

Review

Evaluating the Potential and Pitfalls of AI-Powered Conversational Agents as Humanlike Virtual Health Carers in the Remote Management of Noncommunicable Diseases: Scoping Review

Sadia Azmin Anisha^{1*}, BBusCom, MBusInfSys; Arkendu Sen^{1*}, MBBS, GCHE, DipOphthal, DipForensMed, MSurg, PhD; Chris Bain^{2*}, MBBS, MInfoTech, MACS, PhD

¹Jeffrey Cheah School of Medicine & Health Sciences, Monash University Malaysia, Bandar Sunway, Malaysia

²Faculty of Information Technology, Data Future Institutes, Monash University, Clayton, Australia

* all authors contributed equally

Corresponding Author:

Arkendu Sen, MBBS, GCHE, DipOphthal, DipForensMed, MSurg, PhD
Jeffrey Cheah School of Medicine & Health Sciences
Monash University Malaysia
Jalan Lagoon Selatan
Bandar Sunway, 47500
Malaysia
Phone: 60 3 551 46325
Fax: 60 3 551 46323
Email: arkendu.sen@monash.edu

Abstract

Background: The rising prevalence of noncommunicable diseases (NCDs) worldwide and the high recent mortality rates (74.4%) associated with them, especially in low- and middle-income countries, is causing a substantial global burden of disease, necessitating innovative and sustainable long-term care solutions.

Objective: This scoping review aims to investigate the impact of artificial intelligence (AI)-based conversational agents (CAs)—including chatbots, voicebots, and anthropomorphic digital avatars—as human-like health caregivers in the remote management of NCDs as well as identify critical areas for future research and provide insights into how these technologies might be used effectively in health care to personalize NCD management strategies.

Methods: A broad literature search was conducted in July 2023 in 6 electronic databases—Ovid MEDLINE, Embase, PsycINFO, PubMed, CINAHL, and Web of Science—using the search terms “conversational agents,” “artificial intelligence,” and “noncommunicable diseases,” including their associated synonyms. We also manually searched gray literature using sources such as ProQuest Central, ResearchGate, ACM Digital Library, and Google Scholar. We included empirical studies published in English from January 2010 to July 2023 focusing solely on health care-oriented applications of CAs used for remote management of NCDs. The narrative synthesis approach was used to collate and summarize the relevant information extracted from the included studies.

Results: The literature search yielded a total of 43 studies that matched the inclusion criteria. Our review unveiled four significant findings: (1) higher user acceptance and compliance with anthropomorphic and avatar-based CAs for remote care; (2) an existing gap in the development of personalized, empathetic, and contextually aware CAs for effective emotional and social interaction with users, along with limited consideration of ethical concerns such as data privacy and patient safety; (3) inadequate evidence of the efficacy of CAs in NCD self-management despite a moderate to high level of optimism among health care professionals regarding CAs’ potential in remote health care; and (4) CAs primarily being used for supporting nonpharmacological interventions such as behavioral or lifestyle modifications and patient education for the self-management of NCDs.

Conclusions: This review makes a unique contribution to the field by not only providing a quantifiable impact analysis but also identifying the areas requiring imminent scholarly attention for the ethical, empathetic, and efficacious implementation of AI in NCD care. This serves as an academic cornerstone for future research in AI-assisted health care for NCD management.

Trial Registration: Open Science Framework; <https://doi.org/10.17605/OSF.IO/GU5PX>

KEYWORDS

conversational agents; artificial intelligence; noncommunicable disease; self-management; remote monitoring; mobile phone

Introduction

Burden of Noncommunicable Diseases

Noncommunicable diseases (NCDs), also known as chronic diseases, are medical conditions that are not primarily caused by infectious agents (eg, viruses, bacteria, fungi, or parasites) and cannot be transmitted from one individual to another through close contact [1]. NCDs, such as cancer, cardiovascular diseases (CVDs), chronic obstructive pulmonary diseases, chronic respiratory diseases, chronic kidney diseases, cognitive disorders, metabolic syndrome, diabetes, and hypertension, are rising worldwide, with significantly higher rates in low- and middle-income countries (LMICs) [2-4]. In 2019, NCDs contributed to the highest proportion of total mortality (74.4%), accounting for 7.1 million additional deaths in 2019 as compared to 2009 [5]. Nearly half of the mortalities in Asia are attributable to NCDs, causing >80% of CVD and diabetes deaths, 90% of chronic obstructive pulmonary disease deaths, and two-thirds of all cancer deaths occurring in LMICs, resulting in 47% of the global burden of disease [3,6]. Major risk factors of NCDs include unhealthy dietary habits; physical inactivity; stress; and consumption of drugs, tobacco, and alcohol, which are generally modifiable due to lifestyle choices [5,7]. These chronic diseases are also major causes of long-term disabilities and prolonged costly treatment that may pose serious threats to a country's health care resources and expenditure, especially among lower-income countries where health systems are not sufficiently equipped to tackle the escalating challenges [3,4,8].

In addition to regular visits to health care centers, NCDs require challenging self-care management, including compliance with medications, lifestyle modifications, and constant symptom monitoring to prevent disease progression; nonetheless, adherence to these procedures is generally low, especially among adults with limited health literacy who struggle to comprehend and follow instructions from their health care professionals [9,10]. Moreover, the shortage of health care providers and their limited time availability are substantial causes of patients' being deprived of receiving adequate health education and support to make informed decisions required for the effective self-management of their chronic illnesses [11,12]. In addition, access to proper health services can be limited in many underdeveloped areas and rural communities due to poor health care infrastructure and mobility facilities [13].

The optimal prevention management strategy should incorporate elements of individual lifestyle management, societal health awareness management, national health policy decisions, and global health strategy [7]. However, the escalating prevalence of NCDs worldwide reveals that traditional disease management techniques are not sufficiently effective, thereby indicating an urgent necessity to develop effective supplementary management strategies to mitigate the substantial financial burden imposed by NCDs on many households, particularly in LMICs [2,3,14].

The Role of Telehealth in NCD Management

Telehealth applications have the potential to improve patient self-care and disease-specific knowledge as well as minimize hospitalizations and mortality [15]. There is evidence from several past studies suggesting the significant effectiveness of mobile-based telehealth apps in improving nutritional intake and physical activity with technology intervention, resulting in body weight loss and adoption of recommended lifestyle changes due to their convenience and accessibility [16]. This is because such apps, with the help of existing and emerging technologies, have the ability to assist patients in managing their chronic diseases more effectively by providing constant self-monitoring tools and promoting improved self-management of health problems [17]. Furthermore, the ever-growing ownership of mobile devices worldwide has greatly contributed to the shift toward digital health care services, including assessment, monitoring, and treatment of physical and mental health, thereby indicating the promising ability of mobile health apps in self-monitoring, assessment, and treatment of NCDs at a reduced cost [18-20]. Indeed, the COVID-19 pandemic has expedited the use of telehealth [21,22], which has been proposed as a cost-effective method of delivering better health care services to people with chronic illnesses in a more flexible, personalized, transparent, dynamic, and accessible way [21,23].

Potential Enhancement of Existing Telehealth Apps With Artificial Intelligence

Although mobile-based telehealth apps offer an ideal platform for systems designed to help patients manage their chronic illnesses due to the smartphones' computational power, connectivity, and consistent availability, many individuals struggle with complex user interfaces of existing digital health technologies [10]. While the ubiquity of these apps reduces some acceptance barriers, most apps still overlook many other barriers, such as lack of motivational, psychological, and emotional support [10,24]. Nonetheless, technological advancement involving artificial intelligence (AI) seems to have the potential to further upgrade existing mobile health apps with more user-friendly features that can support individual user needs [25]. For instance, many patients lack the ability to effectively navigate conventional telehealth apps due to limited health-related or computer literacy and disabilities such as visual impairment; hence, such intelligent dialogue systems, specifically termed *conversational agents* (CAs), may help overcome these limitations and improve usability by providing an oral presentation of the apps' contents in plain language [15]. AI-powered CAs are computer systems that can communicate with humans through text, voice, and images on mobile, web-based, or audio-based platforms using AI techniques such as machine learning (ML; a statistical method of training models using data for making predictions based on a variety of features) and natural language processing (machines' ability to detect and interpret humans' verbal and written languages) [26,27]. A CA may serve as an *empathetic* listener to understand patients'

problems as well as aid in monitoring a patient's health 24/7 and notify physicians about an anticipated medical emergency [13].

The popularity of CAs, especially those that use unrestrained natural language, has increased over the last decade as consumers can use their smartphones to interact with CAs for daily tasks [27]. When deployed on mobile devices, CAs have the potential to augment human intelligence and demonstrate multiple benefits such as delivering health education and behavior change for a range of chronic health conditions [10]. Moreover, these AI-based CAs provide additional communication channels that are particularly effective for developing trust and therapeutic alliance to encourage adherence among users [10]. The human-like conversational features of CAs due to advancements in natural language processing, voice

recognition, and AI are increasingly substituting human employees in service encounters, including the health care industry to deliver personalized care and support to individuals with chronic health issues [28].

Types of CAs

In this paper, we broadly categorize intelligent CAs into 3 main types (Table 1): *chatbots* (text based or voice enabled—*voicebots*) without embodiment, computer-based embodied *digital avatars*, and physically embodied *humanoid robots* [29]. Although *robots* are beyond the scope of this paper, AI humanoid or social robots can also be classified as CAs that can be used as human-like health caregivers for managing NCDs. Despite the differences, all CAs, including humanoid robots, aim to enhance relational outcomes through human-like communication [29].

Table 1. Types of conversational agents (CAs).

AI ^a -based CA type (with examples)	Features and functionality	Interaction mode	Advantages	Challenges
Chatbots (eg, ChatGPT)	<ul style="list-style-type: none"> • Computer programs integrated into messaging platforms that interact with users via free text [30] • Evolved from preprogrammed, fixed scripted responses to AI-powered versions enabling more human-like conversations [31-33] 	Text based	<ul style="list-style-type: none"> • Simple user interfaces • Easy and cheaper to design and develop 	<ul style="list-style-type: none"> • Offers mostly chat-based interactions • Lack of personalized emotional tones [34]
Voicebots (or voice assistants; eg, Amazon Alexa and Apple's Siri)	<ul style="list-style-type: none"> • Voice-enabled intelligent chatbots that interact with users and respond to speaker commands primarily through voice [35,36] 	Voice based (or text+voice enabled)	<ul style="list-style-type: none"> • Offers greater flexibility by allowing for hands-free conversations and multitasking [35] • Beneficial for users with limited digital expertise (eg, users with typing disabilities) 	<ul style="list-style-type: none"> • Development of enhanced voice interfaces crucial for a more natural conversation [36] • Require some level of emotion awareness (a significant component of voice conversations) [36] • Prone to speech recognition errors [37]
Digital avatars (eg, Replika, Mitsuku, and Soul Machines)	<ul style="list-style-type: none"> • AI-created anthropomorphic representations of real-world characters in a computer-simulated environment, partially automated in some actions and movements, emulating human behavior [38,39] • Digital AI avatars can generate human-like interactions through a fusion of multimodal features [40,41] • Digital avatars evolved from 2D cartoonlike characters to visually realistic and interactive human faces with 3D imaging, leading to the emergence of highly realistic avatars ("Digital humans") [42] • Companies such as Soul Machines use CGI^b, AI, and NLP^c to create digital human prototypes in metaverse spaces [43] • Digital humans can represent fictional characters or virtual replicas of real humans, requiring digital twin technology for data-driven personalized care [44-46] 	Multimodal (text based, voice activated, and face-to-face)	<ul style="list-style-type: none"> • Ability to mimic natural human interactions by delivering highly personalized responses [40,41] • Capable of face-to-face conversations, displaying physical nonverbal behaviors (eg, facial expressions, hand gestures, nodding, and head and body postures) [47] 	<ul style="list-style-type: none"> • Technical implementation is more complex, time-consuming, and resource intensive [47] • Real-time video interactions with digital AI avatars require high computing power and bandwidth connectivity [48] • Risk of perceived uncanniness due to increased realism, posing ethical concerns [25,49]

AI ^a -based CA type (with examples)	Features and functionality	Interaction mode	Advantages	Challenges
Humanoid (or anthropomorphic or social) robots (eg, Sophia and Ameca)	<ul style="list-style-type: none"> Designed to mimic human-like characteristics both in behavior and physical appearance, including body structure and autonomous movement [50,51] Often used for education, entertainment, assistance, and personal care through various sensor channels such as hearing, sight, and touch [52] Similar to digital avatars, humanoid robots can exhibit human-like communicative behaviors, including social praise, head and torso movements, and nodding, to stimulate more natural conversations [29] 	Physical embodiment	<ul style="list-style-type: none"> Capability to perform difficult or dangerous tasks, provide companionship, and participate in social interactions, especially in circumstances where human interaction is limited [50,53] Physical presence enables more distinct and natural interactions compared to chatbots or virtual avatars as HRI^d differs from traditional HCI^e by involving both linguistic and physical aspects [29,52] 	<ul style="list-style-type: none"> Higher development cost relative to other CAs Lower ease of access for users compared to other CAs [52,54] Higher risk of ethical and safety concerns

^aAI: artificial intelligence.

^bCGI: computer-generated imagery.

^cNLP: natural language processing.

^dHRI: human-robot interaction.

^eHCI: human-computer interaction.

Ethical Concerns of AI Agents

While AI agents have the capability to provide constant health surveillance support, challenges and risks associated with using AI-based CAs in health care remain. These include ethical concerns regarding data collection and interpretability of results, patient safety risks, biases encoded in algorithms, and cybersecurity [55,56]. Furthermore, customers are often reluctant to engage with such AI-based CAs due to many factors such as trust, reliability, learning curve, usability, privacy, and data security that should be addressed in the design, deployment, and use of AI applications [12,29,57-59].

State-of-the-Art Summary

Our study is novel regarding CAs as human-like digital agents for managing NCDs remotely. As AI-based CAs are a relatively new area, limited research has been conducted on applying these emerging technologies in health care. While existing research on CAs as virtual caregivers for NCDs serves as a seminal foundation, such research has other critical limitations. To illustrate, most studies have primarily focused on applications of CAs for mental health conditions, overlooking broader NCDs such as CVDs, metabolic syndrome, or diabetes [60,61]. A recent review [61] that focused on a different set of research questions from ours concluded the following: “A future chatbot could be tailored to metabolic syndrome specifically, targeting all the areas covered in the literature, which would be novel.” Our research looks specifically into this gap.

Furthermore, despite some favorable anecdotal evidence, the effectiveness of CAs in NCD management is seldom explored in large-scale trials, particularly in older adults, who have the highest risk of developing NCDs [62-64]. Some recent previous

studies have conducted systematic reviews [27,65,66] or scoping reviews [61,62,67] on CAs in chronic disease management; however, to the best of our knowledge, no review has been conducted yet on the application of AI-based CAs as human-like digital agents for managing NCDs remotely. Such limitations and gaps as highlighted previously present both challenges and opportunities for research to advance our understanding of how CAs can contribute to managing NCDs.

Aim

This scoping review aimed to provide an overview of the existing evidence and research on using assistive humanoid AI-based CAs in health and social care for managing NCDs. Our primary objective was to explore the impact of AI-based CAs, including embodied avatars, as human-like health carers for the self-management of chronic diseases. By examining the current literature on this topic, we hoped to identify key areas for future research and provide insights into how these technologies can be effectively used in health care and personalize NCD management strategies.

Research Objectives

Our research objectives were as follows: (1) to explore the current state of research on the use of AI-based CAs as human-like virtual health carers for managing NCDs, (2) to identify the potential benefits and challenges associated with the use of these technologies in the health care field, (3) to explore the efficacy of AI-based CAs in the remote management of NCDs, (4) to discover the specific target users primarily studied, and (5) to provide recommendations for future research in this area.

Research Questions

Our research questions were as follows: (1) what is the current state of research on using AI-based CAs as human-like health carers for managing NCDs? (2) What are the limitations or challenges associated with the use of these technologies in health care? (3) What are the potential benefits of using AI-based CAs in managing NCDs, and how can they be effectively used to improve health care delivery and reduce health care burden? (4) What is the efficacy of the CAs in the remote care of NCDs? (5) What are the frequently targeted user groups for such virtual agents (eg, specific age groups and individuals with special needs)?

Methods

Search Strategy

We followed the methodological frameworks proposed by Arksey and O'Malley [68] and Levac et al [69]. Initially, research objectives and questions were formulated, followed by a systematic literature search conducted on July 31, 2023. For primary searching, we used 6 electronic databases that are considered relevant to the research focus—Ovid MEDLINE, Embase, PsycINFO, PubMed, CINAHL, and Web of Science—applying the same set of keywords, such as

“conversational agents,” “artificial intelligence,” and “noncommunicable diseases,” including their associated synonyms (as shown in [Textbox 1](#)). The Boolean operators “*” and “OR” were used to expand and ensure that different word combinations were included. The operator “AND” was used for combining the main search terms to identify articles focusing on AI-based CAs (as health carers) only applicable in the health care field, particularly for managing NCDs or chronic diseases.

Additional studies were identified by hand searching the reference lists of included studies and relevant review articles. Furthermore, a supplementary manual search was conducted to identify specific articles from diverse sources, including ProQuest Central, ResearchGate, ACM Digital Library, and Google Scholar. Some of these articles were identified through pilot hand searching and preselected as they were closely relevant but not retrieved from the selected databases used for primary searching, whereas others were discovered through manual searches targeting specific authors well known in this field. The search strategy underwent refinement following expert recommendations, including insights from a coauthor with digital health expertise, who advised excluding “robots” from the search terms, recognizing it as a distinct field. These methodological steps collectively ensured a comprehensive exploration of the relevant literature.

Textbox 1. Search query example.

Example

- (“conversational agent*” OR “relational agent*” OR “virtual agent*” OR “dialogue agent*” OR “dialogue system*” OR “virtual assistant*” OR “chatbot*” OR “voice assistant*” OR “voicebot*” OR “voice-bot*” or “voice bot*” OR “humanoid *bot*” OR “social *bot*” OR “avatar*” OR “human-like avatar*” OR “anthropomorphic avatar*” OR “digital human*” OR “human digital twin*” OR “virtual human*”) AND (“intelligent” OR “artificial intelligence” OR “AI” OR “AI-based”) AND (“health” OR “healthcare” OR “caregiver” OR “self-management” OR “self-monitor*” OR “non-communicable disease*” OR “noncommunicable disease*” OR “chronic disease*”)

Eligibility Criteria

Our scoping review used comprehensive inclusion and exclusion criteria to ensure the selection of pertinent studies. We did not impose any limitations based on gender or age groups in the selection of articles. The search scope was confined to scholarly articles in English published between January 2010 and July 2023, aligning with the substantial rise in CAs after 2010 [70], notably with the introduction of Apple’s Siri in 2011 [71]. Evidently, most of the selected papers were on recent studies conducted within the last 5 years due to the accelerated technological advancements and the latest evolution of AI. Specifically, we focused on empirical studies exploring CAs applied exclusively to human interaction within the health care context, emphasizing their role in the remote management of NCDs, ideally within home environments.

The exclusion criteria comprised the exclusion of conference abstracts, posters, reviews, protocols, position papers or viewpoints, and certain types of studies such as those involving noninteractive robotic devices. We also excluded studies related to CAs designed for medical education and non-patient-centered applications in hospital settings and those addressing specific health care domains such as surgery, dentistry, pregnancy or maternity, addiction or substance use disorders, and communicable diseases. Furthermore, we excluded studies

centered solely on medical history data storage, telephone monitoring, data set construction methods, or user evaluations of commercial CAs for health care without their practical applications in the remote management of NCDs. Our criteria aimed to streamline the focus on patient-centered CAs contributing to the self-management and remote monitoring of NCDs, eliminating studies with a primary emphasis on clinical interviews, disease prediction, or decision support without a social interaction element.

Screening and Selection

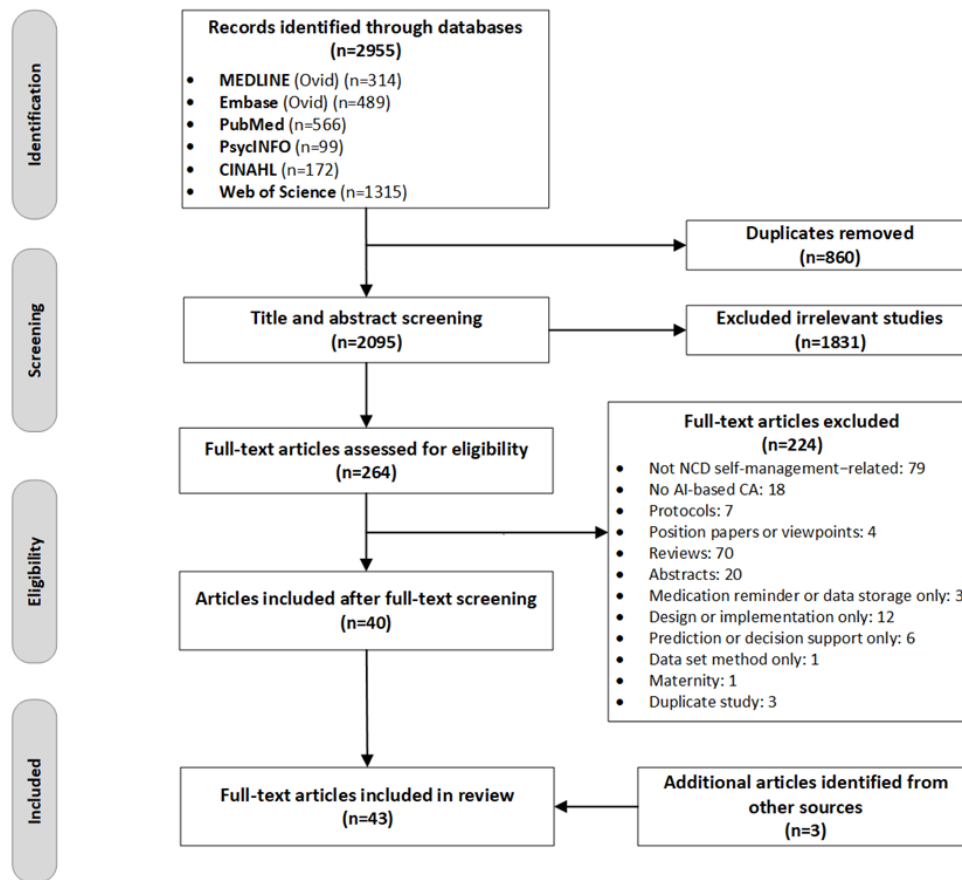
In total, 2 of the authors independently searched each database. Titles and abstracts were screened for inclusion according to the aforementioned criteria, followed by an exclusion of duplicates, unrelated studies, and articles that could not be retrieved. The abstract screening yielded 264 articles eligible for full-text screening, of which 70 (26.5%) were review papers comparing different types of CAs in the health care field and 156 (59.1%) were empirical studies that explored different types of CAs used in the prevention, treatment, or rehabilitation of chronic diseases involving consumers, caregivers, and health care professionals.

Subsequently, the authors screened full texts of the remaining articles independently, and 40 full-text articles that met the inclusion criteria were selected for review. In addition, 10 [10]

specific articles were obtained from hand searches of other sources (Google Scholar, ProQuest, ACM Digital Library, and ResearchGate), of which 3 (30%) [3] relevant ones were selected

upon full-text screening. It was an iterative process; any discrepancies were discussed among the authors. The search and selection process is illustrated in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the search and screening process. AI: artificial intelligence; CA: conversational agent; NCD: noncommunicable disease.



Data Collation and Reporting

An Excel (Microsoft Corp) spreadsheet was initially created to aid the screening and selection process. Following the screening process, a total of 43 articles were eventually selected for synthesis. Quantitative and qualitative data from the included studies were extracted and summarized in a tabular format, including information such as intervention, type of CA, target population, number of participants with their age, methods, study duration, location, measures and outcomes, and limitations. The relevant extracted information was collated and summarized using the narrative synthesis approach, which was deemed appropriate for capturing the breadth of evidence in scoping reviews, identifying themes aligned with the research questions as well as patterns observed across the included studies and relevant reviews. The results were reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. Given the divergent methodologies of the included studies, predominantly using mixed methods and qualitative approaches, no quantitative synthesis or meta-analysis was conducted.

Results

Characteristics of the Included Studies

Most of the included studies (22/43, 51%) were feasibility, acceptability, and usability studies (Table 2). These studies mostly found positive results in terms of feasibility and usability of the proposed CAs. Some studies (4/43, 9%) used the System Usability Scale questionnaire to measure the system’s usability score and predominantly found high usability, with System Usability Scale scores of >70 [72-75]. Alternatively, multiple studies (10/43, 23%) used the subjective ratings of most of the participants to evaluate the feasibility and acceptability of CAs by applying indicators such as perceived usefulness [76-79], ease of use [78,80], user satisfaction [80,81], engagement rate [80,82,83], perceived closeness [73], and Net Promoter Score [84]. However, 2% (1/43) of the studies reported negative usability and acceptability outcomes, indicating unsatisfactory reliability, usability, and goal structure, which hindered health care professionals’ acceptance and trust [85].

Table 2. Characteristics of the included studies.

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
Watson et al [86], 2012	<ul style="list-style-type: none"> Behavior change intervention for obesity self-management Duration: 12 weeks (divided into four 3-week periods—period 1, period 2, period 3, and period 4) 	<ul style="list-style-type: none"> 2D-animated human-like avatar (rule-based AI^b) Platform: desktop and laptop computers (web based) 	<ul style="list-style-type: none"> Adults with overweight or obesity N=62 Participants aged 20-55 years (inclusive) 	<ul style="list-style-type: none"> Pretest-posttest observational study (quantitative)—2-arm RCT^c Location: United States 	<ul style="list-style-type: none"> Step count percentage: <ul style="list-style-type: none"> No significant percentage change in step count between intervention and control arms from start to end (2.9% vs -12.8%, respectively; $P=.07$) Significant difference between percentage change in step count across all study periods in the intervention vs control arms ($P=.02$) Secondary outcomes: no significant changes in secondary outcomes—BMI, 7-day physical activity recall, physical activity stage of change, self-efficacy and exercise benefits and barriers, and program satisfaction (eg, mean decrease in BMI of 0.04 and 0.25 in control and intervention groups, respectively; $P=.44$) Average step count: significant decrease in the average mean step count for the control group from period 1 (7174) to period 4 (6149; $P=.01$) but no significant change for the intervention group from period 1 (6943) to period 4 (7024; $P=.85$) Mean activity level: significant percentage change in mean activity levels between period 1 and period 3 ($P=.02$) but no significant change between period 1 and period 2 ($P=.12$) and between period 1 and period 4 ($P=.07$) 58% of the intervention participants agreed that the virtual coach influenced their increased activity 	<ul style="list-style-type: none"> Participants were primarily White, college-educated women, limiting generalizability of findings to a wider patient population with overweight or obesity Lack of baseline step count data for participants, although survey results showed no substantial baseline activity level differences Initial observed increase in step count after enrollment likely reflects a change from baseline, but without baseline data, this cannot be confirmed

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
Ki-mani et al [80], 2016	<ul style="list-style-type: none"> Lifestyle intervention and patient counseling support for AF^d Duration: not identified 	<ul style="list-style-type: none"> Animated human-like avatar (network-based AI with XML scripting language) Platform: mobile devices or smartphones 	<ul style="list-style-type: none"> Patients with AF N=16 (5 female and 11 male) Participants aged ≥18 years (20-58 years) 	<ul style="list-style-type: none"> Feasibility assessment—mixed methods pilot study with self-report scale measures and a semistructured interview Location: not identified 	<ul style="list-style-type: none"> High overall user satisfaction with the agent (mean 3.45 on a 4-point scale) and the AliveCor heart rhythm monitor (mean 3.54) High ratings for ease of use (mean 3.54) Interaction duration: participants reported a 7- to 10-minute-long interaction with the agent, whereas older participants reported longer interactions due to content relevancy Agent's user satisfaction correlated with the participants' satisfaction with the use of the AliveCor heart rhythm monitor Acceptable feasibility of delivering AF counseling via a smartphone-based humanoid avatar as a virtual agent 	<ul style="list-style-type: none"> Small sample size and limited patient diversity Short-term study duration and no long-term use evaluation No control group for comparison to provide additional insights into the effectiveness of the virtual agent No assessment of the virtual agent's impact on patient outcomes or behavior change
Samkhi et al [87], 2017	<ul style="list-style-type: none"> Behavioral intervention for self-management of stress and depression Duration: 9-21 weeks 	<ul style="list-style-type: none"> 3D-animated human-like avatar (rule-based AI) Platform: touchscreen tablets 	<ul style="list-style-type: none"> Adults with chronic pain and depression N=154 Participants aged ≥18 years 	<ul style="list-style-type: none"> Mixed methods observational study—2-arm RCT Location: United States 	<ul style="list-style-type: none"> Impact on patient performance and satisfaction: significant positive stress management behaviors among the CA intervention participants compared to the control group after 9 weeks (t136=3.74; <i>P</i><.001) Avatar (Gabby) was found to be very useful for group visits, allowing for review of class lessons and detailed information at any time and location Most participants found the meditation, yoga, and mindfulness sessions provided by Gabby to be very "useful" and "relaxing" Participants were very likely to recommend Gabby to others 	<ul style="list-style-type: none"> Relatively small sample size from health centers in the Boston area, limiting generalizability of findings to other populations Inability to isolate the effects of face-to-face weekly group visits from those of the home-based CA Potential usability issues, especially for less computer-literate users; may affect Gabby's user satisfaction
Bickmore et al [10], 2018	<ul style="list-style-type: none"> AF self-management through symptom monitoring and patient counseling Duration: 30 days 	<ul style="list-style-type: none"> 3D-animated human-like avatar (network-based scripted AI) Platform: mobile devices or smartphones 	<ul style="list-style-type: none"> Older adult patients with AF N=120 Participants aged ≥60 years (mean 72.1, SD 9.10 years) 	<ul style="list-style-type: none"> Mixed methods pretest-posttest (observational) study—quasi-experimental demonstration of CA and 2-arm RCT for usability and effectiveness evaluation Location: United States 	<ul style="list-style-type: none"> Change in AFEQT^e score: significantly higher AFEQT score among the intervention group participants after 30 days compared to those in the control group (<i>P</i><.05) Over 89.7% of participants rated the CA positively, whereas 4.4% rated it negatively and 5.9% had mixed opinions Participants praised the personalized interaction with the avatar 	<ul style="list-style-type: none"> Relatively small convenience samples with limited participant details and a singular focus on a specific chronic condition (AF) may impact generalizability Lack of long-term, objectively measured health outcomes Insufficient emphasis on the complexity of self-care management regimens, especially for patients with low health literacy

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
Cheng et al [58], 2018	<ul style="list-style-type: none"> Personalized patient education and medication adherence intervention for improved self-management of T2DM^f Duration: 1 month 	<ul style="list-style-type: none"> AI voice assistant and voicebot (ML^g based) Platform: Google Home devices (API.AI platform) 	<ul style="list-style-type: none"> Older adult patients with T2DM N=10 Participant age not identified 	<ul style="list-style-type: none"> Feature-based comparison between the proposed voicebot and other similar mHealth^h apps and usability evaluation—qualitative assessment Location: not identified 	<ul style="list-style-type: none"> AADEⁱ requirement evaluation: the feature-based comparison indicates that most similar mHealth apps fail to meet the AADE requirements for effective T2DM self-management compared to the proposed voicebot Practical usability evaluation: There were more satisfied users than unsatisfied ones, primarily due to the voicebot's speaker functionality and natural conversation flow Unsatisfied users reported difficulty learning commands and limited answer choices 80% of older adult participants would prefer using Healthy Coping on Google Home over a smartphone 	<ul style="list-style-type: none"> Voicebot's limitations Narration speed concern: older adult users may struggle with application commands due to narration speed; a speed setting could allow users to adjust the pace Accessibility limitations: Healthy Coping's voice interface may not fully support users with hearing or speech disabilities; integrating devices such as Bluetooth-enabled hearing aids could improve accessibility
Suganuma et al [88], 2018	<ul style="list-style-type: none"> Internet-based cognitive behavioral intervention for self-monitoring of mental health Duration: 1 month 	<ul style="list-style-type: none"> AI (rule-based) chatbot via SABORI (mental health) app Platform: web based 	<ul style="list-style-type: none"> Adults with psychological distress and mental health problems N=454 (intervention group: n=191; control group: n=263) Participants' age: mean 38.07 (SD 10.75) years 	<ul style="list-style-type: none"> Feasibility and acceptability (pilot) study—non-randomized prospective study with quantitative pretest-posttest intervention questionnaire Location: Japan 		<ul style="list-style-type: none"> Nonrandomized comparison between control and experimental groups Short-term study duration (1 month) for assessing behavioral activation effects Recruitment of healthy participants in control group may have influenced outcomes, resulting in lack of effect for behavioral activation factors of avoidance and rumination

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
Hus-sain and Athula [89], 2018	<ul style="list-style-type: none"> AI-ML chatbot (VDMS⁴) Platform: computer or mobile devices (web based) 	<ul style="list-style-type: none"> Patients with diabetes and their carers seeking diabetes-related information N=10 Participants aged 20-50 years 	<ul style="list-style-type: none"> Japanese version of the WHO-5^j: a 1-factor scale to measure positive mental health based on physical aspects using a 5-item, 6-point Likert scale On the variable of positive mental health, a significant main effect ($P=.02$) and interaction for time ($P=.02$), including a significant simple main effect for the intervention group ($P=.002$) K10^k: a 1-factor scale to measure negative mental health based on physical aspects using a 10-item, 5-point Likert scale On the variable of negative mental health, a significant main effect ($P=.02$) and interaction for time ($P=.005$), including a significant simple main effect for the intervention group ($P=.001$) BADS^l: a scale to measure 4 factors—"Activation" (BADS-AC^m), "Avoidance/Rumination" (BADS-ARⁿ), "Work/School Impairment" (BADS-WS^o), and "Social Impairment" (BADS-SI^p) 2-way ANOVA test showed a significant trend for behavioral activation, indicating the possibility for a certain degree of effectiveness BADS-AC: a significant main effect for time ($P=.10$), including a significant simple main effect trend in the intervention group ($P=.06$) BADS-AR: no significant main or interaction effect 	<ul style="list-style-type: none"> Evaluation metrics (VDMS vs other sources): Correct replies: VDMS=65%; search engine=80% Satisfactory user satisfaction level of VDMS (very close to that of the search engine) due to VDMS's timely response and correct answers' quality and clarity 		

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
	<ul style="list-style-type: none"> To provide patient education for improved self-management of diabetes Duration: not identified 			<ul style="list-style-type: none"> Quantitative and qualitative study—performance evaluation of VDMS compared to other information sources (eg, search and websites) Location: not identified 		<ul style="list-style-type: none"> Small sample size (10 participants) may restrict diversity and evaluation scope VDMS chatbot's implementation is incomplete and lacks full testing Conversation is mostly controlled by the chatbot, limiting user input Extracted data from Wikipedia may lack reliability as no published work has explored integrating Wikipedia knowledge with a diabetes management chatbot
Inkster et al [90], 2018	<ul style="list-style-type: none"> To deliver positive psychology and mental well-being techniques for improved self-management of depression and anxiety Duration: 8 weeks 	<ul style="list-style-type: none"> Commercial CA: text-based AI chatbot (Wysa) Platform: mobile devices or smartphones 	<ul style="list-style-type: none"> Individuals with self-reported symptoms of depression N=not identified Participants' age not identified 	<ul style="list-style-type: none"> Quasi-experimental (pretest-posttest) design—mixed methods study using the Wysa app's inbuilt assessment questionnaire (for effectiveness evaluation) Location: global (not specified) 	<ul style="list-style-type: none"> Impact (pre-post) analysis outcomes: PHQ-9^f score (within groups) measured using a Wilcoxon signed rank test: significant reduction in PHQ-9 score among both high users ($P<.001$) and low users ($P=.01$), indicating an improvement in depression symptoms from before to after the intervention High users' group showed significantly higher average mood improvement compared with the low users' group (Mann-Whitney $P=.03$, with a moderate effect size of 0.63) User engagement and feedback: 73.6% of users provided at least one response to the in-app feedback questions (indicating high user engagement), and 60.9% of them reported feeling better after app use 	<ul style="list-style-type: none"> Small sample size restricts investigation of user reactions to app design elements Small and unbalanced comparison group sizes undermine the findings' generalizability Limited detailed feedback on users' app experience hinders qualitative analysis Lack of a randomized controlled environment may introduce biases Absence of users' previous health information hampers comprehensive understanding Quasi-experimental design used, slightly lower in design quality compared to interrupted time-series designs
Neerincx et al [85], 2019	<ul style="list-style-type: none"> Behavioral intervention and patient education for self-management of T1DM[†] Duration: 3 months 		<ul style="list-style-type: none"> Children with T1DM N=49 Participants aged 7-14 years 	<ul style="list-style-type: none"> Qualitative study—prototype design and usability evaluation using an iterative, incremental development process Location: Italy and the Netherlands 		

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
		<ul style="list-style-type: none"> Humanoid robot (PAL^u) and its robotlike digital avatar version for home (MyPal; rule based+ML with cloud computing) Platform (avatar): tablets 			<ul style="list-style-type: none"> Unsatisfactory reliability, usability, and goal structure of PAL system, which hindered the acceptance and trust of health care professionals Improvement in children's diabetes knowledge when using the PAL system and increased motivation in performing diabetes-related activities due to enjoyable interaction with the PAL robot and avatar Improvement suggestions: personalization enhancement required to establish patients' adherence Innovative PAL functions (eg, experience-sharing function) were identified and tested with positive results Reduced bonding effect with higher perceived similarity between the robot and its digital avatar Most children stopped using the avatar (MyPAL app) some weeks after the study started 	<ul style="list-style-type: none"> Small sample size may affect the findings' generalizability Limited study duration may not capture the PAL system's long-term effects on T1DM management Comparison between physical robot and avatar function may introduce biases in user preference assessment Insufficient detailed information on AI algorithms or ML techniques used restricts insights into the PAL system's technical aspects
Easton et al [91], 2019	<ul style="list-style-type: none"> Self-management for people with comorbid long-term conditions and mental health problems Duration: not identified 	<ul style="list-style-type: none"> AI chatbot with an animated human-like 2D avatar (Avachat) Platform: computers, tablets, smartphones, and televisions 	<ul style="list-style-type: none"> Older adults with comorbid long-term conditions and mental health problems N=10 Participants aged ≥55 years (56-86 years) 	<ul style="list-style-type: none"> Co-design and acceptability testing involving stakeholders—qualitative study with snowball sampling and workshops for initial user requirement gathering of prototype and user feedback Location: United Kingdom 	<ul style="list-style-type: none"> Acceptability and feasibility: Patients found Avachat to be helpful, informative, and easy to use Health care professionals are optimistic about the avatar's potential to improve patient outcomes and reduce the health care burden Improvement suggestions: enhance personalization of contents and ensure accessibility for users with visual and hearing impairments 	<ul style="list-style-type: none"> Small sample of White, British, medically stable, and regional participants recruited may limit findings No current mental health problems reported by participants, although past instances of low mood or worry were mentioned
Chaix et al [92], 2019			<ul style="list-style-type: none"> Patients with breast cancer and their relatives N=4737 Participants' age not identified 	<ul style="list-style-type: none"> Quantitative study using user-chatbot conversational data Location: not identified 		

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
	<ul style="list-style-type: none"> To provide personalized patient education for improved quality of life and medication adherence in patients with breast cancer Duration: 1 year 	<ul style="list-style-type: none"> Commercial CA—AI chatbot (Vik) Platform: mobile devices or smartphones (iOS or Android) and web browser via Messenger app 			<ul style="list-style-type: none"> Users' interactivity level: average of 132,970 messages exchanged per month, with 147 total average interactions per question (for open-ended questions), resulting in 2.7 interactions per person per question Overall user satisfaction rate: 93.9% (900/958), and 88% (843/958) found Vik helpful in following treatment effectively Average medication adherence rate of patients improved by >20% ($P=.04$) after using Vik for 4 weeks 	<ul style="list-style-type: none"> Absence of a control group for comparison Lack of long-term evaluation of the chatbot's impact on clinical outcomes Reliance on users' self-reported data, introducing potential bias No investigation of the viewpoints of health care providers or other stakeholders in patient care Potential technical issues or limitations of the chatbot not addressed
Stephens et al [76], 2019	<ul style="list-style-type: none"> Behavioral intervention for constant management of obesity and prediabetes symptoms Duration: 6 months 	<ul style="list-style-type: none"> Commercial CA—AI behavioral coaching chatbot (Tess) Platform: mobile devices or smartphones via messaging apps (eg, WhatsApp and Messenger); Google Home or Amazon Alexa (for voice conversations) 	<ul style="list-style-type: none"> Adolescent pediatric patients with obesity and prediabetes symptoms N=23 Participants aged 9-19 years 	<ul style="list-style-type: none"> Feasibility study—mixed methods approach with qualitative interviews, usability testing, and quantitative surveys (SUS^v questionnaire) Location: United States 	<ul style="list-style-type: none"> Positive progress toward goals was reported by the participants 81% of the time 4123 messages were exchanged, with patients rating the chatbot's usefulness 96% of the time, indicating its high perceived usefulness and feasibility among adolescents 	<ul style="list-style-type: none"> Small sample size, causing limited generalizability Gradual program adjustments may introduce variability and potential inconsistency Lack of an experimental design to control for factors Inability to ensure detection of a treatment effect
Balsa et al [72], 2020	<ul style="list-style-type: none"> Behavioral intervention for self-management of T2DM Duration: 8-10 days 	<ul style="list-style-type: none"> Animated human-like avatar (Vitória) in a 3D environment (rule-based AI) Platform: mobile devices or smartphones 	<ul style="list-style-type: none"> Older adults with T2DM N=20 Participants aged ≥65 years (67-80 years) 	<ul style="list-style-type: none"> Usability study—qualitative assessment Location: Portugal 	<ul style="list-style-type: none"> Usability evaluation for T2DM medication adherence and lifestyle improvement Usability (SUS) score: 73.75 (SD 13.31; indicates significantly high usability of Vitória) 	<ul style="list-style-type: none"> Small sample size and skewed participant sample toward expert technology users Reliance on previous experience and available resources for sample size estimation Criticism of questionnaires for yielding only overall measures without addressing specific concerns Potential bias in field-testing compared to laboratory setting for usability tests

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
Gong et al [24], 2020	<ul style="list-style-type: none"> Behavioral intervention with personalized support and motivational coaching for remote self-management of T2DM Duration: 12 months 	<ul style="list-style-type: none"> Human-like (text+voice enabled) avatar (Laura; BCT^w-based prescribed AD) Platform: mobile devices or smartphones 	<ul style="list-style-type: none"> Adults with T2DM N=187 Participants aged ≥18 years (mean 57, SD 10 years) 	<ul style="list-style-type: none"> Pretest-posttest (quantitative) observational study—2-arm, open-label RCT for adoption, usefulness, and effectiveness evaluation Location: Australia 	<ul style="list-style-type: none"> Program adoption and use: Number of valid chats with Laura completed per person: 18.4 (SD 15.0; range 1-53) Total duration of valid chats per person: 242.7 (SD 212.3; range 0-1050) minutes Number of glucose level uploads per person: 181.8 (SD 192.1; range 1-966) Number of clinical alerts: total=297; average per month=13.7 (SD 8.8) Number of technical alerts: total=179; average per month=8.3 (SD 6.5) Number of posts on the web-based discussion forum: total=19; average per month=1.1 Program effectiveness (in terms of coprimary and secondary outcomes)—coprimary outcomes: Statistically significant between-arm difference at 12 months in the mean change in HRQoL^x (AQoL-8D^y utility value: 0.04, 95% CI 0.00-0.07; <i>P</i>=.04) Reduction in HbA_{1c}^z levels during the trial but no statistically significant between-arm difference at 6 months (0.06, 95% CI -0.35 to 0.47; <i>P</i>=.78) or 12 months (-0.04, 95% CI -0.45 to 0.36; <i>P</i>=.84) Significant improvement in HRQoL from baseline to 12 months (mean estimated change in AQoL-8D score: 0.04, 95% CI 0.01-0.06; <i>P</i>=.007) Increase in the score of the physical health and mental health subscales compared with baseline Secondary outcomes: Significant between-arm difference in the mean change in the HADS^{aa} anxiety score at 6 months (-0.89, 95% CI -1.74 to -0.04; <i>P</i>=.04) but not at 12 months or for other secondary outcomes reported 	<ul style="list-style-type: none"> Small sample size may restrict the generalizability of the findings, and subgroup analyses require cautious interpretation Absence of blinding of participants and their GPs^{ab} to the study arm allocation might lead to potential self-report bias and Hawthorne effects Control arm participants showed a higher rate of completed assessments, possibly due to their interest in program access or higher attrition in the intervention arm Subgroup analyses were underpowered, with multiple testing increasing the risk of false positives

Issom et al [77], 2020

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
	<ul style="list-style-type: none"> Health behavioral intervention for self-management of SCD^{ac} Duration: not identified 	<ul style="list-style-type: none"> AI chatbot Platform: mobile devices or smartphones 	<ul style="list-style-type: none"> People with SCD N=19 Participants' age not specified 	<ul style="list-style-type: none"> Preliminary feasibility study—quantitative posttest survey Location: France 	<ul style="list-style-type: none"> 88% of participants rated the following question—"The chatbot contains all the information I need"—with at least 3/4, and its total score was 54/68 58% rated the following question—"The chatbot encouraged me to be more active in order to improve my condition"—with at least 3/4, and its total score was 51/68 Results indicate high perceived usefulness of the chatbot in promoting knowledge and motivation for improved self-care practices 	<ul style="list-style-type: none"> Small sample size: only 17 participants completed the evaluation, potentially limiting generalizability 2 withdrawals due to smartphone issues may have impacted data collection and results Reliance on self-reported data in the posttest survey could introduce bias or inaccuracies Limited scope: evaluation focused solely on perceived usefulness of the chatbot, lacking assessment of long-term impacts on self-care or health outcomes Lack of comparison with other support forms or interventions hinders assessment of relative effectiveness
Anastasiadou et al [93], 2020	<ul style="list-style-type: none"> To offer continuous health education and interaction in English, Spanish, and Bulgarian for self-management of diabetes Duration: 6 months 	<ul style="list-style-type: none"> Multilingual AI-ML chatbot (EVA^{ad}) Platform: mobile devices or smartphones 	<ul style="list-style-type: none"> People diagnosed with diabetes N=not identified Participants' age not identified 	<ul style="list-style-type: none"> Qualitative pilot study—validation and acceptance evaluation by integrating EVA into an mHealth app (CHRODIS PLUS Joint Action) to collect data on user queries and responses provided based on the educational content Location: Greece 	<ul style="list-style-type: none"> Users sent a total of 940 unique messages to EVA Users' common questions related to diabetes varied based on EVA's language: <ul style="list-style-type: none"> English: self-measurement and understanding what diabetes is Spanish: glucose self-monitoring Bulgarian: insulin and high blood pressure A comprehensive analysis of the effectiveness of EVA in improving diabetes management or patient outcomes was not provided 	<ul style="list-style-type: none"> Lack of detailed user demographic information may affect generalizability Study duration limited to 6 months, possibly overlooking long-term user interactions and feedback Absence of specified measures to address potential biases in user interactions or data collection No discussion about potential technical limitations and challenges during EVA system development and testing
Roca et al [94], 2021		<ul style="list-style-type: none"> AI-ML-based chatbot Platform: mobile devices or smartphones 	<ul style="list-style-type: none"> People with comorbid T2DM N=13 Participants aged ≥18 years 	<ul style="list-style-type: none"> Quantitative pilot trial—pretest-posttest observational study (1 arm) Location: Spain 		

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
	<ul style="list-style-type: none"> Medication adherence intervention for improved self-management of T2DM and depressive disorder Duration: 9 months 				<ul style="list-style-type: none"> Medical outcomes: <ul style="list-style-type: none"> Significant improvement in the average HbA1c level and PHQ-9 scores from 7.6 (SD 0.7) to 7.3 (SD 0.8) and from 13.2 (SD 6.0) to 8.6 (3.6), respectively Reduction in the number of physical medical consultations per month from 2.0 (SD 2.6) to 1.3 (SD 1.5) in 30.8% of the patients Health care professionals participating in the study found the chatbot useful in improving medication adherence Chatbot feedback: <ul style="list-style-type: none"> Participants who used the chatbot daily found it useful in fulfilling their medication reminder needs The chatbot's language and vocabulary were appropriate and easy to understand 38% of participants had difficulty learning to use the chatbot, and 15.4% reported the chatbot's inability to understand the users' instructions Almost 70% of the patients (9/13) expressed willingness to continue using the chatbot after the study 	<ul style="list-style-type: none"> Small sample size (13 participants) may limit generalizability of the study findings Some patients did not update medication information after the initial configuration, leading to reminders being stopped Patients require digital literacy or assistance for configuring the messaging platform Reminder sounds may be accidentally disabled by patients, affecting the use of the chatbot
Yao et al [95], 2021	<ul style="list-style-type: none"> Remote monitoring of patients after stroke and automation of suspected stroke screening Duration: not identified 	<ul style="list-style-type: none"> Hyperrealistic human-like 3D avatar (iLAMA; ML based with computer vision algorithms) Platform: mobile devices or smartphones and tablet 	<ul style="list-style-type: none"> Older adult patients after stroke <ul style="list-style-type: none"> N=140 (videos) Participants aged ≥60 years 	<ul style="list-style-type: none"> Quantitative beta testing of the prototype on 140 videos with volunteers via email Location: not identified 	<ul style="list-style-type: none"> The app was able to recognize body parts and extract 68 facial landmarks from the facial videos The app could provide accurate stroke screening results for neurologists and stroke specialists to review No major technical issues or problems with the app during beta testing were reported iLAMA has the potential to improve stroke assessment and care, particularly in areas with limited access to stroke specialists 	<ul style="list-style-type: none"> Participants were regular volunteers (friends and family) rather than patients after stroke Beta testing involved a relatively small sample size of 140 videos The app may struggle to differentiate minor mistakes from actual stroke signs, especially without supervision Screening process takes approximately 5 minutes, which some users may find burdensome
Krishnakumar et al [82], 2021			<ul style="list-style-type: none"> Adults with T2DM possessing Android smartphones N=102 Participants aged ≥18 years 			

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
	<ul style="list-style-type: none"> Lifestyle intervention and self-monitoring of diet, exercise, weight, and blood glucose for T2DM management Duration: 16 weeks 	<ul style="list-style-type: none"> AI chatbot via mHealth therapeutic app (Wellthy Care) Platform: Android smartphones 		<ul style="list-style-type: none"> Retrospective study design—pretest-posttest (quantitative) observational study for effectiveness evaluation Location: India 	<ul style="list-style-type: none"> Primary outcomes: <ul style="list-style-type: none"> Statistically significant reduction in HbA1c levels from baseline to 16 weeks ($P < .001$) Secondary outcomes: <ul style="list-style-type: none"> Mean FBG^{ae} and PPBG^{af} levels decreased significantly from baseline to 16 weeks ($P < .001$) Mean BMI and weight decreased significantly from baseline to 16 weeks ($P < .001$) Average duration of interactions was 106 minutes with the app (87 participants; 95% CI 65-147 minutes) and 88 minutes with the AI chatbot (102 participants; 95% CI 66-110), indicating a significant positive association with program engagement 	<ul style="list-style-type: none"> Short program duration No control group as real-world data were retrospectively analyzed Selection bias from multiple selection approaches (physician recommended and voluntary approach) of participants Reliance on self-reported disease biomarkers Loss of data during follow-up Variation in the number of male and female participants, introducing potential bias
Egede et al [96], 2021	<ul style="list-style-type: none"> To improve mental health (depression and anxiety) assessments and remote counseling via VH^{ag} mediation Duration: not identified 	<ul style="list-style-type: none"> Human-like avatar (Greta; Wizard-of-Oz AI system using prescribed text and human-controlled responses) Platform: not identified 	<ul style="list-style-type: none"> Adults with anxiety and depression N=56 Participants aged ≥ 18 years (18-45 years) 	<ul style="list-style-type: none"> Mixed methods comparison and UX^{ah} evaluation study: Behavioral expressiveness comparison between VH system and text-only system users UX assessment and task ratings using UEQ^{ai} and TAM^{aj} questionnaires Location: Japan 		<ul style="list-style-type: none"> Small sample size may limit depth of analysis, particularly in comparing depression severity classes Combination of moderate and moderately severe depression groups may restrict findings; a 4-class comparison could yield more valuable data Focus on VH mediation may have overshadowed other factors influencing user engagement and task performance Long-term effects of VH-mediated tasks on user engagement and mental health outcomes were unexplored, limiting insights into intervention sustainability

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
Ro-moosi et al [97], 2021		<ul style="list-style-type: none"> AI chatbot (Elomia; NLP^{pk}+ML based) Platform: not identified 	<ul style="list-style-type: none"> Young adults with depression, anxiety, and low mood N=412 (202 female and 210 male) Participants aged ≥18 years (19-23 years) 		<ul style="list-style-type: none"> Stronger visual cues produced by VH avatar guidance compared to non-VH ($P=.04$; 0.036 for 2 facial action units indicating a statistically significant difference in the frequency of activation of these facial action units between the VH and non-VH systems; $P=.005$ and 0.0032 for 2 head pose descriptors indicating a statistically significant difference in head pose characteristics between the VH and non-VH systems) VH users make wider and broader head movements (indicating higher behavior expressiveness) compared to text-only users ($P=.005$ and 0.0032 for 2 head pose descriptors) VH mediation improves users' inclination toward tasks, as indicated by the UEQ and TAM questionnaires Significant differences ($P=.05$) found between the system modes in task ratings for facial expression mimicking ($P=.04$), emotion recall ($P=.01$), thematic apperception ($P=.03$), and mindfulness ($P=.04$) tasks, with VH mode's mean ratings significantly higher than those for the text-only mode except for the emotion recall task, where VH's ratings were lower Groups with moderate depression exhibit more visible activity compared to groups with less depression ($P=.04$ and 0.004 for 2 facial action units) Significant reduction in the high tendency toward depression (up to 28%) Significant reduction in the high tendency toward anxiety (up to 31%) Significant reduction in the high tendency toward negative effects (up to 15%) through the regular use of Elomia 	

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
	<ul style="list-style-type: none"> Cognitive-behavioral intervention for self-management of mental health (anxiety, depression, and low mood) Duration: 4 weeks 			<ul style="list-style-type: none"> Pretest-posttest observational study—effectiveness evaluation using psychological quantitative research methods through a controlled experiment in 3 stages: formation of experimental and control samples, baseline testing, and final testing Location: Ukraine 		<ul style="list-style-type: none"> Relatively small sample size and focus on specific student age group may limit generalizability of study findings to other populations Short 4-week study duration may be insufficient to assess Elomia's long-term effectiveness No follow-up assessment may restrict evaluation of Elomia's sustainability Absence of qualitative analysis of user feedback may limit UX insights Lack of comparison with other web-based psychological services or traditional face-to-face therapy may produce a less comprehensive evaluation of the chatbot's effectiveness
Anan et al [98], 2021	<ul style="list-style-type: none"> Lifestyle intervention to support adherence to exercises for self-management and improvement in musculoskeletal symptoms Duration: 12 weeks 	<ul style="list-style-type: none"> AI chatbot Platform: mobile devices or smartphones via a messaging app (LINE) 	<ul style="list-style-type: none"> Employees with musculoskeletal symptoms (eg, neck and shoulder stiffness and low back pain) and smartphone users N=121 (intervention group: n=61; control group: n=60) Participants' age not specified 	<ul style="list-style-type: none"> Pretest-posttest (quantitative) observational study—2-arm RCT following CONSORT-EHEALTH^{al} guidelines Location: Japan 		<ul style="list-style-type: none"> Limited generalizability due to single-company study setting No information collected on causes, diagnosis, and treatment status of neck and shoulder pain and low back pain Pain improvement observed in the intervention group may be influenced by treatment changes and occupational factors, not only by the intervention. Lack of long-term follow-up to assess intervention sustainability

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
Zisis et al [99], 2021	<ul style="list-style-type: none"> To provide patient education and support self-management of acute decompensated HF^{ao} Duration: 12 weeks 	<ul style="list-style-type: none"> AI-based avatar (HF-Coach; NLP+ML based) Platform: mobile devices or smartphones and tablets 	<ul style="list-style-type: none"> Patients with chronic acute decompensated HF N=200 Participants' age: mean 55 years 	<ul style="list-style-type: none"> RCT—quantitative study using patient medical records and oral interviews Location: Australia 	<ul style="list-style-type: none"> Primary outcomes: <ul style="list-style-type: none"> Statistically significant improvement ($P<.001$) in the average pain level (on a scale of 1 [highest] to 5 [lowest]) of the neck and shoulder stiffness or pain or low back pain in the intervention group (average pain level=3.0, SD 1.1), compared to the control group (average pain level=4.0, SD 0.8) Proportion of participants with severe symptoms significantly decreased from 77% (37/48) to 33% (16/48) in the intervention group, whereas the decrease in the control group was from 76% (33/46) to 67% (31/46) Significant improvements in the severity of the neck and shoulder pain and stiffness and low back pain in the intervention group compared to the control group (OR^{am} 6.36, 95% CI 2.57-15.73; $P<.001$) Subjective assessment: 75% in the intervention group showed symptom improvement, whereas only 3% in the control group showed improvement ($P<.001$) Secondary outcomes: <ul style="list-style-type: none"> The AI-based health intervention was effective in improving the EQ-5D-5L score and the RMDQ^{an} score in the intervention group No significant differences between the intervention and control groups in mood (GAD-7^s and PHQ-9), cognition (MoCA^{ap}), HF knowledge (DHFKS^{aq}), or HRQoL Enrolled participants had better self-care behavior Barriers to implementation identified: lack of interest, inadequate technological and computer literacy, language limitations, anxiety caused by app questions, data retrieval issues, and disabilities (eg, visual, hearing, and cognitive impairments) 	<ul style="list-style-type: none"> Small enrollment numbers Low patient engagement High dropout from the HF app Findings based on experiences from a single trial

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
Stara et al [73], 2021	<ul style="list-style-type: none"> To support self-management of dementia in the home environment Duration: 4 weeks 	<ul style="list-style-type: none"> 3D-animated human-like AI-based avatar (Anne) Platform: mobile device or smartphones Microsoft Surface Pro tablet 	<ul style="list-style-type: none"> Older adults with dementia and their caregivers N=20 (30% male and 70% female) Participants aged ≥65 years (mean 75.5, SD 4.2 years) 	<ul style="list-style-type: none"> Usability and acceptability evaluation study—mixed methods approach Location: Italy 	<ul style="list-style-type: none"> Usability (SUS) score: 67.1 among older adults and 71.4 among caregivers, indicating acceptable usability No significant change in quality of life in older adults with cognitive impairment (QoL-AD^{af}) before or after the study (mean change 0.4, SD 4.6) 42% (8/20) of older adults' perceived closeness with Anne as feeling some overlaps, while 26% (6/20) reported feeling no overlap, and 11% (2/20) felt strong, equal or little overlap 	<ul style="list-style-type: none"> Small sample size and Italian national context and culture may limit result generalization Short 4-week study duration may be insufficient to significantly evaluate acceptability and usability Technical discomfort of participants with automatic speech recognition
Nguyen et al [74], 2021	<ul style="list-style-type: none"> Self-management of diabetes through patient education on diabetes care, blood glucose monitoring, and managing diabetes complications Duration: 8 weeks (52 days) 	<ul style="list-style-type: none"> Multilingual (text+voice enabled) AI bot (AMANDA; ML based using deep learning) Platform: mobile devices or smartphones 	<ul style="list-style-type: none"> Patients with diabetes who need self-care N=20 (nurses and clinicians) Participants' age not identified 	<ul style="list-style-type: none"> Usability and quality evaluation of AMANDA—a quantitative survey questionnaire filled out by 20 judges after listening to the real and TTS^{as}-generated audio samples Location: Singapore 	<ul style="list-style-type: none"> Naturalness: 4.07 Accent uniqueness: 3.98 Information clarity: 3.88 SUS score: 80.625 (above the average score of 68); 70% of the participants gave a score of ≥80 (indicating high usability of the interface) 	<ul style="list-style-type: none"> Small sample size and language restriction (availability in English only) may limit result generalization Laboratory setting evaluation may not reflect real-world use Impact of the CA on users' health outcomes (eg, blood glucose levels and medication adherence) and their emotional well-being or quality of life was not considered in the evaluation
Apergi et al [100], 2021	<ul style="list-style-type: none"> Self-care and communication improvement between patients and health care providers Duration: 90 days 	<ul style="list-style-type: none"> Commercial voice-bot (Alexa) and a voice-enabled, animated avatar Platform: Alexa+: smart speaker device (Amazon) Avatar: tablets 	<ul style="list-style-type: none"> Patients with HF N=55 (Alexa+: n=28; avatar: n=27) Participants' age not identified 	<ul style="list-style-type: none"> Quantitative pilot comparison study—demographic and technology survey and daily questionnaires Location: not identified 	<ul style="list-style-type: none"> Small sample size and focus on specific patient populations may limit generalizability of the findings Potential bias due to exclusion of possible important variables or unaccounted confounding factors in the model Lack of normalization of data for control variables Limited exploration of reasons for differences in technology engagement among Black patients 	

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
					<ul style="list-style-type: none"> • Positive correlation between patients' age and the technology use (coefficient=1.19; $P=.004$) • No statistically significant difference in engagement levels between the avatar and Alexa+ user groups • Decrease in use over time for both technologies, with a sharper decrease observed for Alexa+ participants • Black patients with similar characteristics used the technology 21 fewer times on average compared to non-Black patients (coefficient=-15.96; $P=.08$) • Results indicate that technology design may need to be better tailored for Black patients with HF 	
Kataoka et al [81], 2021	<ul style="list-style-type: none"> • Personalized patient education for self-management of lung cancer symptoms • Duration: 1 month 	<ul style="list-style-type: none"> • AI chatbot (rule-based AI using Google Cloud's Dialogflow and predetermined keywords) • Platform: mobile devices or smartphones and web browser (via LINE app) 	<ul style="list-style-type: none"> • Patients with lung cancer and their caregivers • N=12 (11 patients and 1 caregiver) • Participants' age not identified 	<ul style="list-style-type: none"> • Feasibility and usability study—sequential mixed methods approach through a web-based qualitative survey questionnaire and quantitative alpha and beta testing of chatbot with stakeholders • Location: Japan 	<ul style="list-style-type: none"> • The chatbot was found to be feasible but inadequately used • Mean user satisfaction score was 2.7/5, indicating low user satisfaction • Chatbot was able to deliver appropriate responses to most FAQs^{at} and also identify areas where additional responses were required • 82 categories of FAQs and formulated responses to these FAQs were identified, which were used to develop the chatbot • Potential of chatbot to improve patient knowledge of symptom management was demonstrated 	<ul style="list-style-type: none"> • Small sample size and single-hospital setting may limit applicability of study findings • 8 questions did not match well with responses in phase 5, leading to patient dissatisfaction • Questions for nonexistent categories remained unmatched, indicating the need to add educational categories and responses through further discussion
Rathnaya et al [101], 2022	<ul style="list-style-type: none"> • Behavioral intervention for remote management of mental health issues • Duration: 8 weeks 	<ul style="list-style-type: none"> • AI chatbot (Bunji; rule based+ML) • Platform: mobile devices or smartphones 	<ul style="list-style-type: none"> • Individuals with mental health issues • N=318 • Participants' age not identified 	<ul style="list-style-type: none"> • Mixed methods pilot study—participatory evaluation through quasi-experimental design and qualitative feedback • Location: Australia 	<ul style="list-style-type: none"> • Participatory evaluation (through 3 experimental pilot studies) • Study 1: positive impact in improving users' mood using Bunji's emotional support features • Study 2: positive impact of personalized conversation on improving the effectiveness of Bunji in providing mental health support to users • Study 3: positive feedback on the features of the chatbot and its effectiveness in providing remote mental health support to users 	

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
Kannam-pallil et al [102], 2022	<ul style="list-style-type: none"> To provide personalized and accessible mental health care with realistic and cognitively plausible verbal interaction Duration: not identified 	<ul style="list-style-type: none"> AI voice-based coach and voicebot (Lumen) Platform: Amazon Alexa 	<ul style="list-style-type: none"> Individuals with mild to moderate depression and anxiety seeking behavioral therapy N=26 Participants' age: mean 43.9 years 	<ul style="list-style-type: none"> Mixed methods observational study: Efficacy evaluation—RCT Acceptability and usability evaluation with WAI-TECH^{au} survey Location: not identified 	<ul style="list-style-type: none"> Task load (mental and physical effort required to complete a task): medium workload (higher temporal workload in session 2 [mean Task subscale 5.3, SD 0.9] than in session 1 [mean Task subscale 5.2, SD 0.9]) Work alliance (collaborative relationship between therapist and client): moderately high, indicating well-aligned sessions with participants' needs, potential goals (session 1-mean Goal subscale 5.0, SD 0.9; session 2- mean Goal subscale 5.1, SD 0.9), and mutual liking between participants and voicebot (session 1-mean Bondscales 4.9, SD 1.0; session 2-mean Bondscales 4.7, SD 1.0) Participants highlighted lack of personalization, depth, and emotional engagement in the conversations Overall UX: positive evaluation (values of >0.8) for pragmatic, hedonic, and overall qualities related to UX with Lumen for both sessions 	<ul style="list-style-type: none"> Limited sample size may impact result generalizability Short study duration (8 weeks) may be inadequate to capture long-term effects or user behaviors Reliance on users' self-reported data such as mood scores and survey responses could introduce bias Chatbot's performance could be affected by technical issues such as connectivity or device compatibility Lack of a control group may limit the ability to attribute changes to the chatbot's intervention <ul style="list-style-type: none"> Small sample size of users in a relatively controlled environment, potentially limiting applicability of findings Evaluation limited to 2 sessions, not representing the full 8-session PST^{av} program Potential influence of research coordinator and note taker on participant responses and Lumen use Inability to assess impact of various measures (task load and work alliance) over time due to study constraints Technological limitations of current AI-based voice technology include challenges in understanding natural language and interpreting emotional cues in voice interactions
Rahmanti et al [78], 2022			<ul style="list-style-type: none"> Individuals intending to lose or maintain weight N=10 (100% female) Participants aged 24-34 years 		<ul style="list-style-type: none"> Positive chatbot UX: ease of use; usefulness; and fun attributes such as use of emoticons, stickers, and GIF^{aw} images Negative chatbot UX: slow response time and irrelevant responses 	

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
	<ul style="list-style-type: none"> Lifestyle intervention for weight management through empathetic persuasive conversation flows Duration: 7 days 	<ul style="list-style-type: none"> AI (empathetic) chatbot (SlimMe; rule based+ML using Dialogflow) Platform: mobile devices or smartphones 		<ul style="list-style-type: none"> Mixed methods usability study through simulation trial—UX questionnaire after the trial Location: Taiwan 		<ul style="list-style-type: none"> Small sample size and language restriction (only English) may limit result generalization Nutrition assessment methods rely on user self-reports, which may introduce bias Self-reported anthropometric measurements may be less accurate compared to direct measurements
Alturaiki et al [103], 2022	<ul style="list-style-type: none"> To improve self-management of β thalassemia and communication between patients and health care providers Duration: not identified 	<ul style="list-style-type: none"> AI chatbot (rule based) Platform: web browser on computers or mobile devices 	<ul style="list-style-type: none"> Patients with β thalassemia N=34 Participants' age not identified 	<ul style="list-style-type: none"> Prototype implementation and usability evaluation study through qualitative posttest survey Location: not identified 	<ul style="list-style-type: none"> Usability and perceived utility assessment: Most participants (72%) found the chatbot easy to use Above 90% of participants considered the chatbot beneficial Most participants agreed that the chatbot made managing β thalassemia easier and more efficient Chatbot has the potential to save patients time and money that they usually spend on hospital visits 	<ul style="list-style-type: none"> Small sample size, which may limit generalizability of findings Chatbot was tested in a live chat scenario, possibly not reflecting real-world use accurately Limited data availability, with findings not publicly accessible due to privacy or ethical constraints No specification of statistical analysis or validation methods for assessing the chatbot's effectiveness Long-term usability and effectiveness of the chatbot in managing β thalassemia was unexplored
Zahedi et al [104], 2022		<ul style="list-style-type: none"> Virtual human avatars (AI nurse avatar and digital patient avatars) in a virtual hospital system (Wepital) Platform: computer (laptops and desktop computers) via the Second Life website 	<ul style="list-style-type: none"> Patients with GI problems N=61 Participants' age not identified 	<ul style="list-style-type: none"> Hypothesis experiment and usability study with real patients using the Wepital prototype—mixed methods approach Location: not identified 		

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
	<ul style="list-style-type: none"> To support constant ubiquitous medical care for GI^{ax} disease management through an avatar-based telepresence system Duration: not identified 				<ul style="list-style-type: none"> Significant positive impact of telepresence affordance on patient satisfaction with the Wepital system, with a path coefficient of 0.41 ($P<.001$) Significant positive impact of convenience affordance on patient satisfaction with the Wepital system, with a path coefficient of 0.33 ($P<.001$) Significant positive impact of trust affordance on patient satisfaction with the Wepital system, with a path coefficient of 0.43 ($P<.001$) Significant positive impact of using real avatars with wearable sensors on patient satisfaction with the Wepital system, with a coefficient of 0.11 ($P<.01$) Significant positive impact of patients' predispositions, such as flexibility regarding method of care and understanding information, on their perceptions of affordances in the Wepital system, with coefficients of 0.70, 0.63, and 0.29 ($P<.001$ in all cases), respectively No significant influence on trust due to lack of privacy concern 	<ul style="list-style-type: none"> Small sample size due to limited patient participation because of sensitive context and strict recruitment rules Researchers' restricted access to patients' medical records prevented the integration of real avatars with comprehensive medical information Study focused on patients with GI in a single context, limiting generalizability of findings to other medical fields Patients staying in the Wepital for persistent care were not involved in the study, limiting understanding of patient satisfaction
Eagle et al [37], 2022	<ul style="list-style-type: none"> Evaluation of CA's health-related advice for improved self-management of depression and anxiety Duration: not identified 	<ul style="list-style-type: none"> 6 commercial CAs—chatbots (Wysa and Replika) and voicebots (Google Assistant, Alexa, Cortana, and Siri) Platform: Chatbots: mobile devices or smartphones; Voicebots: smart speaker devices 	<ul style="list-style-type: none"> People experienced using CAs for mental health advice and management N=141 Participants' age not identified 	<ul style="list-style-type: none"> Mixed methods observational study—response quality evaluation using user-bot conversational data, PHQ-9 and GAD-7 survey questionnaires, and interviews with clinicians Location: United States 		

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
Maharjan et al [83], 2022	<ul style="list-style-type: none"> Self-report of depression or bipolar disorder Duration: 4 weeks 	<ul style="list-style-type: none"> Voice assistant and voicebot (Sophia; rule-based AI using Dialogflow) Platform: Google Nest Mini smart speakers 	<ul style="list-style-type: none"> Adults with depression or bipolar disorder N=20 Participants aged ≥18 years (18-34 years) 		<ul style="list-style-type: none"> Generally low quality of advice and recommendations provided by CAs for mental health questions Wide variability in the quality of responses across different CAs—no significant advantage of voice agents over chatbots Text-based chatbots slightly outperformed voice-based agents with improved responses regarding better treatment advice due to the chatbots' superior dialogue capabilities and empathy support, whereas voice assistants were prone to speech recognition errors and were more dependent on simple web searches in generating advice CAs struggled to handle unexpected inputs Anxiety-related responses were slightly better than the responses for depression and crisis situations Extended dialogue context access improved the CAs' quality of responses Design implications included providing clarification of subdialogues and access to extended dialogue context 	<ul style="list-style-type: none"> Limited sample size may restrict the generalizability of findings to a wider population Assessment of CAs' responses limited to mental health questions on anxiety and depression Standardized survey-based conversational probes may not have captured the full range of user queries and responses in real-world settings Quality of responses assessed through subjective participant ratings, introducing potential biases No investigation on the impact of user characteristics (eg, age, gender, or previous mental health service experience) on response quality Long-term effects of CAs' advice on user outcomes such as mental health improvement or treatment adherence were not addressed Ethical implications of using CAs for mental health advice, including privacy concerns and potential harm from inaccurate information, were unexplored CAs' responses were not compared with those provided by human experts, which could have provided insights into effectiveness

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
				<ul style="list-style-type: none"> Mixed methods in-the-wild study—acceptability and usability evaluation using user-agent engagement data, fortnightly completed WHO-5 health questionnaire, and semistructured interviews Location: not identified 	<ul style="list-style-type: none"> Global average engagement rate=75%, indicating a high engagement level with the voicebot Diverse UX with varying personified perceptions, social context, privacy and security concerns, and conversational features Perceived helpfulness of self-reporting practice in daily reflection and through organization, and the process was considered meditative by some participants The voicebot was perceived as usable and acceptable for self-reporting mental health well-being in a naturalistic setting 	<ul style="list-style-type: none"> Study conducted during the pandemic, potentially impacting participants' mental, physical, and social conditions, affecting result generalizability Participants' interaction with the CA may be influenced by lockdowns, social distancing, and travel restrictions, affecting their typical context and experiences Interview method (web-based or in person) could influence how participants conveyed their experiences, potentially influencing study outcomes Privacy and data security concerns among participants, leading to self-censorship during self-reporting Participants' engagement and experiences with the CA may be affected by their varying technology familiarity levels
Meheili et al [105], 2022	<ul style="list-style-type: none"> Cognitive behavioral intervention for self-management of chronic pain and mental health Duration: not identified 	<ul style="list-style-type: none"> Commercial CA—AI mental health chatbot (Wysa) Platform: mobile devices or smartphones 	<ul style="list-style-type: none"> Individuals with self-reported chronic pain seeking digital mental health support N=2194 (real-world data) Participants' age not identified 	<ul style="list-style-type: none"> Mixed methods retrospective observational study—chatbot effectiveness evaluation through user-chatbot engagement data and pretest-posttest assessment questionnaires (PHQ-9 and GAD-7) Location: not identified 	<ul style="list-style-type: none"> Identified themes related to health concerns, socioeconomic concerns, and pain management concerns among users with chronic pain Users with chronic pain showed significantly higher app engagement ($P<.001$) compared to users without chronic pain, with a large effect size (Vargha and Delane $A=0.76-0.80$) Significant reduction in anxiety and depression symptoms among users with chronic pain, as shown by the pretest-posttest assessments using PHQ-9 and GAD-7 scales ($P<.001$), with a medium effect size (Cohen $d=0.60-0.61$) Chatbot was found to be useful in managing challenges, including anxiety, sleep, low energy, motivation, loss, and pain Conversational data revealed perceived needs and experiences of individuals with chronic pain, including the need for personalized and flexible support 	<ul style="list-style-type: none"> Small samples for the third objective may further constrain generalizability of findings, underscoring the need for careful interpretation as preliminary outcomes Nonrandomized sampling design may limit generalizability of findings Data extraction keywords based on guidelines, literature, and clinical experience may have missed relevant pain terms Repeated measurements for efficacy without a control group might risk increasing regression to the mean Users were not mandated to complete assessments, limiting efficacy study to a small subset

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
Henson et al [106], 2022	<ul style="list-style-type: none"> To provide appropriate patient education for self-management of GERD^{ay} Duration: not identified 	<ul style="list-style-type: none"> Commercial CA—AI chatbot (ChatGPT) Platform: multiple devices—mobile devices or smartphones, laptops, and desktop computers (web browser and app) 	<ul style="list-style-type: none"> Patients with GERD N=11 (8 patients and 3 gastroenterologists) Participants' age not identified 	<ul style="list-style-type: none"> Feasibility study (quantitative)—potential utility evaluation of ChatGPT in GERD using 23 GERD management prompts Location: not identified 	<ul style="list-style-type: none"> 63/69 (91.3%) appropriate responses to GERD management queries delivered by ChatGPT Frequent inconsistency in responses to the same prompt, with 16/23 (70%) prompts resulting in variable appropriateness Highest number of appropriate responses (39.4%) was to treatment-related prompts, whereas the highest number of inappropriate responses (14.3%) was to diagnosis and management-related prompts Most responses (78.3%) contained some specific guidance, particularly for diagnosis-related prompts (93.3%) Patients with different educational backgrounds considered the responses understandable and more useful than a search engine ChatGPT failed to recommend consideration of Roux-en-Y gastric bypass for ongoing GERD symptoms with pathological acid exposure in obesity setting Current limitations prevent the integration of ChatGPT into routine clinical practice at present 	<ul style="list-style-type: none"> Small sample size may limit the generalizability of findings ChatGPT was not specifically trained on medical literature, and its ability to address repeated requests for clarification was not evaluated Inappropriate responses with inconsistencies to the same prompt were observed, which may affect results Limited specific guidance and content errors PPI^{az} risks were stated as fact, which lacked balanced consideration of benefits and context ChatGPT often provided overly long responses, diluting the clinical impact of responses in many cases
Babington - Ashaye et al [75], 2023	<ul style="list-style-type: none"> To improve patient knowledge and support symptom monitoring of hemophilia Duration: not identified 	<ul style="list-style-type: none"> AI chatbot (Saytù Hemophilie; NLU^{ba} and ML based) Platform: mobile devices or smartphones (via messaging apps such as WhatsApp, Signal, and Telegram) 	<ul style="list-style-type: none"> People with hemophilia and their families in Senegal N=30 (20 people with hemophilia and 10 family members) Participants' age not identified 	<ul style="list-style-type: none"> Usability study—mixed methods approach with SUS survey Location: Senegal (West Africa) 	<ul style="list-style-type: none"> Average usability (SUS) score: 81.7 (indicates high usability of the system) 75.4% of participants expressed an overall high level of satisfaction with the French version of the chatbot Most participants perceived the proposed AI-based chatbot as a potential solution to manage the symptoms at home while awaiting a physician's consultation 	<ul style="list-style-type: none"> Small sample size restricted the ability to conduct additional statistical tests, limiting result generalizability Potential sample bias as 40.6% of participants recruited were students While the SUS score was used, additional evaluation measures could be used to assess users' overall opinion of the chatbot
Alhmedat and Alotaibi [79], 2023		<ul style="list-style-type: none"> Humanoid robot (SARA) Platform: physical presence 	<ul style="list-style-type: none"> Children with T1DM N=5 (children) Participants aged 5-9 years 			

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
	<ul style="list-style-type: none"> Symptom monitoring and patient education for T1DM management Duration: 5 days 			<ul style="list-style-type: none"> Pilot study—acceptability evaluation of SARA using quantitative questionnaires, observations, and experiments Location: Saudi Arabia 	<ul style="list-style-type: none"> All participants succeeded in all the following measures with different outcomes: <ul style="list-style-type: none"> Total interaction time with SARA Productivity of education task Productivity of the short test task Productivity of listening to stories Moderate worthiness level: 88.2%, indicating a high relative acceptability level User feedback: positive feedback from the patients, their parents, and clinical staff 	<ul style="list-style-type: none"> Very small sample size of the pilot study, limiting generalizability of results SARA framework is still under development, requiring further experiments before finalizing the robotic platform
Calvo et al [84], 2023	<ul style="list-style-type: none"> To estimate the risk of an asthma exacerbation and provide recommendations for improved self-management of asthma Duration: 28 days 	<ul style="list-style-type: none"> Text-based chatbot (rule-based AI) Platform: mobile devices or smartphones (WhatsApp based) 	<ul style="list-style-type: none"> Adults with asthma N=300 (40 for alpha testing and 260 for beta testing) Participants aged ≥18 years 	<ul style="list-style-type: none"> Feasibility and usability study with pretest-posttest quantitative questionnaires Location: United Kingdom 	<ul style="list-style-type: none"> Chatbot was found to be acceptable, usable, and satisfactory to participants as it was helpful in improving their asthma control Task completion rate: 80% of participants completed the risk assessment process, whereas 20% dropped out Net Promoter Score for the chatbot: 8.5 (indicating high likelihood of recommending it to others with asthma) Participants reported that they were satisfied with the chatbot and that the risk measurement by the chatbot was useful Participants suggested improving the chatbot by incorporating more personalized feedback and additional interactive features Chatbot seemed to be effective in promoting motivational quality, engagement, and UX in relation to basic psychological needs High level of alignment between expected and calculated risk, and the calculated risk was perceived as accurate by the user 	<ul style="list-style-type: none"> Limited sample size No intention to find gender or age differences within the current sampling No intention to measure significant asthma control improvements in 28 days Potential for incorrect responses from individuals who incorrectly answered the screening questionnaire Lack of validation for measurable health outcomes
Epalte et al [107], 2023		<ul style="list-style-type: none"> AI chatbot (Vigo; rule based+ML) Platform: smart devices (eg, Apple iPad tablet) 	<ul style="list-style-type: none"> Patients with stroke and their families N=12 Participants' age not identified 	<ul style="list-style-type: none"> Usability study—qualitative semistructured interviews after the intervention Location: not identified 		

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
	<ul style="list-style-type: none"> Personalized exercise program with patient education and counseling support for self-management of post-stroke recovery Duration: 1 month 				<ul style="list-style-type: none"> Participants considered the chatbot transparent, understandable, and handy Overall design of Vigo was rated as good Participants were mostly unsatisfied with the difficulty level and diversity of exercises 5 themes related to the app's effectiveness were identified: flexibility, information, exercises, emotional support, and assessment Chatbot could have a potential positive impact on post-stroke outcomes, particularly in emotional status, social interaction, improved mood, and motivation 	<ul style="list-style-type: none"> Strict inclusion criteria and chatbot's complexity limited recruitment of patients with greater physical limitations, potentially skewing insights Patients with severe functional limitations, aphasia, cognitive deficits, and comorbidities may struggle to use the app independently, reducing the potential user base Chatbot's availability only in Latvian may restrict use for Russian-speaking patients Difficulty in adapting the program for patients due to their different health statuses, personal factors, and preferences
LeRouge et al [108], 2023	<ul style="list-style-type: none"> Personalized lifestyle intervention for self-management of obesity Duration: not identified 	<ul style="list-style-type: none"> AI-based computer-animated cartoon-like 2D avatars Platform: mobile devices or smartphones and desktop computers 	<ul style="list-style-type: none"> Adolescents with obesity or overweight N=70 (for phase 2) Participants' age not identified 	<ul style="list-style-type: none"> Qualitative participatory research study in 2 phases: Phase 1: focus groups with teenagers, provider interviews, and parent interviews Phase 2: midrange prototype assessment by teenagers and providers Location: United States 		<ul style="list-style-type: none"> Limited data sources may limit universal applicability of findings Study scope was limited to a specific audience using qualitative methods, which may limit generalizability Potential coding bias in the data analysis process despite minimization efforts

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
Park et al [109], 2023	<ul style="list-style-type: none"> 2 AI chatbots (human-like vs machine-like; NLP using Dialogflow) Platform: not identified 	<ul style="list-style-type: none"> Adults seeking web-based mental health support N=385 Participants' age not identified 	<ul style="list-style-type: none"> Quasi-experimental design—quantitative UX evaluation of a human-like vs machine-like chatbot for web-based mental health counseling Location: United States 	<ul style="list-style-type: none"> Effectiveness: avatars and virtual agents can be effective in engaging adolescents in weight management programs Avatars and virtual agents can provide personalized support and motivation to adolescents Users would be more motivated to use the virtual agents if they were more fun, creative, human-like, interactive, communicative, and indicative of progress or success Design characteristics: <ul style="list-style-type: none"> Personalize avatars matching user needs and preferences Ability to stimulate 2-way interaction (preferably fun, creative, and more human-like) Support multiple communication modes (eg, text, voice, and video) Provide feedback on user progress and behavior Challenges: <ul style="list-style-type: none"> Implementation of avatars requires technical expertise and resources User acceptance barriers due to trust issues and privacy concerns Integration requirements of avatars with existing health care systems and workflows Development of avatars and virtual agents can be expensive 		

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
	<ul style="list-style-type: none"> Improvement of web-based mental health counseling efficacy through effective conversation between a user and a chatbot agent Duration: not identified 				<ul style="list-style-type: none"> Human-like chatbot apparently yielded higher user intention to comply with health recommendations (mean 5.69, SD 0.61) than the machinelike one (mean 5.34, SD 0.93; $t_{383}=4.41$; $P<.001$) Human-like chatbot resulted in higher compliance with health recommendations mediated by psychological distance Positive correlation between human representation and perceived trust and between perceived trust and compliance with health recommendations Human-like chatbot led to psychological closeness and increased trust, thereby causing higher intention to comply with health recommendations 	<ul style="list-style-type: none"> Convenience sampling could limit generalizability Study focused solely on chatbot's physical representation, excluding other important cues such as conversational and identity cues User characteristics such as involvement and self-efficacy, which could moderate the impact of human-like cues, were unexplored Effect of demographic variables (eg, age, gender, and ethnicity) was not investigated for both users and chatbot agents Using a self-administered web-based questionnaire might introduce limitations such as social desirability bias and response bias while lacking control over participants' environments Long-term effects of chatbot counseling on mental health outcomes were not investigated
Bog-giss et al [110], 2023	<ul style="list-style-type: none"> To deliver self-compassion coping lessons to support T1DM self-management and mental health Duration: not identified 	<ul style="list-style-type: none"> AI self-compassion chatbot with customizable avatar (COMPASS) Platform: mobile devices or smartphones 	<ul style="list-style-type: none"> Adolescents with T1DM N=30 (19 adolescents with T1DM and 11 diabetes health care professionals) Participants aged 12-16 years 	<ul style="list-style-type: none"> Qualitative study—focus groups and interviews Location: New Zealand 	<ul style="list-style-type: none"> COMPASS chatbot was found to be acceptable to adolescents with T1DM and diabetes health care professionals The chatbot apparently had the potential to support the T1DM self-management and mental health of adolescents during the COVID-19 pandemic Participants emphasized the importance of personalization of both content and features of the chatbot (eg, games, apps, background color schemes, and hobbies) and cultural appropriateness to support users with different ethnic backgrounds 	

Study, year	Intervention and study duration	CA ^a type and delivery platform	Target population, sample size, and participants' age	Methods and location	Measures and outcomes	Limitations
						<ul style="list-style-type: none"> • Small sample size may not be representative of the breadth of challenges faced by adolescents with T1DM or additional barriers to standard care access • Potential bias in user feedback due to existing relationships between the facilitator and participants • Feedback was limited as participants only viewed screen recordings of the chatbot, potentially reducing its richness • Primary assessment limited to acceptability, perceived clinical utility, and usability of chatbot features, potentially overlooking the intervention's overall impact • User feedback may be influenced by inequities in access to technology such as smartphones and continuous glucose monitors

^aCA: conversational agent.

^bAI: artificial intelligence.

^cRCT: randomized controlled trial.

^dAF: atrial fibrillation.

^eAFEQT: AF Effect on Quality of Life questionnaire.

^fT2DM: type 2 diabetes mellitus.

^gML: machine learning.

^hmHealth: mobile health.

ⁱAADE: American Association of Diabetes Educators.

^jWHO-5: 5-item World Health Organization Well-Being Index.

^kK10: Kessler Psychological Distress Scale.

^lBADS: Behavioral Activation for Depression Scale.

^mBADS-AC: Behavioral Activation for Depression Scale-Activation.

ⁿBADS-AR: Behavioral Activation for Depression Scale-Avoidance/Rumination.

^oBADS-WS: Behavioral Activation for Depression Scale-Work/School Impairment.

^pBADS-SI: Behavioral Activation for Depression Scale-Social Impairment.

^qVDMS: virtual diabetes management system.

^rPHQ-9: 9-item Patient Health Questionnaire.

^sGAD-7: 7-item Generalized Anxiety Disorder Scale.

^tT1DM: type 1 diabetes mellitus.

^uPAL: personal assistant for a healthy lifestyle.

^vSUS: System Usability Scale.

^wBCT: behavior change theory.

^xHRQoL: health-related quality of life.

^yAQoL-8D: Assessment of quality of life-8 dimensions.

^zHbA_{1c}: glycated hemoglobin.

^{aa}HADS: Hospital Anxiety and Depression Scale.

^{ab}GP: general practitioner.

^{ac}SCD: sickle cell disease.

^{ad}EVA: Education virtual assistant.

^{ae}FBG: fasting blood glucose.

^{af}PPBG: postprandial blood glucose.

^{ag}VH: virtual human.

^{ah}UX: user experience.

^{ai}UEQ: User Experience Questionnaire.

^{aj}TAM: technology acceptance model.

^{ak}NLP: natural language processing.

^{al}CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth.

^{am}OR: odds ratio.

^{an}RMDQ: Roland-Morris Disability Questionnaire.

^{ao}HF: heart failure.

^{ap}MoCA: Montreal Cognitive Assessment.

^{aq}DHFKS: Dutch Heart Failure Knowledge Scale.

^{ar}QoL-AD: Quality of Life in Alzheimer Disease scale.

^{as}TTS: Text-to-Speech.

^{at}FAQ: frequently asked question.

^{au}WAI-TECH: Working Alliance Inventory–Technology Version.

^{av}PST: problem-solving treatment.

^{aw}GIF: graphics interchange format.

^{ax}GI: gastrointestinal.

^{ay}GERD: gastrointestinal reflux disease.

^{az}PPI: proton-pump inhibitor.

^{ba}NLU: Natural Language Understanding.

A few studies (6/43, 14%) focused on evaluating the effectiveness of the CAs used for the self-management of certain chronic illnesses, such as atrial fibrillation [10], type 2 diabetes mellitus [24,82], and mental health conditions such as depression and anxiety [90,96,97]. All these studies found improvement in various parameters in management, such as significant reduction in glycated hemoglobin level ($P < .001$), as well as BMI, and weight ($P < .001$) in patients with type 2 diabetes mellitus [82]; notable improvement in quality of life measured using the Atrial Fibrillation Effect on Quality of Life questionnaire score ($P < .05$) or health-related quality of life utility score ($P = .007$) in patients with atrial fibrillation [10,24]; and substantial reduction in depression (9-item Patient Health Questionnaire score) and anxiety (7-item Generalized Anxiety Disorder Scale score) symptoms [90,97,105].

Types of AI-Based CAs in the Included Studies

Of the 47 AI-based CAs representing digital health assistants identified from the included studies, 9 (19%) were existing commercial CAs available as marketed products, which included voice assistants—Google Assistant, Amazon Alexa, Microsoft Cortana, and Apple's Siri, as well as social chatbots—ChatGPT and Replika (a voice-enabled, avatar-based chatbot), and health care chatbots—Vik, Wysa, and Tess. The remaining 38 CAs were proposed designs or prototypes, where 19 (50%) of them were chatbots, 2 (5%) were voice assistants or voicebots [58,102], 13 (34%) were human-like digital avatars, 1 (3%) was a digital avatar version of a humanoid social robot [85], 1 (3%)

was a text-based voice-enabled bot [74], and 2 (5%) were avatar-based chatbots [91,110]. In addition, 5% (2/38) were CAs as humanoid social robots [79,85], which were included as they particularly emphasized managing NCDs such as diabetes (although the *robots* topic was beyond the scope of this study).

Intelligence Level of the AI-Based CAs

Of the 38 CA prototypes presented, 27 (71%) were found to have indicated their AI algorithm implementation methods to generate appropriate health-related responses. A total of 30% (8/27) of them were implemented using rule-based algorithms [72,81,83,84,86-88,103], 7% (2/27) of the CAs were developed using network-based scripted algorithms [10,80], and 4% (1/27) of the CAs were implemented using prescribed algorithms based on behavior change theory, whereas 33% (9/27) of the CAs were implemented using ML approaches [58,74,75,89,93-95,97,99] and 15% (4/27) of the CAs were developed using a combination of rule-based and ML techniques [78,85,101,107]. However, one of the CAs identified from a study was a Wizard-of-Oz human-like AI avatar with prescribed text and human-controlled responses to evaluate the potential effectiveness of visually embodied avatars in the assessment of anxiety and depression through users' behavioral expressiveness [96].

Types of Health Interventions and Study Duration

Most included studies (24/43, 56%) focused on behavioral or lifestyle interventions and patient education for improved self-management of physical NCDs such as *CVDs* [10,80,99,100], *diabetes* [24,72,74,79,82,85,89,93] or *prediabetes* [76], *cancer* [81,92], *gastrointestinal diseases* [104,106], and obesity [76,86,108].

A reasonable number of studies (13/43, 30%) focused on cognitive behavioral interventions related to the self-management of chronic cognitive impairments (eg, dementia and poststroke) [73,107], including chronic mental health conditions such as anxiety, depression, psychological distress, and bipolar disorders [83,87,88,90,96,97,101,102].

The duration of these studies was reasonably divergent, ranging from 5 days to 1 year. However, a substantial number of studies (14/43, 33%) could be considered long term, such as 8 weeks to 2 months [85,88,90,99,101], 12 weeks to 90 days [86,98,100], 16 weeks to 6 months [76,82,87,93], 9 months [94], and 12 months or 1 year [24,92]. A total of 26% (11/43) of the studies conducted trials for approximately ≤ 1 month [10,73,81,83,84,88,97,107] (ie, within 5-10 days) [72,78,79]. A total of 42% (18/43) of the studies did not mention the study duration.

Study Population and Location

Almost half (20/43, 47%) of the included studies did not specify the target population other than patients with a particular chronic disease that the health intervention was meant for. The studies that specified the target population or their age group (24/43, 56%) were relatively heterogeneous, with older adults [10,58,72,73,91,95] and middle-aged adults [24,99] being the most frequently targeted user group (8/43, 19%). However, some studies (6/43, 14%) also focused on younger adults [97], adolescents [76,108,110], and children [79,85].

The locations of the studies conducted were largely diverse, but most (27/43, 63%) were conducted in higher-income countries. A total of 9 studies were completed in Europe, where 2 (22%) of them were conducted in the United Kingdom [84,91], 2 (22%) studies were conducted in Italy and the Netherlands [73,85], 1 (11%) was conducted in Portugal [72], 1 (11%) was conducted in France [77], 1 (11%) was conducted in Greece [93], 1 (11%) was conducted in Spain [94], and 1 (11%) was conducted in Ukraine [97]. A total of 16% (7/43) of the studies were conducted in the United States [10,37,76,86,87,108,109]. In total, 4 of the studies were carried out in the Australasia region—3 (75%) in Australia [24,99,101] and 1 (25%) in New Zealand [110]. A total of 7 studies were conducted in Asia, where 4 (57%) of them were conducted in Japan [81,88,96,98], 1 (14%) was conducted in Singapore [74], 1 (14%) was conducted in Taiwan [78], and 1 (14%) was conducted in India [82]. A total of 5% (2/43) of the studies were conducted in Saudi Arabia [79] and Senegal in West Africa [75]. The remaining 33% (14/43) of the studies did not specify the study location.

Data Collection Methods of the Included Studies

The included studies mainly used quantitative (17/43, 40%) or mixed methods (17/43, 40%) approaches. A substantial

proportion of these studies (11/43, 26%) were pretest-posttest intervention studies to observe the differences before and after the intervention within the same group of participants (single armed) or between intervention and control groups (2 armed). The study designs of the quantitative and mixed methods studies included observational experiments using 2-armed randomized controlled trials (7/34, 21%) [10,24,86,87,98,99,102], a nonrandomized prospective study (1/34, 3%) [88], retrospective studies (2/34, 6%) [82,105], quasi-experimental studies (4/34, 12%) [10,90,101,109], a single-armed pilot trial [94], and comparative evaluation studies (3/34, 9%) [89,96,100].

The qualitative studies (9/43, 21%) were mostly usability and acceptability assessments of prototypes through qualitative posttest surveys [103], postintervention interviews [107], focus groups [110], a participatory design approach with focus groups and interviews [108], and workshops [91]. A total of 22% (2/9) of the qualitative studies compared the proposed prototypes against similar existing applications [58,89].

Discussion

Principal Findings

The discussion of this scoping review synthesizes the findings in light of the overarching themes emergent from each research question, elucidating the interplay between the current state, challenges, benefits, and the targeted users of AI-based CAs within the scope of the examined literature.

Current State of AI-Powered CAs

Regarding the current state of research on AI-powered CAs as human-like health carers for managing NCDs, empirical studies have revealed the limited ability of existing commercial CAs to provide appropriate lifestyle advice and health recommendations [37,111]. Perhaps this is due to the lack of specified data training for these commercial tools in the medical context, which is currently restricting their integration in clinical and remote care [106].

Functionality-wise, commercial chatbots (eg, Wysa and Replika) tend to slightly outperform voice agents (eg, Google Assistant, Siri, Alexa, and Cortana) with improved response quality due to the former's better dialogue capabilities and empathy support. At the same time, the latter are more susceptible to speech recognition errors, including greater dependency on simple web searches to generate advice [37]. Nonetheless, voice-enabled CAs offer hands-free communication, making them preferable over text-based chatbots for older adult users, as suggested by a recent scoping review conducted by Even et al [70] focusing on the benefits and challenges of CAs in the older adult population (Table 1). In addition, another study revealed that older adults preferred voice-based communications over SMS text messages or free-text entry as voice is regarded as a powerful mode for promoting motivation [112]. In addition, voice agents may also be more accessible compared to digital avatars due to their lower bandwidth requirements, making them suitable for locations with lower bandwidth connectivity or devices with limited computing power [48]. Conversely, digital avatars are generally more likely to be favored by users over other types of CAs for their widely accepted face-to-face

conversation format [10] as they can support both text and audio outputs [47,113] and, therefore, can serve as a valuable alternative complement for remote care assistance [114]. This suggests that digital avatars remain a popular choice for their versatility and interactive capabilities, driving the rapid progression toward the anthropomorphism era as they represent realistic human-like traits in behaviors and appearances for enhanced interaction with the users [115].

Moreover, anthropomorphic CAs are more preferred over nonanthropomorphic CAs, as shown by Park et al [109], indicating that the human-like representations of a chatbot yielded a higher likelihood of intervention compliance than the machinelike chatbot because the former increased closeness and trust. Similarly, another study demonstrated that an augmented human-like embodiment of a digital avatar resulted in greater social closeness among the participants [116]. Correspondingly, strong social presence and emotional closeness are regarded as key facilitators of favorable attitudinal and behavioral outcomes among users [117]. Hence, anthropomorphism positively correlates to users' perceived trust and acceptance of a virtual agent [12,117]. Indeed, digital human faces as intelligent CAs are considerably promising for achieving a human-like presence to enhance human-machine interaction [42]. Consequently, the application of AI-based conversational avatars as digital human-like health carers has been gaining attention recently, leading patients to prefer digital human avatars as medical professionals over chatbots due to the former's perceived human-like and interactive nature, which is essential for maintaining the professionalism and emotional aspects in physician-patient relationships [39].

Nonetheless, this evolution may apparently raise perceived uncanniness, which is considered a genuine ethical concern with realistic anthropomorphic avatars, yet it does not appear to be a universal problem for the avatar appearance or limit users' engagement with the virtual environment [55]. Realistic humanoid avatars representing virtual carers can appear more trustworthy and increase compliance among users [55,109]. However, striking a balance between realistic and stylized avatar appearances can be worthwhile to avoid visual overloading and discomfort among users while representing the required elements accurately to maximize perceived trust, closeness, and acceptability [25].

Challenges and Limitations of CAs

The following challenges with CAs, including limitations in intelligence level, empathy, data privacy, and ensuring patient safety, highlight significant barriers to their adoption and effectiveness in health care.

Intelligence Level of CAs

Most of the proposed AI-based CAs were rule-based or scripted systems with a predetermined set of rules and conditions, which are easier to design and implement but have limited capability to handle complex conversations (unlike the automated ML-based systems) [33,56,118] and, therefore, cannot generate personalized responses in a more human-like manner, which the users would mostly prefer [84,91,108-110]. Furthermore, few studies proposed integrating personality-tailored behavior

in CAs, precisely known as personality-adaptive CAs, which can automatically recognize and adapt to the users' distinctive personality traits accordingly to be able to support patients in a personalized and authentic human-like manner [119,120] as well as effectively persuade users to alter their attitudes and behaviors for healthy lifestyle habits [121].

Empathy of CAs

Studies have emphasized the importance of including the "empathy" element in the design of CAs [122,123], yet only one of the included studies notably aimed to incorporate "empathy" in the proposed chatbot by applying text-based emotion analysis to recognize the user's emotion and respond accordingly using emojis, stickers, and Graphics Interchange Format images [78]. Furthermore, a scoping review on AI-based CAs conducted by Kusal et al [56] has suggested that the major shortcoming of CAs is their lack of competence to make communication with humans seem natural with the aid of empathy and sentiments, mainly due to their failure to comprehend the human context and the users' emotions as their responses are restricted to the queries they are trained for, which can lead to a frustrating user experience. Indeed, a study conducted by Casas et al [124] determined the capability of an empathetic chatbot to surpass a benchmark bot or even humans in predicting the emotional state of textual messages. Therefore, CAs should be able to model their context and select important information to remember as communication depends on past messages [56].

Data Privacy

Surprisingly, the included studies demonstrated limited emphasis on the data privacy of the AI-based CAs while handling patient data, which is a common challenge faced in the design and evaluation of CAs [111,125]. Hence, it is recommended to prioritize data privacy and confidentiality factors during the development of such technologies to gain users' trust and acceptance [63].

Patient Safety

Another area that needs to be focused on is patient safety [126], which was also not highlighted in the included studies. For instance, the inability of digital health tools to recognize the urgency of the patient's health condition may generate inappropriate health advice or lead to delayed treatment, increasing the risk of further complications. In addition, limited validation and medical training for existing CAs (as mentioned previously) poses risk of inaccuracy and raises concerns about patient safety. Therefore, assessing the validation status of these digital health tools is crucial for patient safety and optimal health care outcomes.

Uses and Benefits of AI-Based CAs in Managing NCDs

This review suggests that CAs are primarily used for supporting nonpharmacological interventions such as behavioral or lifestyle medication for managing NCDs such as CVDs, diabetes, obesity, cancer, and gastrointestinal diseases, as well as preventive measures for healthy living, indicating their potential ability to complementarily assist in the long-term remote management of NCDs [114,127,128]. The ubiquitous self-care support through constant symptom monitoring, along with the

personalized behavioral modifications provided by CAs, can possibly rescue patients from minor health concerns and prevent the progression of chronic diseases [10,104]. Health care professionals are optimistic about CAs' potential ability to save patients time and money that they usually spend on physical clinic visits [103], thereby reducing the health care cost and burden and increasing accessibility [91,104]. Augmented reality applications could enhance the training of health care professionals and even patients in using these new technologies effectively [115,129,130].

Efficacy

The efficacy evaluation of CAs for remote care is still not thoroughly explored and well understood. According to Monaco et al [63], inadequate evidence regarding the efficacy of digital health tools may have hindered the potential beneficial outcomes of using such resources to control NCDs. Nonetheless, the few included studies that aimed to identify the effectiveness of CAs in managing the target NCDs found favorable outcomes [10,24,82,90,96,97,105]. Moreover, many reviews have emphasized the efficacy of CAs in health care support. For instance, a systematic review conducted by Milne-Ives et al [26] investigated the effectiveness and usability of AI-based CAs designed for health care and found them generally effective, with positive or mixed evidence, whereas a narrative review by Dingler et al [60] used context-aware voice assistants (eg, Google Home and Amazon Echo), demonstrating the utility of speech-based CAs in providing personalized remote health care support. However, a systematic review conducted by Hossain et al [131] exhibited low positive outcomes of digital health interventions among people with NCDs in India, requiring further exploration of advanced technologies for the equitable and sustainable development of digital health tools.

Targeted Users

Most of the target user groups were based on chronic disease conditions rather than any specific age group. A fair proportion of studies (6/43, 14%) focused on the application of CAs in the older adult population, exhibiting that this vulnerable user group may be most in need of such technologies as older adults have greater risk of developing serious NCDs [28,58,61] and many older adult patients lead an isolated, lonely life and, therefore, require constant remote monitoring as well as a virtual companion for healthy living [132,133]. Hence, in other words, the selected user groups were mostly based on targeted chronic diseases that may be more prevalent in older adults. However, we did not encounter any significant study that modified the CAs to enhance accessibility for older adults, but social networking sites could potentially enhance engagement with such CAs among this user group [134].

Moreover, it is crucial to recognize that the vast majority (41/43, 95%) of the included studies were conducted in high-income countries, underscoring the necessity of further research among underserved populations, particularly in LMICs where individuals may be more vulnerable to developing NCDs due to limited health awareness and access to health care services.

Limitations

Although this is an emerging field of research and we conducted an extensive literature search strategy, the possibility of noninclusion of relevant studies may still exist (eg, missing certain keywords or search terms, studies in languages other than English, studies not listed in the searched databases, and unidentified studies from gray literature). Similarly, our stringent inclusion and exclusion criteria were implemented to refine our focus on patient-centered AI-based CAs tailored for the remote management of NCDs, aiming to ensure the relevancy and applicability of our review to this particular domain of interest. However, this approach may have inadvertently narrowed the scope, potentially overlooking other relevant studies that could provide wider insights into our findings. Despite this, we strived to conduct a fair selection of studies meeting our inclusion criteria to mitigate selection bias and ensure a balanced representation of the literature. Nonetheless, it is important to acknowledge that other potential biases, such as publication bias, may still exist, which are beyond our control. In addition, new articles may have appeared after our search deadline (July 31, 2023), which may cause additional differentiations.

Furthermore, some common limitations were also identified within the included studies that may impact the overall findings of this review: (1) limited sample sizes (a vast majority of the studies had relatively small sample sizes, which could limit the generalizability of the findings), (2) short-term interventions or lack of long-term outcome measures (the duration of interventions in most studies was relatively short [80,82,88,101], potentially overlooking long-term user interactions and health outcomes essential for a comprehensive evaluation of CAs' ability to sustainably assist in NCD management [10,37,77,85,91-93,96-98,103,109]), (3) limited demographic data (some studies lacked detailed demographic information of the participants [37,93,109], which may potentially limit the analysis of how demographic factors can impact the feasibility or effectiveness of the CAs), (4) potential bias (many studies had potential biases resulting from self-reported user data [37,77,78,101,109], sample bias [75,82,105], coding bias [108], and exclusion of variables or unaccounted confounding factors in analyses [100], which could impact the validity of the results), and (5) unaddressed technical aspects (some studies did not discuss the technical aspects of the proposed CAs, including their potential technical limitations or challenges [85,92,93]). Addressing these issues could contribute to the enhancement of the design, development, and usability of CAs in the future [135].

Implications and Future Recommendations

As CAs strive to facilitate personalized and empathetic human-like conversations in the health care field [113], the necessity for context-oriented and domain-specific training emerges as a crucial consideration for improved accuracy and relevance [136]. Although advanced AI systems such as large language models trained on vast textual data (eg, the generative pretrained transformer series and LLaMA) have proven transformative in various sectors such as marketing, education, and customer service, their application in the health care sector remains underexplored, primarily due to a scarcity of relevant

high-quality medical data sets [137]. In enhancing the AI-based CAs for health care, only a few recent studies have aimed to investigate real-world interactions between patients and health care professionals [136-139]. While recognizing the dynamic and nonstandardized nature of historical human interactions in health care, our recommendation to study real-world communications between patients and health caregivers stems from the objective of training CAs to be contextually relevant, particularly in addressing the unique challenges associated with specific chronic diseases. This approach may seek to empower CAs to navigate the complexities of diverse health conditions with greater intelligence and empathy, thereby improving their adaptability through the use of high-quality data sets acquired from real-world scenarios [136] given the substantial amount of data and extensive knowledge base required for domain-specific contextual training [56]. However, to ensure effective training of AI agents in health care, it is essential to define their roles (eg, nurse, health assistant, nutritionist, and pharmacist), which can determine the appropriate level of training required, allowing for the selection of specific data sets tailored to the medical training and evaluation of CAs.

Findings from the literature suggest an overall higher user preference for realistic human-like representations in CAs, specifically in the form of anthropomorphic digital avatars, which can be accessed through mobile devices (unlike physical humanoid robots) [52,54], thus accommodating a broad user base including older adult users [58,140]. Therefore, future research should prioritize investigating the impact of anthropomorphic avatars or virtual humans in managing NCDs, including their potential ethical concerns (eg, uncanny valley) that should be addressed in the design considerations to increase acceptance of CAs for self-care [25,49,123].

Moreover, user preferences related to CA types and features may vary notably among different user groups. For instance, older adults and users with disabilities may require

voice-activated CAs over text-based CAs due to their typing inabilities, and therefore, enhanced speech recognition functionality is required to incorporate into CAs to accommodate voice conversations in multiple languages [70]. In addition to language requirements, factors such as culture, ethnicity, and individual personality traits may also play a significant role in influencing the adoption of AI dialogue systems for remote care purposes. Indeed, the adoption of CAs can be impeded among populations traditionally experiencing health inequities due to the limited availability of such tools with localized and culturally tailored features. These populations, often at higher risk of NCDs, could greatly benefit from more culturally sensitive health care technologies, which could improve health care access and outcomes, thereby necessitating careful consideration in future work.

Conclusions

The review highlighted the promising acceptance of CAs by users for the self-management of chronic conditions, with feedback indicating helpfulness, satisfaction, and ease of use in most of the included studies. AI-based CAs present opportunities for communication and interaction in health care settings. However, understanding and optimizing the communication channels between humans and such CAs is crucial for enhancing their capabilities and potential benefits in health care. While our study confirmed the increasing role of CAs in augmenting self-care and potentially reducing health care costs, it also exposed critical limitations, particularly concerning conversational depth and emotional intelligence. This review also emphasized the lack of reliable and comparable evidence to determine the efficacy of AI-enabled CAs for chronic health conditions. Therefore, while user feedback and acceptance were positive, there is a need for more rigorous studies and standardized reporting to evaluate the effectiveness of AI-based CAs as human-like health carers for managing NCDs.

Acknowledgments

We acknowledge the article processing fee (APF) sponsorship provided by the Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia.

This study would not have been possible without the project's support provided by the Jeffrey Cheah School of Medicine and Health Sciences (Seed grant I-M010-SED-000158).

All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this paper.

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Noncommunicable diseases. World Health Organization. 2021. URL: <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases> [accessed 2024-04-29]

2. Boutayeb A, Boutayeb S. The burden of non communicable diseases in developing countries. *Int J Equity Health*. Jan 14, 2005;4(1):2. [FREE Full text] [doi: [10.1186/1475-9276-4-2](https://doi.org/10.1186/1475-9276-4-2)] [Medline: [15651987](https://pubmed.ncbi.nlm.nih.gov/15651987/)]
3. Islam SM, Purnat TD, Phuong NT, Mwingira U, Schacht K, Fröschl G. Non-communicable diseases (NCDs) in developing countries: a symposium report. *Global Health*. Dec 11, 2014;10:81. [FREE Full text] [doi: [10.1186/s12992-014-0081-9](https://doi.org/10.1186/s12992-014-0081-9)] [Medline: [25498459](https://pubmed.ncbi.nlm.nih.gov/25498459/)]
4. Lee JS, Kang MA, Lee SK. Effects of the e-Motivate4Change program on metabolic syndrome in young adults using health apps and wearable devices: quasi-experimental study. *J Med Internet Res*. Jul 30, 2020;22(7):e17031. [FREE Full text] [doi: [10.2196/17031](https://doi.org/10.2196/17031)] [Medline: [32729838](https://pubmed.ncbi.nlm.nih.gov/32729838/)]
5. Qiao J, Lin X, Wu Y, Huang X, Pan X, Xu J, et al. Global burden of non-communicable diseases attributable to dietary risks in 1990-2019. *J Hum Nutr Diet*. Feb 2022;35(1):202-213. [doi: [10.1111/jhn.12904](https://doi.org/10.1111/jhn.12904)] [Medline: [33834556](https://pubmed.ncbi.nlm.nih.gov/33834556/)]
6. Wagner KH, Brath H. A global view on the development of non communicable diseases. *Prev Med*. May 2012;54 Suppl:S38-S41. [doi: [10.1016/j.ypmed.2011.11.012](https://doi.org/10.1016/j.ypmed.2011.11.012)] [Medline: [22178469](https://pubmed.ncbi.nlm.nih.gov/22178469/)]
7. Budreviciute A, Damiati S, Sabir DK, Onder K, Schuller-Goetzburg P, Plakys G, et al. Management and prevention strategies for non-communicable diseases (NCDs) and their risk factors. *Front Public Health*. 2020;8:574111. [FREE Full text] [doi: [10.3389/fpubh.2020.574111](https://doi.org/10.3389/fpubh.2020.574111)] [Medline: [33324597](https://pubmed.ncbi.nlm.nih.gov/33324597/)]
8. Mohamud WN, Ismail AA, Sharifuddin A, Ismail IS, Musa KI, Kadir KA, et al. Prevalence of metabolic syndrome and its risk factors in adult Malaysians: results of a nationwide survey. *Diabetes Res Clin Pract*. Feb 2011;91(2):239-245. [doi: [10.1016/j.diabres.2010.11.025](https://doi.org/10.1016/j.diabres.2010.11.025)] [Medline: [21146882](https://pubmed.ncbi.nlm.nih.gov/21146882/)]
9. Bickmore T. Relational agents for chronic disease self-management. In: Hayes BM, Aspray W, editors. *Health Informatics: A Patient-Centered Approach to Diabetes*. Cambridge, MA. MIT Pres; 2010:181-204.
10. Bickmore T, Pusateri A, Kimani EN, Paasche-Orlow MK, Trinh H, Magnani JW. Managing chronic conditions with a smartphone-based conversational virtual agent. In: *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 2018. Presented at: IVA '18; November 5-8, 2018:119-124; Sydney, Australia. URL: <https://dl.acm.org/doi/10.1145/3267851.3267908> [doi: [10.1145/3267851.3267908](https://doi.org/10.1145/3267851.3267908)]
11. Jennings A, Powell J, Armstrong N, Sturt J, Dale J. A virtual clinic for diabetes self-management: pilot study. *J Med Internet Res*. Mar 30, 2009;11(1):e10. [FREE Full text] [doi: [10.2196/jmir.1111](https://doi.org/10.2196/jmir.1111)] [Medline: [21821504](https://pubmed.ncbi.nlm.nih.gov/21821504/)]
12. Moriuchi E. Leveraging the science to understand factors influencing the use of AI-powered avatar in healthcare services. *J Technol Behav Sci*. Sep 09, 2022;7(4):588-602. [doi: [10.1007/s41347-022-00277-z](https://doi.org/10.1007/s41347-022-00277-z)]
13. Ahmed F. Beyond patient monitoring: conversational agents role in telemedicine and healthcare support for home-living elderly individuals. arXiv. Preprint posted online March 3, 2018. [FREE Full text]
14. Kankeu HT, Saksena P, Xu K, Evans DB. The financial burden from non-communicable diseases in low- and middle-income countries: a literature review. *Health Res Policy Syst*. Aug 16, 2013;11:31. [FREE Full text] [doi: [10.1186/1478-4505-11-31](https://doi.org/10.1186/1478-4505-11-31)] [Medline: [23947294](https://pubmed.ncbi.nlm.nih.gov/23947294/)]
15. Gingele AJ, Amin H, Vaassen A, Schnur I, Pearl C, Brunner-La Rocca HP, et al. Integrating avatar technology into a telemedicine application in heart failure patients : a pilot study. *Wien Klin Wochenschr*. Dec 2023;135(23-24):680-684. [FREE Full text] [doi: [10.1007/s00508-022-02150-8](https://doi.org/10.1007/s00508-022-02150-8)] [Medline: [36732377](https://pubmed.ncbi.nlm.nih.gov/36732377/)]
16. Toro-Ramos T, Lee DH, Kim Y, Michaelides A, Oh TJ, Kim KM, et al. Effectiveness of a smartphone application for the management of metabolic syndrome components focusing on weight loss: a preliminary study. *Metab Syndr Relat Disord*. Nov 2017;15(9):465-473. [doi: [10.1089/met.2017.0062](https://doi.org/10.1089/met.2017.0062)] [Medline: [29035677](https://pubmed.ncbi.nlm.nih.gov/29035677/)]
17. Jakobsen AS, Laursen LC, Rydahl-Hansen S, Østergaard B, Gerds TA, Emme C, et al. Home-based telehealth hospitalization for exacerbation of chronic obstructive pulmonary disease: findings from "the virtual hospital" trial. *Telemed J E Health*. May 2015;21(5):364-373. [FREE Full text] [doi: [10.1089/tmj.2014.0098](https://doi.org/10.1089/tmj.2014.0098)] [Medline: [25654366](https://pubmed.ncbi.nlm.nih.gov/25654366/)]
18. Gan SK, Koshy C, Nguyen PV, Haw YX. An overview of clinically and healthcare related apps in Google and Apple app stores: connecting patients, drugs, and clinicians. *Sci Phone Appl Mob Devices*. Jul 19, 2016;2(1). [doi: [10.1186/s41070-016-0012-7](https://doi.org/10.1186/s41070-016-0012-7)]
19. Karduck J, Chapman-Novakofski K. Results of the clinician apps survey, how clinicians working with patients with diabetes and obesity use mobile health apps. *J Nutr Educ Behav*. Jan 2018;50(1):62-9.e1. [doi: [10.1016/j.jneb.2017.06.004](https://doi.org/10.1016/j.jneb.2017.06.004)] [Medline: [29325664](https://pubmed.ncbi.nlm.nih.gov/29325664/)]
20. Marshall JM, Dunstan DA, Bartik W. Clinical or gimmickal: the use and effectiveness of mobile mental health apps for treating anxiety and depression. *Aust N Z J Psychiatry*. Jan 2020;54(1):20-28. [doi: [10.1177/0004867419876700](https://doi.org/10.1177/0004867419876700)] [Medline: [31552747](https://pubmed.ncbi.nlm.nih.gov/31552747/)]
21. Ladin K, Porteny T, Perugini JM, Gonzales KM, Aufort KE, Levine SK, et al. Perceptions of telehealth vs in-person visits among older adults with advanced kidney disease, care partners, and clinicians. *JAMA Netw Open*. Dec 01, 2021;4(12):e2137193. [FREE Full text] [doi: [10.1001/jamanetworkopen.2021.37193](https://doi.org/10.1001/jamanetworkopen.2021.37193)] [Medline: [34870680](https://pubmed.ncbi.nlm.nih.gov/34870680/)]
22. Matamala-Gomez M, Bottiroli S, Realdon O, Riva G, Galvagni L, Platz T, et al. Telemedicine and virtual reality at time of COVID-19 pandemic: an overview for future perspectives in neurorehabilitation. *Front Neurol*. 2021;12:646902. [FREE Full text] [doi: [10.3389/fneur.2021.646902](https://doi.org/10.3389/fneur.2021.646902)] [Medline: [33841313](https://pubmed.ncbi.nlm.nih.gov/33841313/)]
23. Krausz M, Ward J, Ramsey D. From telehealth to an interactive virtual clinic. In: Mucic D, Hilty DM, editors. *e-Mental Health*. Cham, Switzerland. Springer; 2016:289-310.

24. Gong E, Baptista S, Russell A, Scuffham P, Riddell M, Speight J, et al. My diabetes coach, a mobile app-based interactive conversational agent to support type 2 diabetes self-management: randomized effectiveness-implementation trial. *J Med Internet Res*. Nov 05, 2020;22(11):e20322. [FREE Full text] [doi: [10.2196/20322](https://doi.org/10.2196/20322)] [Medline: [33151154](https://pubmed.ncbi.nlm.nih.gov/33151154/)]
25. Viana Ó, Terroso M, Serejo C, Vilaça JL. Ascertaining the influence of style on the credibility and appeal of a digital health avatar. In: *Proceedings of the 6th International Conference on Design and Digital Communication*. Cham. Springer; 2022. Presented at: DIGICOM '22; November 3–5, 2022:27; Barcelos, Portugal. URL: https://link.springer.com/chapter/10.1007/978-3-031-20364-0_6 [doi: [10.1007/978-3-031-20364-0_6](https://doi.org/10.1007/978-3-031-20364-0_6)]
26. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. *J Med Internet Res*. Oct 22, 2020;22(10):e20346. [FREE Full text] [doi: [10.2196/20346](https://doi.org/10.2196/20346)] [Medline: [33090118](https://pubmed.ncbi.nlm.nih.gov/33090118/)]
27. Bin Sawad A, Narayan B, Alnefaie A, Maqbool A, Mckie I, Smith J, et al. A systematic review on healthcare artificial intelligent conversational agents for chronic conditions. *Sensors (Basel)*. Mar 29, 2022;22(7):2625. [FREE Full text] [doi: [10.3390/s22072625](https://doi.org/10.3390/s22072625)] [Medline: [35408238](https://pubmed.ncbi.nlm.nih.gov/35408238/)]
28. Balsa J, Neves P, Félix I, Guerreiro MP, Alves P, Carmo MB, et al. Intelligent virtual assistant for promoting behaviour change in older people with T2D. In: *Proceedings of the 19th EPIA Conference on Artificial Intelligence on Progress in Artificial Intelligence*. 2019. Presented at: EPIA '19; September 3-6, 2019:372-383; Vila Real, Portugal. URL: https://link.springer.com/chapter/10.1007/978-3-030-30241-2_32 [doi: [10.1007/978-3-030-30241-2_32](https://doi.org/10.1007/978-3-030-30241-2_32)]
29. Van Pinxteren MM, Pluymaekers M, Lemmink JG. Human-like communication in conversational agents: a literature review and research agenda. *J Serv Manag*. Jun 11, 2020;31(2):203-225. [doi: [10.1108/josm-06-2019-0175](https://doi.org/10.1108/josm-06-2019-0175)]
30. Potor M. Voice bot vs. chatbot: what's the difference, and why does it matter? *Sinch Engage*. URL: <https://engage.sinch.com/blog/voice-bot-vs-chatbot-whats-the-difference/> [accessed 2024-04-29]
31. Singh S, Thakur HK. Survey of various AI chatbots based on technology used. In: *Proceedings of the 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*. 2020. Presented at: ICrito '20; June 4-5, 2020:1074-1079; Noida, India. URL: <https://ieeexplore.ieee.org/document/9197943> [doi: [10.1109/icrito48877.2020.9197943](https://doi.org/10.1109/icrito48877.2020.9197943)]
32. Daher K, Casas J, Khaled OA, Mugellini E. Empathic chatbot response for medical assistance. In: *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. 2020. Presented at: IVA '20; October 20-22, 2020:1-3; Virtual Event. URL: <https://dl.acm.org/doi/10.1145/3383652.3423864> [doi: [10.1145/3383652.3423864](https://doi.org/10.1145/3383652.3423864)]
33. Singh S, Beniwal H. A survey on near-human conversational agents. *J King Saud Univ Comput Inf Sci*. Nov 2022;34(10):8852-8866. [doi: [10.1016/j.jksuci.2021.10.013](https://doi.org/10.1016/j.jksuci.2021.10.013)]
34. Valtolina S, Hu L. Charlie: a chatbot to improve the elderly quality of life and to make them more active to fight their sense of loneliness. In: *Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter*. 2021. Presented at: CHIItaly '21; July 11-13, 2021:1-5; Bolzano, Italy. URL: <https://dl.acm.org/doi/abs/10.1145/3464385.3464726> [doi: [10.1145/3464385.3464726](https://doi.org/10.1145/3464385.3464726)]
35. Oruganti SC. Virtual bank assistance: an AI based voice bot for better banking. *Int J Res*. 2020;9(1):177-183. [FREE Full text] [doi: [10.13140/RG.2.2.21535.10405](https://doi.org/10.13140/RG.2.2.21535.10405)]
36. Ma Y, Drewes H, Butz A. Fake moods: can users trick an emotion-aware VoiceBot? In: *Proceedings of the Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021. Presented at: CHI EA '21; May 8-13, 2021:1-4; Yokohama, Japan. URL: <https://dl.acm.org/doi/10.1145/3411763.3451744> [doi: [10.1145/3411763.3451744](https://doi.org/10.1145/3411763.3451744)]
37. Eagle T, Blau C, Bales S, Desai N, Li V, Whittaker S. "I don't know what you mean by 'I am anxious'": a new method for evaluating conversational agent responses to standardized mental health inputs for anxiety and depression. *ACM Trans Interact Intell Syst*. Jul 20, 2022;12(2):1-23. [doi: [10.1145/3488057](https://doi.org/10.1145/3488057)]
38. Magyar G, Balsa J, Cláudio AP, Carmo MB, Neves P, Alves P, et al. Anthropomorphic virtual assistant to support self-care of type 2 diabetes in older people: a perspective on the role of artificial intelligence. In: *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. 2019. Presented at: VISIGRAPP '19; February 25-27, 2019:323-331; Pragur, Czech Republic. URL: <https://www.scitepress.org/Papers/2019/75724/75724.pdf> [doi: [10.5220/0007572403230331](https://doi.org/10.5220/0007572403230331)]
39. Sestino A, D'Angelo A. My doctor is an avatar! the effect of anthropomorphism and emotional receptivity on individuals' intention to use digital-based healthcare services. *Technol Forecast Soc Change*. Jun 2023;191:122505. [doi: [10.1016/j.techfore.2023.122505](https://doi.org/10.1016/j.techfore.2023.122505)]
40. Shaked NA. Avatars and virtual agents - relationship interfaces for the elderly. *Healthc Technol Lett*. Jun 2017;4(3):83-87. [FREE Full text] [doi: [10.1049/hfl.2017.0009](https://doi.org/10.1049/hfl.2017.0009)] [Medline: [28706725](https://pubmed.ncbi.nlm.nih.gov/28706725/)]
41. Jacob K. What is a digital avatar? *DaveAI*. URL: <https://www.iamdave.ai/blog/what-is-a-digital-avatar/> [accessed 2024-04-29]
42. Seymour M, Yuan L, Dennis AR, Riemer K. Facing the artificial: understanding affinity, trustworthiness, and preference for more realistic digital humans. In: *Proceedings of the 53rd Hawaii International Conference on System Sciences*. 2020. Presented at: HICSS '20; January 7-10, 2020:4673-4682; Maui, HI. URL: <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/07da3db0-4dec-40ae-9bd8-550de5636593/content> [doi: [10.24251/hicss.2020.574](https://doi.org/10.24251/hicss.2020.574)]
43. Bellan R. Are lifelike digital humans the future of customer experience? *Techcrunch*. URL: <https://techcrunch.com/2022/02/14/are-lifelike-digital-humans-the-future-of-customer-experience/> [accessed 2024-04-29]

44. Lukan E. Everything you need to know about digital humans. Firework. URL: <https://www.synthesia.io/post/digital-humans> [accessed 2024-04-29]
45. Okegbile SD, Cai J. Edge-assisted human-to-virtual twin connectivity scheme for human digital twin frameworks. In: Proceedings of the 95th Conference on Vehicular Technology Conference. 2022. Presented at: VTC '22; June 19-22, 2022:1-6; Helsinki, Finland. URL: <https://ieeexplore.ieee.org/document/9860619> [doi: [10.1109/vtc2022-spring54318.2022.9860619](https://doi.org/10.1109/vtc2022-spring54318.2022.9860619)]
46. Lauer-Schmaltz MW, Cash P, Hansen JP, Maier A. Designing human digital twins for behaviour-changing therapy and rehabilitation: a systematic review. Proc Des Soc. May 26, 2022;2:1303-1312. [doi: [10.1017/pds.2022.132](https://doi.org/10.1017/pds.2022.132)]
47. Shubham K, Mukherjee A, Jayagopi DB. Review of realistic behavior and appearance generation in embodied conversational agents: a comparison between traditional and modern approaches. In: Proceedings of the 2022 International Conference on Multimodal Interaction. 2022. Presented at: ICMI '22; November 7-11, 2022:191-197; Bengaluru, India. URL: <https://dl.acm.org/doi/10.1145/3536221.3556592> [doi: [10.1145/3536221.3556592](https://doi.org/10.1145/3536221.3556592)]
48. Yaqoob I, Salah K, Jayaraman R, Omar M. Metaverse applications in smart cities: enabling technologies, opportunities, challenges, and future directions. Internet Things. Oct 2023;23:100884. [doi: [10.1016/j.iot.2023.100884](https://doi.org/10.1016/j.iot.2023.100884)]
49. Lucas GM, Gratch J, King A, Morency LP. It's only a computer: virtual humans increase willingness to disclose. Comput Human Behav. Aug 2014;37:94-100. [doi: [10.1016/j.chb.2014.04.043](https://doi.org/10.1016/j.chb.2014.04.043)]
50. Tulsulkar G, Mishra N, Thalmann NM, Lim HE, Lee MP, Cheng SK. Can a humanoid social robot stimulate the interactivity of cognitively impaired elderly? A thorough study based on computer vision methods. Vis Comput. 2021;37(12):3019-3038. [FREE Full text] [doi: [10.1007/s00371-021-02242-y](https://doi.org/10.1007/s00371-021-02242-y)] [Medline: [34345091](https://pubmed.ncbi.nlm.nih.gov/34345091/)]
51. Mukherjee S, Baral MM, Pal SK, Chittipaka V, Roy R, Alam K. Humanoid robot in healthcare: a systematic review and future research directions. In: Proceedings of the 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing. 2022. Presented at: COM-IT-CON '22; May 26-27, 2022:822-826; Faridabad, India. URL: <https://ieeexplore.ieee.org/document/9850577> [doi: [10.1109/com-it-con54601.2022.9850577](https://doi.org/10.1109/com-it-con54601.2022.9850577)]
52. Linders GM, Vaitonytė J, Alimardani M, Mitev KO, Louwerse MM. A realistic, multimodal virtual agent for the healthcare domain. In: Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents. 2022. Presented at: IVA '22; September 6-9, 2022:1-3; Faro, Portugal. URL: <https://dl.acm.org/doi/abs/10.1145/3514197.3551250> [doi: [10.1145/3514197.3551250](https://doi.org/10.1145/3514197.3551250)]
53. Adams B, Breazeal C, Brooks RA, Scassellati B. Humanoid robots: a new kind of tool. IEEE Intell Syst. Jul 2000;15(4):25-31. [doi: [10.1109/5254.867909](https://doi.org/10.1109/5254.867909)]
54. Kaptein F, Kiefer B, Cully A, Celiktutan O, Bierman B, Rijgersberg-peters R, et al. A cloud-based robot system for long-term interaction: principles, implementation, lessons learned. J Hum Robot Interact. Oct 18, 2021;11(1):1-27. [doi: [10.1145/3481585](https://doi.org/10.1145/3481585)]
55. Garner TA, Powell WA, Carr V. Virtual carers for the elderly: a case study review of ethical responsibilities. Digit Health. 2016;2:2055207616681173. [FREE Full text] [doi: [10.1177/2055207616681173](https://doi.org/10.1177/2055207616681173)] [Medline: [29942577](https://pubmed.ncbi.nlm.nih.gov/29942577/)]
56. Kusal S, Patil S, Choudrie J, Kotecha K, Mishra S, Abraham A. AI-based conversational agents: a scoping review from technologies to future directions. IEEE Access. 2022;10:92337-92356. [doi: [10.1109/access.2022.3201144](https://doi.org/10.1109/access.2022.3201144)]
57. Yaghoubzadeh R, Kramer M, Pitsch K, Kopp S. Virtual agents as daily assistants for elderly or cognitively impaired people: studies on acceptance and interaction feasibility. In: Proceedings of the 13th International Conference on Intelligent Virtual Agents. 2013. Presented at: IVA '13; August 29-31, 2013:29-31; Edinburgh, UK. URL: https://link.springer.com/chapter/10.1007/978-3-642-40415-3_7 [doi: [10.1007/978-3-642-40415-3_7](https://doi.org/10.1007/978-3-642-40415-3_7)]
58. Cheng A, Raghavaraju V, Kanugo J, Handrianto YP, Yi S. Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes. In: Proceedings of the 15th IEEE Annual Consumer Communications & Networking Conference. 2018. Presented at: CCNC '18; January 12-15, 2018:1-5; Las Vegas, NV. URL: <https://ieeexplore.ieee.org/document/8319283> [doi: [10.1109/ccnc.2018.8319283](https://doi.org/10.1109/ccnc.2018.8319283)]
59. Thakur N, Han CY. An approach to analyze the social acceptance of virtual assistants by elderly people. In: Proceedings of the 8th International Conference on the Internet of Things. 2018. Presented at: IOT '18; October 15-18, 2018:1-6; Santa Barbara, CA. URL: <https://dl.acm.org/doi/10.1145/3277593.3277616> [doi: [10.1145/3277593.3277616](https://doi.org/10.1145/3277593.3277616)]
60. Dinger T, Kwasnicka D, Wei J, Gong E, Oldenburg B. The use and promise of conversational agents in digital health. Yearb Med Inform. Aug 2021;30(1):191-199. [FREE Full text] [doi: [10.1055/s-0041-1726510](https://doi.org/10.1055/s-0041-1726510)] [Medline: [34479391](https://pubmed.ncbi.nlm.nih.gov/34479391/)]
61. Lyzwinski LN, Elgendi M, Menon C. Conversational agents and avatars for cardiometabolic risk factors and lifestyle-related behaviors: scoping review. JMIR Mhealth Uhealth. May 25, 2023;11:e39649. [FREE Full text] [doi: [10.2196/39649](https://doi.org/10.2196/39649)] [Medline: [37227765](https://pubmed.ncbi.nlm.nih.gov/37227765/)]
62. Tudor Car L, Dhinakaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res. Aug 07, 2020;22(8):e17158. [FREE Full text] [doi: [10.2196/17158](https://doi.org/10.2196/17158)] [Medline: [32763886](https://pubmed.ncbi.nlm.nih.gov/32763886/)]
63. Monaco A, Palmer K, Holm Ravn Faber N, Kohler I, Silva M, Vatland A, et al. Digital health tools for managing noncommunicable diseases during and after the COVID-19 pandemic: perspectives of patients and caregivers. J Med Internet Res. Jan 29, 2021;23(1):e25652. [FREE Full text] [doi: [10.2196/25652](https://doi.org/10.2196/25652)] [Medline: [33464206](https://pubmed.ncbi.nlm.nih.gov/33464206/)]

64. Guerreiro MP, Angelini L, Rafael Henriques H, El Kamali M, Baixinho C, Balsa J, et al. Conversational agents for health and well-being across the life course: protocol for an evidence map. *JMIR Res Protoc*. Sep 17, 2021;10(9):e26680. [FREE Full text] [doi: [10.2196/26680](https://doi.org/10.2196/26680)] [Medline: [34533460](https://pubmed.ncbi.nlm.nih.gov/34533460/)]
65. Xing Z, Yu F, Qanir YA, Guan T, Walker J, Song L. Intelligent conversational agents in patient self-management: a systematic survey using multi data sources. *Stud Health Technol Inform*. Aug 21, 2019;264:1813-1814. [doi: [10.3233/SHTI190661](https://doi.org/10.3233/SHTI190661)] [Medline: [31438357](https://pubmed.ncbi.nlm.nih.gov/31438357/)]
66. Griffin AC, Xing Z, Khairat S, Wang Y, Bailey S, Arguello J, et al. Conversational agents for chronic disease self-management: a systematic review. *AMIA Annu Symp Proc*. 2020;2020:504-513. [FREE Full text] [Medline: [33936424](https://pubmed.ncbi.nlm.nih.gov/33936424/)]
67. Martinengo L, Jabir AI, Goh WW, Lo NY, Ho MR, Kowatsch T, et al. Conversational agents in health care: scoping review of their behavior change techniques and underpinning theory. *J Med Internet Res*. Oct 03, 2022;24(10):e39243. [FREE Full text] [doi: [10.2196/39243](https://doi.org/10.2196/39243)] [Medline: [36190749](https://pubmed.ncbi.nlm.nih.gov/36190749/)]
68. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. Feb 2005;8(1):19-32. [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
69. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci*. Sep 20, 2010;5(1):69. [FREE Full text] [doi: [10.1186/1748-5908-5-69](https://doi.org/10.1186/1748-5908-5-69)] [Medline: [20854677](https://pubmed.ncbi.nlm.nih.gov/20854677/)]
70. Even C, Hammann T, Heyl V, Rietz C, Wahl HW, Zentel P, et al. Benefits and challenges of conversational agents in older adults : a scoping review. *Z Gerontol Geriatr*. Aug 2022;55(5):381-387. [doi: [10.1007/s00391-022-02085-9](https://doi.org/10.1007/s00391-022-02085-9)] [Medline: [35852588](https://pubmed.ncbi.nlm.nih.gov/35852588/)]
71. Grudin J. ChatGPT and chat history: challenges for the new wave. *Computer*. May 2023;56(5):94-100. [doi: [10.1109/mc.2023.3255279](https://doi.org/10.1109/mc.2023.3255279)]
72. Balsa J, Félix I, Cláudio AP, Carmo MB, Silva IC, Guerreiro A, et al. Usability of an intelligent virtual assistant for promoting behavior change and self-care in older people with type 2 diabetes. *J Med Syst*. Jun 13, 2020;44(7):130. [doi: [10.1007/s10916-020-01583-w](https://doi.org/10.1007/s10916-020-01583-w)] [Medline: [32533367](https://pubmed.ncbi.nlm.nih.gov/32533367/)]
73. Stara V, Vera B, Bolliger D, Rossi L, Felici E, Di Rosa M, et al. Usability and acceptance of the embodied conversational agent Anne by people with dementia and their caregivers: exploratory study in home environment settings. *JMIR Mhealth Uhealth*. Jun 25, 2021;9(6):e25891. [FREE Full text] [doi: [10.2196/25891](https://doi.org/10.2196/25891)] [Medline: [34170256](https://pubmed.ncbi.nlm.nih.gov/34170256/)]
74. Nguyen TT, Sim K, Kuen AT, O'Donnell RR, Lim ST, Wang W, et al. Designing AI-based conversational agent for diabetes care in a multilingual context. *arXiv*. Preprint posted online May 20, 2021. [FREE Full text]
75. Babington-Ashaye A, de Moerloose P, Diop S, Geissbuhler A. Design, development and usability of an educational AI chatbot for people with haemophilia in senegal. *Haemophilia*. Jul 2023;29(4):1063-1073. [doi: [10.1111/hae.14815](https://doi.org/10.1111/hae.14815)] [Medline: [37347648](https://pubmed.ncbi.nlm.nih.gov/37347648/)]
76. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. *Transl Behav Med*. May 16, 2019;9(3):440-447. [doi: [10.1093/tbm/ibz043](https://doi.org/10.1093/tbm/ibz043)] [Medline: [31094445](https://pubmed.ncbi.nlm.nih.gov/31094445/)]
77. Issom DZ, Rochat J, Hartvigsen G, Lovis C. Preliminary evaluation of a mHealth coaching conversational artificial intelligence for the self-care management of people with sickle-cell disease. *Stud Health Technol Inform*. Jun 16, 2020;270:1361-1362. [doi: [10.3233/SHTI200442](https://doi.org/10.3233/SHTI200442)] [Medline: [32570659](https://pubmed.ncbi.nlm.nih.gov/32570659/)]
78. Rahmanti AR, Yang HC, Bintoro BS, Nursetyo AA, Muhtar MS, Syed-Abdul S, et al. SlimMe, a chatbot with artificial empathy for personal weight management: system design and finding. *Front Nutr*. 2022;9:870775. [FREE Full text] [doi: [10.3389/fnut.2022.870775](https://doi.org/10.3389/fnut.2022.870775)] [Medline: [35811989](https://pubmed.ncbi.nlm.nih.gov/35811989/)]
79. Alhmiedat T, Alotaibi M. Employing social robots for managing diabetes among children: SARA. *Wireless Pers Commun*. Mar 14, 2023;130(1):449-468. [doi: [10.1007/S11277-023-10293-8](https://doi.org/10.1007/S11277-023-10293-8)]
80. Kimani E, Bickmore T, Trinh H, Ring L, Paasche-Orlow MK, Magnani JW. A smartphone-based virtual agent for atrial fibrillation education and counseling. In: *Proceedings of the 16th International Conference on Intelligent Virtual Agents*. 2016. Presented at: IVA '16; September 20-23, 2016:120-127; Los Angeles, CA. URL: https://link.springer.com/chapter/10.1007/978-3-319-47665-0_11 [doi: [10.1007/978-3-319-47665-0_11](https://doi.org/10.1007/978-3-319-47665-0_11)]
81. Kataoka Y, Takemura T, Sasajima M, Katoh N. Development and early feasibility of chatbots for educating patients with lung cancer and their caregivers in Japan: mixed methods study. *JMIR Cancer*. Mar 10, 2021;7(1):e26911. [FREE Full text] [doi: [10.2196/26911](https://doi.org/10.2196/26911)] [Medline: [33688839](https://pubmed.ncbi.nlm.nih.gov/33688839/)]
82. Krishnakumar A, Verma R, Chawla R, Sosale A, Saboo B, Joshi S, et al. Evaluating glycemic control in patients of South Asian origin with type 2 diabetes using a digital therapeutic platform: analysis of real-world data. *J Med Internet Res*. Mar 25, 2021;23(3):e17908. [FREE Full text] [doi: [10.2196/17908](https://doi.org/10.2196/17908)] [Medline: [33764306](https://pubmed.ncbi.nlm.nih.gov/33764306/)]
83. Maharjan R, Doherty K, Rohani DA, Bækgaard P, Bardram JE. Experiences of a speech-enabled conversational agent for the self-report of well-being among people living with affective disorders: an in-the-wild study. *ACM Trans Interact Intell Syst*. Jul 20, 2022;12(2):1-29. [doi: [10.1145/3484508](https://doi.org/10.1145/3484508)]
84. Calvo RA, Peters D, Moradbakhti L, Cook D, Rizos G, Schuller B, et al. Assessing the feasibility of a text-based conversational agent for asthma support: protocol for a mixed methods observational study. *JMIR Res Protoc*. Feb 02, 2023;12:e42965. [FREE Full text] [doi: [10.2196/42965](https://doi.org/10.2196/42965)] [Medline: [36729586](https://pubmed.ncbi.nlm.nih.gov/36729586/)]

85. Neerinx MA, van Vught W, Blanson Henkemans O, Oleari E, Broekens J, Peters R, et al. Socio-cognitive engineering of a robotic partner for child's diabetes self-management. *Front Robot AI*. 2019;6:118. [FREE Full text] [doi: [10.3389/frobt.2019.00118](https://doi.org/10.3389/frobt.2019.00118)] [Medline: [33501133](https://pubmed.ncbi.nlm.nih.gov/33501133/)]
86. Watson A, Bickmore T, Cange A, Kulshreshtha A, Kvedar J. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. *J Med Internet Res*. Jan 26, 2012;14(1):e1. [FREE Full text] [doi: [10.2196/jmir.1629](https://doi.org/10.2196/jmir.1629)] [Medline: [22281837](https://pubmed.ncbi.nlm.nih.gov/22281837/)]
87. Shamekhi A, Bickmore T, Lestoquoy A, Gardiner P. Augmenting group medical visits with conversational agents for stress management behavior change. In: *Proceedings of the 12th International Conference on Persuasive Technology: Development and Implementation of Personalized Technologies to Change Attitudes and Behaviors*. 2017. Presented at: PERSUASIVE '17; April 4-6, 2017:55-67; Amsterdam, the Netherlands. URL: https://link.springer.com/chapter/10.1007/978-3-319-55134-0_5
88. Suganuma S, Sakamoto D, Shimoyama H. An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: feasibility and acceptability pilot trial. *JMIR Ment Health*. Jul 31, 2018;5(3):e10454. [FREE Full text] [doi: [10.2196/10454](https://doi.org/10.2196/10454)] [Medline: [30064969](https://pubmed.ncbi.nlm.nih.gov/30064969/)]
89. Hussain S, Athula G. Extending a conventional chatbot knowledge base to external knowledge source and introducing user based sessions for diabetes education. In: *Proceedings of the 32nd International Conference on Advanced Information Networking and Applications Workshops*. 2018. Presented at: WAINA '18; May 16-18, 2018:698-703; Krakow, Poland. URL: <https://ieeexplore.ieee.org/document/8418155> [doi: [10.1109/waina.2018.00170](https://doi.org/10.1109/waina.2018.00170)]
90. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth*. Nov 23, 2018;6(11):e12106. [FREE Full text] [doi: [10.2196/12106](https://doi.org/10.2196/12106)] [Medline: [30470676](https://pubmed.ncbi.nlm.nih.gov/30470676/)]
91. Easton K, Potter S, Bec R, Bennion M, Christensen H, Grindell C, et al. A virtual agent to support individuals living with physical and mental comorbidities: co-design and acceptability testing. *J Med Internet Res*. May 30, 2019;21(5):e12996. [FREE Full text] [doi: [10.2196/12996](https://doi.org/10.2196/12996)] [Medline: [31148545](https://pubmed.ncbi.nlm.nih.gov/31148545/)]
92. Chaix B, Bibault JE, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. *JMIR Cancer*. May 02, 2019;5(1):e12856. [FREE Full text] [doi: [10.2196/12856](https://doi.org/10.2196/12856)] [Medline: [31045505](https://pubmed.ncbi.nlm.nih.gov/31045505/)]
93. Anastasiadou M, Alexiadis A, Polychronidou E, Votis K, Tzovaras D. A prototype educational virtual assistant for diabetes management. In: *Proceedings of the 20th International Conference on Bioinformatics and Bioengineering*. 2020. Presented at: BIBE '20; October 26-28, 2020:999-1004; Cincinnati, OH. URL: <https://ieeexplore.ieee.org/document/9288129> [doi: [10.1109/bibe50027.2020.00169](https://doi.org/10.1109/bibe50027.2020.00169)]
94. Roca S, Lozano ML, García J, Alesanco Á. Validation of a virtual assistant for improving medication adherence in patients with comorbid type 2 diabetes mellitus and depressive disorder. *Int J Environ Res Public Health*. Nov 17, 2021;18(22):12056. [FREE Full text] [doi: [10.3390/ijerph182212056](https://doi.org/10.3390/ijerph182212056)] [Medline: [34831811](https://pubmed.ncbi.nlm.nih.gov/34831811/)]
95. Yao K, Wong KK, Yu X, Volpi J, Wong ST. An intelligent augmented lifelike avatar app for virtual physical examination of suspected strokes. *Annu Int Conf IEEE Eng Med Biol Soc*. Nov 2021;2021:1727-1730. [doi: [10.1109/EMBC46164.2021.9629720](https://doi.org/10.1109/EMBC46164.2021.9629720)] [Medline: [34891620](https://pubmed.ncbi.nlm.nih.gov/34891620/)]
96. Egede JO, Price D, Krishnan DB, Jaiswal S, Elliott N, Morriss R. Design and evaluation of virtual human mediated tasks for assessment of depression and anxiety. In: *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*. 2021. Presented at: IVA '21; September 14-17, 2021:52-59; Virtual Event. URL: <https://dl.acm.org/doi/10.1145/3472306.3478361> [doi: [10.1145/3472306.3478361](https://doi.org/10.1145/3472306.3478361)]
97. Romanovskiy O, Pidbutska N, Knysh A. Elomia chatbot: the effectiveness of artificial intelligence in the fight for mental health. In: *Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems*. 2021. Presented at: COLINS '21; April 22-23, 2021:1-10; Kharkiv, Ukraine. URL: <https://ceur-ws.org/Vol-2870/paper89.pdf>
98. Anan T, Kajiki S, Oka H, Fujii T, Kawamata K, Mori K, et al. Effects of an artificial intelligence-assisted health program on workers with neck/shoulder pain/stiffness and low back pain: randomized controlled trial. *JMIR Mhealth Uhealth*. Sep 24, 2021;9(9):e27535. [FREE Full text] [doi: [10.2196/27535](https://doi.org/10.2196/27535)] [Medline: [34559054](https://pubmed.ncbi.nlm.nih.gov/34559054/)]
99. Zisis G, Carrington MJ, Oldenburg B, Whitmore K, Lay M, Huynh Q, et al. An m-Health intervention to improve education, self-management, and outcomes in patients admitted for acute decompensated heart failure: barriers to effective implementation. *Eur Heart J Digit Health*. Dec 2021;2(4):649-657. [FREE Full text] [doi: [10.1093/ehjdh/ztab085](https://doi.org/10.1093/ehjdh/ztab085)] [Medline: [36713108](https://pubmed.ncbi.nlm.nih.gov/36713108/)]
100. Apergi LA, Bjarnadottir MV, Baras JS, Golden BL, Anderson KM, Chou J, et al. Voice interface technology adoption by patients with heart failure: pilot comparison study. *JMIR Mhealth Uhealth*. Apr 01, 2021;9(4):e24646. [FREE Full text] [doi: [10.2196/24646](https://doi.org/10.2196/24646)] [Medline: [33792556](https://pubmed.ncbi.nlm.nih.gov/33792556/)]
101. Rathnayaka P, Mills N, Burnett D, De Silva D, Alahakoon D, Gray R. A mental health chatbot with cognitive skills for personalised behavioural activation and remote health monitoring. *Sensors (Basel)*. May 11, 2022;22(10):3653. [FREE Full text] [doi: [10.3390/s22103653](https://doi.org/10.3390/s22103653)] [Medline: [35632061](https://pubmed.ncbi.nlm.nih.gov/35632061/)]

102. Kannampallil T, Ronneberg CR, Wittels NE, Kumar V, Lv N, Smyth JM, et al. Design and formative evaluation of a virtual voice-based coach for problem-solving treatment: observational study. *JMIR Form Res*. Aug 12, 2022;6(8):e38092. [FREE Full text] [doi: [10.2196/38092](https://doi.org/10.2196/38092)] [Medline: [35969431](https://pubmed.ncbi.nlm.nih.gov/35969431/)]
103. Alturaiki AM, Banjar HR, Barefah AS, Alnajjar SA, Hindawi S. A smart chatbot for interactive management in beta thalassemia patients. *Int J Telemed Appl*. 2022;2022:9734518. [FREE Full text] [doi: [10.1155/2022/9734518](https://doi.org/10.1155/2022/9734518)] [Medline: [35601050](https://pubmed.ncbi.nlm.nih.gov/35601050/)]
104. Zahedi FM, Zhao H, Sanvanson P, Walia N, Jain H, Shaker R. My real avatar has a doctor appointment in the Wepital: a system for persistent, efficient, and ubiquitous medical care. *Inf Manag*. Dec 2022;59(8):103706. [doi: [10.1016/j.im.2022.103706](https://doi.org/10.1016/j.im.2022.103706)]
105. Meheli S, Sinha C, Kadaba M. Understanding people with chronic pain who use a cognitive behavioral therapy-based artificial intelligence mental health app (Wysa): mixed methods retrospective observational study. *JMIR Hum Factors*. Apr 27, 2022;9(2):e35671. [FREE Full text] [doi: [10.2196/35671](https://doi.org/10.2196/35671)] [Medline: [35314422](https://pubmed.ncbi.nlm.nih.gov/35314422/)]
106. Henson JB, Glissen Brown JR, Lee JP, Patel A, Leiman DA. Evaluation of the potential utility of an artificial intelligence chatbot in gastroesophageal reflux disease management. *Am J Gastroenterol*. Dec 01, 2023;118(12):2276-2279. [doi: [10.14309/ajg.0000000000002397](https://doi.org/10.14309/ajg.0000000000002397)] [Medline: [37410934](https://pubmed.ncbi.nlm.nih.gov/37410934/)]
107. Epalte K, Tomsone S, Vētra A, Bērziņa G. Patient experience using digital therapy "Vigo" for stroke patient recovery: a qualitative descriptive study. *Disabil Rehabil Assist Technol*. Feb 2023;18(2):175-184. [doi: [10.1080/17483107.2020.1839794](https://doi.org/10.1080/17483107.2020.1839794)] [Medline: [33155507](https://pubmed.ncbi.nlm.nih.gov/33155507/)]
108. LeRouge C, Dickhut K, Lisetti C, Sangameswaran S, Malasanos T. Engaging adolescents in a computer-based weight management program: avatars and virtual coaches could help. *J Am Med Inform Assoc*. Jan 2016;23(1):19-28. [FREE Full text] [doi: [10.1093/jamia/ocv078](https://doi.org/10.1093/jamia/ocv078)] [Medline: [26159466](https://pubmed.ncbi.nlm.nih.gov/26159466/)]
109. Park G, Chung J, Lee S. Human vs. machine-like representation in chatbot mental health counseling: the serial mediation of psychological distance and trust on compliance intention. *Curr Psychol*. Apr 20, 2023:1-12. [FREE Full text] [doi: [10.1007/s12144-023-04653-7](https://doi.org/10.1007/s12144-023-04653-7)] [Medline: [37359642](https://pubmed.ncbi.nlm.nih.gov/37359642/)]
110. Boggiss A, Consedine N, Hopkins S, Silvester C, Jefferies C, Hofman P, et al. Improving the well-being of adolescents with type 1 diabetes during the COVID-19 pandemic: qualitative study exploring acceptability and clinical usability of a self-compassion chatbot. *JMIR Diabetes*. May 05, 2023;8:e40641. [FREE Full text] [doi: [10.2196/40641](https://doi.org/10.2196/40641)] [Medline: [36939680](https://pubmed.ncbi.nlm.nih.gov/36939680/)]
111. Kocaballi AB, Quiroz JC, Rezazadegan D, Berkovsky S, Magrabi F, Coiera E, et al. Responses of conversational agents to health and lifestyle prompts: investigation of appropriateness and presentation structures. *J Med Internet Res*. Feb 09, 2020;22(2):e15823. [FREE Full text] [doi: [10.2196/15823](https://doi.org/10.2196/15823)] [Medline: [32039810](https://pubmed.ncbi.nlm.nih.gov/32039810/)]
112. Wiratunga N, Cooper K, Wijekoon A, Palihawadana C, Mendham V, Reiter E, et al. FitChat: conversational artificial intelligence interventions for encouraging physical activity in older adults. *arXiv*. Preprint posted online April 29, 2020. [FREE Full text] [doi: [10.48550/arXiv.2004.14067](https://doi.org/10.48550/arXiv.2004.14067)]
113. Denecke K, Tschanz M, Dorner TL, May R. Intelligent conversational agents in healthcare: hype or hope? *Stud Health Technol Inform*. 2019;259:77-84. [Medline: [30923277](https://pubmed.ncbi.nlm.nih.gov/30923277/)]
114. Kim K, Norouzi N, Losekamp T, Bruder G, Anderson M, Welch G. Effects of patient care assistant embodiment and computer mediation on user experience. In: *Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality*. 2019. Presented at: AIVR '19; December 9-11, 2019:17-177; San Diego, CA. URL: <https://ieeexplore.ieee.org/document/8942267> [doi: [10.1109/aivr46125.2019.00013](https://doi.org/10.1109/aivr46125.2019.00013)]
115. Sen A, Liew SH. Augmented reality and its use in education. In: *Tatnall A, editor. Encyclopedia of Education and Information Technologies*. Cham, Switzerland. Springer; 2018:1719-1726.
116. Kim K, Boelling L, Haesler S, Bailenson J, Bruder G, Welch GF. Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR. In: *Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality*. 2018. Presented at: ISMAR '18; October 16-20, 2018:105-115; Munich, Germany. URL: <https://ieeexplore.ieee.org/document/8613756> [doi: [10.1109/ISMAR.2018.00039](https://doi.org/10.1109/ISMAR.2018.00039)]
117. Li Q, Luximon Y, Zhang J. The influence of anthropomorphic cues on patients' perceived anthropomorphism, social presence, trust building, and acceptance of health care conversational agents: within-subject web-based experiment. *J Med Internet Res*. Aug 10, 2023;25:e44479. [FREE Full text] [doi: [10.2196/44479](https://doi.org/10.2196/44479)] [Medline: [37561567](https://pubmed.ncbi.nlm.nih.gov/37561567/)]
118. Grunitz M. Rule-based AI vs machine learning: what's the difference? *WeAreBrain*. Sep 2021. URL: <https://wearebrain.com/blog/rule-based-ai-vs-machine-learning-whats-the-difference/> [accessed 2024-04-29]
119. Ahmad R, Siemon D, Gnewuch U, Robra-Bissantz S. The benefits and caveats of personality-adaptive conversational agents in mental health care. In: *Proceedings of the 27th annual Americas Conference on Information Systems*. 2021. Presented at: AMCIS '21; August 9-13, 2021; Virtual Event. URL: https://web.archive.org/web/20220804131608id_/https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1225&context=amcis2021
120. Siemon D, Ahmad R, Harms H, de Vreede T. Requirements and solution approaches to personality-adaptive conversational agents in mental health care. *Sustainability*. 2022;14(7):3832. [FREE Full text]

121. Kang Y, Tan AH, Miao C. An adaptive computational model for personalized persuasion. In: Proceedings of the 24th International Joint Conference on Artificial Intelligence. 2015. Presented at: IJCAI '15; July 25–31, 2015:61-67; Buenos Aires, Argentina. URL: <https://www.ijcai.org/Proceedings/15/Papers/016.pdf>
122. Faddoul G. The virtual diabetician: a virtual platform providing diabetes treatment information through storytelling and virtual companions. Center of Information Systems and Technology Claremont Graduate University. 2018. URL: <https://www.proquest.com/openview/6b73f74fdd97109ee35723b289ea2b96/1?pq-origsite=gscholar&cbl=18750> [accessed 2024-04-29]
123. Tironi A, Mainetti R, Pezzerà M, Borghese NA. An empathic virtual caregiver for assistance in exer-game-based rehabilitation therapy. In: Proceedings of the IEEE 7th International Conference on Serious Games and Applications for Health. 2019. Presented at: SeGAH '19; August 5-7, 2019:1-6; Kyoto, Japan. URL: <https://ieeexplore.ieee.org/document/8882477>
124. Casas J, Spring T, Daher K, Mugellini E, Khaled OA, Cudre-Mauroux P. Enhancing conversational agents with empathic abilities. In: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents. 2021. Presented at: IVA '21; September 14-17, 2021:41-47; Virtual Event. URL: <https://dl.acm.org/doi/10.1145/3472306.3478344> [doi: [10.1145/3472306.3478344](https://doi.org/10.1145/3472306.3478344)]
125. May R, Denecke K. Security, privacy, and healthcare-related conversational agents: a scoping review. *Inform Health Soc Care.* Apr 03, 2022;47(2):194-210. [doi: [10.1080/17538157.2021.1983578](https://doi.org/10.1080/17538157.2021.1983578)] [Medline: [34617857](https://pubmed.ncbi.nlm.nih.gov/34617857/)]
126. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc.* Sep 01, 2018;25(9):1248-1258. [FREE Full text] [doi: [10.1093/jamia/ocy072](https://doi.org/10.1093/jamia/ocy072)] [Medline: [30010941](https://pubmed.ncbi.nlm.nih.gov/30010941/)]
127. Ranieri A, Ruggiero A. Complementary role of conversational agents in e-health services. In: Proceedings of the 2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering. 2022. Presented at: MetroXRaine '22; October 26-28, 2022:528-533; Rome, Italy. URL: <https://ieeexplore.ieee.org/document/9967603> [doi: [10.7759/cureus.40922](https://doi.org/10.7759/cureus.40922)]
128. Altamimi I, Altamimi A, Alhumimidi AS, Altamimi A, Tamsah MH. Artificial intelligence (AI) chatbots in medicine: a supplement, not a substitute. *Cureus.* Jun 2023;15(6):e40922. [FREE Full text] [doi: [10.7759/cureus.40922](https://doi.org/10.7759/cureus.40922)] [Medline: [37496532](https://pubmed.ncbi.nlm.nih.gov/37496532/)]
129. Sen A, Chuen CL, Hta AC. Toward smart learning environments: affordances and design architecture of augmented reality (AR) applications in medical education. In: Proceedings of 1st International Conference on Smart System, Innovations and Computing. 2019. Presented at: SSIC '17; April 14-16, 2017:843-861; Jaipur, India. URL: https://link.springer.com/chapter/10.1007/978-981-10-5828-8_80 [doi: [10.1177/2055207619896153](https://doi.org/10.1177/2055207619896153)]
130. Sen A, Leong C. Technology-enhanced learning. In: Tatnall A, editor. *Encyclopedia of Education and Information Technologies.* Cham, Switzerland. Springer; 2019:1-8.
131. Hossain MM, Tasnim S, Sharma R, Sultana A, Shaik AF, Faizah F, et al. Digital interventions for people living with non-communicable diseases in India: a systematic review of intervention studies and recommendations for future research and development. *Digit Health.* 2019;5:2055207619896153. [FREE Full text] [doi: [10.1177/2055207619896153](https://doi.org/10.1177/2055207619896153)] [Medline: [31897307](https://pubmed.ncbi.nlm.nih.gov/31897307/)]
132. Maher CA, Davis CR, Curtis RG, Short CE, Murphy KJ. A physical activity and diet program delivered by artificially intelligent virtual health coach: proof-of-concept study. *JMIR Mhealth Uhealth.* Jul 10, 2020;8(7):e17558. [FREE Full text] [doi: [10.2196/17558](https://doi.org/10.2196/17558)] [Medline: [32673246](https://pubmed.ncbi.nlm.nih.gov/32673246/)]
133. Jegundo AL, Dantas C, Quintas J, Dutra J, Almeida AL, Caravau H, et al. Perceived usefulness, satisfaction, ease of use and potential of a virtual companion to support the care provision for older adults. *Technologies.* Jul 25, 2020;8(3):42. [FREE Full text] [doi: [10.3390/technologies8030042](https://doi.org/10.3390/technologies8030042)]
134. Nagarajan A, Sen A. Can Bloom's higher order thinking skills be achieved by gamified learning through social networking sites (SNS) like Facebook? *Interact Des Archit.* 2022;(53):144-160. [FREE Full text]
135. Sen A, Richardson S. Some controversies relating to the causes and preventive management of computer vision syndrome. In: Proceeding of the 2002 conference on CybErg. 2002. Presented at: CybErg '02; September 2-4, 2002:123-130; Johannesburg, South Africa. URL: <https://research.monash.edu/en/publications/some-controversies-relating-to-the-causes-and-preventive-managememe> [doi: [10.1109/icbir54589.2022.9786451](https://doi.org/10.1109/icbir54589.2022.9786451)]
136. Yang S, Zhao H, Zhu S, Zhou G, Xu H, Jia Y, et al. Zhongjing: enhancing the chinese medical capabilities of large language model through expert feedback and real-world multi-turn dialogue. arXiv. Preprint posted online August 7, 2023. [FREE Full text]
137. Wang J, Yao Z, Yang Z, Zhou H, Li R, Wang X, et al. NoteChat: a dataset of synthetic doctor-patient conversations conditioned on clinical notes. arXiv. Preprint posted online October 24, 2023. [FREE Full text]
138. Kanzaki K, Isahara H. Analysis of dialogue between caregivers and recipients for a communication robot. In: Proceedings of the 7th International Conference on Business and Industrial Research. 2022. Presented at: ICBIR '22; May 19-20, 2022:446-450; Bangkok, Thailand. URL: <https://ieeexplore.ieee.org/document/9786451> [doi: [10.1109/icbir54589.2022.9786451](https://doi.org/10.1109/icbir54589.2022.9786451)]

139. Huang CW, Wu BC, Nguyen PA, Wang HH, Kao CC, Lee PC, et al. Emotion recognition in doctor-patient interactions from real-world clinical video database: initial development of artificial empathy. *Comput Methods Programs Biomed.* May 2023;233:107480. [doi: [10.1016/j.cmpb.2023.107480](https://doi.org/10.1016/j.cmpb.2023.107480)] [Medline: [36965299](https://pubmed.ncbi.nlm.nih.gov/36965299/)]
140. Lewis L. Avatars and robots as social companions in healthcare: requirements, engineering, adoption and ethics. *Int J Enterp Inf Syst.* 2014;10(2):21-39. [doi: [10.4018/ijeis.2014040102](https://doi.org/10.4018/ijeis.2014040102)]

Abbreviations

AI: artificial intelligence

CA: conversational agent

CVD: cardiovascular disease

LMIC: low- and middle-income country

ML: machine learning

NCD: noncommunicable disease

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

Edited by T de Azevedo Cardoso; submitted 08.01.24; peer-reviewed by M Kim, T Su, E Hekler, DLT Steven; comments to author 05.02.24; revised version received 06.03.24; accepted 25.03.24; published 16.07.24

Please cite as:

Anisha SA, Sen A, Bain C

Evaluating the Potential and Pitfalls of AI-Powered Conversational Agents as Humanlike Virtual Health Carers in the Remote Management of Noncommunicable Diseases: Scoping Review

J Med Internet Res 2024;26:e56114

URL: <https://www.jmir.org/2024/1/e56114>

doi: [10.2196/56114](https://doi.org/10.2196/56114)

PMID: [39012688](https://pubmed.ncbi.nlm.nih.gov/39012688/)

©Sadia Azmin Anisha, Arkendu Sen, Chris Bain. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 16.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.