A systematic review of artificial intelligence-powered (AI-powered) chatbot intervention for managing chronic illness

Moh Heri Kurniawan, Hanny Handiyani, Tuti Nuraini, Rr Tutik Sri Hariyati & Sutrisno Sutrisno

RESEARCH ARTICLE

A systematic review of artificial intelligence-powered (AI-powered) chatbot intervention for managing chronic illness

Moh Heri Kurniawan[a,b] (iD), Hanny Handiyani[c] (iD), Tuti Nuraini[c] (iD), Rr Tutik Sri Hariyati[c] (iD) and Sutrisno Sutrisno[b] (iD)

[a]Doctoral Student, Faculty of Nursing, Universitas Indonesia, Depok, Indonesia; [b]Departement of Nursing, Faculty of Health, Universitas Aisyah Pringsewu, Kabupaten Pringsewu, Indonesia; [c]Department of Nursing, Faculty of Nursing, Universitas Indonesia, Depok, Indonesia

ABSTRACT

Background: Utilizing artificial intelligence (AI) in chatbots, especially for chronic diseases, has become increasingly prevalent. These AI-powered chatbots serve as crucial tools for enhancing patient communication, addressing the rising prevalence of chronic conditions, and meeting the growing demand for supportive healthcare applications. However, there is a notable gap in comprehensive reviews evaluating the impact of AI-powered chatbot interventions in healthcare within academic literature. This study aimed to assess user satisfaction, intervention efficacy, and the specific characteristics and AI architectures of chatbot systems designed for chronic diseases.
Method: A thorough exploration of the existing literature was undertaken by employing diverse databases such as PubMed MEDLINE, CINAHL, EMBASE, PsycINFO, ACM Digital Library and Scopus. The studies incorporated in this analysis encompassed primary research that employed chatbots or other forms of AI architecture in the context of preventing, treating or rehabilitating chronic diseases. The assessment of bias risk was conducted using Risk of 2.0 Tools.
Results: Seven hundred and eighty-four results were obtained, and subsequently, eight studies were found to align with the inclusion criteria. The intervention methods encompassed health education ($n = 3$), behaviour change theory ($n = 1$), stress and coping ($n = 1$), cognitive behavioural therapy ($n = 2$) and self-care behaviour ($n = 1$). The research provided valuable insights into the effectiveness and user-friendliness of AI-powered chatbots in handling various chronic conditions. Overall, users showed favourable acceptance of these chatbots for self-managing chronic illnesses.
Conclusions: The reviewed studies suggest promising acceptance of AI-powered chatbots for self-managing chronic conditions. However, limited evidence on their efficacy due to insufficient technical documentation calls for future studies to provide detailed descriptions and prioritize patient safety. These chatbots employ natural language processing and multimodal interaction. Subsequent research should focus on evidence-based evaluations, facilitating comparisons across diverse chronic health conditions.

Introduction

The incidence of chronic illnesses is experiencing a worldwide upsurge, presenting one of the foremost healthcare challenges in the twenty-first century [1]. These enduring conditions have long-lasting ramifications, extending over an individual's entire lifespan, necessitating ongoing management by both patients and healthcare practitioners [2]. The enduring nature of chronic ailments significantly impacts individuals' health related quality life and generates substantial healthcare costs due to disability, recurrent hospitalizations and treatment measures [3].

Digital health interventions have garnered special attention from the World Health Organization (WHO) as a means to improve public healthcare services and attain universal health coverage. In more recent times, the focal point of digital health intervention has veered from conventional domains such as eHealth or mobile health towards the frontiers of advanced computational sciences, notably encompassing big data analytics, genomics and artificial intelligence (AI) [4]. The utilization of AI in the realm of healthcare has garnered extensive traction, encompassing pivotal facets such as early-stage disease detection, interpretation of disease progression, optimization of treatment

regimens, and the advent of pioneering intervention strategies [5].

Digital intervention in healthcare using AI-powered chatbots has become increasingly pervasive as a result of improvements in AI, natural language processing (NLP) and voice recognition. These sophisticated computer programs are meticulously crafted to emulate and effectively process human dialogues, whether they are in written or spoken form. Consequently, users are afforded the unique opportunity to interact with digital devices in a manner that strikingly resembles engaging in a conversation with an authentic human interlocutor [2,6,7]. The most recent developments in AI enable interactions that are more and more like those between people and their computer agent counterparts [8,9]. The simulation of communication between humans and machines has become increasingly intricate and sophisticated [10]. The chatbot industry has witnessed a substantial growth in sectors such as e-commerce, travel, tourism and healthcare, which are aiming to achieve interactions with a human-like quality [11].

AI-powered chatbots have demonstrated significant benefits across various industries, with the healthcare sector being a prime example. They have been employed to provide cost-effective, scalable medical support solutions that can be accessed anytime through online platforms or smartphone applications [12,13]. For instance, chatbots providing assistance and monitoring to adults receiving cancer treatment led to reduced levels of anxiety, eliminating the need for the involvement of a healthcare professional [14]. Chatbots can play a crucial role in enhancing consultations by assisting patients and clinicians, supporting individuals in modifying their behaviour, and aiding senior citizens in their homes [15]. In addition, chatbots possess the capacity to play a crucial role in the achievement of particular objectives, such as self-monitoring and surmounting barriers to self-management. These functions hold significant importance in the realm of chronic illness management [2,16].

Numerous research has reported various advantages associated with the implementation of AI-powered chatbots in diverse healthcare contexts. These advantages encompass but are not limited to promoting behaviour modification, delivering guidance for adopting a healthy lifestyle, offering assistance to individuals diagnosed with breast cancer, and enabling self-reporting of medical histories among therapy patients [17,18].

Prior investigations primarily emphasized the technical facets of AI-powered chatbot development, often involving a limited participant pool or adopting a pilot study framework [19]. Furthermore, three reviews on AI-powered chatbots interventions predicated on conversational modalities within healthcare settings encompassed studies characterized by diverse methodologies, case studies, uncontrolled clinical trials and single-group designs [19–21]. Some reviews adopted a singular and focused intervention approach, such as exclusively examining social robots within psychosocial interventions [22], or concentrated solely on specific outcomes, particularly mental well-being [23].

There is a conspicuous absence of systematic reviews that have systematically assessed the evidence derived from randomized controlled trials (RCTs) pertaining to interventions employing AI-powered chatbot within the healthcare domain. The matter of information quality concerning AI-powered chatbot in the context of healthcare interventions remains a facet insufficiently addressed by prior systematic reviews. Consequently, this systematic review is poised to specifically scrutinize RCT concerning the implementation of AI-powered chatbot interventions within the healthcare setting. Notably, this review will encompass a rigorous evaluation and discourse on the information quality associated with the examined AI-powered chatbot. It will undertake a comprehensive and thorough examination of the existing body of evidence related to AI-powered chatbot healthcare interventions to date. The review will delve into aspects encompassing the user satisfaction, effectiveness and patient safety of these interventions. Furthermore, it will culminate in the formulation of recommendations for prospective research endeavours and the pragmatic utilization of AI-powered chatbot interventions in healthcare settings.

Methods

Reporting standards

As researchers, we affirm our adherence to the PRISMA guidelines for conducting systematic reviews [24]. We have ensured compliance with these guidelines and have included a completed flowchart/table in accordance with the PRISMA recommendations. The protocol was submitted for registration to Prospero registration (CRD42023405505).

Search strategy

In February 2023, an extensive search was performed on English-language articles published from 2013 to 2023, using multiple reputable databases including PubMed MEDLINE, CINAHL, EMBASE, PsycINFO, ACM

Digital Library and Scopus. Various search terms were employed and applied consistently across all databases. These terms included: (chatbot or 'conversational agent' or 'social bot*' or 'softbot*' or 'virtual agent' or 'automated agent' or 'automated bot' or 'virtual therap*') AND (randomi*) AND (clinical or stud* or trial) AND (health or nurs* or disease* or illness*). The detailed search strategy for PubMed can be found in Appendix 1. Similar combinations of search terms were utilized for the other databases.

Study selection criteria

Incorporated within this study were original research investigations meeting the following criteria: (1) incorporation of elements pertaining to AI-powered chatbots within their intervention framework, regardless of therapist involvement; (2) intention to assess the impact of the intervention on outcomes relevant to healthcare; (3) classification as primary studies employing a RCT, encompassing both pilot and feasibility studies; (4) focal point on a chronic illness condition, with participants being a minimum of 18 years of age. Conversely, studies were excluded if they: (1) were solely available in abstract form without access to the full text; (2) constituted study protocols or trial registrations lacking the dissemination of research outcomes; (3) were studies not published in the English language.

Screening, data extraction and data synthesis

After the searches were concluded, all identified citations were obtained. Then, reference management software like Endnote© and Mendeley©, used under proper copyright authorization, was utilized to remove any duplicate entries. Subsequently, the titles and abstracts of each article were extracted from the reference manager and imported into an Excel spreadsheet for additional analysis.

Prior to commencing the screening process, preliminary screenings were conducted. Subsequently, an investigating filter based on the information available in the article titles and abstracts was applied as the initial filtering criterion. This screening procedure was independently carried out by two evaluators. Additionally, the full-text screening was performed by two impartial evaluators. In order to address any disagreements regarding the exclusion of specific articles, two independent reviewers engaged in a Zoom meeting for deliberation. Subsequently, four reviewers meticulously extracted pertinent information from

each study, including the first author, publication year, study site, chronic condition, objectives, study design and methodologies, participant characteristics, evaluation metrics and key findings. Evaluation metrics were drawn from three categories: user satisfaction, health-related measures and patient safety.

The evaluation of user satisfaction in relation to chatbots entailed an examination of the system's properties or components from the perspectives of users. Both quantitative and qualitative techniques were employed to assess user satisfaction. In addition to examining health outcomes such as diagnostic accuracy or symptom alleviation, considerations were given to health-related metrics and patient safety within the included studies. The assessed characteristics of the chatbots identified in the selected studies remain pertinent. A meta-analysis was precluded by the diversity observed in the types of interventions, outcomes and settings across the studies under review.

Study risk of bias assessment

Two expert reviewers (HH and TN) individually assessed the quality of eligible studies. The RCTs underwent evaluation using the Cochrane Risk of Bias 2.0 Tool [25]. This tool scrutinized various aspects such as random allocation concealment, sequence generation, blinding of participants and personnel, incomplete outcome data, blinding of outcome assessment, selective reporting and other potential biases. Reviewers were instructed to render a judgment of 'yes' (indicating low bias), 'no' (indicating high bias) or 'unclear' (suggesting either a lack of pertinent information or uncertainty regarding bias) for each criterion. Any disagreements arising from the quality assessment conducted by the two reviewers (HH and TN) were resolved by a third reviewer (RT).

Results

In this study, an extensive search was carried out in six databases, resulting in a total of 784 articles. Following the removal of duplicates, 247 unique articles were retained. These articles underwent initial assessment based on their titles and abstracts, leading to the identification of 97 relevant studies. After a thorough examination of the full texts, 97 articles were excluded. Furthermore, 89 studies were eliminated based on the predefined criteria, ultimately resulting in eight articles qualified for included studies. A visual depiction of this selection process is presented in Figure 1.

Description of included studies

According to the data presented in Table 1, a comprehensive total of 10 studies were included in this review. Out of the eight studies, five focused primarily on patients, providing support for education and self-care [27–31]. Three studies offered assistance to both patients and healthcare professionals in utilizing AI-powered chatbots for treatment and education [14,26,32].

Among the chronic illness investigated, cancer, including breast cancer, post-cancer treatment and geriatric oncology, was the subject of investigation in four studies [14,26,27,32]. One studying type 2 diabetes [29]. One studying irritable bowel syndrome [31]. Other conditions studied chronic pain [30], and hypertension [28].

Regarding methods, all studies used RCTs [14,26–32].

Description of AI-powered chatbots interventions

Table 2 presents an overview of the technologies utilized to support chatbots, including independent platforms, web or mobile applications. Out of the eight AI-powered chatbots considered in this review, all studies were classified as chatbots, which are software systems designed to emulate human conversation through voice or text-based interactions. Three studies were not specified in AI methods [14,28,30]. Table 3 presents the characterization of AI-powered chatbots as described in the reviewed papers. The chatbots discussed in these papers employed a range of AI techniques, such as speech recognition and NLP, to facilitate their functionality.

The encompassed studies featured various theoretical approaches, such as health education ($n = 3$), behaviour change theory ($n = 1$), stress and coping ($n = 1$), cognitive behaviour therapy ($n = 2$) and self-care behaviour ($n = 1$). The duration of interventions utilizing AI-powered chatbots varied, ranging from short-term (four weeks) to long-term (one or two years), contingent on the intended purpose and targeted health concerns. Some studies ($n = 3$) allowed for flexible usage with unrestricted access to the conversational agent throughout the intervention phase,
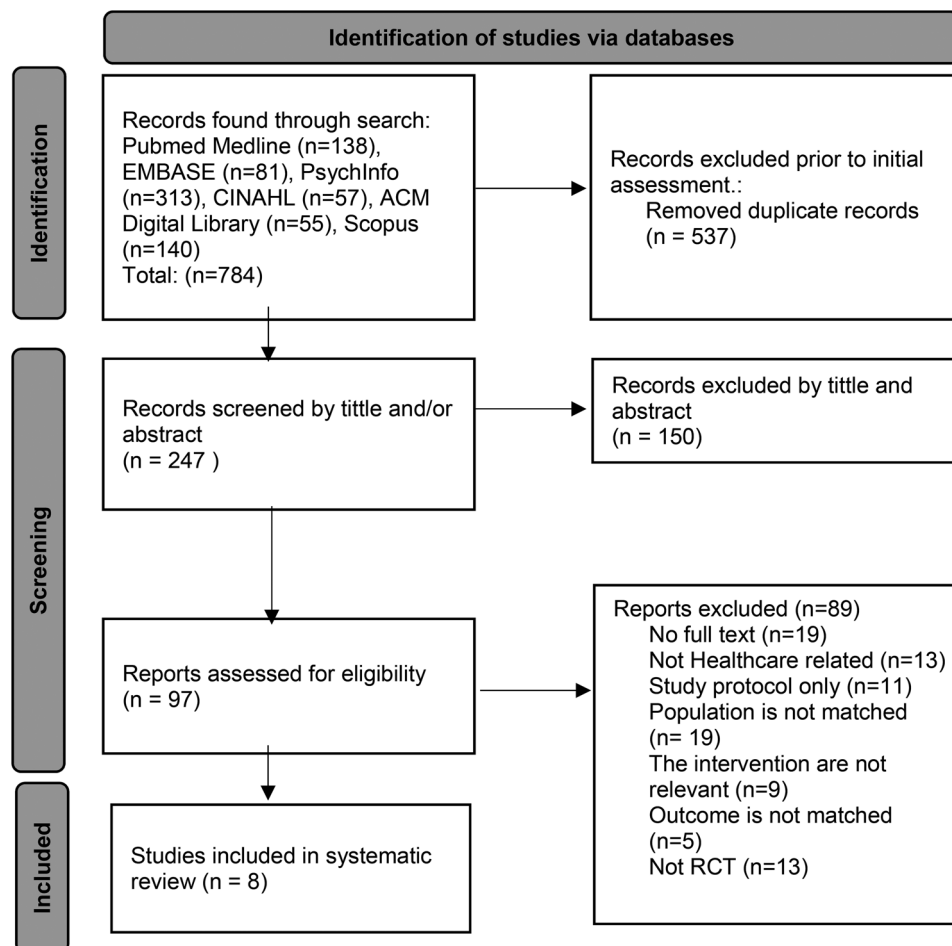

Figure 1. The study selection flowchart.

Table 1. The characteristics of included studies.

Study	Country setting	Chronic illness	Aim of studies	Methods	Study population	User satisfaction	Outcomes — Health related measures	Outcomes — Patient safety
Al-Hilli et al. [26]	USA	Breast cancer	Education	RCT	IG = 19; CG = 18; women with stage 0–III breast cancer	No significant difference in median satisfaction score between Chatbot and traditional counseling (30 vs. 30, $p = .19$)	Not reported	Not reported
Bibault et al. [27]	France	Breast cancer	Self-management	RCT	The study included female subjects aged 18 years or older who were either undergoing treatment for breast cancer or in remission. These participants demonstrated nonopposition and possessed internet literacy.	The findings indicate that among the patients, 59% expressed a desire for more information compared to 65% in the control group. In the study, 85% of the patients reported finding the information provided by the chatbot to be valuable, and the same percentage expressed satisfaction with the quantity of information received. This stands in contrast to the control group, where 83.1% and 77% reported similar levels of usefulness and satisfaction, respectively. Moreover, when comparing the ratings of chatbot responses to those provided by physicians in real-time, patients rated the chatbot responses more favourably. Overall, the study demonstrated noninferiority in the perceived quality of responses between the chatbot and physician-provided responses.	Not reported	The research exhibits a deficiency in transparent reporting concerning various facets of adverse events. These include the absence of specific data on the quantity of adverse events communicated through the electronic clinical assistant (ECA) and directly to the clinic by the patient, the duration taken to identify and rectify adverse events, as well as the rate of false alarms associated with adverse events reported through the ECA and directly to the clinic by the patient.
Echeazarra et al. [28]	Spain	Hypertension	Self-management	RCT	Individuals experiencing elevated blood pressure Total = 112; intervention group (IG) = 55; control group (CG) = 57	The participants gave the bot a highly favourable rating, found it to be valuable and user-friendly, and maintained its use even after the intervention period.	The bot group demonstrated comparable adherence to the blood pressure check schedule (bot group mean = 12.0, control group mean = 13.4, $p = .109$). Additionally, the intervention group (IG) significantly enhanced knowledge of proper blood pressure self-monitoring practices compared to the control group (CG) (IG mean = 24.126, CG mean = 17.591, $p = .03737$).	Not reported

(Continued)

Table 1. Continued.

Study	Country setting	Chronic illness	Aim of studies	Methods	Study population	Outcomes		
						User satisfaction	Health related measures	Patient safety
Gong et al. [29]	Australia	Type 2 diabetes mellitus	Self-management	RCT	Adult with type 2 diabetes mellitus, $N = 187$; IG = 93; CG = 94	92 samples (98.9%) engaged in at least one chat with chatbot, and 83 participants (89.2%) uploaded their blood glucose levels to the app at least once. These 92 participants had a combined total of 1942 chats over the 12-month program, averaging 243 min per person (SD = 212).	At 12 months, there was a notable difference between the two groups: quality of life (difference: 0.04, 95% CI [0.00, 0.07]; $p = .04$), not in HbA1c levels. In comparison to baseline, both groups experienced a reduction in mean estimated HbA1c levels at 12 months (IG: mean estimated change: -0.33%; $p = .03$ and control arm: -0.28%; $p = .05$).	Not reported
Greer et al. [14]	USA	Post cancer treatment	After cancer treatment	Pilot RCT	The research encompassed a cohort of 45 young adults, segregated into two distinct groups. Group 1, constituting 25 young adults, was designated as the experimental cohort, while group 2, comprising 20 young adults, functioned as the control cohort. At the outset, all participants underwent a comprehensive survey assessment.	Their assessments indicated a positive reception, with an average rating of 2.0 out of 3 (SD 0.72) for perceived helpfulness, and an expressed likelihood to recommend it to others, with a mean score of 6.9 out of 10 (SD 2.6).	The intervention group showed an average anxiety reduction of 2.58 standardized t-score units. Results showed a nearly significant group–time interaction ($p = .09$), resulting in an effect size of 0.41. Additionally, the IG experienced greater anxiety reduction with more frequent engagement sessions ($p = .06$). However, there were no significant differences between the groups in terms of changes in depression, positive or negative emotion.	Not reported
Hauser-Ulrich et al. [30]	Germany and Switzerland	Chronic pain	Self-management of chronic pain	Pilot RCT	The study involved 102 participants recruited from the SELMA app. Their average age was 43.7 years. Of these, 14 were male and 88 were female. The duration of their participation in the study spanned two months.	Acceptance: 63% of users found the app fun, 47% found it useful, and 84% found it easy to use among 38 respondents. Adherence: the regular adherence percentage was 71% for 200 conversations initiated on Instagram.	There was a significant positive intention to alter behaviour concerning impairment and pain intensity ($p = .01$). There was no statistically significant alteration in pain-related impairment ($p = .01$) observed in the IG compared CG.	Not reported

(Continued)

Table 1. Continued.

Study	Country setting	Chronic illness	Aim of studies	Methods	Study population	Outcomes		
						User satisfaction	Health related measures	Patient safety
Hunt et al. [31]	USA	Irritable bowel syndrome (IBS)	Education and self-management	RCT	121 adult with IBS, IG = 62, CG = 59	Completion rate 94%	Immediate treatment group improved significantly compared to wait-list CG in: Gastrointestinal symptom severity (Cohen's $d = 1.02$, $p < .001$) QoL (Cohen's $d = 1.25$, $p < .001$) GI-specific catastrophizing (Cohen's $d = 1.47$, $p < .001$) Fear of food (Cohen's $d = 0.62$, $p = .001$) Visceral anxiety (Cohen's $d = 1.07$, $p < .001$) Depression (Cohen's $d = 1.07$, $p = .002$) However, anxiety did not show a significant result for the IG.	Not reported
Tawfik et al. [32]	Egypt	Breast cancer	Education, self-management	RCT	150 participants (50 per group), aged 20+, recently diagnosed with early-stage non-metastatic breast cancer (stage 0–III), scheduled for first or second chemotherapy, able to read/write, and own a smartphone.	Easy to use: 94% Useful and informative responses: 94% Understanding: 72% Welcoming setup: 88% Women find it navigable: 70% Realistic and engaging personality: 72% Some reported mistakes: 76%	Participants showed significantly fewer, milder and less distressing symptoms, along with better self-care ($p > .001$), compared to nurse-led education and routine care groups.	Not reported

Table 2. Details of the AI-powered chatbots intervention in the included studies.

Study	Theory or core element	AI Methods	Duration	Type	Input	Output	Dialogue management	Dialogue Initiative	Task-Oriented	Therapist involved	Comparison
Al-Hilli et al. [26]	Health education and counseling	NLP	Flexible	Chatbot	Written	Written	Frame based	System	No	Yes	Traditional counselling
Bibault et al. [27]	Health education	Machine learning	Flexible	Web based or smartphone application	Written	Written	Frame based	System	Yes	No	Physician
Echeazarra et al. [28]	Education	Not reported	Daily, for 2 years	Smartphone app, chatbot	Written	Written	Finite state	System	No	Yes	Traditional writing
Gong et al. [29]	Behaviour change theories	NLP	Weekly up to one year	Smartphone app, chatbot	Spoken, and written	Spoken, and written	Frame based	System	Yes	Yes	Usual care
Greer et al. [14]	Stress and coping theory	Not reported	4 weeks exposure	Facebook messenger; chatbot.	Written	Written, visual	Finite state	System	Yes	No	Waitlist control
Hauser-Ulrich et al. [30]	Cognitive behaviour therapy, psychological therapy	Not reported	Every day, for 8 weeks	Smartphone app; chatbot	Written	Written, visual	Finite state	System	No	No	Education
Hunt et al. [31]	Cognitive behaviour therapy	NLU	8 weeks	Smartphone app; chatbot	Written	Written	Frame based	System	No	No	Waiting list control
Tawfik et al. [32]	Self-care behaviour	NLP	Flexible	WhatsApp; chatbot	Written	Written	Frame based	System	No	Yes	Nurse-led education and routine care

while others ($n = 5$) restricted access frequencies (one or two times per day) or imposed a maximum time limit. Half of the studies, therapist involvement complemented the AI-powered chatbot system, whereas in others ($n = 4$), the intervention was solely facilitated by the AI-powered chatbot. Comparison groups primarily received usual care, treatment as usual, or self-guided interventions administered by healthcare professionals.

Risk of bias included studies

Among eight RCTs, only one [32] exhibited a low risk, two [29,30] exhibited a high-risk of missing outcome data, a significant portion of the data was not available, while five [14,26–28,31] raised some concerns regarding potential bias in outcome measurement. This was attributed to either the omission of blinding in the assessment process or the utilization of an unsuitable method to evaluate the intervention's effects. Detailed results are presented in Figure 2.

Evaluation measures

Evaluation measures were categorized into three groups: user satisfaction, health-related measures and patient safety. Comprehensive details regarding the evaluation of these interventions are presented in Table 1. Regarding technical performance, two studies consistently reported positive performance measures for the conversational agents, including accuracy, precision, sensitivity, specificity and *F*-measure, demonstrated high rates of message response. All studies employing descriptive methodologies, reported a moderate to high level of participant satisfaction. Participants acknowledged several perceived benefits of AI-powered chatbots, including attributes such as being 'useful', 'communicative', 'responsive', 'inquisitive', 'valuable', 'user-friendly', 'personalized', 'adoptive', 'helpful', 'recommendable', 'accessible', 'efficient', 'gratifying' and 'beneficial'. Nevertheless, a subset of studies identified challenges associated with utilizing conversational agents. These challenges encompassed difficulties in comprehending user input, issues of repetitiveness, technical glitches, perceived lack of warmth, disruptions in natural flow, time commitment required for the intervention, and considerations regarding the quality of the AI-powered chatbots. The evaluation of patient safety received limited attention in the studies included in our analysis. Only one study explicitly addressed safety concerns, shedding light on

Table 3. Characterization of AI-powered chatbots [19].

Dialogue management	Finite-state	Users are guided through a predetermined series of steps or states in the dialogue.
	Frame-state	Users are presented with inquiries to fill in placeholders within a template, allowing for task-specific execution based on their input.
	Agent-based	Agent-based systems facilitate advanced communication between multiple agents, enabling them to reason about their own actions and beliefs while considering others'. The dialogue model in these systems adapts dynamically, incorporating contextual information at each interaction step.
Dialogue initiative	User	Users assume an active role in guiding and directing the ongoing conversation.
	System	The conversation is steered and directed by the system, with the system taking the lead in guiding the interaction.
	Mixed both	Both the user and the system have the capacity to initiate and guide the conversation, assuming the role of conversation leaders interchangeably.
Input modality	Spoken	Users interact with the system using spoken language.
	Written	Users engage with the system through written language for interaction purposes.
Output modality	Spoken, written and visual modes of communication encompass various channels of expression, including verbal interaction, written text and visual cues such as facial expressions and body movements.	
Task-oriented	Yes	The system is designed for a specific task, engaging in concise conversations to acquire essential information and accomplish objectives, like scheduling consultations.
	No	The system's objective is not solely focused on quickly achieving a predetermined outcome like casual chatbots do.

Figure 2. Risk of bias summary.

potential issues such as the false alarm rate of adverse events reported through the chatbot and directly to the clinic by the patient.

The delivery of healthcare counselling sessions through AI-powered chatbots to notable significant improvements in disease-specific knowledge and comprehension of pertinent information [27,28]. Two study demonstrated noteworthy advancements in alleviating depressive symptoms [14,31], while an additional two studies underscored significant effects of interventions driven by AI-powered chatbot in mitigating symptoms of anxiety [14,29]. Participants reported substantial reductions in negative cognitions and emotions following interactions with interventions guided by AI-powered chatbots [14,31].

The bot group displayed comparable adherence to blood pressure checks compared to the control group. In terms of knowledge improvement regarding blood pressure self-monitoring procedures, the intervention group showed significant advancement over the control group [28]. At 12 months, a notable difference in mean change in quality of life was observed, while no notable disparity observed in HbA1c levels [29]. The experimental group exhibited a substantial reduction in anxiety levels, particularly with more frequent engagement sessions. However, no significant differences were noted between groups in terms of changes in positive or negative and depression [14]. Additionally, positive intentions for behaviour change were observed in relation to impairment and pain intensity. However,

there was no statistically significant alteration in pain-related impairment observed in the intervention group compared to the control group [30]. Additionally, participants in the intervention group demonstrated improved treatment adherence, self-care management, heightened awareness of symptoms and triggers, fewer, milder and less distressing symptoms, and overall better self-care compared to the nurse-led education and routine care groups [32]. Some studies were not reported health related measures [26,27].

One study noted notable enhancements in quality of life [29]. Additionally, one study suggested that interventions employing AI-powered chatbots could furnish disease-related information with a satisfaction level comparable to responses from clinicians [27].

Discussion

This study represents the inaugural systematic review dedicated to evaluating the user satisfaction of interventions employing AI-powered chatbots within the healthcare sector. The comprehensive findings affirm the feasibility, acceptability and effectiveness of interventions incorporating AI-powered chatbots in improving various healthcare outcomes, including health-related metrics and patient safety. These results have significant implications for the potential integration of AI-powered chatbots in future healthcare interventions. However, it is essential to approach these findings with caution due to certain trials having modest sample sizes, and a notable proportion of included studies displaying a heightened risk of bias. Furthermore, this review addresses a notable gap in prior research by evaluating the quality of information, augmenting the existing literature on the subject.

The results of this review highlight that interventions utilizing AI-powered chatbots offer notable advantages in terms of accessibility and engagement, potentially contributing to the observed low attrition rates and high adherence to intervention protocols. However, qualitative insights from the review also shed light on challenges faced in the adoption of AI-powered chatbot interventions. Future research should focus on enhancing these interventions by proactively addressing issues such as technical glitches and refining artificial methods through more comprehensive and insightful training of the AI systems.

The interventions incorporating AI-powered chatbots, as examined in the reviewed studies, demonstrated significant impacts across various health and nursing dimensions. These encompassed improvements in physical function, adoption of healthier lifestyle practices, enhancement of mental health, promotion of psychosocial well-being, and improvements in pain management and related parameters. Importantly, some of the reviewed studies suggested that AI-powered chatbot interventions demonstrated efficacy comparable to interventions led by healthcare professionals or physicians, consistent with findings from previous reviews [19]. These positive outcomes underscore the potential of AI-powered chatbots in addressing the growing healthcare demands, particularly in providing support during and after hospital visits or patient appointments.

A recent systematic review assessed the effects of automated telephone communication systems, which lack natural language understanding, on preventive healthcare and chronic condition management. This review by Posadzki et al. [33] indicated that these systems have the potential to enhance specific health behaviours and improve health outcomes. Notably, advancements in dialogue management and NLP techniques have surpassed the rule-based approaches prevalent in the studies included in our investigation [14,28,30]. While rule-based approaches in finite-state dialogue management systems have a straightforward construction process and are suitable for well-structured tasks, they restrict user input to predefined words and phrases, limiting the user's ability to initiate dialogues and correct misrecognized elements [34]. In contrast, frame-based systems provide enhanced capabilities for system and mixed-initiative interactions, allowing for a more flexible dialogue structure [34].

Both finite-state and frame-based methodologies can manage tasks by requesting user input through form completion, but frame-based systems excel in accommodating user responses in a non-linear fashion, enabling users to provide additional information beyond what is strictly required by the system. In this context, the AI-powered chatbot effectively maintains and organizes essential information, adapting its inquiries accordingly. Through our comprehensive review, we identified five AI-powered chatbots (consisting of three task-oriented and two non-task-oriented) that utilized the frame-based approach to manage dialogues. These agents primarily focused on facilitating self-management tasks and gathering data [26,27,29, 31,32].

Agent-based systems, in comparison to finite-state and frame-based systems, exhibit notable capacity for effectively managing complex dialogues, empowering users to initiate and guide the conversation. Dialogue management techniques in agent-based systems often employ statistical models trained on authentic human–computer dialogues, offering advantages such as enhanced speech recognition, improved performance,

greater scalability and increased adaptability [34,35]. Recent advancements in machine learning, along with a resurgence of interest in neural networks, have significantly advanced the development of advanced and efficient conversational agents [35,36]. Interestingly, the utilization of agent based dialogue management techniques seems to be limited in healthcare application programs. Within the scope of our review, no studies were found that examined the implementation of such AI-powered chatbots specifically within the healthcare domain.

The reviewed studies demonstrate that interventions utilizing AI-powered chatbot have yielded significant effects on a range of health and nursing outcomes. These encompass enhancements in physical function, improvements in quality of life, psychosocial well-being, as well as outcomes related to pain. Moreover, it was observed that in certain studies, AI-powered chatbot interventions yielded results on par with those achieved through physical interventions or those led by healthcare worker. This corroborates findings from previous reviews which consistently reported positive impacts of AI-powered chatbot interventions on healthcare outcomes [19]. The positive outcomes of AI-powered chatbot interventions underscore their potential in meeting the expanding healthcare demands, especially in the context of health intervention, including the need for assistance during and after hospital visits or hospital appointments [37]. These forms of support encompass consultations regarding disease information, management of medications, post-hospital discharge self-care at home, and provision of emotional support [15]. Furthermore, AI-powered chatbot interventions offer clients the prospect of accessing healthcare services in geographically isolated or remote areas, obtaining support earlier in the progression of a disease, and curbing the expenses associated with physician visits [38]. Additionally, health professionals stand to gain from these interventions as they enable more efficient access to patient data, ultimately leading to time savings for physicians. This, in turn, allows them to allocate more attention to critical cases demanding urgent treatment.

The utilization of AI-powered chatbots in automating healthcare activities and promoting consumer self-care is projected to grow as they become more capable and dependable [21]. However, these gains necessitate thorough and ongoing examination. The impact of automation on human activities has serious safety consequences, with the hazards varying depending on the extent of automation and the individual automated functions used [39]. Therefore, it is imperative to exercise careful monitoring of the utilization of AI-powered chatbots with unrestricted input capabilities of natural language, as well as other AI applications, in the healthcare domain [40]. By closely monitoring these advancements, potential safety concerns can be addressed and mitigated effectively.

Remarkably, existing research on AI-powered chatbots lacks a comprehensive social-systems analysis, a gap that has also been identified in previous literature examining AI applications [41]. Currently, there is a lack of agreement on the approaches utilized to evaluate the lasting impacts of this technology on human populations. It is essential to acknowledge the potential for biased design in these applications, as they can reinforce stereotypes or have disproportionate effects on already marginalized groups, influenced by factors such as gender, race or socioeconomic status. Therefore, it is vital to consistently incorporate a consideration of the social implications of AI-powered chatbot at all stages, from their inception to their practical implementation. Neglecting this aspect may lead to adverse outcomes for the health and well-being of certain populations.

Limitations

However, it is crucial to acknowledge some limitations in the reviewed studies. First, while each study adhered to a RCT design, certain studies exhibited a notable degree of bias risk, potentially influencing the internal validity of their findings. Furthermore, the diversity among the AI-powered chatbot interventions examined in this review prevented the possibility of conducting a meta-analysis. Moreover, the included studies lacked comprehensive technical performance details, which impedes the replicability and comparability of the findings.

Implication of this study for nursing practice

The integration of AI-powered chatbots holds tremendous potential in augmenting patient care and outcomes across a spectrum of healthcare domains. These interventions have demonstrated notable feasibility, acceptability and effectiveness in improving diverse healthcare metrics, ranging from patient safety to mental health. The accessibility and engagement offered by AI-powered chatbots can significantly enhance patient adherence and participation in intervention protocols. Moreover, the study suggests that these interventions can yield outcomes on par with those led by healthcare professionals. This indicates a

transformative shift in the way nursing care can be delivered, potentially allowing nurses to allocate their expertise and time towards more critical cases, while AI-powered chatbots handle routine consultations and information dissemination. However, it is crucial for nurses to approach these interventions with a discerning eye, recognizing the potential limitations highlighted in the study, such as the need for ongoing monitoring and mitigation of safety concerns. In navigating this evolving landscape, nurses have a pivotal role in championing the responsible integration of AI-powered chatbots, ensuring they complement and enhance the quality of care provided to patients.

Conclusions

The evaluated research offered valuable insights into the effectiveness and usability of AI-powered chatbots in managing diverse chronic conditions. Generally, users exhibited promising acceptance of AI-powered chatbots for self-management of chronic illnesses, with all of the included studies reporting positive user feedback regarding perceived helpfulness, satisfaction and ease of use. To address this knowledge gap, future research should strive to provide comprehensive and explicit descriptions of the technical aspects of the AI-powered chatbots employed, supported by the development of a clear and comprehensive taxonomy specific to healthcare AI-powered chatbots. Furthermore, the aspect of safety in AI-powered chatbots has been largely overlooked and should be considered as a fundamental consideration in the design process.

Acknowledgements

We would like to extend our sincere gratitude to Yayasan Aisyah Lampung, BPI, BPPT and LPDP for the invaluable support.

Author contributions

Moh Heri Kurniawan led in project administration, conceptualization, methodology and original draft creation, with a strong hand in editing. Hanny Handiyani excelled in data curation, formal analysis and took charge of the initial draft. Tuti Nuraini contributed significantly in data curation, formal analysis, and both drafting and editing. Rr Tutik Sri Hariyati contributed in conceptualization, data curation and formal analysis along with supervision. Sutrisno Sutrisno provided project administration and editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Moh Heri Kurniawan http://orcid.org/0000-0001-6847-6484
Hanny Handiyani http://orcid.org/0000-0002-1746-267X
Tuti Nuraini http://orcid.org/0000-0003-1170-0398
Rr Tutik Sri Hariyati http://orcid.org/0000-0003-4496-4795
Sutrisno Sutrisno http://orcid.org/0009-0000-8700-5864

Data availability statement

The data underpinning the results of this study can be obtained from the corresponding author upon a reasonable request.

References

[1] Hajat C, Stein E. The global burden of multiple chronic conditions: a narrative review. Prev Med Rep. 2018;12:1–15. doi: 10.1016/j.pmedr.2018.10.008.
[2] Schachner T, Keller R, Wangenheim FV. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. J Med Internet Res. 2020;22(9):e20701. doi: 10.2196/20701.
[3] Holman HR. The relation of the chronic disease epidemic to the health care crisis. ACR Open Rheumatol. 2020;2(3):167–173. doi: 10.1002/acr2.11114.
[4] Papagiannidis S, Harris J, Morton D. WHO led the digital transformation of your company? A reflection of IT related challenges during the pandemic. Int J Inf Manage. 2020;55:102166. doi: 10.1016/j.ijinfomgt.2020.102166.
[5] Nadarzynski T, Miles O, Cowie A, et al. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health. 2019;5:2055207619871808. doi: 10.1177/2055207619871808.
[6] Allouch M, Azaria A, Azoulay R. Conversational agents: goals, technologies, vision and challenges. Sensors. 2021;21(24):8448. doi: 10.3390/s21248448.
[7] Kocaballi AB, Quiroz JC, Rezazadegan D, et al. Responses of conversational agents to health and lifestyle prompts: investigation of appropriateness and presentation structures. J Med Internet Res. 2020;22(2):e15823. doi: 10.2196/15823.
[8] Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Appl. 2019;129:56–67. doi: 10.1016/j.eswa.2019.03.054.
[9] Suta P, Lan X, Wu B, et al. An overview of machine learning in chatbots. Int J Mech Eng Robot Res. 2020;9:502–510. doi: 10.18178/ijmerr.9.4.502-510.
[10] Guzman AL, Lewis SC. Artificial intelligence and communication: a human–machine communication research agenda. New Media Soc. 2020;22(1):70–86. doi: 10.1177/1461444819858691.
[11] Ivanov S, Webster C. Economic fundamentals of the use of robots, artificial intelligence, and service automation in travel, tourism, and hospitality. Ivanov, S. and Webster, C. (Ed.). In: Robots, artificial intelligence, and service automation in travel, tourism and hospitality. Emerald Publishing Limited; 2019. p. 39–55. doi: https://doi.org/10.1108/978-1-78756-687-320191017.
[12] Bickmore TW, Kimani E, Trinh H, et al. Managing chronic conditions with a smartphone-based conversational

virtual agent. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents. New York (NY): Association for Computing Machinery; 2018. p. 119–124. doi: 10.1145/3267851.3267908.

[13] Pereira J, Díaz Ó. Using health chatbots for behavior change: a mapping study. J Med Syst. 2019;43(5):135. doi: 10.1007/s10916-019-1237-1.

[14] Greer S, Ramo D, Chang Y-J, et al. Use of the chatbot "vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR Mhealth Uhealth. 2019;7(10):e15018. doi: 10.2196/15018.

[15] Xu L, Sanders L, Li K, et al. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. JMIR Cancer. 2021;7(4):e27850. doi: 10.2196/27850.

[16] Griffin AC, Xing Z, Khairat S, et al. Conversational agents for chronic disease self-management: a systematic review. AMIA Annu Symp Proc. 2020;2020:504–513.

[17] Kang J, Thompson RF, Aneja S, et al. National Cancer Institute Workshop on artificial intelligence in radiation oncology: training the next generation. Pract Radiat Oncol. 2021;11(1):74–83. doi: 10.1016/j.prro.2020.06.001.

[18] McGreevey JD, Hanson CW, Koppel R. Conversational agents in health care—reply. JAMA. 2020;324(23):2444–2445. doi: 10.1001/jama.2020.21518.

[19] Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc. 2018;25(9):1248–1258. doi: 10.1093/jamia/ocy072.

[20] Bendig E, Erb B, Schulze-Thuesing L, et al. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health – a scoping review. Verhaltenstherapie. 2022;32(Suppl. 1):64–76. doi: 10.1159/000501812.

[21] Bin Sawad A, Narayan B, Alnefaie A, et al. A systematic review on healthcare artificial intelligent conversational agents for chronic conditions. Sensors. 2022;22(7):2625. doi: 10.3390/s22072625.

[22] Robinson NL, Cottier TV, Kavanagh DJ. Psychosocial health interventions by social robots: systematic review of randomized controlled trials. J Med Internet Res. 2019;21(5):e13203. doi: 10.2196/13203.

[23] Abd-Alrazaq AA, Rababeh A, Alajlani M, et al. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res. 2020;22(7):e16021. doi: 10.2196/16021.

[24] Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;350(1):g7647. doi: 10.1136/bmj.g7647.

[25] Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's Tool for assessing risk of bias in randomised trials. BMJ. 2011;343(2):d5928. doi: 10.1136/bmj.d5928.

[26] Al-Hilli Z, Noss R, Dickard J, et al. A randomized trial comparing the effectiveness of pre-test genetic counseling using an artificial intelligence automated chatbot and traditional in-person genetic counseling in women newly diagnosed with breast cancer. Ann Surg Oncol. 2023;30(10):5997–5998. doi: 10.1245/s10434-023-13888-4.

[27] Bibault J-E, Chaix B, Guillemassé A, et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. J Med Internet Res. 2019;21(11):e15787. doi: 10.2196/15787.

[28] Echeazarra L, Pereira J, Saracho R. TensioBot: a chatbot assistant for self-managed in-house blood pressure checking. J Med Syst. 2021;45(4):54. doi: 10.1007/s10916-021-01730-x.

[29] Gong E, Baptista S, Russell A, 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. 2020;22(11):e20322. doi: 10.2196/20322.

[30] Hauser-Ulrich S, Künzli H, Meier-Peterhans D, et al. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth. 2020;8(4):e15806. doi: 10.2196/15806.

[31] Hunt M, Miguez S, Dukas B, et al. Efficacy of Zemedy, a mobile digital therapeutic for the self-management of irritable bowel syndrome: crossover randomized controlled trial. JMIR Mhealth Uhealth. 2021;9(5):e26152. doi: 10.2196/26152.

[32] Tawfik E, Ghallab E, Moustafa A. A nurse versus a chatbot – the effect of an empowerment program on chemotherapy-related side effects and the self-care behaviors of women living with breast cancer: a randomized controlled trial. BMC Nurs. 2023;22(1):102. doi: 10.1186/s12912-023-01243-7.

[33] Posadzki P, Mastellos N, Ryan R, et al. Automated telephone communication systems for preventive healthcare and management of long-term conditions. Cochrane Database Syst Rev. 2016;12(12):CD009921. doi: 10.1002/14651858.CD009921.pub2.

[34] López-Cózar R, Callejas Z, Espejo G, et al. Enhancement of conversational agents by means of multimodal interaction. In: Perez-Marin D, Pascual-Nieto I (editors). Conversational agents and natural language interaction. IGI Global; 2011. p. 223–252. doi: 10.4018/978-1-60960-617-6.ch010.

[35] Radziwill NM, Benton MC. Evaluating quality of chatbots and intelligent conversational agents. arXiv, abs/1704.04579; 2017.

[36] Young S, Gasic M, Thomson B, et al. POMDP-based statistical spoken dialog systems: a review. Proc IEEE. 2013;101(5):1160–1179. doi: 10.1109/JPROC.2012.2225812.

[37] Aggarwal A, Tam CC, Wu D, et al. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res. 2023;25:e40789. doi: 10.2196/40789.

[38] Bedi G, Carrillo F, Cecchi GA, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr. 2015;1(1):15030. doi: 10.1038/npjschz.2015.30.

[39] Nishida T, Nakazawa A, Ohmoto Y, et al. Conversational informatics. Japan: Springer; 2014.

[40] Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA. 2017;318(6):517–518. doi: 10.1001/jama.2017.7797.

[41] Crawford K, Calo R. There is a blind spot in AI research. Nature. 2016;538(7625):311–313. doi: 10.1038/538311a.

Appendix 1. Search terms

1. Search strategy for PubMed
   (chatbot or 'conversational agent' or 'social bot*' or 'softbot*' or 'virtual agent' or 'automated agent' or 'automated bot' or 'virtual therap*') AND (randomi*) AND (clinical or stud* or trial) AND (health or nurs* or disease* or illness*)
2. Search strategy for MEDLINE
   "Conversational agent*" OR "conversational system*" OR "dialog system*" OR "dialogue system*" OR "assistance technology" OR "assistance technologies" OR "relational agent*" OR chatbot* AND "Chronic illness" OR "Chronic condition"
3. Search strategy for EMBASE
   Conversational agent*.mp OR conversational system*.mp OR dialog system*.mp OR dialogue system*.mp OR assistance technology.mp OR assistance technologies.mp OR relational agent*.mp OR chatbot*.mp AND Chronic Illness*
4. Search strategy for PsycINFO
   Conversational agent*.mp OR conversational system*.mp OR dialog system*.mp OR dialogue system*.mp OR assistance technology.mp OR assistance technologies.mp OR relational agent*.mp OR chatbot*.mp AND Chronic*
5. Search strategy for ACM Digital Library
   "Conversational agent*"
   "conversational system*"
   "dialog system*"
   "dialogue system*"
   "relational agent*"
   chatbot*
6. Search strategy for Scopus
   "Conversational agent*" OR "conversational system*" OR "dialog system*" OR "dialogue system*" OR "assistance technology" OR "assistance technologies" OR "relational agent*" OR chatbot* AND 'Chronic illness' OR 'chronic condition'