Supplementary appendix 1. Databases and queries

1. Database: SCOPUS

Search date: June 17, 2025

#	Key words	Results
1	(TITLE-ABS-KEY ("chat bot") OR TITLE-ABS-KEY (chatbot) OR TITLE-ABS-KEY (bot) OR TITLE-ABS-KEY ("virtual chatbot") OR TITLE-ABS-KEY ("online robot") OR TITLE-ABS-KEY ("conversational computer program") OR TITLE-ABS-KEY ("Human - bot conversation") OR TITLE-ABS-KEY ("virtual assistant") OR TITLE-ABS-KEY ("messaging bot") OR TITLE-ABS-KEY ("software robot") OR TITLE-ABS-KEY ("conversational agent") OR TITLE-ABS-KEY ("chat base interface") OR TITLE-ABS-KEY ("chatbot services") OR TITLE-ABS-KEY ("dialogue system") OR TITLE-ABS-KEY ("relational agent") OR TITLE-ABS-KEY ("Conversational user interface") OR TITLE-ABS-KEY ("Human-machine communication") OR TITLE-ABS-KEY ("embodied conversational agent") OR TITLE-ABS-KEY ("instant messaging") OR TITLE-ABS-KEY (avatar) OR TITLE-ABS-KEY ("virtual agent") OR TITLE-ABS-KEY (chatterbot) OR TITLE-ABS-KEY ("chatter bot") OR TITLE-ABS-KEY ("conversational bot") OR TITLE-ABS-KEY ("conversational system")) AND (LIMIT-TO (LANGUAGE , "English"))	73,850 results
2	(TITLE-ABS-KEY ("Chronic disease*") OR TITLE-ABS-KEY ("chronic condition*") OR TITLE-ABS-KEY ("chronic status") OR TITLE-ABS-KEY ("chronic illness*") OR TITLE-ABS-KEY ("chronic pathy") OR TITLE-ABS-KEY ("chronic disorder*") OR TITLE-ABS-KEY ("long-term condition")) AND (LIMIT-TO (LANGUAGE , "English"))	487,253 results
3	#1 AND #2	297 results

2. Database: PubMed

Search date: June 17, 2025

#	Key words	Results
1	("chat bot"[All Fields] OR "chabot"[All Fields] OR "bot"[All Fields] OR (("virtual"[All Fields] OR "virtuality"[All Fields] OR "virtualization"[All Fields] OR "virtualized"[All Fields] OR "virtualizing"[All Fields] OR "virtuals"[All Fields]) AND "chatbot"[All Fields]) OR ("online"[All Fields] AND ("robot"[All Fields] OR "robot s"[All Fields] OR "robotically"[All Fields] OR "robotics"[MeSH Terms] OR "robotics"[All Fields] OR "robotic"[All Fields] OR "robotization"[All Fields] OR "robotized"[All Fields] OR "robots"[All Fields])) OR (("conversant"[All Fields] OR "conversants"[All Fields] OR "conversation"[All Fields] OR "conversational"[All Fields] OR "conversations"[All Fields] OR "conversed"[All Fields] OR "conversing"[All Fields]) AND ("computer programme"[All Fields] OR "software"[MeSH Terms] OR "software"[All Fields] OR ("computer"[All Fields] AND "program"[All Fields]) OR "computer program"[All Fields])) OR (("human s"[All Fields] OR "humans"[MeSH Terms] OR "humans"[All Fields] OR "human"[All Fields]) AND "bot"[All Fields] AND ("conversant"[All Fields] OR "conversants"[All Fields] OR "conversation"[All Fields] OR "conversational"[All Fields] OR "conversations"[All Fields] OR "conversed"[All Fields] OR "conversing"[All Fields])) OR "virtual assistant"[All Fields] OR (("message"[All Fields] OR "message s"[All Fields] OR "messaged"[All Fields] OR "messages"[All Fields] OR "messaging"[All Fields]) AND "bot"[All Fields]) OR "software robot"[All Fields] OR "conversational agent"[All Fields] OR ("chat"[All Fields] AND "base"[All Fields] AND ("interface"[All Fields] OR "interface s"[All Fields] OR "interfaced"[All Fields] OR "interfaces"[All Fields] OR "interfacing"[All Fields])) OR "chatbot services"[All Fields] OR "dialogue system"[All Fields] OR "relational agent"[All Fields] OR "Conversational user interface"[All Fields] OR "Human-machine communication"[All Fields] OR "embodied conversational agent"[All Fields] OR "instant messaging"[All Fields] OR ("avatar"[All Fields] OR "avatar s"[All Fields] OR "avatars"[All Fields]) OR "virtual agent"[All Fields] OR (("chatter"[All Fields] OR "chattering"[All Fields] OR "chatters"[All Fields]) AND "bot"[All Fields]) OR "chatterbot"[All Fields] OR (("conversant"[All Fields] OR "conversants"[All Fields] OR "conversation"[All Fields] OR "conversational"[All Fields] OR "conversations"[All Fields] OR "conversed"[All Fields] OR "conversing"[All Fields]) AND "bot"[All Fields]) OR "conversational system"[All Fields]) AND (english[Filter])	37,368 results
2	("chronic disease*"[All Fields] OR "chronic condition*"[All Fields] OR "chronic status"[All Fields] OR "chronic illness*"[All Fields] OR (("chronic"[All Fields] OR "chronical"[All Fields] OR "chronically"[All Fields] OR "chronicities"[All Fields] OR "chronicity"[All Fields] OR "chronicization"[All Fields] OR "chronics"[All Fields]) AND ("pathies"[All Fields] OR "pathy"[All Fields])) OR "chronic disorder*"[All Fields] OR "long-term condition"[All Fields]) AND (english[Filter])	364,821 results
3	#1 AND #2	189 results

3. Database: Web of Science

Search date: June 17, 2025

#	Key words	Results
1	**"chat bot"** (All Fields) or **chatbot** (All Fields) or **bot** (All Fields) or **"virtual chatbot"** (All Fields) or **"online robot"** (All Fields) or **"conversational computer program"** (All Fields) or **"Human - bot conversation"** (All Fields) or **"virtual assistant"** (All Fields) or **"messaging bot"** (All Fields) or **"software robot"** (All Fields) or **"conversational agent"** (All Fields) or **"chat base interface"** (All Fields) or **"chatbot services"** (All Fields) or **"dialogue system"** (All Fields) or **"relational agent"** (All Fields) or **"Conversational user interface"** (All Fields) or **"Human-machine communication"** (All Fields) or **"embodied conversational agent"** (All Fields) or **"instant messaging"** (All Fields) or **avatar** (All Fields) or **"virtual agent"** (All Fields) or **chatterbot** (All Fields) or **"chatter bot"** (All Fields) or **"conversational bot"** (All Fields) or **"conversational system"** (All Fields) and **English** (Languages)	597,159 results
2	**"Chronic disease*"** (All Fields) or **"chronic condition*"** (All Fields) or **"chronic status"** (All Fields) or **"chronic illness*"** (All Fields) or **"chronic pathy"** (All Fields) or **"chronic disorder*"** (All Fields) or **"long-term condition"** (All Fields) and **English** (Languages)	186,759 results
3	#1 AND #2	489 results

4. Database: IEEE

Search date: June 17, 2025

#	Key words	Results
1	(((((All Metadata:bot) OR (All Metadata:"conversational bot") OR (All Metadata:"conversational system"))) OR ((No Keywords Specified))) AND ((All Metadata:"chat bot") OR (All Metadata:chatbot) OR (All Metadata:"virtual chatbot") OR (All Metadata:"online robot") OR (All Metadata:"conversational computer program") OR (All Metadata:"Human - bot conversation") OR (All Metadata:"virtual assistant") OR (All Metadata:"messaging bot") OR (All Metadata:"software robot") OR (All Metadata:"conversational agent") OR (All Metadata:"chat base interface"))) OR ((All Metadata:"chatbot services") OR (All Metadata:"dialogue system") OR (All Metadata:"relational agent") OR (All Metadata:"Conversational user interface") OR (All Metadata:"Human-machine communication") OR (All Metadata:"embodied conversational agent") OR (All Metadata:"instant messaging") OR (All Metadata:avatar) OR (All Metadata:"virtual agent") OR (All Metadata:chatterbot) OR (All Metadata:"chatter bot"))	14,337 results
2	("All Metadata":"Chronic disease*") OR ("All Metadata":"chronic condition*") OR ("All Metadata":"chronic status") OR ("All Metadata":"chronic illness*") OR ("All Metadata":"chronic pathy") OR ("All Metadata":"chronic disorder*") OR ("All Metadata":"long-term condition")	3,743 results
3	#1 AND #2	19 results

5. Database: Google Scholar

Search date: June 17, 2025

#	Key words	Results
1	allintitle: "chat bot" OR chatbot OR bot OR "virtual chatbot" OR "online robot" OR "conversational computer program" OR "Human bot conversation" OR "virtual assistant" OR "messaging bot" OR "software robot" OR "conversational agent" OR "chat base interface" OR "chatbot services" OR "dialogue system" OR "relational agent" OR "Conversational user interface" OR "Human-machine communication" OR "embodied conversational agent" OR "instant messaging" OR avatar OR "virtual agent" OR chatterbot OR "chatter bot" OR "conversational bot" OR "conversational system"	64,500 results
2	allintitle: "Chronic disease" OR "chronic condition" OR "chronic status" OR "chronic disorder" OR "chronic pathy" OR "long-term condition" OR "chronic illness"	21,300 results
3	#1 AND #2	41

Supplementary appendix 2. Quality appraisal of the studies using MMAT (Mixed Methods Appraisal Tool), version 2018

Qualitative									
Authors/years	S1. Are there clear research questions?	S2. Do the collected data allow to address the research questions?	1.1. Is the qualitative approach appropriate to answer the research question?	1.2. Are the qualitative data collection methods adequate to address the research question?	1.3. Are the findings adequately derived from the data?	1.4. Is the interpretation of results sufficiently substantiated by data?	1.5. Is there coherence between qualitative data sources, collection, analysis and interpretation?	Score (%)	**Quality rankings**
Pienkowska et al. (2023)	YES	YES	YES	YES	YES	YES	YES	100	High
Quantitative randomized controlled trials									
Authors/years	S1. Are there clear research questions?	S2. Do the collected data allow to address the research questions?	2.1. Is randomization appropriately performed?	2.2. Are the groups comparable at baseline?	2.3. Are there complete outcome data?	2.4. Are outcome assessors blinded to the intervention provided?	2.5 Did the participants adhere to the assigned intervention?	Score (%)	**Quality rankings**
MacNeill et al. (2024)	YES	YES	YES	YES	YES	No	YES	80	High
Fallah et al. (2024)	YES	YES	YES	YES	YES	Can't Tell	YES	80	High
Bruijnes et al. (2023)	YES	YES	YES	YES	YES	YES	YES	100	High
Gong et al. (2020)	YES	YES	YES	YES	YES	Can't Tell	YES	80	High
Quantitative descriptive									
Authors/years	S1. Are there clear research questions?	S2. Do the collected data allow to address the research questions?	4.1. Is the sampling strategy relevant to address the research question?	4.2. Is the sample representative of the target population?	4.3. Are the measurements appropriate?	4.4. Is the risk of nonresponse bias low?	4.5. Is the statistical analysis appropriate to answer the research question?	Score (%)	**Quality rankings**
Gollapalli and Ng (2025)	YES	YES	Can't Tell	YES	YES	Can't Tell	YES	60	Moderate
Kelly et al. (2025)	YES	YES	Can't Tell	YES	YES	Can't Tell	YES	60	Moderate
Chatzimina et al. (2024)	YES	YES	YES	Can't Tell	YES	YES	YES	80	High
Armijos et al. (2024)	YES	YES	Can't Tell	YES	YES	Can't Tell	YES	60	Moderate
Arthur Bézie et al. (2022)	YES	YES	YES	YES	YES	YES	YES	100	High
Chatzimina et al. (2021)	YES	YES	YES	Can't	YES	YES	NO	60	Moderate
Joana Lobo et al. (2017)	YES	YES	YES	Can't Tell	YES	YES	YES	80	High
Mixed methods									
Authors/years	S1. Are there clear research questions?	S2. Do the collected data allow to address the research questions?	5.1. Is there an adequate rationale for using a mixed methods design to address the research question?	5.2. Are the different components of the study effectively integrated to answer the research question?	5.3. Are the outputs of the integration of qualitative and quantitative components adequately interpreted?	5.4. Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?	5.5. Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?	Score (%)	**Quality rankings**
Maia et al. (2023)	YES	YES	YES	YES	YES	YES	YES	100	High
AwaBabington-Ashaye et al. (2023)	YES	YES	YES	YES	YES	YES	YES	100	High
Magnolia Cardona et al. (2023)	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**100**	**High**

Wang-Chin Tsai et al. (2022)	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**Can't Tell**	**80**	**High**
Robert Mash et al. (2022)	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**100**	**High**
Kowatsch et al. (2021)	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**100**	**High**
Surya Roca et al. (2021)	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**100**	**High**
Ter Stal et al. (2021)	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**100**	**High**
Easton K et al. (2019)	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**100**	**High**
Timothy W. Bickmore et al. (2018)	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**100**	**High**

Each article was classified as low (≤ 59%), moderate (60 to 79), or high (≥ 80) according to its study quality.

Supplementary appendix 3 -Table 1. Clinical insights of included studies.

Study ID, study location, study design	Chronic Disease	Applications	Measured outcomes	Measured tools	Key findings	Effectiveness measurement methods
Lobo et al. (35) (2017), Portugal, Mixed-method*	Heart failure	To support self-management	System design, user task accomplishment, and user satisfaction	Questionnaire consisted of 16 statements based on the System Usability Scale (SUS)	- The system that was created was recognized for its straightforwardness, the quality of information provided, the level of user engagement, its coherence, and its design focused on improving modularity. -Home-monitoring and treatment efficacy can be enhanced. - The financial strain and elevated costs associated with heart failure readmissions and death rates can be diminished.	Not mentioned
Bickmore et al.(38) (2018), US, Mixed-method	Atrial fibrillation	To support self-care management	Participant Reactions, Perception of virtual agent, appropriate, Interaction patterns, Interaction motivation	Qualitative analysis of the interview transcripts using thematic analysis techniques	-The study participants were satisfied with the agent. - Participants saw significant improvements in their self-reported quality of life.	Randomized pilot evaluation
Easton et al. (40) (2019), UK, Mixed-method	Chronic Obstructive Pulmonary Disease (COPD)	To support self-management	Examine the appropriateness of both the structure and the substance of a digital virtual agent equipped with natural language processing abilities.	Video-Based Scenario Testing of Acceptability, System Usability Scale (SUS) questionnaires	Self-management assistance provided through an independent virtual agent was found to be acceptable by the participants.	Undone
Gong et al. (29) (2020), Australia, Randomized controlled trial (RCT)	Diabetes mellitus (Type 2)	To support self-management	1.Clinical and psycho-behavioral outcomes of program. 2.Anxiety and depressive symptoms	RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance) framework. The efficacy of the program was evaluated based on clinical and psycho-behavioral results. 1. HbA1c levels (expressed as % and mmol/mol) were assessed using a pathology blood test.2. The health-related quality of life (HRQoL) was evaluated through participants' completion of the AQoL-8D scale for the Assessment of Quality of Life. 3. Symptoms of anxiety and depression were evaluated using the	- The initiative explores the application of app-based interactive CAs in a residential environment and their effectiveness in aiding diabetes self-management. -The MDC program was feasible and shown to be effective in improving participants' HRQoL	Statistical method

Study ID, study location, study design	Chronic Disease	Applications	Measured outcomes	Measured tools	Key findings	Effectiveness measurement methods
				Hospital Anxiety and Depression Scale (HADS). Diabetes-specific distress was measured through the Problem Areas in Diabetes scale, along with clinical assessments of body weight.		
Chatzimina et al. (49) (2021), Greece, Mixed-method	Cancer (hematologic malignancies)	To support self-management	User experience elements (attractiveness, stimulation, novelty) and usability aspects (dependability, perspicuity, efficiency).	UEQ (User Experience Questionnaire)	-CA usability evaluation showed that pragmatic, hedonic and attractiveness score are in positive range. - The suggested method is attractive to users and may offer an alternative option for symptom reporting that will be accessible to users as a continuous service.	Undone
Kowatsch et al. (51) (2021), Switzerland, Mixed-method	Asthma	To support disese-management	To assess the: 1. reach of MAX 2.CA patient working alliance 3. **acceptance of MAX** 4. intervention completion rate 5. cognitive and behavioral outcomes	1. The effectiveness of the intervention was evaluated by the comparison between the number of participants who were approached and those who began engaging with the CAMAX within the mobile application.2. The working relationship between the patient and the CA MAX was evaluated using a German-adapted edition of the Session Alliance Inventory.3.**MAX Acceptance**-TAM2 4. The completion rate of the intervention was determined by calculating the ratio of participants who completed the intervention within 60 days to those who began engaging with the CA MAX. 5. Knowledge about asthma (ie, cognitive skill) was assessed by a health literacy quiz for children with asthma has been validated, with scores ranging from 0 (indicating no knowledge) to 11 (indicating good knowledge) 6. The inhalation methods used by each patient (i.e., their behavioral skills) were carefully evaluated by the healthcare professional in charge, utilizing established assessment criteria.	The findings from this initial feasibility study suggest a generally favorable assessment regarding the effectiveness of the intervention's outreach. - A solid collaborative relationship was established between patients and the MAXCA . - There was broad acceptance of the intervention among all pertinent stakeholders. - The MAX CA frequently succeeded in encouraging family members to assist the young patients when requested. - MAX was not only embraced in this supportive "team player" capacity but also demonstrated the potential to enhance health-related outcomes in the management of chronic diseases.	Statistical method

Study ID, study location, study design	Chronic Disease	Applications	Measured outcomes	Measured tools	Key findings	Effectiveness measurement methods
Roca et al. (41) (2021), Spain, Mixed-method	Diabetes Mellitus (Type 2) and Depressive Disorder	To support self-management**	1. The clinical efficacy of the virtual assistant (the amount of glycosylated hemoglobin (HbA1c),2. The patients' use and acceptance of the virtual assistant (acceptance and real use)	1. The efficacy of the virtual assistant in a medical context was assessed through the glycosylated hemoglobin (HbA1c) levels in patients The severity of depression was assessed using the patient health questionnaire (PHQ-9), which tracks both the intensity of depression and the effectiveness of treatment. Medication adherence was measured through the medication possession ratio (MPR), with an 80% threshold set as the standard. The impact on healthcare resources was gauged by counting the number of medical appointments each patient had per month. The patients' utilization and acceptance of the virtual assistant were evaluated in the following ways: (a) Daily interactions with the virtual assistant were quantified to determine usage; the type of interaction was analyzed to identify patient preferences for numeric versus text-based communication. (b) The use of tools was quantified by tracking the number of features utilized each day. (c) The effectiveness of reminders was assessed by counting how many reminders were responded to. (d) Acceptance was determined by the number of patients who chose not to uninstall Signal. (e) The perceived usefulness of the assistant was evaluated based on how often patients felt misunderstood by the virtual assistant.	-Improvement of HbA1c -Improvements in the HbA1c measure. -The patients depression significantly reduced - The interaction with the virtual assistant appears to have been quite favorable. - Most patients were in favor of continuing to use the virtual assistant after the study. - Healthcare professionals considered the virtual assistant beneficial for patients. - Utilizing virtual assistants can be advantageous and effective in enhancing patient medication adherence.	Statistical method
Ter Stal et al. (36) (2021), Netherlands, Mixed-method	COPD and CHF (Congestive Heart Failure)	To support Self-management	Patients evaluate the design of the ECA as time progresses, considering its traits (such as friendliness,	1.self-reported questionnaires: Five characteristics of Sylvia (the agent in the MATCH self-management app): friendliness, trustworthiness, involvement, expertise, and authority. • The importance of these 5	-The patient's perception of friendliness, expertise, reliability, involvement, and authority of the ECA did not change over time. -The research indicated that a lifelike agent might increase the likelihood of users adhering to the agent's recommendations, in contrast to a non-moving cartoon. -Most users showed little	Undone

Study ID, study location, study design	Chronic Disease	Applications	Measured outcomes	Measured tools	Key findings	Effectiveness measurement methods
			trustworthiness, engagement, expertise, and authority), small talk exchanges, and the likelihood of adhering to the agent's recommendations.	characteristics of an ECA for self-management in general. • The probability of heeding Sylvia's agent's recommendation. All items were evaluated using a 7-point Likert scale. 2.The initial questionnaire included queries concerning the patient's attributes. 3. An analysis was conducted on small talk interactions using log data 4.The perceptions of the patient's view on the agent's traits, the probability of heeding the agent's recommendations, the casual conversation, and various other design elements were collected through five semi-structured interviews	interest in the agent's casual conversation. - Patients displayed a decreasing tendency to comply with the agent's suggestions as time progressed.	
Mash et al. (24) (2022), South Africa, Mixed-method	Chronic kidney disease (CKD)	To support self-health management	To conduct diet management and patient education	N/A Semi-structured interviews	Enable users to seek information about nutrition, assist individuals with chronic illnesses in enhancing their understanding of medical care, promote healthier eating behaviors, and enhance overall quality of life.	Undone
Tsai et al. (42) (2022), Taiwan, Mixed-method	Diabetes mellitus (Type 2)	To support self-management	Acceptability, adoption, appropriateness, feasibility, fidelity, cost, coverage, effects, and sustainability	Descriptive exploratory semi-structured interviews	The CA has significant potential to enhance conventional health care methods for individuals with diabetes while also aiding in more thorough patient education.	Undone
Bézie et al. (52) (2022), France, quantitative descriptive	Headaches	Control of medication-overuse headache (MOH) in the adult population	Impact of MOH on the patient's life	Headache Impact Test (HIT-6) score	The findings indicated that excessive use of medication was frequently associated with a rise in headache occurrences, and that specific prescription medications, particularly triptans and opioids, were the most commonly overused among the group studied. This underscores the potential impact healthcare providers could have in addressing these medications to prevent medication overuse and its consequences.	Undone
Pienkowska et al. (43)(2023), Singapore, Qualitative	Diabetes mellitus (Type 2)	Educational App	Participants' perceptions of living with T2D, attitudes toward digital diabetes education, acceptance and	health information technology acceptance model (HITAM) and the Mobile Application Rating Scale (MARS)	Evaluations of the survey and interview data identified three key themes: the difficulties associated with managing T2D; the validation, acceptability, and usability of the diabetes education app prototype; and viewpoints on digital diabetes education.	Thematic mapping

Study ID, study location, study design	Chronic Disease	Applications	Measured outcomes	Measured tools	Key findings	Effectiveness measurement methods
			usability of the prototype			
Maia et al. (25) (2023), Portugal, Mixed-method	Dementia	To support self-care management	Accuracy and usability of *GECA*'s functionalities	Based on two Scenarios testing	-GECA has demonstrated a success rate in delivering accurate responses. - GECA was evaluated as usable and Efficient.	Statistical method
Babington-Ashaye et al. (53) (2023), Switzerland, Mixed-method	Hemophilia	Educational App	Usability	System Usability Scale (SUS)	The system was evaluated as highly usable, achieving a SUS score of 81.7, which is above the average. Even though they were quite satisfied with the French version, 42% expressed a preference for the Wolof version.	Statistical method
Bruijnes et al. (44)(2023), Netherlands, Randomized controlled trial (RCT)	Diabetes mellitus (All types)	To support self-management	Social diabetes Distress, attitude towards the intervention, Feeling of Being Heard (FBH) and Usability	Diabetes Distress Scale (DDS), the Client Satisfaction Questionnaire (CSQ-8), Feeling of Being Heard (FBH), and System Usability Scale (SUS)	The results indicate that a CA can effectively identify social diabetes distress in individuals with diabetes and provide suitable suggestions to alleviate that distress (H1).	Pre-post testing
Cardona et al. (37) (2023), Australia, Mixed-method	Heart failure	Assist in the process of preparing for shared decision-making with older adults considered to be in their final year of life.	Usability and acceptability	System Usability Scale (SUS)	Healthcare professionals, patients, and caregivers expressed their approval of its usability and were mostly receptive to its content and layout.	Undone
Chatzimina et al. (50) (2024), Greece, Quantitative descriptive	Hematologic Malignancies	self-care management (improve their care management and patient results)	user experience elements (attractiveness, stimulation, novelty) and usability components (dependability, perspicuity, efficiency).	User Experience Questionnaire (UEQ)	Facilitating more individualized care strategies Boost patient involvement in a healthcare environment	Undone
MacNeill et al. (46) (2024), Randomized controlled trial (RCT)	arthritis and diabetes	Self-care management includes mental health assistance for individuals with chronic illnesses like	Depression,Anxiety, Stress, and user feedback (user experience)	The Patient Health Questionnaire–9 (PHQ-9) was used to evaluate depression. Anxiety levels were measured using the Generalized Anxiety Disorder Scale–7 (GAD-7).	The findings of this research indicate that mental health CAs can serve as a valuable resource for mental health support for those with chronic conditions, particularly for individuals with arthritis or diabetes. Participants in the intervention group reported a reduction in both depression and anxiety severity after utilizing a mental health CAfor four weeks. Users expressed appreciation for the app's features and	Randomized Controlled Trial

Study ID, study location, study design	Chronic Disease	Applications	Measured outcomes	Measured tools	Key findings	Effectiveness measurement methods
		arthritis and diabetes.			functions, its overall design, and their experience using it.	
Armijos et al. (45) (2024), quantitative descriptive	Diabetes mellitus	self-care management (enhancing the safety of medication usage in individuals with diabetes	drug-food and drug-drug interactions	Chatbot Usability Questionnaire (CUQ)	the system detects interactions between drugs and between drugs and food, offering helpful suggestions as advised by specialists.	undone
Fallah et al.(39) (2024), Randomized Controlled Trial	Atrial Fibrillation	Self-care management (Investigating the Capabilities of Virtual Agents in Atrial Fibrillation Management) Management)	Electrocardiogram (ECG)	The impact on quality of life is measured using the AF Effect on Quality of Life (AFEQT) questionnaire. Health literacy is evaluated through both the Newest Vital Sign (NVS).	The promotion of the heart rhythm sensor by the agent proved effective, as participants in the intervention group recorded significantly more heart rhythm measurements than those in the control group.	Randomized Controlled Trial
Gollapalli et al.(47) (2025), quantitative descriptive	Diabetes mellitus type 2	self-care management (reducing Psychological Insulin Resistance (PIR) in patients with T2D)	Psychological Insulin Resistance (PIR)	Not specified	PIRsuader can enhance patients' openness to trying insulin while addressing their specific concerns with empathy.	Undone
Kelly et al. (48), (2025), quantitative descriptive	Diabetes mellitus type 2	to enhance health literacy regarding type 2 diabetes mellitus (T2DM)	health literacy	All assessments were conducted using simulated or artificial queries.	The results indicate that such CAs could assist in patient education, encourage self-management, and can be easily adapted to various health contexts.	undone

* Mixed method (both quantitative and qualitative)

Supplementary Appendix 4 – Table 2. Technical insights of included studies.

Study ID	CA Name	Platform	CA languages	Response generation	Dialogue initiative	Input modality	Output modality	CA design software	Integration with specific external sources	Evaluation of CAs	Ethics or Privacy Discussion	Study limitation(s)
Lobo et al. (35) (2017)	CARMIE	Android	Portuguese	Artificial intelligence	User	voice	Both voice and text	N/A	N/A	Simplicity, Portability,Usability	N/A	-The research was not carried out with the ultimate target age group. Feedback from the questionnaire concerning the overall usability of the system might vary for older users.
Bickmore et al. (38) (2018)	No name	iOS	English	Rull-based	User	both voice and text	Both voice and text	N/A	N/A	Usability	N/A	1.Small sample size 2. the lack of long-term, objectively-measured health outcomes.
Easton et al. (40) (2019)	Avachat	prototype Avachat system	English	Artificial intelligence	user	Text and speech	Text and speech	N/A	N/A	The acceptability of both the structure and the substance of a digital virtual agent.	N/A	Small sample size
Gong et al. (29) (2020)	Laura	Android and iOS	English	N/A	CA	Voice	Text and speech	N/A	N/A	Adoption, Use, and Effectiveness	Ethical approval and informed consent have obtained.	1. A limited number of participants were included in the study. 2. Participants and their general practitioners could not be blinded to

Study ID	CA Name	Platform	CA languages	Response generation	Dialogue initiative	Input modality	Output modality	CA design software	Integration with specific external sources	Evaluation of CAs	Ethics or Privacy Discussion	Study limitation(s)
												the allocation of study arms, which may result in self-report bias and the influence of the Hawthorne effect. 3. A greater proportion of participants in the control group completed the assessments compared to those in the intervention group.
Chatzimina et al. (49) (2021)	CA	mobile and web application	English	Artificial intelligence	user	Text	Text and Voice	RASA Framework	N/A	Usability and Usefulness	Ethical approval and informed consent have obtained.	N/A
Kowatsch et al. (51) (2021)	MAX	Android and iOS (for patients) and web-based (for health care professional)	German	Artificial intelligence	CA	Email, mobile chat, SMS text (text and speech)	Text and speech	open-source software platform MobileCoach	N/A	Feasibility of Acceptance	Ethical approval have obtained.	1. Limited sample size. 2. The inductive open coding of the interviews was conducted solely by one author (SH), which introduced a potential bias into the qualitative findings. 3. There is an issue with

Study ID	CA Name	Platform	CA languages	Response generation	Dialogue initiative	Input modality	Output modality	CA design software	Integration with specific external sources	Evaluation of CAs	Ethics or Privacy Discussion	Study limitation(s)
												the reliability of the web-based MAX interface for healthcare professionals linked to HIS or the information provided. 4. Given that the assessment of social support was based on self-reporting, it is likely that the involvement of family members was overestimated. 5. Only the costs and efforts associated with the MAX intervention were reported, without any discussion of cost-effectiveness implications.
Roca et al. (41) (2021)	Elena	Android and iOS	Spanish	Artificial Intelligence	CA	text or voice	Text or voice	Artificial Intelligence Markup Language (AIML) based on HL7	within messaging platforms	Acceptance and Real use	Ethical approval and informed consent have been secured. Privacy and user rights were addressed in	1. The virtual assistant is intended to inquire about the overall duration of the medication;

Study ID	CA Name	Platform	CA languages	Response generation	Dialogue initiative	Input modality	Output modality	CA design software	Integration with specific external sources	Evaluation of CAs	Ethics or Privacy Discussion	Study limitation(s)
								(FHIR) standard			accordance with national data protection law LO 03/2018 and the European GDPR. To ensure security, end-to-end encryption and the privacy of central servers were utilized.	however, we have noticed that certain patients fail to update this information following the initial setup. In these instances, the virtual assistant ceases to send medication reminders, leading to patients often forgetting to utilize the virtual assistant. 2. Patients must possess digital literacy or have individuals nearby who can assist them.
Ter Stal et al. (36) (2021)	Sylvia	eHealth platform (an app on a tablet)	Dutch	Artificial intelligence	CA	Text	Text	N/A	N/A	Understand the traits of the agent and their casual conversation, as well as the probability of adherence to the agent's recommendations.	Ethical approval and informed consent have obtained.	1. Small sample size 2. Participants utilized a Fitbit, a smart scale, and a smart sensorized inhaler alongside the self-management application.

Study ID	CA Name	Platform	CA languages	Response generation	Dialogue initiative	Input modality	Output modality	CA design software	Integration with specific external sources	Evaluation of CAs	Ethics or Privacy Discussion	Study limitation(s)
												3. Numerous participants expressed dissatisfaction regarding the malfunctioning of the sensors. 4. The interviews addressed all aspects of the self-management intervention, rather than concentrating solely on the agent's design. 5. Not every participant contributed information pertinent to the research question of this study; consequently, we must exercise caution when interpreting the interview results..
Mash et al. (24) (2022)	No name	Android and iOS	English	N/A	User	Text	Text	LINE APP is utilized to create the system architecture for the chat robot.	N/A	N/A	N/A	1. Given that individuals suffering from chronic kidney disease are predominantly elderly and may have impaired

Study ID	CA Name	Platform	CA languages	Response generation	Dialogue initiative	Input modality	Output modality	CA design software	Integration with specific external sources	Evaluation of CAs	Ethics or Privacy Discussion	Study limitation(s)
												vision, it is essential for the design of CAsto take into account both the font size on the interface and the need to simplify the text as much as possible. 2. Substituting text with images or audio messages can help patients with chronic kidney disease to comprehend the information more effectively, thereby enhancing the accuracy of their actions.
Tsai et al. (42) (2022)	GREAT4Diabetes	Android and iOS	English, Afrikaans, or Xhos	N/A	User	Text	Voice messages and graphics	Aviro Health on whatsapp	N/A	Acceptability, Adoption, Appropriateness, Feasibility, Fidelity, Cost, Coverage, Effects, and Sustainability	Ethical approval and informed consent have obtained.	Additional investigation is required to thoroughly examine the patient's interaction with the CAand assess its efficacy within this context.

Study ID	CA Name	Platform	CA languages	Response generation	Dialogue initiative	Input modality	Output modality	CA design software	Integration with specific external sources	Evaluation of CAs	Ethics or Privacy Discussion	Study limitation(s)
Bézie et al. (52) (2022)	Vik Migraine	Android and iOS	French	Artificial intelligence	User	Text	Text	Wefigh	N/A	N/A	Ethical approval and informed online consent have obtained.	Subsequent research should involve a comparison of the intensity and occurrence of headaches prior to and following the cessation of medication, particularly by engaging neurologists to identify Medication Overuse Headache (MOH) in patients.
Pienkowska et al. (43)(2023)	No name	Android	English	Rull-based	User	Text	Text	The CAdialogues were created in part using Adobe XD and in part utilizing Quriobot to facilitate active involvement in testing.	N/A	Usability	Ethical approval and informed consent have obtained.	1.Small sample size 2. the lack of familiarity among participants regarding the concept of a flat prototype that offers limited interaction options and a predetermined flow.
Maia et al. (25) (2023)	GECA	Android	English or European Portuguese	Artificial intelligence	User	both voice and text	both voice and text	Python and Kotlin programming languages	The GECA CAis capable of accessing a patient's prescribed medications and health device	Functional testing	OpenID protocol was applied to ensure Patient privacy and security	N/A

Study ID	CA Name	Platform	CA languages	Response generation	Dialogue initiative	Input modality	Output modality	CA design software	Integration with specific external sources	Evaluation of CAs	Ethics or Privacy Discussion	Study limitation(s)
									measurements.			
Babington-Ashaye et al. (53) (2023)	Saytù Hemophilie	Android and iOS	French and Wolof	Artificial intelligence	CA	Voice	Voice	React Native,a JavaScript-based mobile app framework	N/A	Usability	Ethical approval and informed consent have obtained.	1. Limited sample size 2. Employment of alternative usability assessment techniques 3. While the research protocol adhered to the Declaration of Helsinki and the study obtained internal approval from the Blood Transfusion Centre of Senegal (CNTS) committee, it may have been beneficial to seek further guidance from an independent ethics committee.
Bruijnes et al. (44)(2023)	No name	N/A	English	Artificial intelligence	CA	Text	Text	Rasa toolkit	N/A	Usability	Ethical approval and informed consent have obtained.	1. The duration of the intervention was somewhat restricted, consisting of three distinct

Study ID	CA Name	Platform	CA languages	Response generation	Dialogue initiative	Input modality	Output modality	CA design software	Integration with specific external sources	Evaluation of CAs	Ethics or Privacy Discussion	Study limitation(s)
												sessions conducted over a three-week timeframe. A more extended intervention, lasting several months or even a year, or the inclusion of additional sessions, could potentially enhance the effect size. 2. There is insufficient information regarding the type of diabetes that the participants have, as well as whether they are currently undergoing treatment for diabetes-related distress.
Cardona et al. (37) (2023)	CHAT	Web-based	English	N/A	User	Text	Text	front-end using React JS and Redux.. backend service upon the HL7 , (FHIR)	hospital electronic medical records	Usability	Ethical approval and written informed consent have obtained.	1.The questionnaire's effectiveness has not been validated 2.CHAT did not include video or

Study ID	CA Name	Platform	CA languages	Response generation	Dialogue initiative	Input modality	Output modality	CA design software	Integration with specific external sources	Evaluation of CAs	Ethics or Privacy Discussion	Study limitation(s)
												other visual aids
Chatzimina et al. (50) (2024),	No name	Mobile application	Greek	Not specified	CA	text	text		N/A	usability and user experience	Ethical approval and informed consent have obtained.	The system struggles to effectively address open-ended inquiries.
MacNeill et al. (46) (2024)	Wysa	Mobile application	Not specified	artificial intelligence–enabled	CA	text	text	N/A	N/A	Feedback Questions (2 open-ended questions	The study was cataloged on ClinicalTrials.gov (ID NCT04620668) and reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials)–EHEALTH guidelines.	The applicability of the findings is also uncertain. These programs have their constraints and might not be ideal for everyone.
Armijos et al. (45) (2024)	No name	Mobile application (Ios and Android)	English	artificial intelligence–enabled	CA	Text and voice	text	Flutter	N/A	usability evaluation	N/A	It's crucial to understand that certain platforms, like Android, may encounter difficulties with voice functionality due to specific technical limitation.

Study ID	CA Name	Platform	CA languages	Response generation	Dialogue initiative	Input modality	Output modality	CA design software	Integration with specific external sources	Evaluation of CAs	Ethics or Privacy Discussion	Study limitation(s)
Fallah et al.(39) (2024)	WebMD	Mobile application	English	Not specified	CA	text	text	hierarchical transition network-based scripting language with template-based text generation	N/A	N/A	N/A	While all participants were diagnosed with AF and recruited from a single metropolitan area, our results may not be applicable to all regions of the US or elsewhere
Gollapalli et al.(47) (2025)	PIRsuader	Desktop	English	artificial intelligence	CA	text	text	using LLM APIs from OpenAI	N/A	Conversation Quality	This study was carried out in accordance with the ACM Code of Ethics.	we did not explore local models for the dialog agent, which could potentially be more cost-effective and present fewer privacy concerns.
Kelly et al. (48), (2025)	No name	web browsers on any device	English	RAG-based artificial intelligence	CA	text	text	Python	can be effortlessly incorporated into various applications or onto a website.	Question Evaluation, Consultation Evaluation	No patient information was utilized in this research. Ethics approval was not required. All evaluations were performed using simulated or artificial queries.	The evaluation was carried out using a curated collection of one-off questions and a simulated patient consultation. Although these methods are relevant, align with previous research methodologi

Study ID	CA Name	Platform	CA languages	Response generation	Dialogue initiative	Input modality	Output modality	CA design software	Integration with specific external sources	Evaluation of CAs	Ethics or Privacy Discussion	Study limitation(s)
												es, and offer valuable insights, they might not entirely reflect the complexity and variability of questions that individuals encounter in real-world situations.