Clinical and Translational Radiation Oncology 16 (2019) 55–59

Contents lists available at ScienceDirect

Clinical and Translational Radiation Oncology

journal homepage: www.elsevier.com/locate/ctro

Review Article

Healthcare ex Machina: Are conversational agents ready for prime time in oncology?

Jean-Emmanuel Bibault [a,*], Benjamin Chaix [b,c], Pierre Nectoux [b], Arthur Pienkowski [b], Arthur Guillemasé [b], Benoît Brouard [b]

[a] *Department of Radiation Oncology, Hôpital Européen Georges Pompidou, AP-HP, Paris, France*
[b] *WeFight Inc., Institut du Cerveau et de la Moelle épinière, Hôpital Pitié-Salpêtrière, Paris, France*
[c] *ENT Department and University Montpellier 1, Hôpital Gui-de-Chauliac, Montpellier, France*

ARTICLE INFO

Article history:
Received 8 February 2019
Revised 1 April 2019
Accepted 3 April 2019
Available online 4 April 2019

Keywords:
Conversational agent
Digital assistant
Chatbot
Oncology
Cancer

ABSTRACT

Chatbots, also known as conversational agents or digital assistants, are artificial intelligence–driven software programs designed to interact with people in a conversational manner. They are often used for user-friendly customer-service triaging. In healthcare, chatbots can create bidirectional information exchange with patients, which could be leveraged for follow-up, screening, treatment adherence or data-collection. They can be deployed over various modalities, such as text-based services (text messaging, mobile applications, chat rooms) on any website or mobile applications, or audio services, such as Siri, Alexa, Cortana or Google Assistant. Potential applications are very promising, particularly in the field of oncology. In this review, we discuss the available publications and applications and the ongoing trials in that setting.

Contents

1. Introduction

In 1950, Alan Turing envisioned a future where a computer would be able to express itself with a level of sophistication that would render it indistinguishable from humans [1]. The famous "Turing Test" was born: a trial-by-fire where a computer and a human are asked various questions while a third party attempts to distinguish between the two. Today, chatbots can imitate human conversation by using a field of artificial intelligence (AI) known as Natural Language Processing. They are widely available

* Corresponding author at: Radiation Oncology Department, Hôpital Européen Georges Pompidou, 20 rue Leblanc, Paris, France.
E-mail address: jean-emmanuel.bibault@aphp.fr (J.-E. Bibault).

https://doi.org/10.1016/j.ctro.2019.04.002
2405-6308/

as text or voice-based assistants on smartphones or computers. They provide information and create a dynamic interaction between the agent and the user, without human back-end intervention. In healthcare, the first example of a computer program used as a conversational agent was Joseph Weizenbaum's ELIZA, a program that mimicked a Rogerian psychotherapist that was able to rephrase the patient's sentences as questions and provide prerecorded answers [2]. In 1991, Dr. Sbaitso was created as an AI speech synthesis program created for MS DOS Personal Computers. In this software, Dr Sbaitso was a psychologist, with very limited possibilities [3]. Four years later, A.L.I.C.E (Artificial Linguistic Internet Computer Entity) included 40.000 knowledge categories that was later awarded the Loebner Prize thrice [4]. In 2001, SmarterChild was made available as a bot distributed across SMS networks and is now considered as a precursor to Apple's Siri, which was released on iPhones in 2010.

Chatbots are now widely used in several forms as voice-based agents, such as Siri (Apple, Palo Alto, California, USA), Google Now (Google, Mountain View, California, USA), Alexa (Amazon, Seattle, Washington, USA) or Cortana (Microsoft, Seattle, Washington, USA). But they are not used in healthcare yet. Text-based chatbots are available as Messenger (Facebook, Menlo Park, California, USA) agents or as stand-alone mobile or web applications. Patients can now use chatbots to check for symptoms and to monitor their health [5]. But their relevance and validity has rarely been assessed [6]. In this review we examine the current landscape of the use of conversational agents, or chatbots, in oncology. We address available literature that demonstrates the power of chatbots and discuss the few examples of their applications in oncology. We also describe the ongoing trials evaluating chatbots in oncology and finally discuss the challenges that remain to be addressed for a successful implementation.

2. Methods

2.1. Systematic search

This review focuses on articles published in peer-reviewed journals and conference proceedings. The review is in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

2.2. Search strategy

We conducted a comprehensive search of PubMed, Embase and Google Scholar for relevant peer-reviewed publications from 1980, to October 2018. A comprehensive list of MeSH terms and keywords was used to query Medline; category 1: "conversational agents", "chatbots", and category 2: "healthcare", "oncology", "cancer". Each term of category 1 was researched separately and as an association with the second category. The search strategy also included screening of reference lists of relevant publications ("snowball" search technique).

ClinicalTrial.gov was also searched for ongoing trials using chatbots in healthcare.

2.3. Eligibility criteria

Publications were selected if they met the following criteria: clinical studies reporting on the use of conversational agents or chatbots in healthcare, randomized studies comparing chatbots to standard interventions, ongoing studies registered on ClinicalTrials.gov. Studies reporting on the use of AI systems for image segmentation and classification, diagnosis and outcome prediction were excluded.

3. Results

3.1. Published studies

The search returned 25 studies that were individually screened. Six studies were included in the final analysis. The PRISMA flowchart is shown in Fig. 1. They are summarized in Table 1.

3.1.1. Cancer screening

The iDecide chatbot is a conversational agent that delivers information about prostate cancer to patients (epidemiology, risk factors, treatment options and their side-effects). A study was designed to evaluate the effects of iDecide on prostate cancer knowledge, informed decision-making self-efficacy, technology use self-efficacy, and intention to engage in informed decision-making among African American men [7]. This metrics were measured on the complete cohort before and after use of the chatbot. Participants were African American men, aged 40 years and over, without a prior prostate cancer diagnosis ($n = 354$). Participants experienced significant improvements in their prostate cancer knowledge ($p \leq 0.001$), informed decision-making self-efficacy ($p \leq 0.001$), and technology use self-efficacy ($p \leq 0.001$), post intervention. Sixty-seven percent of participants reported an intention to engage in informed decision-making.

3.1.2. Mental health

As much as 40% of cancer patients will face severe anxiety or depression [8,9]. In that setting, chatbots could be used to engage patients and provide them support at any time. Woebot is a Web-based cognitive-behavioral therapeutic (CBT) that was evaluated by Fitzpatrick et al.[10] The objective their study was to determine the feasibility, acceptability, and preliminary efficacy of a fully automated conversational agent to deliver a self-help program for people with symptoms of anxiety and depression. In an unblinded trial, 70 individuals were recruited online from a social media site and were randomized to receive either 2 weeks (up to 20 sessions) of self-help content from CBT in a conversational format with a Woebot ($n = 34$) or were directed to the National Institute of Mental Health ebook, "Depression in College Students," as an information-only control group ($n = 36$). The participants completed Web-based versions of the 9-item Patient Health Questionnaire (PHQ-9), the 7-item Generalized Anxiety Disorder scale (GAD-7), and the Positive and Negative Affect Scale at baseline and 2–3 weeks later (T2). Participants in the Woebot group engaged with the conversational agent an average of 12.14 (SD 2.23) times over the study period. No significant differences existed between the groups at baseline, and 83% (58/70) of participants provided data at T2 (17% attrition). Patients in the Woebot group significantly reduced their symptoms of depression over the study period as measured by the PHQ-9 ($F = 6.47$; $P = 0.01$), in an intent-to-treat univariate analysis, while those in the information control group did not. Another chatbot, Koko, was designed to express empathy, which could be positively leveraged in cancer-patients [11]. In this study, Morris et al report on the design of a conversational agent that could express empathic support in ways that might approach, or even match, human capabilities. The authors also assessed how users might appraise this system. The authors used responses from an existing pool of online peer support data that were repurposed by the chatbot and presented to the user. Information retrieval techniques and word embeddings were used to select historical responses that best matched a user's concerns. Ratings from 37,169 users were collected to evaluate the system. A controlled experiment ($N = 1284$) to test whether the alleged source of a response (human or machine) might change user perceptions was performed: the majority of responses created

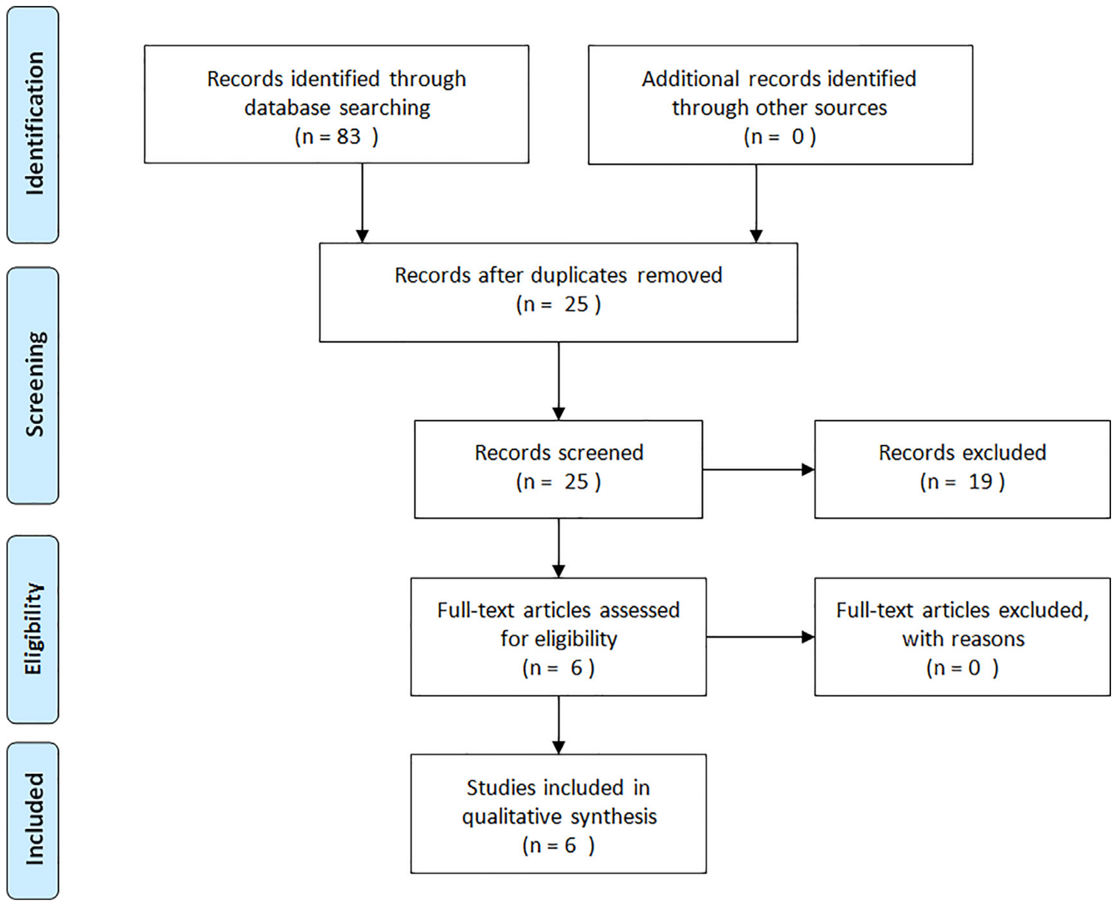

Fig. 1. PRISMA flowchart for studies selection.

by the agent (2986/3770, 79.20%) were deemed acceptable by users. But, users still made the difference between the chatbot's and the human answers and significantly preferred the efforts of their peers ($p < 0.001$). This effect was maintained in a controlled study ($p = 0.02$). Only one randomized controlled trial has been published [12]. In this study, Ly et al assessed the effectiveness and adherence to a smartphone app delivering strategies used in positive psychology to improve happiness and reduce negative symptoms, via an automated chatbot (Shim) for a non-clinical population. A total of 28 participants were randomized to either receive the chatbot intervention (n = 14) or to a wait-list control group (n = 14). Findings revealed that participants who adhered to the intervention (n = 13) showed a significant interaction effects with the chatbot compared to the wait-list control group.

Expression of caring and understanding is valued in supportive human communications. A study recently published by Liu et al. assessed whether humans could perceive empathy from a chatbot: in the first group of people (N = 158), participants simply read a dialogue between a chatbot and a human user. In a second group, participants (N = 88) interacted with a real chatbot. Three types of empathic expression were tested: sympathy, cognitive, and affective empathy, on the individuals' perceptions of the service and the chatbot. Data revealed that expression of sympathy and empathy is favored over unemotional provision of advice, in support of the Computers are Social Actors (CASA) paradigm [13].

3.1.3. Lifestyle change

FitTrack was developed to assess the ability of an Embodied Conversational Agents (ECA) called "Laura" to establish a long-term therapeutic alliance with users, and to determine if these relationships could be used to induce health behavior change [14]. In the FitTrack system, the ECA used nonverbal behavior to convey information in addition to the speech channel. Laura played the role of an exercise advisor who motivated sedentary persons to obtain the minimum level of physical activity recommended by current public health guidelines [15] over a two-month period of time. The system ran on standard home desktop computers so that participants could interact with the system on a daily basis. The system used the BEAT text-to-embodied-speech translator [16] to generate nonverbal behavior for the agent. FitTrack was successfully used in two randomized clinical trials, one involving MIT students and the second an urban, older adult population [17,18].

3.2. Ongoing trials

A search on ClinicalTrials.gov returns four trials evaluating chatbots in healthcare. In the United Kingdom, a non-randomized trial is being performed by the National Health Service to compare the Babylon chatbot to the non-emergency 111 telephone number [19]. Patients can interact with an automatic agent in order to describe their symptoms. Advices and information are given in return by the chatbot. Randomized studies demonstrating the superiority (or at least non-inferiority) of chatbots, compared to an intervention performed by a physician, do not exist. If chatbots are to be safely used by a large number of patients, they must be evaluated like a medical device, or even a drug.

The Buddy Study (NCT02742740) evaluates an ECA Oncology Trial Advisor for Cancer Trials that acts as an advisor to patients on chemotherapy regimens, promoting protocol adherence and retention, providing anticipatory guidance and answering ques-

Table 1
Studies evaluating conversational agents.

Tool	Intervention	Inclusion criteria	Number of patients	Endpoints
iDecide	Population information to all cohort	- African-American - Men - 40 years-old or older - With or without prostate cancer diagnosis	354	- Prostate cancer knowledge, - Informed decision-making self-efficacy, - Intention to engage in informed decision-making
Woebot (Web-based cognitive-behavioral therapeutic)	2 weeks of self-help content from CBT in a conversational format with a Woebot (n = 34)or National Institute of Mental Health ebook, "Depression in College Students," as an information-only control group (n = 36)	Online recruitment on social media	70	Feasibility, acceptability, and preliminary efficacy of the tool
Koko	Answers from a database of 37,169 real users vs answers from the conversational agent	Online recruitment on Koko platform	1284	User perceptions of a response depending on the alleged source of a response (human or machine)
Shim	Chatbot intervention (n = 14) or to a wait-list control group (n = 14)	Online recruitment on Shim platform	28	Effectiveness and adherence to a smartphone app delivering strategies used in positive psychology to improve happiness and reduce negative symptoms
Chatfuel	Group 1 (N = 158) participants simply read a dialogue between a chatbot and a human user. Group 2: participants (N = 88) interacted with a real chatbot	Online recruitment through Amazon Mechanical Turk	246	Empathy perception through the use of a conversational chatbot
FitTrack	Exercise advisor who motivated sedentary persons to obtain the minimum level of physical activity recommended by current public health guidelines	Students recruited from Massachussets Institute of Technology	101	working alliance inventory questionnaire (physical activity score)

tions. The chatbot also serves as a conduit to capture information about complaints or adverse events. Usability metrics will include session time, satisfaction, and error rates. Subjects will be identified from among patients on chemotherapy regimens at the Boston Medical Center [20]. All subjects will be enrolled for 2 months and randomized to the chatbot group or control group. Baseline data will be collected after enrollment and before randomization and will include: sociodemographic data, mental and social health, medical diagnosis (cancer type and stage), the Short Assessment of Health Literacy and the Short Portable Mental Status Questionnaire (SPMSQ) for cognitive impairment. The exit interview will be conducted 2 months after baseline. Data collected at the exit interview will include: mental and social health and the satisfaction with ECA Review of the subject's medical chart will be performed bi-weekly while the subject is enrolled. Chart review will be performed by a staff member who is blinded to the subject's group (intervention vs control). Primary outcome will be treatment protocol adherence, defined by the number of treatment visits attended/number of treatment visits scheduled. The secondary outcome will measure: subject satisfaction, number of adverse events as reported through the ECA and directly to clinic by patient, time to detect and resolve adverse events as reported through the ECA and directly to clinic by patient, adverse event false alarm rate, as reported through ECA and directly to clinic by patient.

A third study, the RAISE project (NCT01458002) [21], is designed to promote exercise and sun protection. The primary aims are to develop and assess the effectiveness of a tailored internet intervention on a national sample, to develop and assess the effectiveness of the internet intervention enhanced by a relational agent, and to determine if the intervention with the relational agent can outperform the regular tailored internet intervention. The study will include three groups (Control, Internet, Internet plus Relational Agent). A representative national sample of 1639 individuals at risk for both behaviors will be recruited.

A fourth study, INCASE, a randomized, controlled, blind study will be launched in France, comparing the information given by the Vik chatbot vs a multidisciplinary group of physicians (from medical, radiation and surgical oncology) to breast cancer patients (NCT03556813). One hundred and forty patients will be randomized in this non-inferiority trial. The EORTC QLQ-INFO25 questionnaire, that was validated to assess cancer patients' information [22], will be used to compare the quality of the information provided to the two groups of patients by the physician or the chatbot.

3.3. Applications in radiation oncology

In the field of Radiation Oncology (RO), chatbots could be leveraged in each step of the process: from cancer screening, diagnosis to patient information on treatment modalities and potential toxicities. They could also be used to seamlessly collect patient-reported outcomes (PRO) during and after treatments in a convenient manner. Algorithms could then be used on curated data to analyze in real-time PRO in order to red-flag the most at-risk patients or trigger an additional consultation with the physician and adapt supportive care treatments. Considering the documented discrepancies in available RO professionals in Europe, this kind of tool could have a significant interest [23,24].

4. Conclusion

The scarcity of clinical trials evaluating conversational agents in healthcare in general and oncology in particular is in sharp contrast with their potential benefits for the patients and the healthcare system. Health chatbots will need to be used by many and have access to rich data sets in order to increase their knowledge of medical terms, symptoms, and treatments. These systems will not replace the physicians and should be considered as a resource to enhance the efficacy of healthcare interventions [25]. If chatbots are shown to be effective and safe, they could be prescribed like any drug to improve patient information, monitoring, or treatment adherence. The best conversational agents may save patients with minor health concerns from a visit to the doctor. This could allow clinicians to spend more time to treat patients who need a consultation the most. Consultations for whose symptoms that don't necessitate an actual consultation could be avoided, potentially saving a significant amount of money and resources. But if the quality of these computer programs is not rigorously assessed, they could be unable to actually detect the difference between minor and major symptoms, without anyone knowing. The lack of objective evidence for the relevance and efficacy of this kind of applications is concerning since they are poised to be used by more patients. There are several differences between a conversational agent and a drug. For example, a standard drug formulary is locked down, evaluated, and then distributed. Conversational agents can effectively learn and adapt while deployed such that a conversational agent today deployed to millions of users will likely be vastly different in response to its operation within weeks. Significant hurdles currently exist in the widespread application of chatbots at this time. One major limitation stems from noncompliance of many commercially available systems with the Health Insurance Portability and Accountability Act.

Conflicts of interest

Pierre Nectoux and Benoit Brouard are respectively Chief Technical Officer and Chief Executive Officer in WeFight Inc.

References

[1] Turing AM. Computing machinery and intelligence. Mind 1950;49:433–60.
[2] Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 1966;9:36–45.
[3] Dr. Sbaitso [Internet]. Wikipedia, 2018[cited 2018 Oct 23] Available from: https://en.wikipedia.org/w/index.php?title=Dr._Sbaitso&oldid=865309185.
[4] AISB – The Society for the Study of Artificial Intelligence and Simulation of Behaviour – Loebner Prize [Internet][cited 2018 Oct 23] Available from: https://www.aisb.org.uk/events/loebner-prize.
[5] Miner AS, Milstein A, Schueller S, et al. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med 2016;176:619–25.
[6] Brouard B, Bardo P, Bonnet C, et al. Mobile applications in oncology: is it possible for patients and healthcare professionals to easily identify relevant tools? Ann Med 2016;48:509–15.
[7] Owens OL, Felder T, Tavakoli AS, et al. Evaluation of a computer-based decision aid for promoting informed prostate cancer screening decisions among african american men: iDecide. Am J Health Promot 2018. 890117118786866.
[8] Kadan-Lottick NS, Vanderwerker LC, Block SD, et al. Psychiatric disorders and mental health service use in patients with advanced cancer. Cancer 2005;104:2872–81.
[9] Mitchell AJ, Chan M, Bhatti H, et al. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. Lancet Oncol 2011;12:160–74.
[10] Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Mental Health 2017;4:e19.
[11] Morris RR, Kouddous K, Kshirsagar R, et al. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J Med Internet Res 2018;20:e10148.
[12] Ly KH, Ly A-M, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Intervent 2017;10:39–46.
[13] Liu B, Sundar SS. Should machines express sympathy and empathy? Experiments with a health advice chatbot. Cyberpsychol., Behav., Soc. Network. 2018;21:625–36.
[14] Bickmore TW: Relational Agents: Effecting Change through Human-Computer Relationships 284.
[15] Pate RR, Pratt M, Blair SN, et al. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. JAMA 1995;273:402–7.
[16] Cassell J, Vilhjálmsson HH, Bickmore T. BEAT: the Behavior Expression Animation Toolkit [Internet]. In: Prendinger H, Ishizuka M, editors. Life-Like Characters: Tools, Affective Functions, and Applications. Berlin, Heidelberg: Springer Berlin Heidelberg; 2004. p. 163–85 [cited 2018 Oct 23] Available from: https://doi.org/10.1007/978-3-662-08373-4_8.
[17] Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput-Hum Interact 2005;12:293–327.
[18] Bickmore T, Giorgino T. Health dialog systems for patients and consumers. J Biomed Inform 2006;39:556–71.
[19] Babylon Health: Online Doctor Consultations & Advice [Internet]. Babylon Health, 2018 [cited 2018 May 25]. Available from: https://www.babylonhealth.com.
[20] Study Buddy (an ECA Oncology Trial Advisor) for Cancer Trials – Full Text View – ClinicalTrials.gov [Internet] [cited 2018 Oct 23]. Available from: https://clinicaltrials.gov/ct2/show/NCT02742740.
[21] Online Tailored Interventions & Relational Agents for Exercise and Sun Protection – Full Text View – ClinicalTrials.gov [Internet] [cited 2018 Oct 23]. Available from: https://clinicaltrials.gov/ct2/show/NCT01458002.
[22] Arraras JI, Greimel E, Sezer O, et al. An international validation study of the EORTC QLQ-INFO25 questionnaire: an instrument to assess the information given to cancer patients. Eur J Cancer 2010;46:2726–38.
[23] Dunscombe P, Grau C, Defourny N, et al. Guidelines for equipment and staffing of radiotherapy facilities in the European countries: final results of the ESTRO-HERO survey. Radiother Oncol 2014;112:165–77.
[24] Lievens Y, Defourny N, Coffey M, et al. Radiotherapy staffing in the European countries: final results from the ESTRO-HERO survey. Radiother Oncol 2014;112:178–86.
[25] Goldstein IM, Lawrence J, Miner AS. Human-machine collaboration in cancer and beyond: the centaur care model. JAMA Oncol 2017;3:1303–4.