White Paper:
Considerations for Conversational Agents in mHealth Apps (Chatbots)

March, 2021
Philip A. Smeltzer, PhD

Executive Summary

Healthcare costs and patient satisfaction challenges combined with low to moderate improvements in safety and quality provide an opportune environment for improvement. The advent of Chatbots into mainstream use for patient interactions may yield gains in efficiency, patient satisfaction and improved health for the patients. Chatbots and artificial intelligence were initially created over 40 years ago.

A Chatbot can range from straightforward fixed designs to intelligent complex designs with robust artificial intelligence engines. Artificial intelligence (AI) attempts to mimic the human-to-human interaction and conversation. When free text or speech is part of the Chatbot design, natural language processes including deep learning techniques to power the conversational machine. The free-flowing conversation requires a large and robust data base with conversational context for the Chatbot to operate. A single fixed question design with five discrete responses for each follow-up questions grows exponentially to a required library of 125 feedback comments after only three levels.

The user interface and experiences are the differentiating set of features for a Chatbot compared to a conventional survey set of questions and answers. mHealth Chatbots operate in a more linear fashion compared to the open navigation of an app. The majority of healthcare Chatbots are text based, fixed design. A Chatbot can be configured for delivery through multiple media channels such as SMS text, in-app push notification, voice enabled digital assistant, voice-phone, interactive voice response, and browser-based interface. The Chatbot design features include the personality or profile characteristics of the Chatbot. A younger female doctor is the highest rated Chatbot profile for health conversations.

Patients routinely track personal health data. Physical activity minutes, steps, nutrition intake, and body weight are common tracking elements. Patients with long term conditions may track health data through peripheral devices including blood pressure, oxygen saturation levels, and glucose. Patient generated data, peripheral device data and historical data from electronic medical records provides a rich environment to improve the tailoring and personalization of conversational agents.

Chatbot accuracy and performance for assessment and decision making in clinical diagnosis has been found to compare with human physicians, if not more accurate in some reported use cases. Physicians are skeptical of the applied use of Chatbots due to the impersonal nature of the interactions and lack of empathy for the unique patient needs. Wider acceptance is garnered for administrative actions, such as making an appointment to gain efficiency and lower costs for all stakeholders. Privacy and data security is an issue cited by patients as a concern of Chatbot generated data.

Chatbot use can be expected to expand. The management of peripheral devices, collection of symptoms and administrative processes are the most likely areas to benefit from a Chatbot. Where large sets of data require management, Chatbots are the most likely to meet these needs and gain cost and productivity efficiencies as well as improved health of the patients.

The Need for Chatbots in Healthcare

The current structure of healthcare services in the United States is evolving. Clinical interactions with a patient seek to improve and maintain health. A variety of headwinds are present within the complicated multi-party healthcare ecosystem found in the United States. The cost inefficiency and low-quality performance of healthcare services are challenges [1-3]. There is a perceived lack of personalization and holistic approach based on data. Access is not considered equitable among many patients. [4] The advent of electronic medical records, big data and data mining in healthcare seeks to solve these issues of high cost/low efficiency and quality of care.

The consumers of healthcare, commonly serving as patients report low satisfaction when interacting with the current healthcare delivery organizations [5]. Increased personal care leads to better communication, more patient involvement, and overall, better quality of care [6, 7]. Patient satisfaction increases with the use of nurse telephone follow-up calls [8] while patient satisfaction decreases with longer waiting times [9]. Transitioning from provider-focused to patient-focused care improves patient experience [10]. The sculpting of the current healthcare landscape to increase personal care and reduce redundant and manual processes is expected to improve patient care and satisfaction. This concept of replacing clinical tasks with machines is referred to as transhumanism [11]. There appears to be a desire for improved human interactions for complex health management and automated solutions for the more routine issues and big data needs of health.

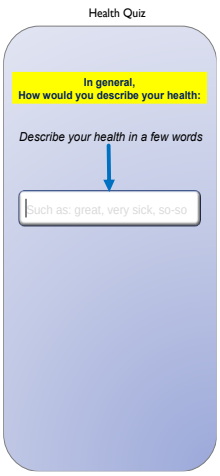
The overarching goals of population health and the triple aim seek to improve the quality of care, deliver higher satisfaction for patients and lower the total cost of care [12]. This backdrop of gaps in care and satisfaction may justify increased automation for routine clinical interactions. These challenges, problems and inefficiencies are the basis for Chatbot development. The use of conversational agents, conversational robots or Chatbots is a potential solution [13].

Chatbot Types and Structure

The spectrum of chatbot designs range from simple or fixed to complex or intelligent. There are multiple variations and hybrid choices for the chatbot designer to consider. Chatbots can be autonomous or embedded as intelligent software within a mobile app or other software or web platform [14].

Simple or fixed chatbots are rigid. These conversational agents follow an algorithm with designated branching logic or rules, and paths to follow. The customer of a scripted chatbot must select from predetermined or discrete, closed ended queries. There is no free play or open text options for the customer to respond and steer the conversation [13, 15]. An example of a fixed Chatbot interaction is illustrated here.

Intelligent chatbots are fluid interactions that seek to mimic better the normal human to human conversations. The goal of the conversation is defined, yet the path is guided by the responses of the customer with artificial intelligence (AI) engine follow-up questions. The customer provides voice or free form text to drive the conversation. The chatbot response is tailored to the customer input. The complexity of the intelligent chatbot is tethered to the knowledge base of conversations for the AI engine. True AI engines learn from the conversation. The most complex AI engines will learn from the conversation with each customer and tailor responses within the contextual framework of the current interactions [13, 15]. A voice activated Chatbot that considers the pitch, frequency, energy and rhythm to gauge emotion of the customer adds a new dimension of complexity, and personalization. An example of an intelligent Chatbot interaction is illustrated here.



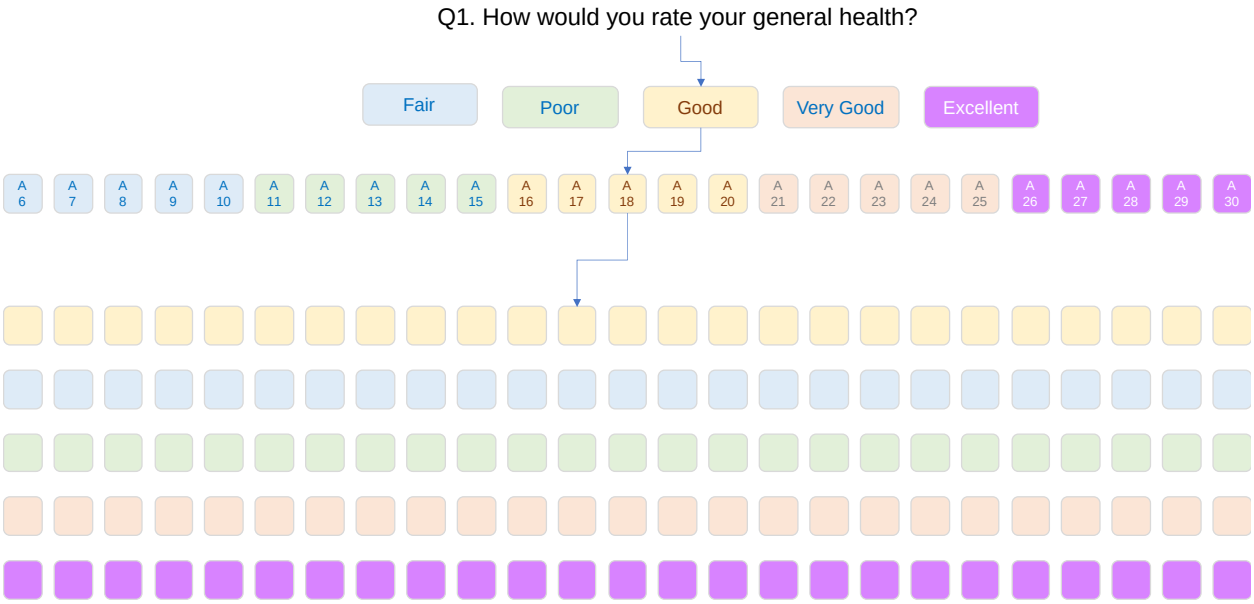
Artificial Intelligence

Artificial intelligence (AI) Chatbot features attempt to simulate a human-to-human conversation. A limitation is the amount of flexibility available to a Chatbot algorithm. The human brain neurons have access to numerous neurons within reasonable proximity yet the Chatbot machine is built with artificial neural networks with discrete layers, connections and directions of data flow [16]. The concept of AI is creating a capability by a machine or computer that can complete a task normally associated with intelligent actions of human beings [17]. AI is an overarching term for computers following complex algorithms with inputs processed to produce a specific outcome with specific results [18]. The first individuals reported to identify AI as a scientific discipline were Marvin Minsky and John McCarthy in 1956 [19]. An AI Chatbot response will deliver recommendations or suggestions. A simple or rule based Chatbot follows a fixed path, (if option 2 is the answer then deliver response B2). AI is a free-flowing set of interactions and learnings from previous interactions.

The capability of a Chatbot to emulate human conversations requires the processing of the customers natural speech. This natural language processing (NLP) is a technology with linguistic analysis and deep learning techniques to develop knowledge from the free-flowing text of the patient [20].

The integration of external data sources as a supplement to the conversation can improve the depth and richness of the conversation and improve the performance of the AI engine. The greatest challenge is the discovery and access to health data with specific knowledge of the topic for training the AI algorithms [17]. This can help reduce the translation errors that are typical with patient provided text inputs [21]. Electronic medical records, and social determinants of health data are examples of external data of use in AI decisions.

AI has different approaches. There are three categories or approach starting the more rigid machine learning, progressing to cognitive computing and finally, deep learning [22]. Machine learning can be as simple as forwarding a message to a nurse manager if the patient reports a blood pressure that exceed the established control limits. Even a simple stem and leaf rigid approach to a single question can generate a large body of potential responses. A single question beginning and 5 potential responses for each subsequent answer builds to a library of 125 responses within 3 layers as illustrated below.



1 question, 5 responses, each with 5 responses and 5 responses for each: 1 x 5 x 5 x 5 = 125 potential Responses

Machine learning can provide follow up questions to the out-of-control blood pressure data. Follow up questions may include medication intake history for 1-3 days, related symptoms, dietary intake with last 4 hours, and current stress or anxiety levels. These responses combined with the blood pressure data may or may not forward an alert message to the nurse manager. This latter scenario and the machine learning algorithm are designed to mimic an expected follow up in a human-to-human examination.

The level of cognitive computing increases the complexity of the coding to include additional factors [23]. Staying with the blood pressure example, the Chatbot might integrate data points such as poly-morbid diagnoses, pharmacologic agents known to influence blood pressure, available nutritional intake of the patient with caffeine or high sodium content. These new variables can trigger a new path of questions for the patient. An explanation as simple as the patient has had 4 cups of coffee can be included in the alert to the nurse manager for review. An additional path of questions may be started if the patient triggered on the anxiety query. The flag for a new path such as anxiety may not have been a result of a patient response, but the knowledge base feedback indicating patients with similar demographic characteristics, poly-morbid profile and nutrition intake patterns have increased odds of depression. Data mining in this example is used as a component of the AI process.

Deep learning includes the features above with an element of computer-generated evolution. Data is harvested and insights are developed as predictions of expected outcomes or expected associations. The general health question previously cited is an example. Older adults who self-report general health as fair or pair are more likely to require a hospitalization within the next 24 months [24]. Thus, deep learning techniques would flag this data point and trigger more detailed Chatbot questions. Additional questions may include depression screening, frailty scores, poly-pharmacy use, mobility issues.

The performance level of artificial intelligence and language processing requires a robust knowledge base. Every AI program must be trained and educated on the conversational database [21]. The levels of a Chatbot structure may be segmented into; data, information, knowledge and service [25]. This AI limitation requires time and in-depth knowledge of the subject matter. The attempt to create a simulated human-like experience for the patient parallels the need to educate and train a healthcare clinician. There is no standard approach in the marketplace. The sources of conversations used in training any chatbot can be non-generalizable if too specific or homogenous in nature. Thus, the danger of Chatbot bias is high when the patients are dissimilar to the characteristics of individuals in the knowledge base.

The knowledge-base will typically include two basic segments; basic data drawn from questions and symptom related questions [21]. The content of conversations needs to extract and processed. The processing steps include first reducing the extraneous or noise in the data.

Chit-chat such as discussing the weather, children or insurance issues are removed to lower the noise of the data. Sentences will be split to create discrete word bundles or phrases. The document sentences will be identified with a token for the AI processing and referencing [21].

User Interface and Experience (UI/IX)

The Chatbot design creates a new and unique experience for the human. The intent is to mimic the human-to-human interactions in the Chatbot design. Similar to first impressions when meeting a new person, Chatbot engagement is linked to the initial visual appearance and artificial personality. The ongoing dynamic behaviors and characteristics will influence the customer perception and likelihood to continue engagement [26]. Chatbots are distinguished by a linear flow of interactions. The most intelligent Chatbots have the capability to manage multiple paths, yet the linear flow is consistent. This contrasts to typical mHealth app features where the patient can navigate at will to a different component [13].

Any Chatbot can be delivered through the media channels of choice [27]. The lowest development effort is required for a Chatbot to interact within a mobile phone app. A text – SMS message conversation is also very straightforward to design. Chatbots can also be delivered in a browser environment as computer and web options. Chatbot features can include any single channel, combination of channels or all media channels as customer options. The use of a graphical interface to tap the response choice is another design option, rather than having the customer type in the response (A or option 1). Most Chatbots are developed as text-based interactions. The use of voice activated interactions is not common in the healthcare industry. The evaluation of the user experience is not well informed by the research literature [13].

Healthcare Chatbots can be configured to follow a unique personality. The Chatbot can be very fact based with little judgment, opinions or observations. Similar to the demographic characteristics of the Chatbot, the experience can be tailored to the culture of the patient. Colloquial terms, the Chatbot name and if voice generated the accent of the Chatbot can be tailored to the culture of the patient, if desired. A summary of Chatbot articles identified eight characteristic variables[13].

- Gender
- Chatbot as the identity or human-like
- Counselor, coach
- Tailored to a culture
- Fact-based without judgments
- Clinical professional such as nurse, doctor
- Knowledgeable with interactions informed by experts

Personality of the healthcare Chatbot is another dimension influencing patients. The weighted preferences were very close on these characteristics across a narrow range [26].

- Friendly
- Sympathetic
- Accessible
- Professional
- Reliable

Chatbot Uses in Healthcare

Healthcare Chatbot uses present problems common to other industries and unique issues to resolve as well. Chatbots are not as common in healthcare as other industries [27]. The first Chatbot Eliza was introduced in 1966 with a conversation between a psychotherapist and patient [28]. The common interactions between patients (customers) and healthcare professionals are opportunities for Chatbots. These conversational agents may be advantageous for delivering surveys or assessments, monitoring peripheral device data, health service appointment reminders, care plan regimen reminders, administrative processes such as appointments or payment updates [18]. Patients routinely track their weight, diet or exercise routine (approximately 60%). The use of peripheral devices by patients to manage long term conditions with blood pressure devices, glucometers, sleep trackers is common [29]. These computer driven conversations occur along five general categories [13]. Chatbot categories include:

- Service Connections
- Education
- Lifestyle Behavior Change
- Treatment and Monitoring
- Diagnosis

Symptoms such as feelings and self-reported behavior submitted by the patient can be inaccurate. There are multiple dimensions that may degrade the patient input. Patients may inaccurately report behavior. The reporting by a human interacting with a Chatbot requires translational adequacy, fidelity and fluency [30]. Physical activity [31] and nutrition [32] have been reported as two categories of poor recall and widely ranging interpretations of categorical behavior. Nutritional intake is commonly under reported. Physical activity as defined by a research scientist does not correlate to common reporting by patients. Physical activity terms including exercise, vigorous activity, moderate activity, and sedentary can be interpreted and reported differently even if the actual level of exertion is identical.

The most common preference for the Chatbot personality and demographic characteristic in healthcare interactions was a younger female doctor. The age and gender preference was similar for both the general population and older adults [26]. The weighting of preferences for the healthcare Chatbot were as follows.

- Age
- Appearance
- Clothing
- Gender

A conversational agent (Babylon) performance and diagnosis accuracy was compared to health practitioners in the United Kingdom. The capability of triage and diagnosis accuracy of the Chatbot was comparable to practitioner performance on standardized case studies. The Chatbot was found to be more accurate, faster and provided safer triage with diagnosis compared to the human doctors [13].

Although the advent of Chatbots and integrated AI features may yield advances in medical care, physicians are not universally supportive. There is a concern by some physicians that the Chatbot machine might interfere with the physician-patient relationship. The unknown architecture and Chatbot designs may create a black-box phenomenon and the basis for varying physician support. There is clinician support for time saving and monitoring services driven by Chatbots. The literature is informative on physician understanding of Chatbots. Generally, there was a lack of understanding, and technical functions of the computer driving the conversations and data to guide the patient interaction. Physicians have expressed concerns on the accuracy of Chatbots to interpret patient input or feedback and express appropriate emotional intelligence in patient interactions [18]. Chatbot functions with the most promise reported by physicians included scheduling appointments, navigation to healthcare locations, providing medical information [33]. Qualitative interviews reported physicians referring to AI as if it was a person or independent entity [17, 18]. The Chat bot interest of various stakeholders is summarized below [17, 29].

	Clinicians	Administrators	Patients	Researchers
Priorities	Patient Care	Efficiency	Improve Health, Individual rights	Generate Research Results
Forces	Deliberate Applied Use, Change in Training	Will to Develop AI, AI State of Art in Healthcare	Health Improvement	Encouraging Results
Vigilance	Want Proof	-	Information, Privacy, Social Justice	-
Obstacles	Liability, Responsibility	Liability, Responsibility	Liability, Responsibility	Funding Need

The development and applied use of a health care Chatbot is only the beginning. Comparable to other health delivery services the device should be evaluated and trended for quality assurance and quality improvement. Evaluation instruments are available to provide a more accurate assessment of the Chatbot as a device. The bilingual evaluation understudy tool was developed to evaluate Chatbot translational accuracy [30]. A well-established 10 question survey to document usability or satisfaction of mobile apps is the system usability scale (SUS) [34]. Levels of engagement, satisfaction, symptom accuracy, and costs to complete transactions are worthy of investigation. The use of standardized satisfaction, usability and condition specific survey instruments can assist in the Chatbot evaluation process.

The lack of deep outcome studies to document the impact on patient health and healthcare in general is a concern. The challenges of acquiring a database to serve as the knowledge base for natural language processing combined with the complex communication processes between patient and practitioner may dictate a conservative approach to rapid growth of AI driven Chatbots. A Chatbot with a fixed design although not as innovative or novel has advantages. Fixed responses allow for more accurate interpretation and symptom collection in a healthcare environment. Safety and efficacy evaluations of Chatbots are prudent before proceeding to the more complex AI Chatbots.

References

1. Institute of Medicine . Committee on Quality of Health Care in, A., *Crossing the quality chasm : a new health system for the 21st century*. 2001, National Academy Press: Washington, D.C.
2. McGlynn, E.A., et al., *The quality of health care delivered to adults in the United States.* The New England journal of medicine, 2003. **348**(26): p. 2635-45.
3. Whittington, J.W., et al., *Pursuing the Triple Aim: The First 7 Years.* The Milbank quarterly, 2015. **93**(2): p. 263-300.
4. Latif, S., et al., *How 5G Wireless (and Concomitant Technologies) Will Revolutionize Healthcare?* Future internet, 2017. **9**(4): p. 93.
5. Luxford, K., D.G. Safran, and T. Delbanco, *Promoting patient-centered care: a qualitative study of facilitators and barriers in healthcare organizations with a reputation for improving the patient experience.* International journal for quality in health care, 2011. **23**(5): p. 510-515.
6. Branson, C., B. Badger, and F. Dobbs, *Patient satisfaction with skill mix in primary care: a review of the literature.* PRIMARY HEALTH CARE RESEARCH AND DEVELOPMENT, 2003. **4**(4): p. 329-339.
7. Cleary, P. and B. McNeil, *Patient Satisfaction as an Indicator of Quality Care.* Inquiry, 1988. **25**(1): p. 25.
8. Schimpf, M., et al., *Patient Satisfaction With Nurse-Led Telephone Follow-up in an Ambulatory Setting.[Article].* Female Pelvic Medicine & Reconstructive Surgery, 2016. **22**(6): p. 3.
9. Anderson, R.T., F.T. Camacho, and R. Balkrishnan *Willing to wait?: The influence of patient wait time on satisfaction with primary care*. BMC Health Services Research, 2007. **7**, DOI: 10.1186/1472-6963-7-31.
10. Luxford, K., D.G. Safran, and T. Delbanco, *Promoting patient-centered care: a qualitative study of facilitators and barriers in healthcare organizations with a reputation for improving the patient experience.* Int J Qual Health Care, 2011. **23**(5): p. 510-5.
11. Huxley, J., *Transhumanism.* ETHICS IN PROGRESS, 2015. **6**(1): p. 12-16.
12. Berwick, D.M., T.W. Nolan, and J. Whittington, *The triple aim: care, health, and cost.* Health affairs (Project Hope), 2008. **27**(3): p. 759-69.
13. Tudor Car, L., et al., *Conversational Agents in Health Care: Scoping Review and Conceptual Analysis.* J Med Internet Res, 2020. **22**(8): p. e17158.
14. Ruttkay, Z., C. Dormann, and H. Noot, *The blind men and the elephant revisited evaluationg interdisciplinary ECA research*, in *From brows to trust evaluating conversational agents*, Z. Ruttkay and C. Pelachaud, Editors. 2004, Springer: Dordrecht, Netherlands. p. 3-26.
15. *What is a chatbot? All you need to know about chatbots!* Learn 2021 [cited 2021 February 18, 2021]; Available from: botpress.com/learn/what-is-a-chatbot.
16. Coperich, K., E. Cudney, and H. Nembhard. *Continuous improvement study of chatbot technologies using a human factors methodology*. in *Proceedings of the 2017 industrial and systems engineering conference*. 2017.
17. Laï, M.C., M. Brian, and M.F. Mamzer, *Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France.* Journal of translational medicine, 2020. **18**(1): p. 14-14.
18. Nadarzynski, T., et al., *Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study.* Digital health, 2019. **5**: p. 205520761987180-2055207619871808.
19. Winston, P.H., *Marvin L. Minsky : a founding father of artificial intelligence.* Nature (London), 2016. **530**(7590): p. 282.
20. Doan, S., et al., *Building a Natural Language Processing Tool to Identify Patients With High Clinical Suspicion for Kawasaki Disease from Emergency Department Notes.* Academic emergency medicine, 2016. **23**(5): p. 628-636.
21. Omoregbe, N.A.I., et al., *Text Messaging-Based Medical Diagnosis Using Natural Language Processing and Fuzzy Logic.* Journal of healthcare engineering, 2020. **2020**.

22. Nguyen, T.L. and T.T.H. Do. *Artificial Intelligence in Healthcare: A New Technology Benefit for Both Patients and Doctors*. in *2019 Portland International Conference on Management of Engineering and Technology (PICMET)*. 2019.
23. Sivalenka, V., et al., *Exploiting artificial intelligence to enhance healthcare sector.* Materials Science and Engineering 2020. **981**(022061): p. 7.
24. Smith, S.M., et al., *Interventions for improving outcomes in patients with multimorbidity in primary care and community settings.* The Cochrane database of systematic reviews, 2016. **3**: p. CD006560.
25. Chung, K., et al., *Chatbot-based heathcare service with a knowledge base for cloud computing.* Cluster computing, 2019. **22**(S1): p. 1925-1937.
26. ter Stal, S., et al., *Who Do You Prefer? The Effect of Age, Gender and Role on Users' First Impressions of Embodied Conversational Agents in eHealth.* International journal of human-computer interaction, 2020. **36**(9): p. 881-892.
27. Laranjo, L., et al., *Conversational agents in healthcare: a systematic review.* Journal of the American Medical Informatics Association : JAMIA, 2018. **25**(9): p. 1248-1258.
28. Dale, R., *The return of the chatbots.* Natural language engineering, 2016. **22**(5): p. 811-817.
29. Dinh-Le, C., et al., *Wearable Health Technology and Electronic Health Record Integration: Scoping Review and Future Directions.* JMIR Mhealth Uhealth, 2019. **7**(9): p. e12861.
30. Papineni, K., et al., *BLEU: a method for automatic evaluation of machine translation*, in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. 2002, Association for Computational Linguistics: Philadelphia, Pennsylvania. p. 311–318.
31. USDHHS, U.S.D.o.H.a.H.S., *Physical Activity and Health: A Report of the Surgeon General*, P. USDHHS, Office of the Surgeon General, Editor. 1996, National Center for Chronic Disease Prevention and Health Promotion: Atlanta, GA.
32. Harris, J.E., et al., *An introduction to qualitative research for food and nutrition professionals.* Journal of the American Dietetic Association, 2009. **109**: p. 11.
33. Palanica, A., et al., *Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey.* Journal of medical Internet research, 2019. **21**(4): p. e12887-e12887.
34. Brooke, J., *SUS - A quick and dirty usability scale*, in *Usability Evaluation in Industry*, B.T. P.W. Jordan, B.A. Weerdmeester, McClelland, Editor. 1996, Taylor & Francis: London. p. 189-194.