

Designing Depression Screening Chatbots

G GIUNTI^{a,b,1}, M ISOMURSU^a, E GABARRON^c and Y SOLAD^d

^aUniversity of Oulu, Oulu, Finland

^bTU Delft, Delft, Netherlands

^cNorwegian Centre for E-health Research, Tromsø, Norway

^dYale New Haven Health, New Haven, Connecticut, USA

Abstract. Advances in voice recognition, natural language processing, and artificial intelligence have led to the increasing availability and use of conversational agents (chatbots) in different settings. Chatbots are systems that mimic human dialogue interaction through text or voice. This paper describes a series of design considerations for integrating chatbots interfaces with health services. The present paper is part of ongoing work that explores the overall implementation of chatbots in the healthcare context. The findings have been created using a research through design process, combining (1) literature survey of existing body of knowledge on designing chatbots, (2) analysis on state-of-the-practice in using chatbots as service interfaces, and (3) generative process of designing a chatbot interface for depression screening. In this paper we describe considerations that would be useful for the design of a chatbot for a healthcare context.

Keywords. mHealth, chatbots, conversational agents, depression, healthcare, artificial intelligence, machine learning, natural language processing

1. Introduction

Conversational agents (chatbots) like Apple's Siri or Amazon's Alexa, are systems that mimic human dialogue interaction through the use of text or voice-based means [1]. Current trends foreshadow rapid adoption of conversational agents and some have even postulated that in the future, chatbots could be a solution in advanced medical scenarios like reducing overcrowding in hospitals [2] and offering mental health services [3]. However there is little literature that specifically targets the design of the interactions and behaviors that compose the human experience around AI models [4].

The World Health Organization points to depression as the leading cause of disability worldwide, with over 300 million people affected [5]. However, despite its prevalence, depression is underrecognized and underdiagnosed in many patient groups like breast cancer [6] or multiple sclerosis [7], leaving them untreated [8]. Chatbots have a strong potential for healthcare since they, unlike other software programs, can engage directly with people in a more natural way, and there is even evidence to suggest that people respond to them psychologically as though they are human beings [3]. A recent systematic review [1] found that chatbots are tentatively being used to provide counseling on birth control, medication adherence, exercise promotion and the search for suitable clinical trials.

¹ Corresponding Author, Guido Giunti; E-mail: drguidogiunti@gmail.com.

Currently, no guidelines or best practices are currently available that would help the design of chatbots for healthcare scenarios. This paper attempts to shed some light on key concepts that potential healthcare chatbot designers and researchers could benefit from knowing.

2. Methods

We followed a generative Research through design (RtD) process [9], where we implemented a chatbot interface for depression screening. The chatbot was designed to be used as part of a larger multiple sclerosis self-management solution. RtD is a process where knowledge is created through the exploration of issues that rise from the interplay of the design observations and experiences, and existing body of knowledge. The approach has already been used for designing digital health solutions [10].

As a design artifact of our RtD process, we created a chatbot that would be helpful in the screening for depression symptoms based on the Patient Health Questionnaire (PHQ9) [11]. The PHQ-9 is the depression module, which scores each of the nine DSM-IV criteria for depression as “0” (not at all) to “3” (nearly every day). It has been validated for use in primary care.

The chatbot, codenamed IGOR, would guide the user through the PHQ9 questions and calculate the severity of depression following the PHQ9 scoring. The results were sent to a designated professional for final assessment and not displayed to the user. IGOR chatbot was developed using DialogFlow platform with custom-developed backend application (NodeJS and Firebase). Application usability was tested with 10 willing participants who were staff members of the University of Oulu following inspection methods and Nielsen’s usability heuristics [12] and user tasks.

The observations and experiences from the design process were combined with the findings of the literature and state-of-the-art review through a qualitative hermeneutic data analysis process. The researchers involved in the analysis process were experts in healthcare, service design and software development.

3. Results

We categorized our findings into four broad categories, which highlight design considerations relevant to developing chatbot interfaces in the healthcare context.

3.1. *Ethical Implications*

Due to the sensitive nature of the healthcare context, it is imperative that steps be taken to explore potential uses of the system to minimize the risks that they are used in a non-ethical manner. There is always a risk that a chatbot interface is not 100% reliable. It might misinterpret the human input, or it might get input it is not able to correctly process. Users of health chatbots are not likely to know what the full, detailed capabilities of the conversational agent are, either concerning their medical expertise or the aspects of natural language dialogue it can handle. In our case, labeling our chatbot a tool for depression screening did not prevent users from going “off-topic” into areas the conversational agent has no expertise in, like medical emergencies. Regardless of

domain, users can also easily exceed any conversational assistant's natural language processing capabilities, leading to potentially harmful actions [4]. The primary concern should always be the users' safety.

3.2. Personality

We often attribute human characteristics to machines and tend to interact with them similarly as we would with real people [13]. It is then unsurprising that we expect chatbots not only to engage in human-like dialogue but have a human-like attribute and personality traits. Designing a healthcare chatbot requires that we also consider what kind and how much of a personality it will have. Selecting the proper personality and tone can make or break the user experience.

In audio interfaces, the user experience might be affected by design parameters such as perceived gender of the chatbot voice [14]. The design team needs to consider how the target group perceives different personality characteristics of the chatbot. An interesting aspect to consider are the gender stereotypes [15] affecting user experience, and whether the designer wants to reinforce existing stereotypes, or aim for more neutral choices.

3.3. Conversational Flow

Designing a successful conversational assistant requires first and foremost creating a well-thought conversational flow diagram. Typically, interactions with conversational agents begin with the chatbot greeting the user and quickly moving to more utilitarian conversational phases. During these phases it will likely be trying to acquire basic user information, identify relevant intents, and executing the pertinent fulfilment actions.

In a way, designing a chatbot is a combination of creating a decision tree and a movie script. The goal is to map out the potential conversational pathways that can lead the user to a target outcome, while providing clear and meaningful information. With chatbots, dialogue often substitutes user interface so a clean and clutter free experience is expected. Flowcharts are well suited to visually represent the conversation and can help us understand which steps are needed and which are redundant.

3.4. Predictability

Unconstrained user input allows for conversational flexibility but also comes with a higher risk for potential errors, which is a very important aspect in healthcare as it can never be fully predicted and may lead to patient safety issues. This lack of predictability can also create challenges in getting the chatbot-based technology approved to be used in the healthcare context, as validating their reliability might be challenging with current mechanisms. For example, if the chatbot utilizes learning algorithms or uses contextual parameters to personalize the conversation, it is impossible to predict all possible interaction flows for validation purposes.

Hard-coded rule-based systems fail to tackle complex combinations and sequences of symptoms. Hence, the costs of maintaining and extending these systems rise exponentially with time. The use of chatbots in health applications seem to lag behind those used in other areas, where dialogue management and natural language generation methods have advanced beyond the rule-based approaches. Regardless, thanks to technological advancements like NLP, chatbots are moving from being an odd curiosity to become a potentially valuable tool to be used in the not so distant future.

4. Discussion

Chatbots are causing significant disruption and media coverage on them seems to regard them as the next big thing for healthcare. The nature of these interactions is complex and different, arising from independent, evolving systems that need ongoing tending and maintenance by end users. Chatbot agents are usually created following a generic logic, with no regard to specific needs or preferences, and users receive the same responses regardless of cultural, socioeconomic or digital literacy level. There are practices in software development like user-centered design [16] that may be useful to address these problems. The goal of is to design solutions specific to the characteristics and tasks of the users [16].

In retail services, there is evidence that feeling of social pressure can negatively affect customer experience, and be a factor positively affecting adoption of self-service technologies [17]. Similar factors could be relevant in designing chatbot interfaces. They can be used to alter the service delivery situation so that the patient does not feel social pressure for using the shortest possible time to interact because of other patients waiting or the care personnel being busy. This can lead to better experience and more reliable answers, for example, in screening situations.

As discussed in this study, allowing users to provide unconstrained input may lead to unexpected and potentially risky situations. Users may start probing the limits of the chatbot by asking random questions, attempting casual conversation or going far beyond the intended uses of the chatbot. How the conversational agent responds to these situations in constructive and engaging form is critical, because data shows that the conversational style can greatly influence user behavior [18]: expressing sadness to Siri might cause it to reply something like “I’m sorry to hear that” or, “from our deepest sadness springs our deepest joy”. This type of response might be forgivable for a commercial use agent, however, expectations from a health-oriented chatbot are likely to be different [13].

In the case of mental conditions such as depression, the fact that chatbots do not “think” or form “judgments” may lead to people feeling more comfortable disclosing personal issues to them compared to a person or they may not be willing to share at all [19]. It may be that when it comes to empathic expression, conversational agents might be at a perpetual disadvantage.

Limitations: The presented work should be interpreted in the context of its limitations. There are inherent limitations to the research through design methodology. The design considerations use experiences that stem from a single design case. This was due to the context of the research project which influenced the actions that could be embarked on within the overall project scope. However, the construction of the recommendations was based on the available scientific literature and related work and, as it follows empirical evidence, its results are still valid.

5. Conclusions

The use of chatbots in healthcare is an emerging field, full of potential for development, but it must be approached with special care due to the sensitive nature of its context. The proposed design considerations are not intended to be considered as definitive, and rather they aim to be a mere starting point. Chatbot designers and developers should continue to explore and expand our knowledge around the effective design of conversational

agents as they address the burning challenges of modern healthcare. This is just the start of the conversation.

References

- [1] Laranjo L, Dunn AG, Tong HL et al. Conversational agents in healthcare: A systematic review. *J. Am. Med. Informatics Assoc.* 0 (2018) 1-11. doi: 10.1093/jamia/ocy072
- [2] Graham B, Bond R, Quinn M, Mulvenna M. Using data mining to predict hospital admissions from the emergency department. *IEEE Access.* 6 (2018) 10458-10469. doi: 10.1109/ACCESS.2018.2808843
- [3] Miner AS, Milstein A, Hancock JT. Talking to machines about personal mental health problems, *JAMA - J. Am. Med. Assoc.* 318 (2017) 1217-1218. doi: 10.1001/jama.2017.14151
- [4] Bickmore TW, Trinh H, Olafsson S et al. Patient and consumer safety risks when using conversational assistants for medical information: An observational study of Siri, Alexa, and Google Assistant. *J. Med. Internet Res.* 20 (2018) e11510. doi: 10.2196/11510
- [5] World Health Organization. Depression and other common mental disorders: global health estimates. World Heal. Organ. (2017). doi: CC BY-NC-SA 3.0 IGO
- [6] Reich M, Lesur A, Perdrizet-Chevallier C. Depression, quality of life and breast cancer: A review of the literature. *Breast Cancer Res. Treat.* 110 (2008) 9-17. doi: 10.1007/s10549-007-9706-5
- [7] Siegart RJ. Depression in multiple sclerosis: a review. *J. Neurol. Neurosurg. Psychiatry.* 76 (2005) 469-475. doi: 10.1136/jnnp.2004.05463
- [8] Gabarrón Hortal E, Vidal Royo JM, Haro Abad JM, Boix Soriano I, Jover Blanca A, Arenas Prat M. Prevalence and detection of depressive disorders in primary care. *Aten. Primaria.* 29 (2002) 329-337.
- [9] Koskinen I, Zimmerman J, Binder T, Redstrom J, Wensveen S. Design research through practice: From the lab, field, and showroom. 2011.
- [10] Giunti G, Mylonopoulou V, Rivera Romero O. More stamina, a Gamified mHealth Solution for persons with multiple sclerosis: Research through design. *JMIR MHealth UHealth.* 6 (2018) e51. doi: 10.2196/mhealth.9437
- [11] Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2. *Med. Care.* 41 (2003) 1284-1292. doi: 10.1097/01.MLR.0000093487.78664.3C
- [12] Nielsen J. Usability inspection methods. *Conf. Companion Hum. Factors Comput. Syst. - CHI '94.* 25 (1994) 413-414. doi: 10.1145/259963.260531
- [13] Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: System design and user perceptions. *J. Med. Internet Res.* 20 (2018) e10148. doi: 10.2196/10148
- [14] McDonnell M, Baxter D. Chatbots and gender stereotyping. *Interact. Comput.* 31 (2019) 116-121. doi: 10.1093/iwc/iwz007
- [15] Gustavsson E. Virtual servants: Stereotyping female front-office employees on the Internet. *Gender, Work Organ.* 12 (2005) 400-419. doi: 10.1111/j.1468-0432.2005.00281.x
- [16] Dabbs AD, Myers BA, McCurry KR, Dunbar-Jacob J, Hawkins RP, Begey A et al. User-centered design and interactive health technologies for patients. *CIN Comput. Informatics, Nurs.* 27 (2009) 175-183. doi: 10.1097/NCN.0b013e31819f7c7c
- [17] Dahm M, Wentzel D, Herzog W, Wiecek A. Breathing down your neck! *J. Retail.* 94 (2018) 217-230. doi: 10.1016/j.jretai.2018.04.002
- [18] Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Intern. Med.* 176 (2016) 619. doi: 10.1001/jamainternmed.2016.0400
- [19] Ho A, Hancock J, Miner AS. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *J. Commun.* 00 (2018) 1-22. doi: 10.1093/joc/jqy026