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

Jean-Emmanuel Bibault¹, Pierre Nectoux², Benjamin Chaix², Benoît Brouard²

1 – Department of Radiation Oncology, Hôpital Européen Georges Pompidou, AP-HP, Paris, France
2 – WeFight Inc., Institut du Cerveau et de la Moelle épinière, Hôpital Pitié-Salpêtrière, France

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 humans¹. 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 known as Natural Language Processing. They are widely available 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 answers². Patients can now use chatbots to check for symptoms and to monitor their mental health³,⁴. But their relevance and validity has rarely been assessed. A comprehensive search of MEDLINE shows that quality data on their use is still lacking in the literature: only one randomized controlled trial has been published⁵. In this study, Ly et al assessed the effectiveness and adherence of 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 significant interaction effects of the chatbot compared to the wait-list control group.

A search on ClinicalTrials.gov returns only five 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 number⁶. 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. In the fall of 2018, 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, will be used. This study will be the first of its kind. Vik is an Artificial Intelligence (AI) designed to empower patients and relatives with cancer. To understand the users’ messages and send personalized answers, the conversation with the patient goes through three stages: the first stage analyses the sentence and detect intent and entities, using word detection. The second stage activates modules according to the intents and entities detected by the first stage, and the third stage aggregates the answers of all activated module to build the answer sent to the user and saves the conversation on the user’s profile.

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. But the purpose of these AI is not to replace the clinician: it should be considered as a resource to enhance patient empowerment. In the end, if chatbots are shown to be effective and safe, they could be prescribed like any drug to improve patient information 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.
References


6. Babylon Health ; http://www.webcitation.org/70pe6dfHV