personnel specialized in psychological treatment [23][24].
- Effects of CA-mediated interaction are comparable to traditional face-to-face psychotherapy [25].
- CAs can provide richer information from patients, as they can act as attention recipients rather than attention providers [12].
- The availability of CAs enables their introduction beyond clinical settings, helping to detect depression in isolation or in comorbidity, a clear example being the burnout effect, where prevention is a vital factor in avoiding suicides [26].

So far, no CA exhibits all these functionalities together, and when (some of them) are exhibited, they do so with varying degrees of effectiveness.

Delving into the beginnings of CAs, the first documented attempt occurred in 1966, with ELIZA, a software application able to reproduce a text-based patient-psychotherapist dialogue [27]. Since then, the research and quest for an application that, through dialogue, would support all or some of the processes of diagnosis, evaluation, monitoring, and treatment of depression continued to evolve.

Another evidence of the relevance that this topic and the role that mental health acquired in recent years is that we can identify related research initiatives within the framework of international projects such as Empathic or Gatekeeper2.

---

[2] empathic-project.eu, gatekeeper-project.eu

The aim of this paper is to provide a curated classification and analysis of existing CAs, identifying their maturity and psychometric properties, and their different application fields, to provide robust evidence for decision making, both to researchers and practitioners.

2 - Methodology

This study is based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [28]. Basically, PRISMA defines a protocol to guide the development of a systematic literature review and meta-analysis. This protocol is defined as a sequence of tasks covering the review process, from the preparation of the review and the selection of document sources to the actual meta-analysis of the selected papers (cf. Sect. 4.3).

2.1 - Search Strategy

Based on previously published systematic reviews available in the literature, and on their popularity among the scientific community, credibility, recognition, and availability, seven databases were selected for the search process, namely Scopus, PubMed, IEEE Xplore, APA PsycINFO, Cochrane, and Web of Science (WoS). These databases were considered the most appropriate in view of our field of study since they contain the most relevant contributions and cover the entire spectrum of our target topic.

The search process aimed to find proposals that (1) address conversational solutions, i.e., chatbots or CAs that rely on natural language, either spoken or written, as the interaction mechanism. Another selection criterion (2) was to include contributions aimed at or related to the study of the neuropsychiatric disorder of depression in isolation or in combination with other associated disorders or comorbidities such as anxiety or mild cognitive impairment, among others. Finally, (3) proposals should include an analysis of the contributed CA to evaluate the above problems or evaluate a sample of users with respect to these problems.

According to the above requirements for the search process, a query was designed consisting of the logical product of 3 main disjunctive terms expressing the above 3 requirements (i.e., all 3 requirements must be fulfilled for all retrieved documents). Each disjunctive term is composed of several alternative but related terms (i.e., linked with the inclusive OR operator), which together represent the main concept pursued by each term. One could argue whether more terms should be added to improve the outcome and thus obtain additional documents, but we wanted to keep the raw numbers of query results manageable while guaranteeing an acceptable recall rate. To do this, we tried to carefully select the terms that best matched each identified search requirement.

The final query is as follows:

(depres* OR dysthymia)

AND

("conversational agent" OR "virtual agent" OR "conversational assist*" OR "virtual assist*" OR chat* OR "speech agent" OR "speech assist*" OR "virtual spe*" OR "voice agent" OR "voice assist*" OR "virtual voice" OR "dialog agent" OR "dialog assist*" OR "virtual bot" OR "smart spe*" OR "smart voice" OR "smart dialog")

AND

(detect* OR screen* OR triag* OR dianos* OR filter*)

Quotation marks in multi-word query terms indicate that an exact match is sought to prevent search engines from including other combinations of the words contained in the query. In addition, asterisks are used as wildcards to search for documents that include any word with the same stem. No year of publication restrictions were applied.

Appendix A includes a list of the databases used in this research. The original query string was adapted to the specific syntax of each database while representing the same query in all cases. The number of documents initially retrieved (up to April 2023) is also included. After submitting the query to each selected database, all retrieved references were manually loaded into the Mendeley reference manager [29]. Finally, we performed an additional search using alternative methods guided by the authors' previous experience. First, we investigated other recently published literature reviews on the topic, checking whether the reviewed articles were also included in the results of our search process. Secondly, for the final selection of the articles in this review, all their references were retrieved and checked for inclusion, as well as the most recent articles citing them using the online tool available in ResearchGate [30]. In addition, reference lists of relevant articles and gray literature identified in those databases were also considered for selection.

2.2 - Eligibility Criteria

The analysis included documents published in the most prestigious journals and therefore within the first citation quartiles. Both articles in reference journals and communications to relevant congresses were considered for the study, given the growing interest of the scientific community in the use of CAs in the field of mental health, and more specifically in depression. However, some exclusion criteria were established to discard some articles such as those that:

(1) Were not related to the mental disorder of depression.

(2) Were not aimed at the detection, study of depression.

(3) Their analysis was not based on a rigorous methodology, as is the case of studies whose data were obtained from forums or social networks.

(4) The contributed device relied on other technologies such as Augmented Reality, and therefore could not be considered a CA.

(5) It was not possible to access the full article.

2.3 - Selection Process

Mendeley functionalities were utilized to merge all references from all databases queried [29]. Then, we applied an integrated matching tool to identify duplicates, which were discarded. In a first screening phase, we carried out an initial selection of papers based on their relevance to the research question posed. For this selection, we used the titles and abstracts of the papers to detect whether any of the five exclusion criteria in Sect. 2.3 were met. In cases where a decision was not straightforward, the article was marked as a candidate for the next phase. Next, in the eligibility phase, the full paper of all works that passed the screening phase was thoroughly studied. Based on the content of the full paper, the main author scored each article with a weight between 0 (not relevant at all) and 2 (fully relevant to the topic). Articles labelled with a score of 2 were relevant to answering the research question and were therefore selected for further analysis. Those that received a score of 1 were reviewed in a second round, in which a final weight of 0 or 2 was proposed. Appendix B includes a list of the papers that were included in review. This last decision has been made by the rest of the authors individually, assigning a weight they considered and reaching a consensus among the authors.

2.4 - How data was classified

Eventually, 36 articles were selected to be thoroughly analyzed by the authors, with the objective of extracting and classifying information on (1) the specific aim(s) of each contribution; (2) the user group targeted by the conversational system and, in particular, their age, but also gender and other characteristics of the target audience; (3) the actual CA technological instrument involved in the implementation or application of the standard clinical procedure used to detect depression; (4) the validation protocol performed, including the standardized or psychometrically validated clinical tests used, as well as the number of actual users involved; (5) the mode of interaction by which the dialogue between the user and the CA is established; (6) the type of document and the year of publication.

3 - Results

The main objective of this work is to identify in existing literature relevant contributions addressing the detection of depression by means of conversational assistants, as reflected in the research question posed.

3.1 - Selection Process

As depicted in Fig. 2, which graphically shows the identification and screening steps of the PRISMA methodology, 986 papers were compiled from the 6 databases queried, which were analyzed by the authors to identify and characterize such contributions. In addition, the analysis of backward and forward references in the original documents provided 7 additional relevant references. After duplicate removal, the authors examined the titles and abstracts of 754 documents. The outcome of this process was a collection of 81 articles potentially eligible to answer the research question.

3.2 - Search Strategy

The full text of each of the 81 selected papers was checked against the exclusion criteria enumerated above. The papers were classified into three groups, namely those that were considered relevant and were directly included in the review for further analysis; those that were discarded as not relevant; and those whose relevance was borderline or unclear, which were reviewed by an additional researcher. Finally, the relevance of the articles in the third group was resolved by consensus among the researchers who reviewed the article.

Of the 81 articles that entered the eligibility phase, 34 were marked as relevant after reading the full text, while 47 were marked as unclear. Of the latter, after discussion among the authors who participated in the process, 2 were finally selected for analysis. The remaining 45 were discarded (17 borderline and 28 not relevant). The main reason for exclusion (17 articles) was the fact that the solution discussed was not conversational, that is, the user could not interact in natural language, either spoken or written, with the system.

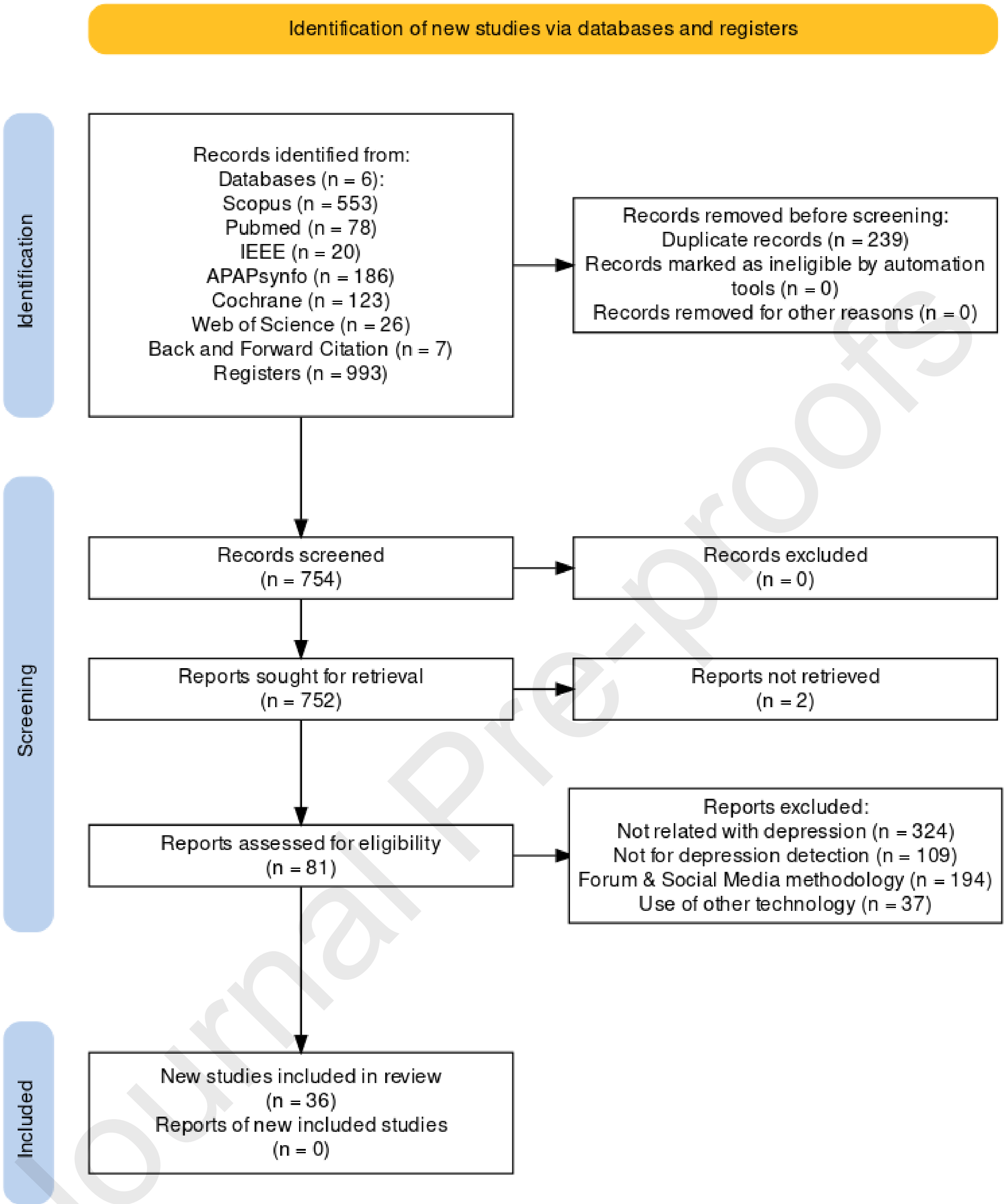


Figure 2. PRISMA Flowchart.

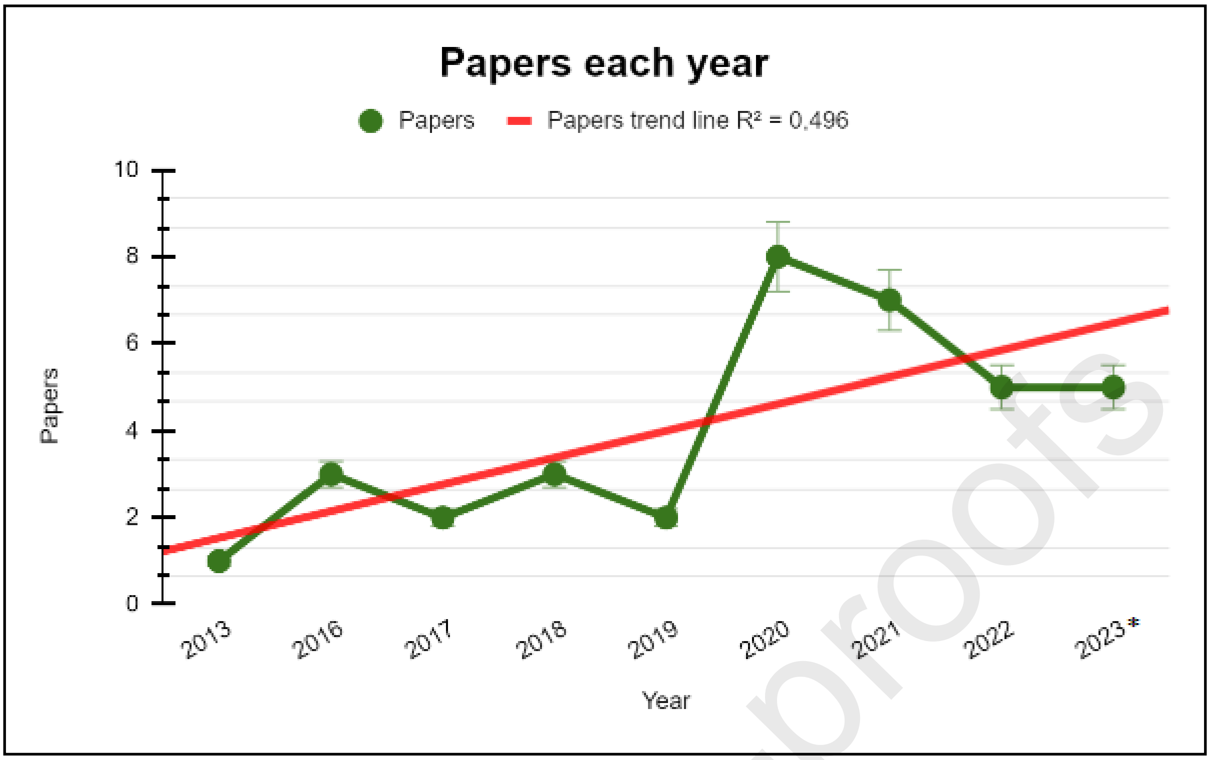


Figure 3: Number of papers per year. *Only up to and including April. Trend line calculated according to the least squares method.

The scientific relevance of this topic is confirmed by the fact that most relevant papers were published in recent years [31], as illustrated in Fig. 3. It should be noted that, despite the last paper included in the study being published in April 2023, this year's publications already exceeded the number of publications for the whole of 2022. Figure 3 also shows a positive trend in the number of publications on screening for depression using CA. As discussed below, in many cases these publications address the performance and reliability of CA agents for the detection of depression.

Table 1 briefly summarizes the results of the meta-analysis of each of the articles that passed the eligibility phase. This table is intended to serve as a tool to identify the key concepts of the reviewed contributions, and to provide a starting point for researchers interested in the application of CA for depression screening.

In all the studies, depression is the main condition addressed, but in many cases, it is treated in relation to other conditions or co-morbidities, or because of other disorders such as chronic illnesses, postpartum situations, as a probable cause of a subsequent episode of anxiety or in the most dramatic case of suicide. The latter case is really alarming among young people, both in terms of frequency and social impact [32]–[34].

4 - Discussion

This section analyzes in detail the relevant characteristics of the contributions retrieved. For this, we perform a meta-analysis of each contribution (Sect. 4.1), where we split the analysis into four different dimensions: clinical test which is based on, types of interaction (i.e., communication channel) used, target audience of the CA and quality of the validation performed. Additionally, the limitations encountered that prevent a more precise analysis of the application and implementation of the CAs in clinical settings are also identified (Sect. 4.2).

4.1 – Critical appraisal

The analysis of the selected works was carried out with the aim of (I) understanding the feasibility, efficacy, perception and satisfaction on the part of the patient and/or caregiver of the different proposals; (II) studying the use of CA as a tool to implement psychometric validation tests and their comparison with the results of the traditional medical detection procedures; and (III) drawing conclusions from pilots or essays in which the detection of depression or the study of the evolution of depression is carried out, both in isolation and related to other diseases, such as chronic diseases.

Table 1: List of selected articles ordered by year and title.

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
A chatbot-based intervention with ELME to improve stress and health-related parameters in a stressed sample: Study protocol of a randomized controlled trial [35].	Journal Article	2023	130 participants	Unknown	Adult Population	male & female	BPQ, GAD-7, ERQ	Germany	English	ELME	SD[C] Web-based	Yes	NA[D]	Chat
Early detection of depression using a conversational AI bot: A non-clinical trial [36].	Journal Article	2023	50 participants	average age of 34.7 years	Patients	male & female	SIGH-D, IDS-C	Australia	English	DEPRA	SD DialogFlow	Yes	NA	Chat
Supporting mental health self-care discovery through a chatbot [37].	Journal Article	2023	80 participants	higher education students	Population	male & female	Unknown	Finlandia	English	BotStar	SD Web-based	Not	NA	Chat

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
User experience with a parenting chatbot micro intervention [38].	Journal Article	2023	170 participants	between the age of 2 to 11 years	Population	male & female	Unknown	Argentina	Spanish	Mila	SD TESS, X2AI	Not	NA	Voice
Vicky Bot, a chatbot for anxiety-depressive symptoms and work-related burnout in primary care and healthcare professionals: development, feasibility, and potential effectiveness studies [39].	Journal Article	2023	40 participants	between 18 and 65 years of age	Carer	male & female	GAD-7, PHQ-9, MBI	Barcelona Spain	Spanish	Vicky Bot	SD APP[H]	Yes	NA	Chat / Voice
Conversational AI Therapist for Daily Function Screening in Home Environments [40].	Conf.[B] paper	2022	7 subjects	aged between 21 and 30	Population	male & female	DLA-20, SF-36	Columbia University, EEUU	English	Alexa	SD APP[H]	Not	Yes	Voice

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
Doctor-Bot: AI Powered Conversational Chatbot for Delivering E-Health [41].	Journal Article	2022	Unknown	Unknown	Population	male & female	Unknown	Nashik (India)	English	Doctor-Bot	SD Python	Not	NA	Chat
Evaluating the Therapeutic Alliance with a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study [42].	Journal Article	2022	G[A] 1, 1205 participants G[A] 2, 226 participants	Unknown	Population	male & female	PHQ-4	NY, EEUU	English	Wysa	SD APP[H]	Yes	NA	Chat
HIGEA: An Intelligent Conversational Agent to Detect Caregiver Burden [43].	Journal Article	2022	10 participants	aged 20-59 years and 1 of 15 years	Carer	male & female	Zarit Test	Granada, Spain	English	HIGEA	SD APP[H], DialogFlow	Yes	NA	Chat

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
Validity of Chatbot Use for Mental Health Assessment: Experimental Study [44].	Journal Article	2022	146 young adults	mean age 24	Population	male & female	Kessler PD PAU	Germany	English	Undefined	SD Microsoft Bot Framework	Not	NA	Chat
Clinical advice by voice assistants on postpartum depression: Cross-sectional investigation using Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana [45].	Journal Article	2021	Unknown	Unknown	Carer	female	Fisher test	Columbus, EEUU	English	Alexa, Siri, Google Assistant, Cortana	SD Alexa, Siri, Cortan, Google Assistant	Not	NA	Four Voice Assistants
DEPRA: An Early Depression Detection Analysis Chatbot [46].	Conf. [B] paper	2021	40 participants	between 18 and 80 years	Patients	male & female	SIGH-D, IDS-C	Victoria University, Melbourne, Australia	English	DEPRA	SD DialogFlow	Yes	NA	Chat

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
Design and Evaluation of Virtual Human Mediated Tasks for Assessment of Depression and Anxiety [47].	Conf. [B] paper	2021	56 participants	aged between 18 and 45	Patients	male & female	PHQ-9, GAD	University of Nottingham UK	English	VHT	SD VALUSKA platform	Yes	Yes	Chat / Video
Designing Depression Screening Chatbots [48].	Journal Article	2021	10 participants	Unknown	Patients	male & female	PHQ-9	Finlandia	English	IGOR	SD DialogFlow, NodeJS, Firebase	Not	NA	Chat
Pandora's Bot: Insights from the Syntax and Semantics of Suicide Notes [49].	Journal Article	2021	Unknown	Unknown	Population	male & female	Unknown	Australia	English	Pandora's Bot	SD Web-based	Not	NA	Chat / Voice
Screening accuracy of a 14-day smartphone ambulatory assessment of	Journal Article	2021	178 participants	average age of 28.56	Patients	male & female	PHQ-9	Berlin Germany	English	Moodpath	SD APP[H]	Yes	NA	Chat

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
depression symptoms and mood dynamics in a general population sample: Comparison with the PHQ-9 depression screening [50].														
Smart Homes as Enablers for Depression Pre-Diagnosis Using PHQ-9 on HMI through Fuzzy Logic Decision System [51].	Journal Article	2021	Unknown	Unknown	Population	male & female	PHQ-9	Monterrey, México	Spanish	Alexa	SD Alexa Skill	Yes	NA	Voice / Chat / Video
116: effect of an automated conversational agent on postpartum mental health: a randomized,	Journal Article	2020	139 women	postpartum women	Patients	female	EPDS, PHQ-9, GAD-7	California, EEUU	English	Undefined	SD APP[H]	Yes	NA	Chat

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
controlled trial [52].														
74: acceptability of postnatal mood management through a smartphone-based automated conversational agent [53].	Journal Article	2020	96 participants	postpartum women	Patients	female	Unknown	California, EEUU	English	Undefined	NA	Not	NA	Chat
Artificial Intelligence Chatbot for Depression: Descriptive Study of Usage [54].	Journal Article	2020	354 users	Unknown	Population	male & female	PHQ-9	EEUU	English	Tess	SD APP[H]	Yes	NA	Chat
Co-designing Strategies to Provide Telecare Through an Intelligent Assistant for	Conf. [B] paper	2020	Student, Nurse, Professor	student, nurse, professor	Carer	male & female	GDS, MMSE, RMBPC	Fluminense Brazil	Brazilian	MarIANA	SD APP[H]	Yes	NA	Chat / Voice

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
Caregivers of Elderly Individuals [55].														
Le chatbot, outil d'accompagnement thérapeutique de la dépression chez les patientes atteintes d'un cancer du sein [56].	Journal Article	2020	74 participants	between 26 and 78 years	Patients	male & female	PHQ-9	Montpellier, France	Francaise	Vik	SD APP[H]	Not	NA	Chat / Voice
Motivating PhD candidates with depression symptoms to complete thoughts-strengthening exercises via a conversational agent [57].	Journal Article	2020	174 participants	Students	PhD Students	male & female	Unknown	The Netherlands	English	RASA	SD APP[H], RASA	Yes	NA	Chat / Voice

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
Perla: un Agente Conversacional para la Detección de Depresión en Ecosistemas Digitales. Diseño, Implementación y Validación [58].	Conf. [B] paper	2020	108 participants	average of 37,21 years	Population	male & female	PHQ-9, BDI	Madrid, Spain	Spanish	PERLA	SD DialogFlow, Kommunicate	Yes	NA	Chat
Study of a smartphone-delivered, therapist-supported mindfulness-based therapy program for depression in Finnish university students [59].	Journal Article	2020	120 patients	Students	Patients	male & female	PHQ-9	San Francisco, EEUU	English	ASCEND	SD APP[H]	Yes	NA	Chat
Vik: A Chatbot to Support Patients with Chronic Diseases [60].	Journal Article	2020	140 patients	over 42 years	Patients	female	EORTC QLQ-INFO25	Montpellier, France	Francaise	Vik	SD APP[H]	Not	NA	Chat

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour [61].	Journal Article	2019	12 participants	Unknown	Patients	male & female	SAD PERSONS / RSIS / HAM-A	Montebello, México	English	HelPath	SD APP[H]	Yes	NA	synthetic voices
NEMSI: A Multimodal Dialog System for Screening of Neurological or Mental Conditions [62].	Conf. [B] paper	2019	Unknown	Unknown	Population	male & female	Unknown	India	English	NEMSI	SD APP[H]	Not	NA	Voice / Chat / Video
Depression Detection from Short Utterances via Diverse Smartphones in Natural	Conf. [B] paper	2018	887 participants	Unknown	Population	male & female	PHQ-9	Sydney, Australia	English	Sonde Health	SD APP[H] VoiceBOX, ope Smile	Not	NA	Voice

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
Environmental Conditions [63].														
Digital Psychiatry - Curbing Depression using Therapy Chatbot and Depression Analysis [64].	Conf. [B] paper	2018	Unknown	Unknown	Population	male & female	Unknown	Hamidpur, Delhi India	English	Woebot	SD Python	Yes	NA	Chat
Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial [65].	Journal Article	2018	74 participants	average age of 22.9 years	Students	male & female	PHQ-9, GAD-7, PANAS	EEUU	English	Tess	SD TESS	Yes	NA	Chat
Screening for risky behaviour and mental health in young people: the	Journal Article	2017	young patients	Adolescents	Students	male & female	PHQ-9, GAD-7	Auckland, New Zealand	English, Maori	YouthCHAT	NA	Yes	NA	Chat

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
YouthCHAT programme [66].														
Virtual human as a new diagnostic tool, a proof-of-concept study in the field of major depressive disorders [67].	Journal Article	2017	179 outpatients	mean age 46.5 years	Patients	male & female	BDI-II	Bordeaux France	Francaise	SANPSY	NA	Yes	Yes	Voice / Video
Pilot randomized controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression [68].	Journal Article	2016	21 participants	age 18-64	Patients	male & female	Beck II, PHQ-9	UK	English	Help4Mood	SD Desktop-based	Yes	NA	Chat
Usability and acceptability assessment of an empathic virtual agent to prevent	Journal Article	2016	60 participants	students, professors & researchers	University Students	male & female	PHQ-9	Valencia, Spain	Spanish	PrevenDep	SD Desktop-based	Yes	NA	Voice / Video

Title	Type	Year	Validation	Age	Target Audience	Genders	Gold standard	Place data acquisition	CA Language	CA	Tech[E]	Emotion module	Bio[G]	Interaction
major depression [69].														
Improving symptom communication through personal digital assistants: the CHAT (Communicating Health Assisted by Technology) project [70].	Journal Article	2013	50 participants	Mean age of 51.0 years	Patients	female	BPI—SF, CES-D, HRQoL	Kansas, EEUU	English	Undefined	NA	Yes	NA	Chat

Notes: [A]Group, [B]Conference, [C]Self-Developed (SD), [D]Not Available (NA), [E]Technology (Tech), [G]Biomarker (Bio), [H]Smartphone APP (APP)

4.1.1 - CA implementation of existing clinical tests

Table 2 provides a classification of the CAs identified according to the clinical tests used as a basis for their implementation. A common feature of patients with depression is a depressed mood, along with mood swings as well as some level of cognitive loss. Some proposals focus their exploration on samples of patients who have some kind of previous disorder such as cancer, stress disorder or postpartum syndrome [55], [64], [72], [74], [78], [88], which may act as triggers for depressive disorder. In these cases, additional tests focused on those disorders are also considered in their design and validation (e.g. [EORTC QLQ-INFO25, GDS, HRQoL, MMSE, RMBPC])[3]

There are also works addressing patients already diagnosed with depression. In these cases, CAs are introduced to study patients' evolution, for whom depression may act as a trigger or initial state of other conditions. Consequently, existing clinical tests addressing depression evolution are used as a gold standard. This is the case of [SIGH-D, IDS-C, Zarit Test, SAD PERSONS, RSIS, HAM-A, MBI, PANAS][1] [35], [36], [48]–[52], [54], [56]–[59], [37], [61], [63]–[69], [38], [39], [42]–[44], [46], [47]

Table 2: Clinical tests integrated into CAs.

Disorder	Gold Standard	Items	References
Illness, previous disorders associated with depression.			
ECNT / Cáncer	BPI—SF, CES-D, HRQoL, GDS, MMSE, RMBPC, EORTC QLQ-INFO25	3	[55], [60], [70]
Stress	BPQ, GAD-7, ERQ	1	[35]
Postpartum	Fisher test	2	[45], [53]
Post disorder, an event associated with depression.			

[3] Acronyms for validation tests are explained in the list of abbreviations for Psychometrically Validated Clinical Tests.

ADHD	BDI-II, PHQ 9, SIGH-D, IDS-C, Zarit Test	20	[35], [36], [54], [56]–[59], [63], [64], [66]–[68], [37], [69], [38], [42], [43], [46], [48], [50], [51]
Suicide	GDS, MMSE, RMBPC, Kessler PD, PAU, SAD PERSONS, RSIS, HAM-A	3	[44], [49], [61]
Anxiety	EPDS, PHQ-9, GAD-7, MBI, PANAS	4	[39], [47], [52], [65]
Association physical state to depression.			
Well being	DLA-20, SF-36	1	[40]
Others	Unknown	2	[41], [62]

This second group addressing depression exclusively is of particular concern, representing 75% of the results. Among them, depression is identified in 55.6% of the cases as a trigger for attention deficit hyperactivity disorder (ADHD). This may be because depression, in addition to apathy and disregard about self-image, has an impact on the direct capacity for attention and therefore a limited effort to try to retain information. Moreover, its relevance is even greater, as this disorder is directly related to the younger population, a group where suicide ideation (8.3% of papers) and anxiety (11.1%) are also relevant.

Finally, three articles address depression along physical wellbeing [40] or independently of their nature of pre- or post-depressive conditions [41], [62].

The 82.35% of studies are focused on patients who were already diagnosed with depression, which can be considered evidence of the research community paying less attention to prevention and care. However, prevention and mental health education are essential to address depression at its roots. Greater investment in research is required to promote awareness, sensibilization and early identification of risk factors. This will allow for a more holistic and balanced approach to addressing depression and other mental disorders.

4.1.2 - Interaction in the CA

With respect to the type of interaction implemented by the CA (cf. Table 3), 58.3% of all contributions rely solely on text interaction at the time when the study was performed 58,3% [35], [36], [52]–[54], [58], [59], [64], [65], [68], [70], [37], [41]–[44], [46], [48], [50], while 13,9% of them use exclusively voiced dialogue [38], [40], [45], [61], [63] The rest of CAs surveyed (27.8%) implement both text and voiced interactions [36], [39], [69], [47], [49], [51], [55]–[57], [60], [67].

In the case of text interaction, in most cases the corresponding dialogue engine was developed ad-hoc, in some cases by the researchers themselves. Besides, in many cases they correspond to pioneering contributions to computerized dialogue generation in this field of study. In addition, the higher number of papers using text interaction (58.3%), is because it was the first technology available to implement computerized dialogue.

On the other side, voiced CAs are typically based on centralized and cloud-based voice engines, such as Alexa, Siri, Cortana, or Google Assistant. Note that, as depicted in Fig. 1, voiced interaction requires additional processing and consequently additional computational resources, as the processing needs are much higher when integrating the whole process in Fig. 1. In other words, if in chat-only interfaces the NLU, DM and AG steps are sufficient, in voice interfaces these steps must be integrated with ASR and TTS. This means that more computing resources are needed, as well as a flexible and scalable architecture that allows the introduction of new machine learning techniques. Note also that there is a trend towards the standardization of output and input formats with the integration of both text and voice, in line with the popularization of commercial devices implementing voice recognition and generation, such as smart speakers. These new devices paved the way for the introduction of voiced interaction in a broad range of situations, including health application and more specifically mental health study and detection.

The interaction module, either text or voice-based, provides an automated entry point for participants' responses. However, in all cases the authors acknowledge the need to collect additional evidence during the CA sessions, with the participation of health professionals. These include, for example, facial expressions or bodily reactions that will contribute to a better study of the patient and therefore to a better diagnosis. In addition, the anamnesis[4] procedures will contribute to obtain a comprehensive patients' understanding that allows for a better diagnosis.

Table 3: Grouping of CAs by type of interactive dialog.

CA	Interaction	Items	References

[4] Process of clinical examination that is executed through the interrogation to personally identify the individual, learn about his current ailments, obtain a retrospective of him, and determine the relevant family, environmental and personal elements.

Isolated interaction			
ASCEND, BotStar, DEPRA, Doctor-Bot, ELME, Help4Mood, HIGEA, IGOR, Moodpath, PERLA, Tess, Wysa, Woebot, YouthCHAT [5]	Chat	20	[35], [36], [52]–[54], [58], [59], [64]–[66], [68], [70], [37], [41]–[44], [46], [48], [50]
Alexa, Cortana, Google Assistant, HelPath, Mila, Siri	Voice	5	[38], [40], [45], [61], [63]
Combined interaction			
MarIANA, Pandora's Bot, RASA, Sonde Health, Vicky Bot, Vik	Chat / Voice	6	[39], [49], [55]–[57], [60]
VHT	Chat / Video	1	[47]
SANPSY	Voice / Video	1	[67]
Alexa & HMI, NEMSI, PrevenDep	Voice / Chat / Video	3	[51], [62], [69]

[3]A short description of each CA is provided in Appendix C

Time constraints and extended timelines typical of health research projects, when compared to the rapid pace of evolution of smart agent technologies, can lead to already obsolete software solutions. It is imperative to develop more adaptable methodologies and protocols to guarantee the long-term impact of technologies on mental health. For example, typical three-phase clinical essays that would take years were never considered in any work surveyed, and therefore clinical-grade validation. On the other hand, relying on off-the-shelf technologies and standardized devices and services to develop new health applications, can

[5] CA details are provided in Appendix C.

help to fill this gap, as standardization would ensure that advancements remain effective and relevant over time.

4.1.3 - Target audience and gender.

Table 3 summarized the contributions surveyed in relation to their target audience. The papers explored show that the vast majority (91.3%) consider both male and female subjects, with very little percentage variation between genres. Studies addressing postpartum depression (8.3%) focus only on female subjects [52], [53]. However, other common conditions associated with women and depression (e.g., postmenopausal or miscarriage depression) are not addressed yet, which may be an interesting niche for future proposals. None of the studies analyzed address male subjects exclusively.

In the case of contributions addressing both male and female individuals, a further classification can be identified according to target audience. 38.9% of studies are targeted to the general population [35], [37], [58], [62]–[64], [38], [40]–[42], [44], [49], [51], [54], whereas 33.3% are aimed to depression patients [36], [46], [68], [70], [47], [48], [50], [56], [59]–[61], [67], students (11.1%) [57], [65], [66], [69] or caregivers (8.3%) [55], [43], [39]. The early detection of depression in caregivers, and depression follow-up, are key due to the role that caregivers play from a public health standpoint, and its social impact.

Table 4: Grouping of studies by target audience and genre.

Genders	Target Audiences	Items	References
male & female			
	Population	14	[35], [37], [58], [62]–[64], [38], [40]–[42], [44], [49], [51], [54]
	Patients	12	[36], [46], [68], [70], [47], [48], [50], [56], [59]–[61], [67]
	Students	4	[57], [65], [66], [69]
	Carer	3	[39], [43], [55]

female			
	Carer	1	[45]
	Patients	2	[52], [53]

Although detection of mental disorders and depression using conversational agents is a promising area of research, there is a lack of studies that address these technologies early and differentiated based on gender. While some resources explored these questions at specific life stages and demographic groups, there is a lack of studies focusing on early care or on identifying gender biases, when these two aspects are key to understanding the impact and evolution of mental illnesses [71]. These studies would contribute to a more personalized and effective approach in the identification of mental disorders, and more specifically depression.

4.1.4 - CA validation

Validation is addressed in most of the contributions analyzed (94.4%, cf. Table 2). In most cases, validation relies on already validated clinical tests as the gold standard, both by the scientific community and clinical practice. In the case of CAs designed to detect depression, the use of classical questionnaires such as BDI-II, PHQ 9, SIGH-D, IDS-C or GDS tests stands out (66% of the articles) compared to 34.4% that use specific-purpose tests such as RSIS for detecting signs of suicide [61], MBI for detecting burnout syndrome symptoms [39], PAU aimed at detecting alcohol addiction [44]. Table 5 below shows the distribution of these tests, together with the number of participants in each of the studies reviewed.

Table 5: Grouping of studies by number of participants.

Number of participants	Gold Standard (Nº)[A]	CA	Items
< 41	BDI-2, Beck II, DLA-20, GAD-7, GDS, HAM-A, IDS-C, MBI, MMSE, PHQ-9 (3), RMBPC, RSIS, SAD PERSONS, SF-36, SIGH-D, Zarit Test,	Alexa, DEPRA, HelPath, Help4Mood, HIGEA, IGOR, MarIANA, Vickybot	[39], [40], [43], [46], [48], [55], [61], [68]

< 100	BPI—SF, CES-D, FSI, GAD-7 (2), HRQoL, IDS-C, PANAS, PHQ-9 (4), SIGH-D,	BotStar, DEPRA, PrevenDep, Tess, VHT, Vik	[36], [37], [47], [53], [56], [65], [69], [70]
< 200	BDI- II (2), BPQ, EORTC QLQ-INFO25, EPDS, EORTC QLQ-INFO25, GAD-7 (2), Kessler-18, PAU, PHQ-2, PHQ-9 (4), Test-3	ASCEND, ELME, Mila, Moodpath, PERLA, RASA, SANPSY, Vik	[35], [38], [45], [50], [52], [57]–[60], [67]
200 +	PHQ-4, PHQ-9 (2)	Sonde Health, Tess, Wysa	[42], [54], [63]
Not defined	GAD-7, Fisher exact test, PHQ-9 (2)	Alexa, Cortana, Google Assistant, Doctor-Bot, NEMSI, Pandora's Bot, Siri, Woebot, YouthCHAT	[41], [45], [49], [51], [62], [64], [66]

(Nº) ᴬ , indicates the number of test interactions used by the different CAs.

PHQ-9 is utilized regardless of the number of participants, which indicates that it is a well-established test and therefore a reference for its application in future studies. The next two most significant tests used are GAD-7 and BDI-II although involving samples smaller than 200 participants ([52] involving 139 participants and [67] 179 participants). PHQ-9 and BDI-II are used in 65% of the studies surveyed.

Restrictions on the inclusion criteria for validation proposed in some studies [68], [39] are strict, which makes obtaining large samples a difficult endeavour. This is the case of [39], [68] where they excluded participants with criteria such as (1) initial BDI-2 lower than 10 or higher than 30; (2) significant thoughts of self-harm or suicide; (3) history of recent self-harm, (4) substance abuse, (5) bipolar disorder, and (5) currently receiving specialized psychological therapy, among others. These requirements are reflected in Table 5, where samples larger than 200 participants represent only 8.3% [54], [63], [42], compared to samples involving less than 200 individuals, which represent 72.2% of the studies.

Finally, some studies involved a limited number of participants (i.e., less than 41 individuals, 22.2% of the articles analyzed). [55] discussing the MarIANA CA is tested with just 2 researchers and 2 nurses.

Although it is not detailed in the technical specifications of most CAs, it can be inferred that one of the limitations of the surveyed works is the lack of scalability and the ability to store critical data. Again (cf. Sect. 4.1.2 above), the use of off-the-shelf devices and standardized solutions would facilitate the development of open and more scalable ecosystems, especially in the face of analytical processing of data that directly influences the provision of more comprehensive and personalized care.

We can find in the literature previous systematic reviews related to the introduction of conversational agents in health-related settings, and more specifically in the field of mental conditions (cf. Appendix D). In most cases, these studies do not focus on depression, but deal with a broad spectrum of conditions [17], [20], [21], [31], [72]–[77]. In the only case that a previous study focuses on depression [78], screening or diagnosis is not addressed, but the effectiveness of chatbot-delivered psychotherapy.

5 - Limitations

This work is focused on the detection of depression by applying the technology provided by CAs. As a general remark, it is important to make it clear that not all the CAs developed to detect depression were included in this survey, but only those in which the same validation process is reported according to basic standards (e.g., comparison with a golden standard generally accepted by the health community). However, despite the specificity of the studies, there are some notorious omissions with respect to the information provided by them. For example, when discussing the validated clinical tests used as foundation or golden standard, the date of administration is not clearly stated.

Besides, in many cases it was not possible to identify in surveyed contributions, besides text or oral interactions, other information about the subject's interaction session that would be instrumental for the detection or diagnosis process (e.g., gestures, facial expressions, comments beyond the answers to the questions included in the questionnaires). In addition, relevant demographic information such as the characteristics of the place of residence (e.g., rural/urban, institutionalized, living alone, etc.), socialization details (e.g., basic characteristics of a subject's network of contacts, if any), or details about the patient's self-perception, were also missing in most cases.

6 - Conclusions

This research collects the 30 CA referenced in the literature relevant to the detection of depressive disorders. This provides an insightful snapshot about both the scientific and technical progress in this field, as well as the interest among the research community and health practitioners in implementing CAs as a facilitating tool for depression detection. In this sense, the state of the art, confirms CAs as a ubiquitous neuropsychological tool that can overcome the barriers of traditional practice, such as (a) the white coat effect, (b) the dispersion of information consequence of the multidisciplinary nature of this condition, by facilitating data collection and consolidation, (c) the complexity of patients' follow-up, by providing support to the continuous improvement of treatment cycles, and (d) the shortage of health professionals, which became evident with the COVID-19 pandemic.

Analysis of the literature indicates relevant CAs for the detection of depression advance the state-of-the-art in three complementary directions, namely (i) the potential of cloud-based virtual assistants in healthcare; (ii) the potential of automated dialogue to complement or even replace classical diagnosis procedures, and (iii) to study and evaluate their performance in different scenarios, particularly their detection capabilities.

Although no restriction on the year of publication was introduced, most of the relevant work is concentrated in the last years, as expected, largely due to the recent popularization of CAs. However, solutions are generally isolated, and there is no common model for the CA-based detection of depression. In other words, a generally accepted protocol or standardized application framework was not yet proposed, apart from some tentative approaches [51].

Incidentally, two diverging trends can be identified in relation to the strategies for depression detection. The first one is characterized by depression being considered the trigger or genesis of other disorders, such as suicide ideation, while the second one considers depression as the consequence or result of a pre-existing condition, such as stress, postpartum disorder, or a chronic illness. This influences the classical clinical instruments used as a reference or golden standard and the very detection protocols implemented.

Among the multiple aspects considered, the authors highlighted studies, such as [48], [51], [69], [79] on the perception of synthesized speech and voice modulation as part of the personality of the as conversational agent as the most valuable factor. Empathy and the agent's ability to establish a strong therapeutic alliance have been shown to be crucial in promoting client trust and fostering emotional openness. This emotional connection between the user and the agent is essential for effective therapy and support in the field of mental health, as it creates an environment conducive to the expression of problems and adherence to treatment. Ultimately, the agent's personality and ability to forge a trusting therapeutic relationship emerge as the most essential component to maximizing the positive impact of conversational agents in mental health care.

PHQ-9 and its variants (e.g., PHQ-4, PHQ-8) are the recurrent golden standards. In many cases, relevant information about the population sample used for validation is missing. Although the preferred target for CAs are patients, in some cases caregivers play a central role, with the implementation of a specific test for them (i.e., the Zarit Test).

Providing user-friendly interfaces to facilitate the administration of clinical tests and clinical procedures is a research line with relevant potential [80]. In the diagnosis and follow-up of depression, the experience and "clinical eye" of the professional are key factors. This was a recurring remark in surveyed documents, pointing out the limitations of CAs in the emotional understanding of users.

7 - Authors' contributions

Iván Otero-González: Original draft preparation, Conceptualization, Methodology, Collecting the data, Writing, Investigation and Visualization Moises Pachecho-Lorenzo and Manuel

Fernández-Iglesias: Writing-review, Writing-curation, and Supervision. Luis Anido-Rifón: Writing-reviewing and Supervision. All authors contributed to the article and approved the submitted version.

8 - Acknowledgments

This research was funded by the Spanish Ministry of Economy, Industry and Competitiveness Grant PID2020-115137RB-I00: Servicios y Aplicaciones para un Envejecimiento Saludable (SAPIENS). We would also like to acknowledge the assistance provided by Noelia Lago-Priego regarding the definition and theories of clinical depression.

9 – Summary Table

What was already known on the topic?

- *Urgent need for innovative and scalable approaches to aid in early detection and intervention of mental health disorders.*
- Conversational agents, powered by artificial intelligence (AI) and natural language processing (NLP), emerged as a promising tool in the field of mental health for the detection and monitoring of clinical depression.
- *Prevalence of depression continues to rise globally.*

What this study added to our knowledge?

- *It sheds light on the technological advancements and challenges associated with the use of conversational agents in mental health care.*
- *It is a resource for understanding the current landscape of conversational agents for the detection of clinical depression.*
- *It identifies relevant advances, knowledge gaps, and areas that require further investigation.*
- *It compiles the conversational agents implemented in scientific papers.*
- *It makes an analysis of conversational agents, in the field of depression, classifying them by target age, the type of interaction, validated clinical trial on which it is based, as well as the size of the work groups with which it has been tested.*

10 - Conflicts of Interest

Disclose any personal financial interests related to the subject matters discussed in the manuscript here. For example, authors who are owners or employees of Internet companies that market the services described in the manuscript will be disclosed here. If none, indicate with “none declared”.

11 - Abbreviations

ADHD: Attention Deficit Hyperactivity Disorder

AIML: Artificial Intelligence mark-up language

APA: American Psychiatric Association

ASR: Automatic Speech Recognition

BCT: Behaviour Change Techniques

CA: Conversational Agent

CBT: Cognitive Behavioural Therapy

CCMD-3: Chinese Classification of Mental Disorders, Third Edition

CES-D: Centre for Epidemiologic Studies–Depression Scale

DSM: Diagnostic and Statistical Manual of Mental Disorders

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition

ICD-10: International Classification of Diseases, Tenth Edition

ICD-11: International Classification of Diseases, Eleventh Edition

MCI: Mild Cognitive Impairment

MDD: Major Depressive Disorder

ML: Machine Learning

MMSE: Mini-mental State Examination

ND: Neurodegenerative Disorder

NIH: National Institutes of Health

NLP: Natural Language Processing

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta Analyses

WHO: World Health Organization

List of Psychometrically Validated Clinical Tests

BDI-II: Beck Depression Inventory-II

BPI—SF: Brief Pain Inventory—Short Form

BPQ: Body Perception Questionnaire

CES-D: Centre of Epidemiological Studies of Depression Scale

DLA-20: Daily Living Activities–20

EPDS: Edinburgh Postnatal Depression Scale

EORTC QLQ-INFO25: European Organisation for Research and Treatment of Cancer Quality of Life Group information questionnaire

ERQ: Emotion Regulation Questionnaire

Fisher test: Statistical significance test used in the analysis of contingency tables.

FSI: Fatigue Symptom Inventory

GDS: Geriatric Depression Scale

GAD-7: Generalized Anxiety Disorder Assessment

HAM-A: Hamilton Anxiety Rating Scale

HAM-D: Hamilton Depression Rating Scale

HRQoL: Health-Related Quality of Life

IDS-C: Inventory of depressive symptomatology

Kessler PD: Kessler Psychological Distress Scale

MBI: Burnout Maslach Burnout Inventory

MMSE: Mini-mental state examination

PANAS: Positive and Negative Affect Scale

PAU: Problematic Alcohol Use

PHQ: Patient Health Questionnaire

PHQ-9: Patient Health Questionnaire - 9

PSA: Public Speaking Anxiety

PTSD: Post Traumatic Stress Disorder

RMBPC: Revised Memory and Behaviour Problem Checklist

RSIS: Roberts' Suicide Ideation Scale

SPs: SAD PERSONS scale

SF-36: 36-Item Short-Form Health Survey

SIGH-D: Hamilton Depression Scale

Zarit Test: Test for measurement the burden of the caregiver

12 - References

[1] D. F. Santomauro *et al.*, "Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic," *Lancet*, vol. 398, no. 10312, pp. 1700–1712, Nov. 2021, doi: 10.1016/S0140-6736(21)02143-7.

[2] EFE. MADRID, "La depresión será la primera causa de discapacidad en 2030," *HERALDO*, Feb. 15, 2015. https://www.heraldo.es/noticias/salud/2015/02/17/la-depresion-sera-primera-causa-discapacidad-2030-340315-2261131.html (accessed May 22, 2023).

[3] N. Rojín and H. Fernández, "La depresión será en 2030 la primera causa de discapacidad en jóvenes y adultos," *EFESALUD*, Jan. 13, 2020.

[4] R. D. Hester, "Lack of access to mental health services contributing to the high suicide rates among veterans," *Int. J. Ment. Health Syst.*, vol. 11, no. 1, p. 47, Dec. 2017, doi: 10.1186/s13033-017-0154-2.

[5] E. Ouldred, "Screening for dementia in older people," *Br. J. Community Nurs.*, vol. 9, no. 10, pp. 434–437, Oct. 2004, doi: 10.12968/bjcn.2004.9.10.16117.

[6] G. P. Couser, "Challenges and opportunities for preventing depression in the workplace: a review of the evidence supporting workplace factors and interventions," *J. Occup. Environ. Med.*, pp. 411–427, 2008, [Online]. Available: http://www.jstor.org/stable/44998668

[7] E. Brown, R. M. Kaiser, and Z. D. Gellis, "Screening and assessment of late-life depression in home healthcare: Issues and challenges," *Ann. Long Term Care*, vol. 15, no. 10, p. 27, 2007, [Online]. Available: https://www.researchgate.net/publication/265273689_Screening_and_Assessment_of_Late-Life_Depression_in_Home_Healthcare_Issues_and_Challenges

[8] A. M. Sheehan and H. McGee, "Screening for depression in medical research: ethical challenges and recommendations," *BMC Med. Ethics*, vol. 14, no. 1, p. 4, Dec. 2013, doi: 10.1186/1472-6939-14-4.

[9] E. Gundersen and B. Bensadon, "Geriatric Depression," *Prim. Care Clin. Off. Pract.*, vol. 50, no. 1, pp. 143–158, Mar. 2023, doi: 10.1016/j.pop.2022.10.010.

[10] C. Bugallo-Carrera *et al.*, "Assessment of validity and comparison of two Spanish versions of the Geriatric Depression Scale," *Front. Psychol.*, vol. 14, p. 1101886, May 2023, doi: 10.3389/fpsyg.2023.1101886.

[11] S. A. and D. John, "Survey on Chatbot Design Techniques in Speech Conversation Systems," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 7, 2015, doi: 10.14569/IJACSA.2015.060712.

[12] A. Ahmed *et al.*, "Chatbot features for anxiety and depression: A scoping review," *Health Informatics J.*, vol. 29, no. 1, p. 146045822211467, Jan. 2023, doi: 10.1177/14604582221146719.

[13] M. S. Pathan, R. Jain, R. Aswani, K. Kulkarni, and S. Gupta, "Anti-Depression Psychotherapist Chatbot for Exam and Study-Related Stress," in *Applied Machine Learning for Smart Data Analysis*, CRC Press, 2019, pp. 21–40. Accessed: Apr. 19, 2023. [Online]. Available: https://www.taylorfrancis.com/chapters/edit/10.1201/9780429440953-2/anti-depression-psychotherapist-chatbot-exam-study-related-stress-mohd-shafi-pathan-rushikesh-jain-rohan-aswani-kshitij-kulkarni-sanchit-gupta

[14] T. W. Bickmore, K. Puskar, E. A. Schlenk, L. M. Pfeifer, and S. M. Sereika, "Maintaining reality: Relational agents for antipsychotic medication adherence," *Interact. Comput.*, vol. 22, no. 4, pp. 276–288, Jul. 2010, doi: 10.1016/j.intcom.2010.02.001.

[15] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "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. Heal.*, vol. 4, no. 2, p. e19, Jun. 2017, doi: 10.2196/mental.7785.

[16] P. Philip *et al.*, "Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders," *Sci. Rep.*, vol. 7, no. 1, p. 42656, Feb. 2017, doi: 10.1038/srep42656.

[17] A. N. Vaidyam, H. Wisniewski, J. D. Halamka, M. S. Kashavan, and J. B. Torous, "Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape," *Can. J. Psychiatry*, vol. 64, no. 7, pp. 456–464, Jul. 2019, doi: 10.1177/0706743719828977.

[18] P. Carlbring, G. Andersson, P. Cuijpers, H. Riper, and E. Hedman-Lagerlöf, "Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis," *Cogn. Behav. Ther.*, vol. 47, no. 1, pp. 1–18, Jan. 2018, doi: 10.1080/16506073.2017.1401115.

[19] A. Fadhil and S. Gabrielli, "Addressing challenges in promoting healthy lifestyles," in *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, May 2017, pp. 261–265. doi: 10.1145/3154862.3154914.

[20] L. Laranjo *et al.*, "Conversational agents in healthcare: a systematic review," *J. Am. Med. Informatics Assoc.*, vol. 25, no. 9, pp. 1248–1258, Sep. 2018, doi: 10.1093/jamia/ocy072.

[21] C. Bérubé *et al.*, "Voice-based conversational agents for the prevention and management of chronic and mental health conditions: Systematic literature review," *J. Med. Internet Res.*, vol. 23, no. 3, p. 14, Mar. 2021, doi: https://doi.org/10.2196/25933.

[22] G. M. Lucas *et al.*, “Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers,” *Front. Robot. AI*, vol. 4, p. 51, Oct. 2017, doi: 10.3389/frobt.2017.00051.

[23] B. D. Kiluk, K. Serafini, T. Frankforter, C. Nich, and K. M. Carroll, “Only connect: The working alliance in computer-based cognitive behavioral therapy,” *Behav. Res. Ther.*, vol. 63, pp. 139–146, Dec. 2014, doi: 10.1016/j.brat.2014.10.003.

[24] K.-J. Oh, D. Lee, B. Ko, and H.-J. Choi, “A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation,” in *2017 18th IEEE International Conference on Mobile Data Management (MDM)*, May 2017, pp. 371–375. doi: 10.1109/MDM.2017.64.

[25] S. Mattsson, S. Alfonsson, M. Carlsson, P. Nygren, E. Olsson, and B. Johansson, “U-CARE: Internet-based stepped care with interactive support and cognitive behavioral therapy for reduction of anxiety and depressive symptoms in cancer - a clinical trial protocol,” *BMC Cancer*, vol. 13, no. 1, p. 414, Dec. 2013, doi: 10.1186/1471-2407-13-414.

[26] G. Anmella *et al.*, “Vickybot, a Chatbot for Anxiety-Depressive Symptoms and Work-Related Burnout in Primary Care and Health Care Professionals: Development, Feasibility, and Potential Effectiveness Studies,” *J. Med. Internet Res.*, vol. 25, p. e43293, Apr. 2023, doi: 10.2196/43293.

[27] S. Natale, “If software is narrative: Joseph Weizenbaum, artificial intelligence and the biographies of ELIZA,” *New Media Soc.*, vol. 21, no. 3, pp. 712–728, Mar. 2019, doi: 10.1177/1461444818804980.

[28] N. R. Haddaway, M. J. Page, C. C. Pritchard, and L. A. McGuinness, “PRISMA2020 : An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis,” *Campbell Syst. Rev.*, vol. 18, no. 2, p. e1230, Jun. 2022, doi: 10.1002/cl2.1230.

[29] “Mendeley Reference Manager,” *Accessed*, 2023. https://www.mendeley.com/reference-manager/library/all-references (accessed May 23, 2023).

[30] “ResearchGate,” *Accessed*, 2023. https://www.researchgate.net/ (accessed May 23, 2023).

[31] M. F. McTear, “Spoken dialogue technology,” *ACM Comput. Surv.*, vol. 34, no. 1, pp. 90–169, Mar. 2002, doi: 10.1145/505282.505285.

[32] K. Hawton, N. T. M. Hill, M. Gould, A. John, K. Lascelles, and J. Robinson, “Clustering of suicides in children and adolescents,” *Lancet Child Adolesc. Heal.*, vol. 4, no. 1, pp. 58–67, Jan. 2020, doi: 10.1016/S2352-4642(19)30335-9.

[33] P. Thorn *et al.*, “Developing a Suicide Prevention Social Media Campaign With Young People (The #Chatsafe Project): Co-Design Approach,” *JMIR Ment. Heal.*, vol. 7, no. 5, p. e17520, May 2020, doi: 10.2196/17520.

[34] C. Rodway, S.-G. Tham, S. Ibrahim, P. Turnbull, N. Kapur, and L. Appleby, “Children and young people who die by suicide: childhood-related antecedents, gender differences and service contact,” *BJPsych open*, vol. 6, no. 3, p. e49, 2020.

[35] C. Schillings, D. Meissner, B. Erb, D. Schultchen, E. Bendig, and O. Pollatos, “A chatbot-based

intervention with ELME to improve stress and health-related parameters in a stressed sample: Study protocol of a randomised controlled trial," *Front. Digit. Heal.*, vol. 5, Mar. 2023, doi: 10.3389/fdgth.2023.1046202.

[36] P. Kaywan, K. Ahmed, A. Ibaida, Y. Miao, and B. Gu, "Early detection of depression using a conversational AI bot: A non-clinical trial," *PLoS One*, vol. 18, no. 2, p. e0279743, Feb. 2023, doi: 10.1371/journal.pone.0279743.

[37] J. Moilanen, N. van Berkel, A. Visuri, U. Gadiraju, W. van der Maden, and S. Hosio, "Supporting mental health self-care discovery through a chatbot," *Front. Digit. Heal.*, vol. 5, Mar. 2023, doi: 10.3389/fdgth.2023.1034724.

[38] G. A. Entenberg *et al.*, "User experience with a parenting chatbot micro intervention," *Front. Digit. Heal.*, vol. 4, Jan. 2023, doi: 10.3389/fdgth.2022.989022.

[39] G. Anmella *et al.*, "Vickybot, a chatbot for anxiety-depressive symptoms and work-related burnout in primary care and healthcare professionals: development, feasibility, and potential effectiveness studies. (Preprint)," *J. Med. Internet Res.*, Oct. 2022, doi: 10.2196/43293.

[40] J. Nie, H. Shao, M. Zhao, S. Xia, M. Preindl, and X. Jiang, "Conversational AI Therapist for Daily Function Screening in Home Environments," in *Proceedings of the 1st ACM International Workshop on Intelligent Acoustic Systems and Applications*, Jul. 2022, pp. 31–36. doi: 10.1145/3539490.3539603.

[41] P. S. A. Khaire, V. Shahane, P. Borse, A. Jundhare, and A. Tatu, "Doctor-Bot: AI Powered Conversational Chatbot for Delivering E-Health," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 4, pp. 2461–2464, Apr. 2022, doi: 10.22214/ijraset.2022.41856.

[42] C. Beatty, T. Malik, S. Meheli, and C. Sinha, "Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study," *Front. Digit. Heal.*, vol. 4, Apr. 2022, doi: 10.3389/fdgth.2022.847991.

[43] E. Castilla, J. J. Escobar, C. Villalonga, and O. Banos, "HIGEA: An Intelligent Conversational Agent to Detect Caregiver Burden," *Int. J. Environ. Res. Public Health*, vol. 19, no. 23, p. 16019, Nov. 2022, doi: 10.3390/ijerph192316019.

[44] A. Schick, J. Feine, S. Morana, A. Maedche, and U. Reininghaus, "Validity of Chatbot Use for Mental Health Assessment: Experimental Study," *JMIR mHealth uHealth*, vol. 10, no. 10, p. e28082, Oct. 2022, doi: 10.2196/28082.

[45] S. Yang, J. Lee, E. Sezgin, J. Bridge, and S. Lin, "Clinical Advice by Voice Assistants on Postpartum Depression: Cross-Sectional Investigation Using Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana," *JMIR mHealth uHealth*, vol. 9, no. 1, p. e24045, Jan. 2021, doi: 10.2196/24045.

[46] P. Kaywan, K. Ahmed, Y. Miao, A. Ibaida, and B. Gu, "DEPRA: An Early Depression Detection Analysis Chatbot," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2021, vol. 13079 LNCS, pp. 193–204. doi: 10.1007/978-3-030-90885-0_18.

[47] J. O. Egede *et al.*, "Design and Evaluation of Virtual Human Mediated Tasks for Assessment of Depression and Anxiety," in *Proceedings of the 21th ACM International Conference on Intelligent Virtual Agents*, Sep. 2021, pp. 52–59. doi: 10.1145/3472306.3478361.

[48] G. Giunti, M. Isomursu, E. Gabarron, and Y. Solad, "Designing Depression Screening Chatbots," in *Studies in Health Technology and Informatics*, vol. 284, IOS Press BV, 2021, pp. 259–263. doi: 10.3233/SHTI210719.

[49] D. Ireland and D. K. Bradford, "Pandora's Bot: Insights from the Syntax and Semantics of Suicide Notes," in *Studies in Health Technology and Informatics*, vol. 276, 2021, pp. 26–31. doi: 10.3233/SHTI210006.

[50] S. Burchert, A. Kerber, J. Zimmermann, and C. Knaevelsrud, "Screening accuracy of a 14-day smartphone ambulatory assessment of depression symptoms and mood dynamics in a general population sample: Comparison with the PHQ-9 depression screening," *PLoS One*, vol. 16, no. 1, p. e0244955, Jan. 2021, doi: 10.1371/journal.pone.0244955.

[51] J. I. Méndez *et al.*, "Smart Homes as Enablers for Depression Pre-Diagnosis Using PHQ-9 on HMI through Fuzzy Logic Decision System," *Sensors*, vol. 21, no. 23, p. 7864, Nov. 2021, doi: 10.3390/s21237864.

[52] S. Suharwardy *et al.*, "116: effect of an automated conversational agent on postpartum mental health: a randomized, controlled trial," *Am. J. Obstet. Gynecol.*, vol. 222, no. 1, p. S91-, 2020, doi: 10.1016/j.ajog.2019.11.132.

[53] M. Ramachandran *et al.*, "74: acceptability of postnatal mood management through a smartphone-based automated conversational agent," *Am. J. Obstet. Gynecol.*, vol. 222, no. 1, p. S62-, 2020, doi: 10.1016/j.ajog.2019.11.090.

[54] B. S. and J. N. C. and C. C. and E. M. and B. E. L. Dosovitsky Gillyand Pineda, "Artificial Intelligence Chatbot for Depression: Descriptive Study of Usage," *JMIR Form Res*, vol. 4, no. 11, p. e17065, Nov. 2020, doi: 10.2196/17065.

[55] M. Monteiro, L. Salgado, F. Seixas, and R. Santana, "Co-designing Strategies to Provide Telecare Through an Intelligent Assistant for Caregivers of Elderly Individuals," in *Human Aspects of IT for the Aged Population. Healthy and Active Aging: 6th International Conference, ITAP 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22*, 2020, pp. 149–166. doi: 10.1007/978-3-030-50249-2_12.

[56] B. Chaix, G. Lobre, S. Mahboub, G. Delamon, J.-E. Bibault, and B. Brouard, "Le chatbot, outil d'accompagnement thérapeutique de la dépression chez les patientes atteintes d'un cancer du sein," *Psycho-Oncologie*, vol. 14, no. 1–2, pp. 17–21, Mar. 2020, doi: 10.3166/pson-2020-0113.

[57] M. Al Owayyed, "Motivating PhD candidates with depression symptoms to complete thoughts-strengthening exercises via a conversational agent," Delft University of Technology, Holland, 2020. Accessed: May 24, 2023. [Online]. Available: https://repository.tudelft.nl/islandora/object/uuid:c5ed5bde-e118-4b73-9f4b-85531920f19a

[58] R. Arrabales, "Perla: un Agente Conversacional para la Detección de Depresión en Ecosistemas Digitales. Diseño, Implemen-tación y Validación," 2020, Accessed: Apr. 20, 2023. [Online]. Available: https://arxiv.org/abs/2008.12875v1

[59] ISRCTN17156687, "Study of a smartphone-delivered, therapist-supported mindfulness-based therapy program for depression in Finnish university students," *https://trialsearch.who.int/Trial2.aspx?TrialID=ISRCTN17156687*, 2018, [Online]. Available:

https://www.cochranelibrary.com/central/doi/10.1002/central/CN-01904604/full

[60] B. Chaix, A. Guillemassé, P. Nectoux, G. Delamon, and B. Brouard, “Vik: A Chatbot to Support Patients with Chronic Diseases,” *Health (Irvine. Calif).*, vol. 12, no. 07, pp. 804–810, 2020, doi: 10.4236/health.2020.127058.

[61] J. Martínez-Miranda *et al.*, “Assessment of users’ acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour,” *J. Med. Syst.*, vol. 43, no. 8, p. 246, Aug. 2019, doi: 10.1007/s10916-019-1387-1.

[62] D. Suendermann-Oeft *et al.*, “NEMSI: A Multimodal Dialog System for Screening of Neurological or Mental Conditions,” in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, Jul. 2019, pp. 245–247. doi: 10.1145/3308532.3329415.

[63] Z. Huang, J. Epps, D. Joachim, and M. Chen, “Depression detection from short utterances via diverse smartphones in natural environmental conditions,” in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, Sep. 2018, vol. 2018-Septe, pp. 3393–3397. doi: 10.21437/Interspeech.2018-1743.

[64] B. Sharma, H. Puri, and D. Rawat, “Digital Psychiatry - Curbing Depression using Therapy Chatbot and Depression Analysis,” in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Apr. 2018, pp. 627–631. doi: 10.1109/ICICCT.2018.8472986.

[65] A. and G. B. and L. L. and R. M. Fulmer Russelland Joerin, “Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial,” *JMIR Ment Heal.*, vol. 5, no. 4, p. e64, Dec. 2018, doi: 10.2196/mental.9782.

[66] F. Goodyear-Smith, R. Martel, M. Darragh, J. Warren, H. Thabrew, and T. C. Clark, “Screening for risky behaviour and mental health in young people: the YouthCHAT programme,” *Public Health Rev.*, vol. 38, no. 1, 2017, doi: 10.1186/s40985-017-0068-1.

[67] P. Philip *et al.*, “Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders,” *Sci. Rep.*, vol. 7, no. 1, p. 42656, Feb. 2017, doi: 10.1038/srep42656.

[68] C. Burton *et al.*, “Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression,” *J. Telemed. Telecare*, vol. 22, no. 6, pp. 348–355, Sep. 2016, doi: 10.1177/1357633X15609793.

[69] A. Bresó, J. Martínez-Miranda, C. Botella, R. M. Baños, and J. M. García-Gómez, “Usability and acceptability assessment of an empathic virtual agent to prevent major depression,” *Expert Syst.*, vol. 33, no. 4, pp. 297–312, Aug. 2016, doi: 10.1111/exsy.12151.

[70] D. M. Post *et al.*, “Improving symptom communication through personal digital assistants: the CHAT (Communicating Health Assisted by Technology) project,” *J. Natl. Cancer Inst. Monogr.*, vol. 2013, no. 47, p. 153-161, 2013, doi: 10.1093/jncimonographs/lgt027.

[71] I. Zarragoitía Alonso, “Sex and gender differences in depression - proclivity in women,” *Medwave*, vol. 13, no. 03, pp. e5651–e5651, Apr. 2013, doi: 10.5867/medwave.2013.03.5651.

[72] M. R. Pacheco-Lorenzo, S. M. Valladares-Rodriguez, L. E. Anido-Rifon, and M. J. Fernandez-

Iglesias, “Smart conversational agents for the detection of neuropsychiatricdisorders: A systematic review,” *J. Biomed. Inform.*, vol. 113, Jan. 2021, doi: 10.1016/j.jbi.2020.103632.

[73] J. L. Z. Montenegro, C. A. da Costa, and R. da Rosa Righi, “Survey of conversational agents in health,” *Expert Syst. Appl.*, vol. 129, pp. 56–67, Sep. 2019, doi: 10.1016/j.eswa.2019.03.054.

[74] A. B. Kocaballi *et al.*, “The Personalization of Conversational Agents in Health Care: Systematic Review,” *J. Med. Internet Res.*, vol. 21, no. 11, p. e15360, Nov. 2019, doi: 10.2196/15360.

[75] S. M. Huq, R. Maskeliūnas, and R. Damaševičius, “Dialogue agents for artificial intelligence-based conversational systems for cognitively disabled: a systematic review,” *Disabil. Rehabil. Assist. Technol.*, pp. 1–20, Nov. 2022, doi: 10.1080/17483107.2022.2146768.

[76] A. E. Chung, A. C. Griffin, D. Selezneva, and D. Gotz, “Health and Fitness Apps for Hands-Free Voice-Activated Assistants: Content Analysis,” *JMIR mHealth uHealth*, vol. 6, no. 9, p. e174, Sep. 2018, doi: 10.2196/mhealth.9705.

[77] J. Pereira and Ó. Díaz, “Using Health Chatbots for Behavior Change: A Mapping Study,” *J. Med. Syst.*, vol. 43, no. 5, p. 135, May 2019, doi: 10.1007/s10916-019-1237-1.

[78] S. M. Lim, C. W. C. Shiau, L. J. Cheng, and Y. Lau, “Chatbot-Delivered Psychotherapy for Adults With Depressive and Anxiety Symptoms: A Systematic Review and Meta-Regression,” *Behav. Ther.*, vol. 53, no. 2, pp. 334–347, Mar. 2022, doi: 10.1016/j.beth.2021.09.007.

[79] J. Martinez-Miranda *et al.*, “Assessment of users’ acceptability of a mobile-based embodiedconversational agent for the prevention and detection of suicidalbehaviour,” *J. Med. Syst.*, vol. 43, no. 8, Aug. 2019, doi: 10.1007/s10916-019-1387-1.

[80] T. Beinema, H. op den Akker, H. J. Hermens, and L. van Velsen, “What to Discuss?—A Blueprint Topic Model for Health Coaching Dialogues With Conversational Agents,” *Int. J. Human–Computer Interact.*, vol. 39, no. 1, pp. 164–182, Jan. 2023, doi: 10.1080/10447318.2022.2041884.

Highlights

Summary Table of hightlithts

What was already known on the topic?

- *Urgent need for innovative and scalable approaches to aid in early detection and intervention of mental health disorders.*
- Conversational agents, powered by artificial intelligence (AI) and natural language processing (NLP), emerged as a promising tool in the field of mental health for the detection and monitoring of clinical depression.
- *Prevalence of depression continues to rise globally.*

What this study added to our knowledge?

- *It sheds light on the technological advancements and challenges associated with the use of conversational agents in mental health care.*
- *It is a resource for understanding the current landscape of conversational agents for the detection of clinical depression.*
- *It identifies relevant advances, knowledge gaps, and areas that require further investigation.*
- *It compiles the conversational agents implemented in scientific papers.*
- *It makes an analysis of conversational agents, in the field of depression, classifying them by target age, the type of interaction, validated clinical trial on which it is based, as well as the size of the work groups with which it has been tested.*

Author statement

Iván Otero-González: Original draft preparation, Conceptualization, Methodology, Collecting the data, Writing, Investigation and Visualization **Moises Pachecho-Lorenzo and *Manuel Fernández-Iglesias***: Writing-reviewin, Writing-curation and Supervision. **Luis Anido-Rifón**: Writing-reviewing and Supervision.

All authors contributed to the article and approved the submitted version.

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Ivan Otero reports financial support was provided by AtlantTIC University of Vigo.