Designing Everyday Conversational Agents for Managing Health and Wellness

A Study of Alexa Skills Reviews

Ji Youn Shin
Media and Information
Michigan State University
East Lansing, Michigan, USA
shinji5@msu.edu

Jina Huh-Yoo
College of Computing
Drexel University
Philadelphia, Pennsylvania, USA
jinahuhyoo@drexel.edu

ABSTRACT

Conversational agents have been developed for supporting a wide array of areas, including autonomous vehicles, decision making, and health behavior change. In the last few years, conversational agents increasingly became available as everyday technologies. This phenomenon enables opportunities for finding novel ways to support health and wellness in everyday contexts. By conducting a content analysis of 433 user reviews of Amazon Alexa's Skills, the goal of this study is two-fold: (1) Extract users' perceived strengths of conversational agents in everyday health and wellness management, (2) develop design heuristics for developing conversational agents for health and wellness. We found that the role of trustworthy content providers is critical during the adoption. The Skills enabled people to overcome logistical barriers to improving daily health and wellness routines. The findings also revealed the importance of transparency in the limitations of the Skill and how to better design command dialogues. We present the design heuristics of conversational agents, building on Nielsen's Usability Heuristics, and discuss implications for designing conversational agents that support health and wellness.

CCS CONCEPTS

• **Human-centered computing~ User studies**

KEYWORDS

Conversational agent; voice user interfaces; voice assistants; smart speakers; Amazon Alexa; user experience; design; design heuristics; health

1 Introduction

Conversational agents became pervasive in everyday contexts through personal smartphones and home devices. By 2020, eMarketer predicted more than 76.5 million users in the U.S. will adopt at least one conversational agent among Amazon Alexa, Google Assistant, Microsoft Cortana, and Apple Siri [20]. These devices offer diverse conversational agent applications (e.g., 'Skills' in Amazon Alexa and 'Actions' in Google Assistant) that span from entertainment, work assistance, and purchasing items to controlling smart devices [39,60]. For instance, Amazon provides more than 50,000 Skills to worldwide users [1] in the areas of Business and Finance, Connected Car, Smart Home, and Health and Fitness. Recent articles highlighted that conversational agents are one of the powerful solutions for individuals who need cognitive behavioral therapy [56] or older adults who live alone [13]. Because agents can seamlessly and closely be integrated into users lives and interact with users at all hours, users can take advantage of them in their daily routines, similar to chatbots being able to be carried around with users' personal devices and integrated within their home environments [3,5,24].

With the advent of off-the-shelf conversational agents making these devices more prevalent, recent studies expanded the scope of contexts where conversational agents can play a significant role. However, ways to improve sustained use of such devices are under-studied. While the conversational agents became prevalent in home environments, it did not automatically lead to successful user adoption, long-term use, and active user involvement [15]. Conversational agents need much improvement in mimicking natural, sophisticated conversational exchanges among humans [4,34,46,49].

Building on existing efforts, researchers continue to discuss how conversational agents can be better designed. In the healthcare environment, conversational agents offer a potential to serve as a personal health assistant by finding healthcare resources, having sympathy and brief advice for users' emotional concerns [9,28,38], monitoring sleep patterns [50], or assisting dementia patients as they can repeatedly answer the same questions with patience [62].

However, these findings have mainly been situated in a lab setting with a specialized conversational agent, not a Skill of off-the-shelf products [28]. We know little about how everyday conversational agents for health and wellness have been perceived and used.

To address this gap, we conducted a content analysis of 433 reviews on 60 conversational agent applications, called "Skills," provided by the Amazon Alexa product. We examined the reviews of the Skills Amazon has classified under the Health and Fitness category. We investigated the types of health-related everyday conversational agents available to users, how and why users incorporated conversational agents into their daily routines, and users' perceived barriers and opportunities of using the conversational agents.

This is the first study to capture users' perceived benefits and barriers of using health-related conversational agents from the reviews of the users who voluntarily adopted the Skill. From the findings, we identified implications for designing conversational agents and the design heuristics that can apply to conversational agents supporting health and wellness and beyond. Our study broadens the knowledge of conversational agent design, building on the increased interest in this area.

2 Related Work

We use the term "Conversational Agent" to refer to the form of a speech dialogue system embedded in personal devices [34]. In 1966, the first chatbot, named ELIZA, enabled users to have a text-based conversation with a psychotherapist [59]. Several decades later, Natural Language Understanding [55] and ontology-based pattern recognition engines were integrated with personal smart devices, such as smartphones and smart speakers, which are off-the-shelf products. During this evolution of conversational agents, researchers investigated how the agents can support users' various tasks and needs and be embedded in everyday contexts through evaluating the agents' usability and barriers of adoption.

2.1 User Experiences with Everyday Conversational Agents

In the last ten years, researchers investigated user dynamics around in-home conversational agents, including why and how such products are used and the factors that led to positive user experiences. For example, Luger et al. examined users' motivations and limitations that affected the ongoing use of conversational agents in everyday life and proposed future design implications. Examples included the revelation of system intelligence, the creation of humorous engagement, and supported ongoing user engagement through the agent design [34]. By 2014, conversational agents became available in the form of a smart home speaker, presenting opportunities to be collectively used by family members. Lovato et al. studied 18 families to investigate the questions children asked their conversational agents and how the agent provided answers [31]. They found the importance of simplifying the answers, understanding the contexts in which children ask questions, and adapting answers to users' repetitive

questions. Also, Purington et al. conducted a content analysis of 587 reviews of Amazon customer reviews of Alexa [47]. They found that how people refer to their Alexa device (i.e., personification of Alexa) is positively related to user satisfaction. Similarly, Gao et al. found that many reviewers of Amazon Alexa personified their device as a friend or family member [21]. The results showed that personifying the conversational agents, either as friend or a family member, predicted more positive emotions compared to those who simply referred to their conversational agent as a 'device'. While previous literature highlighted what enabled conversational agents to be adopted at home and further in family settings, researchers also found the barriers to adopting everyday conversational agents.

2.2 Barriers of Using Everyday Conversational Agents

Although users had high expectations when they first purchased conversational agents, user satisfaction towards conversational agents drastically decreased once they started using the devices [15,34]. Due to the conversational nature of voice assistants, users often established high expectations for system capabilities [34]. To reduce the gap between users' expectations and actual experiences in communicating with conversational agents, researchers investigated Communication Repair Strategies of conversational agents. These strategies refer to users' efforts to modify dialogues to resolve misinterpretations [4]. Cho et al. further looked at the obstacles that users face when they use conversational agents. The team conducted a qualitative study for 12 months with first-time Amazon Alexa users to track user experiences and identify how user experiences changed over time [15]. Findings showed that users expected emotional exchanges, relationship building, and interactions (e.g., conversations) similar to those that they would have with humans. Although there were more than 50,000 conversational agents provided by Amazon Alexa, the study participants showed decreased motivation to use their conversational agent. None of Cho et al.'s study participants were willing to purchase the agent after 12 weeks. To address these challenges, researchers investigated ways to improve communication breakdowns and speech recognition errors [14]. Study results suggested types of tactics users employ to overcome barriers, including hyperarticulation (exaggeration) [10,43], increased volume [10], and the use of different utterances or simplified words [4,44]. Similarly, researchers described repair strategies for using conversational agents, which refers to reformulation strategies, such as addition or substitution, removal, and re- ordering of words [27,30]. By doing interviews, Luger et al. developed how agents can reveal humanlike cues, such as humorous conversations and intelligence [34]. Sciuto et al. also described how households integrated in-home conversational agents into their everyday lives with work-arounds to enable effective use [52]. The study found the importance of the placement of the agent, which promotes consistent use. Furthermore, users integrated conversational agents into their daily routines but often discontinued their use due to the user's instable routine or lost reliability.

2.3 Health Management Through Conversational Agents

A large body of healthcare literature shows the possibilities of using conversational agents to aid health management [55]. For example, Rick et al. designed a chatbot, called SleepBot, to monitor sleep patterns [50]. Using sleep monitoring technology, the SleepBot prompted users with a set of questions on sleep habits. Based on identified problems, the intervention aimed to craft successful sleep behaviors. Bickmore et al. developed a web-based search interface targeted for people with low health and computer literacy [9]. The results showed that conversational agents implemented in the web-based search interface had a positive impact on finding health-related information, particularly appropriate clinical trials. Results also showed that they helped to decrease recruitment bias against marginalized populations. Other studies found positive impacts of conversational agents on individuals with mental illnesses. For instance, Vaidyam et al. conducted a literature review on how conversational agents and related technology could be effective and enjoyable solutions in psychiatric treatments [57]. The results showed that conversational agents can increase benefits in psychoeducation and self-adherence. Similarly, another study identified the barriers in implementing clinical therapy for patients with Post-Traumatic Stress Disorder using Amazon Alexa to provide more effective therapy [41]. The results highlighted the importance of short dialogues and interactivity with regards to developing conversational agents. This work could be extended to provide insights for developing conversational agents that aim to deliver therapeutic content. To support chronic disease management, studies also suggested design guidelines for a smartphone-based conversational agent [7]. Simulating face-to-face health counseling between patients and healthcare providers, conversational agents were designed to provide counseling for patients with chronic conditions. The results from a randomized trial showed that the intervention led to enhanced self-reported quality of life. Similarly, virtual agents implemented in an anxiety and depression LiCBT (Low-Intensity Cognitive Behavioral Therapy) and the Heart Failure Companion were developed in the mental health and heart failure context [33]. The preliminary study results showed improvements in stress levels and health knowledge. However, the participants of the study showed declined use of the coach over time. Chung et al. examined the characteristics of 309 health and fitness Skills for Amazon Alexa and Google Assistant [17]. They found that Skills were targeted for general audiences and that the majority of the Skills focused on delivering health education. Although the study did not provide how users reviewed these Skills, the results suggested a quantitative, descriptive account of how many Skills provided what kind of content and its user ratings. They found that limited Skills were available for patients, caregivers, or medical professionals. Security and privacy issues were mentioned as a potential barrier to advancing the active use of Skills in the healthcare environment.

Existing literature highlighted the opportunities, communication barriers, and strategies in designing conversational agents. To our knowledge, no studies have looked into users' perceived usefulness and effectiveness of conversational agents for everyday health and wellness management that users voluntarily adopted. Studies in conversational agents supporting self-care of health were from a lab-based, experimental setting [55]. Building on the literature, by investigating users' reviews of conversational agent use, we discover the design guidelines of conversational agents for health and wellness.

3 Methods

The data used in this study were publicly available user reviews on Alexa Skills, and the researchers of this study did not interact with the users to obtain any data. Accordingly, the authors' universities' Institutional Review Boards determined the study as non-human subjects research. Alexa Skills are equivalent to 'apps' for mobile phones, where users can download and activate different applications to be used with Alexa [1]. We chose to study Alexa's Skills because it presented 60% of market share in smart speaker systems in 2018 [19]. We collected the reviews in February of 2019.

3.1 Data Collection

The goal of our study was to understand users' perceived experiences in using conversational agents for health and wellness maintenance in their daily lives and to develop design heuristics for designing conversational agents for health and wellness. To do so, we examined the user reviews of Amazon Alexa's Skills. As of February of 2019, Amazon provided 23 categories of Skills, including Health and Fitness, Business and Finance, Education and Reference, and Kids. For the Skills to meet the inclusion criteria, they should be classified by Amazon in the "Health and Fitness" category. The Skills under Health and Fitness provide health information monitoring, information on locations of providers, instructions for exercises, daily tips on healthy living, etc. We initially collected 1,635 Skills that were classified under the Health and Fitness category of Amazon Alexa Skills. We removed one Skill because it was a duplicate. The average number of reviews that 1,634 Skills received was 11. 892 Skills did not receive any reviews, and 742 Skills received at least one review from the users. The 742 Skills with at least one review, we filtered down to 130 Skills that received at least 11 reviews (the average number of total reviews the 1,634 Skills received). The mean and standard deviation start rating score of the 130 Skills were 3.8 and 0.72 respectively.

From 130 selected Skills, 19 Skills received higher star rating score than average (3.8+0.72) and 21 Skills received lower star rating score than average (3.8-0.72). From the remaining 40 skills, we identified 20 randomly selected Skills with average star rating score (3.8±0.72). The final list of 60 Skills were in the areas of Yoga Instruction, Home Exercise, Health Education, Medication Tracker, Daily Health Tips, Meditation Guidance, Sleep Assistance, and Stress Management. Each Skill's profile on the Amazon website provided 'the most frequent keywords' of the reviews. Examples of those keywords from the collected Skills

included "gives me peace," "life saver," "soothing and relaxing," "better choices," "nutrition advice," "goal weight," "quick workout," "simple and easy," "every day," "start the day," "gets you moving," "stopping," "trouble accessing," etc. From each Skill, we collected information about the Skill developer (vendor), reviewer ID, title of the reviews, number of reviews, and number of ratings that each Skill received. We then collected example commands that users can use, the Skill descriptions, and frequently used words that were featured from the reviews. To reduce noise in the data, from each of the Skills we selected, we collected the 'Top Reviews' that were listed on the first page of each Skill review for the final analysis. 'Top Reviews' were a subset of reviews that were rated as 'helpful' based on other customers' votes. From each of 60 Skills selected in total, six to eight 'Top Reviews' were included for the final analysis, and this resulted in a total of 433 reviews.

3.2   Data Analysis

From 433 reviews, we focused on extracting information on the following overall themes of inquiry: (1) Reasons for reviewers using the Skills, (2) impact of the Skills on their everyday health and wellness routines, (3) barriers or facilitators for using the Skills, (4) workarounds for using the Skills effectively, and (5) evaluation criteria of the Skills. The answers to these themes will provide us with implications for how to better design everyday conversational agents for health and wellness domains. Adopting Strauss and Corbin's grounded theory framework [54], we chose thematic analysis to identify the broader patterns throughout the data that answered the above themes. To do so, we initially selected 40 random reviews (approximately 10% of the entire reviews), and the first author conducted initial open coding exercises and presented the results to the second author. The first author then continued to conduct coding for the rest of 400 reviews using Nvivo Pro 11. The research team met weekly to discuss newly emerged codes to maintain agreement as a team. The coding structure thus was modified and refined throughout these iterative cycles of discussions. After the coding was completed, the authors collectively conducted affinity diagramming [6] to identify patterns and high-level themes that represent the rest of the codes. This process enabled grouping and refining of the rest of the subtheme codes according to the high-level themes we developed.

4   Findings

In this section, we report the thematic analysis findings of the reviews: (1) What made people want to use the Skill; (2) How users improved their health and wellness routines through conversational agents; and (3) What can be improved about existing conversational agents for health and wellness. Below, we report our findings, supported by reviewer quotes, and describe their experiences with using the Skills. We refer to each review using a combination of the Skill's ID and the reviewer ID: S1_R1, where S1 refers to the Skill ID and R1 to the reviewer ID.

4.1   What Made People Want to Use the Skill

In addition to the default functions that Amazon Alexa provides when users purchase the device, the company offers an additional 50,000 Skills to be used. Activating these additional Skills is an extra step for the users, so users would generally need some motivation and expectation for activating the Skills. We found that: (1) the Skill developer's positive reputation, (2) the favorable personality of content providers, and (3) personal motivations for improving health were the major themes for why reviewers chose to activate the Skill.

4.1.1 Developers and Content Providers' Reputations' Influence on Perceived Usefulness of the Skill. The Skill developer's history and reputation was one of reviewers' motivations to activate the Skill. For instance, reviewers had encountered the content in the past offered by the same developer from different technology platforms, such as mobile apps and websites. Enabling the well-known content in Alexa was useful for them:

> What a useful Skill. [The Skill], I've been following for over a year now and now at the convenience of Alexa. Always seems to have the best [nutrition] advice exactly when I need it. -S8_R7
>
> I found this app relatively simple to use with minimal guidance. I am familiar with [the general health information Skill] so I found using the Amazon Alexa version easy to use. -S79_R5.

Similarly, S14_R8, S54_R5 and R9 mentioned that they used the Skill because they have been following the content provider as a personal trainer. Another reviewer (S12_R5) used a Skill that provided the sound they enjoyed for having comfort. Because the Skill provided dynamic nature sounds, the reviewer used the Skill to cover box fan sounds. This reviewer then activated other Skills that provided similar effects.

4.1.2 Perceptions toward Content Provider's Characteristics. Another main evaluation criterion was perceptions of content providers. The term "content providers" refer to people who guide exercise or provide information for users through voice in the Skills. In the reviews, reviewers showed positive feelings for content providers, including comments such as "inspirational" (S13_R8), "honesty and openness" (S13_R8), "warm" (S112_R2), "reliable, relatable and accessible" (S8_R1), and "genius" (S2_R3).

Reviewers further pointed out how they were satisfied with the knowledge level of the content providers and how these experiences increased trust towards the content provider and reinforced continued use of the Skill. Reviewers considered the Skills as a source to find quick health information, such as information on pregnancy (e.g., "This is a great resource for quick info, size of baby updates" -S46_R3) or diet. Particularly, if the Skills were intended to deliver knowledge, such as information about meditation or nutrition, reviewers perceived the knowledgeability of the Skill as critical criteria to review:

> It is a great source of information of the basic knowledge that we can do and use with the different oils. -S62_R2a

[The content provider's] answers to my very specific questions about back pain were spot on. -S79_R8

Reviewers also confirmed that the characteristics of the content providers influenced their evaluations of the Skill. For instance, in the case of the Skills that aimed to enhance users' health behaviors or promote internal growth, having content providers who shared their personal experiences and attitudes was perceived as beneficial:

> This podcast type skill has a great narrator and a positive message every day. Sometimes it's humorous, sometimes emotional but it is always encouraging and touching. [The content provider] shares anecdotes and life lessons with his listeners, grounds each person to the reality of their value, strength, and purpose in this big wide world. He endures his own struggles with chronic illness and shares relatable tales about his ups and downs. If you are looking for hope, acceptance and entertainment, you should give this skill a try. -S13_R3

> [The content provider] really gives a warm message and her guidance helped me fall asleep quicker than I have in months! Highly recommend this meditation. I'm going to use it again and again. I can't wait to get more meditations from her! -S112_R2

The way reviewers reviewed the Skills showed that they perceived them as communication channels with content providers, not as a device. This finding showed that the Skills were no longer perceived as an artificial intelligent device, but as a gateway to connect with the content providers.

4.1.3 Skills as a Way to Jump Start Behavior Change. Reviewers also showed personal motivations for improving health that led them to activate the Skills. They talked about wanting to improve health-related routines and daily habits, such as water intake or sleep patterns, critical for establishing healthy lifestyles. They wanted to enhance these routines and habits with the Skill:

> It makes me really think about my health. The first day was about remembering to drink water which I did and actually noticed I was less hungry and felt better. I am looking forward for more tips on wellness. It's like having a virtual health coach for free! -S66_R5

> I've been having trouble sleeping for a while now so I decided to try this meditation when I saw it. -S112_R2

R8 described how they activated the Skill to "jump start" their motivation to change unhealthy behaviors. R8 activated the cardio Skill to begin the fitness routines and expressed satisfaction:

> I haven't worked out in months so [the Skills] was perfect just to get me started. Planning to do it at least a few times a day moving forward! -S15_R8

Reviewers activated the Skills motivated by multiple factors, including the Skill developer's good reputation, the favorable personality of content providers, reviewers' positive feelings

toward them, and personal motivations for improving health. Next, reviewers shared how these motivations were connected to adjusting their health and wellness management routines.

4.2 How Users Improved Their Health and Wellness Routines Through Conversational Agents

The reviews showed that using the Skills helped reviewers better integrate work, life, and wellness routines throughout the day. This integration was a challenge for the reviewers due to time and other logistical constraints. We found that reviewers actively used the Skills to maintain mental and physical health. They overcame time limitations and physical space constraints for working out or seeking health-related information freely throughout the day.

4.2.1 Enhancing Mental and Physical Wellness with Skills. 49 reviewers shared how the Skills they reviewed were used for improving their mental wellness by maintaining stress levels. For instance, R5 shared satisfaction toward the Skill as it affects stress-relieving practices:

> [The Skill is] just what I needed for my overworked, super stressed, extremely tired life" -S9_R5.

Reviewers shared their experiences with practicing mindfulness through Skills. For instance, they shared Skills "provoke positive messages" (S13_R2), "help me relax after work" (S10_R5), "keep me accountable and motivated" (S57_R1), and "give me just what I need to stay productive" (S37_R3). These Skills provided users with podcast-like small talks, spa music playlists, meditation, yoga, and workouts. Reviewers also shared how the Skills provided a good source of white noise and a better environment for concentration. They commented the Skills "can be used to cover all the road noise sounds at night" (S7_R4); "sounds created a wonderful ambience" (S1_R2); "provided relaxation with the reviewer" (S37_R7).

38 reviewers pointed out that the Skills contributed to their enhanced sleep patterns that involved improved duration and quality: "Very tranquil and consoling to help for sleep and relaxation" (S43_R8); "effective in putting our kids to sleep" (S18_R7). Examples of these Skills included 6 Skills that provide simple and quick meditations, 14 Skills that play ambient sounds or music, and 5 Skills offering sleep aids, such as insomnia cure and sleep cycle tracker.

The reviewers expressed that the Skills not only assisted the reviewers in relaxing and being mindful, but also changed their health behaviors. R3 and R7 mentioned how the Skills were helpful in maintaining their physical health:

> It's helpful to hear tips, recipes, etc. while I'm actually preparing and thinking about food! The best thing about [Skill name] and [content expert] is that the focus is on making "better" choices, and not worrying about being perfect all the time! -S8_R3

[The planks exercise Skill] encourages you to make sure you complete daily. -S14_R7

Because Skills provided verbal instructions, R8 pointed out the convenience of following exercises without confusion:

Love being coached through so I don't have to crane my neck to watch a stop watch or mess with one in between positions. Short and sweet but you feel the burn. -S14_R8

4.2.2 Overcoming Time and Spatial Limitations with Skills. Skills helped to overcome time and spatial limitations for reviewers in carrying out a healthy daily routine. More than half of 60 Skills provided a convenient way to obtain health-related goals quickly and effortlessly. Thirteen reviewers talked about being able to be debriefed about news while having their coffee or to learn useful health tips quickly as a way to start the day. Examples included brief health-related information (e.g., "As part of my everyday morning routine I've added [the Skills] and he's the ONE I have to hear as part of my Flash Briefing in order to start my day off on the proper footing" -S13_R1), and exercise strategies, such as yoga, meditations, and cardio instructions (e.g., "We've been doing [cardio] along with a couple other skills in the morning" -S15_R5). R7 commented that she was able to use the health tip Skill at the beginning of the day because of its short, efficient, and useful content:

This is one of my favorite skills added to my morning Flash Briefing. The [health] tips are brief and useful. I would love a little more variety in the tips but overall, I think they're useful. -S66_R7

One reviewer preferred receiving advice from Skills over using screens from smartphones or laptops, which would result in spending more time looking up information:

I've been trying to take the path of healthy living, and this Skill helps answer all my questions without the use of [a major search engine]! Anytime I'm in the kitchen and want a healthier alternative ingredient, [Skill name] gives me options. -S8_R2

Skills also allowed the reviewers to exercise despite their time constraints. These Skills helped reviewers use a short amount of time efficiently without going to the gym. For example, reviewers performed simple workout sessions using a 15-minute exercise Skill whenever they needed to during the middle of the day:

So far I'm happy of this new Skill. I don't have time to go to the gym but the 15-minute workouts suit my mornings well. The intensity of the sessions is tailored to my modest standards too. -S65_R4

Seriously, who can't step back from the desk for 5 minutes for a quick, effective workout? So far, all of the exercise can be done, in a small area. Highly recommended. -S15_R2

As such, reviewers considered the Skills to have contributed to enhancing mental and physical wellness and better integration of work, life, and wellness routines throughout the day.

4.3 What Can Be Improved About Existing Conversational Agents for Health and Wellness

Reviewers evaluated how usable and useful the Skills were depending on their experiences now versus their initial expectations when choosing to activate the Skill. We found that users' experiences and perceived usability differed depending on: (1) whether experiences in using the Skills matched with users' initial expectations of the Skill, (2) the methods used in commercializing the agent, and (3) the quality of agents' instructions and available commands.

4.3.1 Matching Initial User Expectations and Skills Being Transparent about Their Limitations. Ten reviewers commented that the Skills they used provided limited information, unlike their initial expectations. Some reviewers did not want to continue using the Skill as a result. For instance, a Skill was advertised to provide location-specific information, such as air quality reports and environmental information around surfing. When users provided their zip code, the Skill should have provided regional information. However, the Skill did not provide accurate information for people living in non-major cities. R2 expressed the frustration in not being able to get air quality information in non-major regions as advertised, going against their initial expectations:

I've noticed that it is difficult to get NOAA [National Oceanic and Atmospheric Administration] and to provide Air Quality Reports for anything other than major cities. -S87_R2

Another reviewer pointed out the Skill's low ability to be proactive in medication adherence. R5 wanted to have a pill reminder, which could indicate their daily schedule of taking medications. However, the Skill continued to ask for the information from the reviewer while not providing helpful reminders:

I was hoping for something to help remind me when I last took a dose. It just repetitively asked me the medication name, how many times a day I take it, and what time I took it. -S115_R5

Similarly, S109_R8 expressed frustrations with the Skill that tracks babies' bowel movements. The reviewer stated that the Skill was "worthless for the family" because the Skill provided a limited function (e.g., record the time of bowel movement). The reviewer further requested advanced functions, such as types of bowel movement and quantity.

These results show the importance of providing information on what to expect from conversational agents. When reviewers perceived unmet expectations, they experienced frustration, and it lowered user satisfaction.

4.3.2 Methods of Commercialization. Thirteen reviews expressed that including commercial features within the Skills increased negative perceptions. Examples included advertisements being inserted while Skills are being used and charging additional fees for advanced functions along the way (e.g., "[The Skill] wants me to buy a TV channel" -S6_R7). One reviewer commented to skip the

Skill to avoid listening to unnecessary advertisement because (s)he did not want to hear a commercial across the kitchen while drinking morning coffee. (S)he urged the Skill to remove the advertisement:

> The "daily updates" are good, skip the "daily tips" unless you like umbilical cord commercials every day. If they add advertising to the "daily tips" we will stop using the skill all together and come back to leave it one star. REMOVE THE ADS, you RUINED half of a good skill. -S46_R1

Some Skills mandated the reviewers to provide five-star reviews in return for free use. Users did not have a choice to keep using the Skills otherwise. R5 did not see this from the Skill description at the time of activation and complained that they had to write the reviews before they had a chance to further use the Skill:

> I like it a lot, but what I don't like is that this requires a five-star review to continue listening for free. I can't see anywhere in the description that suggests that you can only listen to this for free a few times. All I see is that it's free to enable which implies that this should be free to listen to no matter how many stars you give it in a review or whether you review it at all. -S5_R5

However, not all advertisements were perceived negatively by the reviewers. When R2 had already established trust towards the content provider, the advertisement could result in the product purchase. R2 commented that they purchased new products following the content provider's recommendation:

> I put this on my Skills and I am so glad I did. [The content provider] gives very good information and content that you forget to think about, like taking care of yourself inside and out. She got me into a new natural deodorant from Amazon. She has helped me find new products, try new vitamins and reminding me to keep checking in on myself. -S66_R2

Similarly, reviewers were willing to pay for advanced functions if they were satisfied with the Skill. One reviewer mentioned that (s)he paid the additional cost because of perceived usefulness of the Skill:

> I am going back into my account and signing up for the $5 per month option. -S17_R3

The reviews highlighted that commercializing the agent influenced the reviewer's perception towards the Skills. When reviewers were directly asked to provide high score ratings or activate another Skill on their device, they showed negative reactions. On the other hand, if reliable content providers delivered information, they found it useful and followed the recommendations.

4.3.3 Use Case of Peripheral Devices. Reviewers also showed desires to connect the Skills with peripheral devices to use the Skills' functions better. Because Skills can be used either independently or linked together with peripheral devices, reviewers showed a range of use cases to complete their tasks. One reviewer, for instance, used his wireless Bluetooth speaker to generate more realistic sound effects of a fireplace:

> I play these sounds through a wireless Bluetooth speaker placed directly in front of my gas insert fireplace; sounds very realistic! Now it feels just like a real wood-burning fireplace. -S1_R1

Similarly, the reviewers also mentioned being able to connect the Skills with wireless devices or smartphone apps can help manage personal health tracking (e.g., weight) better. S129 allows users to record health tracking information regularly, but R1 did not find usefulness in doing so, when the weight scale itself can record it:

> I really wanted to use Alexa to manage my weight with the scale. Asking it to add a weight is of no value because I've set up the scale to track my weight… [The Skill] should know that my current weight is, not my goal weight. What's the sense of having this skill if it is so inaccurate -S129_R1

Findings showed reviewers' strong desires to connect the Skills with peripheral devices, such as Bluetooth weight scales, speakers, smartphones, or tablets. However, not many services are provided to support compatibility with such devices.

4.3.4 Quality of Agents' Instructions and Commands. Amazon website provides basic information on how to use Skills, including supported languages, few commands to control the Skill on each Skill page. When users activate the Skill, the page shows resources, such as vendor websites. However, many reviews expressed frustrations about limited information being provided by the developers. Reviewers shared shortcomings of voice-based commands: lack of visual cues and unintuitive forms of command phrases. They also mentioned how characteristics of agents' utterances affect perceived usability of the agent.

Lack of visual cues. In the context of exercise or health information tracking Skills, reviewers expressed the need to use visual cues for learning to correct body postures (e.g., "There is not examples in the app for people to see how to do the plank" -S14_R6). Reviewers compared the Skill with the tablet app which they have used. They mentioned that the lack of visual references makes the Skill less worthwhile (e.g., "An attractive app format on my tablet but the voice entry of weight from Alexa doesn't work" -S129_R2). One reviewer discussed the need to have complementary or alternative ways to have better outcomes from the Skill:

> I doubt anyone who has never done yoga could learn anything without a visual. Go to a class or watch a video first, then use this skill. -S124_R5

The reviewers of medication tracking Skills commented about when visual cues are needed in health management Skills. They described that once their medication intake was recorded, users should be able to visually refer to the previously recorded information to infer about the trends. However, the Skill only recorded the most recent time when the reviewers had medication and did not provide a way to track back (e.g., "Users should be able to either see or hear a list of what it has recorded for that day" -S99_R1).

These results show the importance of how other forms of cues, other than conversational and voice-based cues, should be integrated with the Skills in order to provide wholesome, useful experiences for the users.

Instructions on using commands. Similarly, other reviewers complained about how the product website lacked instructions on using the Skills. Reviewers had a hard time remembering the commands, suggesting the addition of all available commands or advanced functions that could be navigated by the users. For instance, R3 expressed their desire to have better commands that could easily be memorized:

> The first reviewer got it right.... how on earth would a beginner have the necessary vocabulary to request a pose?! Isn't there a cheat sheet for beginners? -S124_R3

While reviewers established better strategies to remember commands by making notes and placing them in easily accessible places, the reviewers commented that commands should be easier. R1 shared their strategy to memorize the Skill commands. (S)he asked to develop shorter commands to keep this Skill on the device:

> We finally wrote it down on a piece of paper and we keep it on our nightstand. Can you change the name or add a shorter command to ask Alexa for this amazing skill? - S112_R1

S122_R2 suggested to have a controlling function over agents' utterances, where users can tell Alexa "to go back one minute or two minutes". Other reviewers also pointed out the need for more control in designing the commands to maneuver through the content:

> I do wish that the sounds were each an individual command to make them quick to start. This would get 5 stars if that was tweaked. That being said, It is worth the extra time to hear this great rain. -S12_R4

> I like [the Skill] but every time I launch this I have to say "Alexa, next" to get to it. Please add the ability to select the sound at the beginning. -S37_R5

These findings showed that voice-controlled commands can be challenging —the commands require users to recall rather than recognize.

Timing, speed, and characteristics of agent utterance. Thirteen reviewers emphasized the importance of Skills estimating an appropriate amount of time that it will take users to complete a task, especially those that require step-by-step instructions. Reviewers expressed that if they encountered a delay or a rush before Alexa played the next content, they had a hard time focusing on the task. Reviewers then asked for clear signs that would indicate the starting and endpoints of the content when they are played in sequences. For example, because there were too many pauses between instructions in exercise Skills, a reviewer lost focus with the exercise:

> Whether it has you do 5 or 20 push-ups it gives you about 20 secs to complete before starting the 'rest' countdown.

> Plenty of time for the day 5 or lower challenges but completely useless beyond that. -S125_R7

On the other hand, if content was played too fast, that also caused decreased satisfaction. R5 felt rushed:

> I liked [the Skill] and I thought it was effective, but it felt a bit rushed. Maybe a little more breathing time between areas, and speak a bit more slowly. Also, at the end a few more minutes of music so you have time to fall asleep. - S116_R5

Furthermore, Skills with a higher level of voice quality and attractive voice were perceived as more useful. If the voice feature, including speed, clarity, and tone did not fit with the Skill content, reviewers found it undesirable. Reviewers expressed Skills that are designed to help users with relaxation, such as sleep aid or yoga Skills, should consider the importance of the voice characteristics. Reviewers expressed perceived limitations of the Skills' content being provided by "robotic voice" (S120_R8) or "Alexa's voice" (S19_R2). They wanted to adjust the voice tone or speed:

> [The content provider] needs to talk slower and more distinctly. Not a fan of the voice but not being clear makes it worse. The voice is really not soothing, you have to work to get used to it. -S116_R3

> [Alexa's] voice has very low volume while music is too loud, really poor mix. - S116_R4

As many Skills applied actual human voice in delivering the content, R1 suggested advanced functions and revisions to the Skill's voice:

> Can the voice speak more slowly or quietly? Not sure what can be controlled by the app but it's worth asking. - S123_R1

As such, time-related design—timing and speed—was important to the reviewers in forming their user experience. Findings suggested that the Skills offered simplicity and convenience as critical benefits over their current health and wellness practices. Moreover, the content provider's attractive characteristics, such as personality and tone of voice, and content expertise improved the users' experiences. Below, we discuss how the findings translate into design implications for conversational agents.

5 Discussion

Capturing 433 reviews of 60 Skills under the Health and Fitness category helped us to understand everyday use scenarios of conversational agent Skills for health and wellness, and what gaps and opportunities exist in improving user experiences.

5.1 Conversational Agents as Personified, Relatable Expert

Users felt the Skill was effective if the agent is made up of content providers whose voices they respect and trust. In the information seeking literature, quality and trust of the information and its source

is important for health-related information seeking [25]. According to our findings, the Skills were often used as a source to find quick health information, such as information on pregnancy, nutrition, or weight management. If the Skills were intended to deliver knowledge, such as information about meditation or nutrition, reviewers' perceived trust of the Skill was a critical piece of criteria for its review. Similarly, research on commercialized conversational agents showed users personify the agents [47]. When we design conversational agents supporting health and wellness, choice and presentation of content providers is a critical factor in users perceiving the Skill as an effective one. Personifying agents gave users the perception of increased access to their favorite content providers as opposed to other forms of media, such as books or TV. The reviewers remembered and called the Skills by the content providers' names and followed them as trustworthy mentors. Beyond personification of agents, sharing the content providers' personal experiences, such as anecdotes about their family relevant to the Skill's purposes, further increased perceived satisfaction and effectiveness of the Skill. In this case, reviewers perceived the content providers to be relatable and friendly mentors who could guide them through their long-term health management. The Skills were no longer perceived as an artificial intelligence system or a machine. Rather, it functioned as a route to encourage connection with the content providers and to obtain reliable information from them.

Previous literature had investigated the effects of trust on conversational agents that were highly anthropomorphized [8]. Presenting honesty [35], making small talk [8,51], and restraining from acting as less encouraging agents [32] were the effective ways to offer simple advice. These findings give lessons to how conversational agents can be designed, especially for health-supporting agents. Agents could be designed with clinician-like characteristics, expertise, and beliefs known to play an effective role for certain health and wellness contexts. For instance, studies show clinicians' gender affects patient trust in the context of outpatient consultation [2]; physicians' beliefs play a critical role in patient decision-making [48]; and studies of Patient Expertise [22] delineated the difference between patient-peer supporters and clinical experts. Peer mentors were strong influencers in end-of-life decision making, for instance [45]. This evidence can be used as cues for designing personified conversational agents that support health and wellness. For agents delivering daily tips for expecting parents, agents could represent content providers or present themselves as those who have raised young children—as peer mentors—and share humorous anecdotes or recommendations on how to navigate new life situations. This will foster a more authentic experience with the agents than those that simply deliver knowledge on what it means to be a good parent.

5.2 Designing Timing and Speed in Conversational Agents

Health and wellness Skills will often require users to follow agents' step-by-step instructions, such as yoga, physical exercises, or healthy recipes. 20 of the 60 Skills we examined were Skills that

prompted users to follow instructions. Accordingly, designing the timing and speed of these health-related instructions and users' responses to them will be crucial. One of the frequent complaints was reviewers not being able to complete a task, especially those that include sequential instructions. Prior studies suggested that temporal concerns, such as fluency or response delays, are one of the major problems that lowered the usability of interactive technology [23,26,29,37,58]. Shiwa et al. also found users expected optimal speed and response time in interacting with user interfaces to be a problem [53]. These points were consistent with Alexa reviewers' opinions on the importance of timing, speed, and pace of the content. When reviewers met delays before Alexa played the next content, reviewers had a hard time focusing on the task because they lost their pace. When Skills advanced to the next steps faster than the reviewers' pace, reviewers felt rushed and lost motivation to follow the remaining content. It is critical to accommodate the varied paces of users in these target activities. The presence of other people during the workout provides better adherence to fitness regimens, a higher level of enjoyment, and an enhanced level of intensity [36]. Because there are beginner exercisers as well as seasoned ones in the same class, finding the best strategies to engage both groups in a session is considered to be critical in group exercises [61]. Choi et al. suggested tactile feedback with sound effects, such as narrations and background music, for encouraging participants in group exercises [16]. While users conduct target activities alone at home, agents could apply these strategies and motivate them not to lose their pace of activity. Even though agents cannot precisely calculate each user's pace, these effects would mitigate individual differences and be effective in reducing user frustration.

5.3 The Design Heuristics for Conversational Agents

Designing conversational agents for health and wellness involves coordinating multiple components. These components include the sources of content, agents' personifying possibilities, the speed, style, and timing of agents' utterances, users' commands, and many more. Below, we present the Design Heuristics for Conversational Agents (DHCA), built on our findings and adapted Nielsen's heuristics [42], one of the foundational principles for usability. DHCA recommends 11 design heuristics for conversational agents. Although not complete, these heuristics can be applied to conversational agents on health and wellness and broader contexts. DHCA is built on the assumption the CA allows the use of verbal dialogues to interact with users with a possibility of adding visual or other cues through peripheral devices, such as tablets or mobile applications.

5.3.1 Visibility of system status refers to "how well the state of the system is conveyed to its users" [42]. Agents should give clear indicators that inform users about the start and end of the content. For instance, for step-by-step instructions, for content such as physical exercises or meditations, keep users informed about when the activity starts, ends, and transitions to the next steps.

5.3.2 Match between system and the real world refers to how "the system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms" [42]. One of the core strengths of conversational agents is their natural interaction with users. The findings showed, however, that commands were often unnatural and difficult to remember. The user commands should consist of easy-to-remember vocabularies that are familiar to users. Timing and speed of the dialogue should also dynamically adapt to users' conversation styles. Although providing natural interaction is critical in all kinds of conversational agents, it is particularly highlighted for agents used on a daily basis, such as everyday health Skills.

5.3.3 User control and freedom refers to "support undo and redo" [42]. For conversational agents, users have little control over maneuvering through various menu items, particularly for Skills reviewed in our study that involve physical and mental activities. Users are forced to listen until the end of the instructions or messages. Providing options that indicate go back, pause, or jump to a specific menu item is critical, especially when the agent is used for activities requiring step-by-step instructions.

5.3.4 Aesthetic and minimalist design refers to how "dialogues should not contain irrelevant or rarely needed information" [42]. Provided content should be simple, minimal, and easily integrated at any time of the day as habitual health behaviors. This was reviewers' primary motivation for using the health and wellness Skills in our study.

5.3.5 Flexibility and efficiency of use refers to "allowing users to tailor frequent actions" [42]. Agents should address tailored and varied levels of user competence in target activities. Allowing users to add tailored shortcuts of their choice, or identifying and calling the most frequently used content from each Skill, as well as across the Skills, can help. For example, if a user wants to create an individualized home training content that is frequently used over others, the system should support it.

5.3.6 Consistency and standards refers to "following platform conventions" [42]. Conversational agents are still in the early stage of adoption. Accordingly, the platform convention might not be available yet. Ensuring consistency across the Skills and following conventions across the agents will help. The findings showed each Skill had different forms of commanding the agent, regardless of their similarity in content.

5.3.7 Recognition rather than recall refers to "minimizing the user's memory load by making objects, actions, and options visible" [42]. Reviewers expressed the difficulties of memorizing commands. Providing available options for commands as shortcuts or commands on a visual interface will help. As health and wellness supporting agents often involve health information and useful wellness tips, providing ways to check recorded health information would be useful. For example, users should be able to either see or hear a list of medications recorded for that day.

5.3.8 Error prevention refers to eliminating error-prone conditions and presenting users with a confirmation option before they commit to the action [42]. If users give wrong commands, error messages should indicate the problem and suggest alternative commands in simple and easy language. If users are in an emergency situation, such as the moment when agents are required to make emergency alarms, agents should be able to recognize users' incomplete commands and notify pre-recorded contacts.

5.3.9 Personified agents. Our findings and empirical research show personifying agents can provide increased accessibility, user satisfaction, and emotional connection [11,12]. A study shows ways to anthropomorphize a virtual agent [18]. Borrowing these strategies, conversational agents should find tailored ways to further personify itself, such as integrating content providers' voice. By using different voice characteristics, users should be able to achieve health-related goals more effectively depending on different contexts and situations (e.g., nighttime relax, morning exercise on weekend, vs. emergency alarm).

5.3.10 Trustful sources of content. Well-designed content providers served as an effective peer mentor in health contexts, such as weight management, in a long-term perspective. Knowledge of content and expertise should be carefully designed and presented using trustworthy sources, particularly for agents delivering health and wellness content.

5.3.11 Tailored utterance speed, tone, and characteristics. Users should be able to accelerate or decelerate the speed of the utterance and choose their preferred voice tone. From the results, these were particularly highlighted in stress management contents, such as sleep aids, meditation, and yoga.

6   Conclusion

Our findings brought implications for design heuristics and recommendations for conversational agents that support health and wellness and beyond. We formalized the 11 heuristics for conversational agents, called the Design Heuristics of Conversational Agents, that future developments can adopt and use to evaluate the usability of the agents and their skills. The study relied on user reviews on Skills of one conversational agent product (Amazon Alexa), limiting the analysis to what was stated about the Skills rather than users' actual behavior. Because we conducted stratified sampling according to users' star rating scores, the data are not representative of the Skills that received only a few reviews. Future study should investigate heuristics based on real use cases and behaviors. Given the increasing interest on designing conversational agents, our study contributes to improving usability and user satisfaction of conversational agents for everyday health management and beyond.

REFERENCES

[1] Amazon. Alexa. Retrieved September 15, 2019 from https://alexa.amazon.com/spa/index.html#skills/?ref-suffix=nav_nav

[2] Amitav Banerjee and Debmitra Sanyal. 2012. Dynamics of doctor–patient relationship: A cross-sectional study on concordance, trust, and patient enablement. *J. Fam. Community Med.* 19, 1 (2012), 12.

[3] Mary Bates. 2019. Health Care Chatbots Are Here to Help. *IEEE Pulse* 10, 3 (2019), 12–14. DOI:https://doi.org/10.1109/MPULS.2019.2911816

[4] Erin Beneteau, Julie A Kientz, Olivia K Richards, Jason Yip, Mingrui Zhang, and Alexis Hiniker. 2019. Communication Breakdowns Between Families and Alexa. (2019). DOI:https://doi.org/10.1145/3290605.3300473

[5] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottrjdge. 2018. Understanding the long-term use of smart speaker assistants. *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.* 2, 3 (2018), 91.

[6] Hugh Beyer and Karen Holtzblatt. 1997. *Contextual design: defining customer-centered systems.* Elsevier.

[7] Timothy W Bickmore, Everlyne Kimani, Ha Trinh, Alexandra Pusateri, Michael K Paasche-Orlow, and Jared W Magnani. 2018. Managing Chronic Conditions with a Smartphone-based Conversational Virtual Agent. In *IVA*, 119–124.

[8] Timothy W Bickmore and Rosalind W Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput. Interact.* 12, 2 (2005), 293–327.

[9] Timothy W Bickmore, Dina Utami, Robin Matsuyama, and Michael K Paasche-Orlow. 2016. Improving access to online health information with conversational agents: a randomized controlled experiment. *J. Med. Internet Res.* 18, 1 (2016), e1.

[10] Holly P Branigan, Martin J Pickering, Jamie Pearson, and Janet F McLean. 2010. Linguistic alignment between people and computers. *J. Pragmat.* 42, 9 (2010), 2355–2368.

[11] Justine Cassell. 2000. More than just another pretty face: Embodied conversational interface agents. *Commun. ACM* 43, 4 (2000), 70–78.

[12] Justine Cassell and Kristinn R Thorisson. 1999. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Appl. Artif. Intell.* 13, 4–5 (1999), 519–538.

[13] CBC. Edmonton Siri with sensitivity: Emotionally intelligent chatbot could ease elder loneliness.

[14] Yi Cheng, Kate Yen, Yeqi Chen, Sijin Chen, and Alexis Hiniker. 2018. Why Doesn't It Work? Voice-Driven Interfaces and Young Children's Communication Repair Strategies. *Idc 2018* (2018), 337–348. DOI:https://doi.org/10.1145/3202185.3202749

[15] Minji Cho, Sang-su Lee, and Kun-Pyo Lee. 2019. Once a Kind Friend is Now a Thing: Understanding How Conversational Agents at Home are Forgotten. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, 1557–1569.

[16] Woohyeok Choi, Joohyun Kim, Jeungmin Oh, Darren Edge, and Uichin Lee. 2015. Designing group fitness swimming exergames: a case study. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, 221–224.

[17] Arlene E Chung, Ashley C Griffin, Dasha Selezneva, and David Gotz. 2018. Health and fitness apps for hands-free voice-activated assistants: content analysis. *JMIR mHealth uHealth* 6, 9 (2018), e174.

[18] Doris M Dehn and Susanne Van Mulken. 2000. The impact of animated interface agents: a review of empirical research. *Int. J. Hum. Comput. Stud.* 52, 1 (2000), 1–22.

[19] Forbes. What Amazon Is Doing To Keep Alexa In The Lead. Retrieved from https://www.forbes.com/sites/andriacheng/2018/07/26/what-amazon-is-doing-to-keep-alexa-in-the-lead/#581f57767fc6

[20] Forbes. 2018. Alexa (And Cortana And Siri) Get Ready For The 9-to-5. Retrieved from https://www.forbes.com/sites/insights-intelai/2018/09/21/alexa-and-cortana-and-siri-get-ready-for-the-9-to-5/#49e752ee1a3c

[21] Yang Gao, Zhengyu Pan, Honghao Wang, and Guanling Chen. 2018. Alexa, My Love: Analyzing Reviews of Amazon Echo. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 372–380.

[22] Andrea Hartzler and Wanda Pratt. 2011. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. *J. Med. Internet Res.* 13, 3 (2011), e62.

[23] Michael Hildebrandt, Alan Dix, and Herbert A. Meyer. 2004. Time design. *Conf. Hum. Factors Comput. Syst. - Proc.* (2004), 1737–1738. DOI:https://doi.org/10.1145/985921.986208

[24] Matthew B Hoy. 2018. Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Med. Ref. Serv. Q.* 37, 1 (2018), 81–88.

[25] Yifeng Hu and S Shyam Sundar. 2010. Effects of online health sources on credibility and behavioral intentions. *Communic. Res.* 37, 1 (2010), 105–132.

[26] Chung Ching Huang and Erik Stolterman. 2011. Temporality in interaction design. *DPPI'11 - Des. Pleasurable Prod. Interfaces, Proc.* c (2011). DOI:https://doi.org/10.1145/2347504.2347572

[27] Jiepu Jiang, Wei Jeng, and Daqing He. 2013. How do users respond to voice input errors?: lexical and phonetic query reformulation in voice search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 143–152.

[28] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, and Annie Y S Lau. 2018. Conversational agents in healthcare: a systematic review. *J. Am. Med. Informatics Assoc.* 25, 9 (2018), 1248–1258.

[29] Lassi A. Liikkanen and Paula Gómez Gómez. 2013. Designing interactive systems for the experience of time. *Proc. 6th Int. Conf. Des. Pleasurable Prod. Interfaces, DPPI 2013* (2013), 146–155. DOI:https://doi.org/10.1145/2513506.2513522

[30] Manja Lohse, Katharina J Rohlfing, Britta Wrede, and Gerhard Sagerer. 2008. "Try something else!"—When users change their discursive behavior in human-robot interaction. In *2008 IEEE International Conference on Robotics and Automation*, 3481–3486.

[31] Silvia B Lovato, Anne Marie Piper, and Ellen A Wartella. 2019. Hey Google, Do Unicorns Exist?: Conversational Agents as a Path to Answers to Children's Questions. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, 301–313.

[32] Gale M Lucas, Nicole Krämer, Clara Peters, Lisa-Sophie Taesch, Johnathan Mell, and Jonathan Gratch. 2018. Effects of Perceived Agency and Message Tone in Responding to a Virtual Personal Trainer. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, 247–254.

[33] Martin H Luerssen and Tim Hawke. 2018. Virtual Agents as a Service: Applications in Healthcare. In *IVA*, 107–112.

[34] Ewa Luger and Abigail Sellen. 2016. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297.

[35] Nikolas Martelaro, Victoria C Nneji, Wendy Ju, and Pamela Hinds. 2016. Tell me more: Designing hri to encourage more trust, disclosure, and companionship. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, 181–188.

[36] Matthew Mauriello, Michael Gubbels, and Jon E Froehlich. 2014. Social fabric fitness: the design and evaluation of wearable E-textile displays to support group running. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2833–2842.

[37] Herbert A. Meyer and Michael Hildebrandt. 2002. Towards time design: Pacing of hypertext navigation by system response times. *Conf. Hum. Factors Comput. Syst. - Proc.* (2002), 824–825.

[38] Adam S Miner, Arnold Milstein, Stephen Schueller, Roshini Hegde, Christina Mangurian, and Eleni Linos. 2016. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Intern. Med.* 176, 5 (2016), 619–625.

[39] Richard Mitev, Markus Miettinen, and Ahmad-Reza Sadeghi. 2019. Alexa Lied to Me: Skill-based Man-in-the-Middle Attacks on Virtual Assistants. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*, 465–478.

[40] Tova Most. 2002. The use of repair strategies by children with and without hearing impairment. *Lang. Speech. Hear. Serv. Sch.* (2002).

[41] Nasim Motalebi and Saeed Abdullah. 2018. Conversational Agents to Provide Couple Therapy for Patients with PTSD. In *Proceedings of the*

12th EAI International Conference on Pervasive Computing Technologies for Healthcare, 347–351.

[42] Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 249–256.

[43] Sharon Oviatt, Jon Bernard, and Gina-Anne Levow. 1998. Linguistic adaptations during spoken and multimodal error resolution. *Lang. Speech* 41, 3–4 (1998), 419–442.

[44] Hannah R M Pelikan and Mathias Broth. 2016. Why that nao?: How humans adapt to a conventional humanoid robot in taking turns-at-talk. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4921–4932.

[45] Erica Perry, June Swartz, Stephanie Brown, Dylan Smith, George Kelly, and Richard Swartz. 2005. Peer mentoring: a culturally sensitive approach to end-of-life planning for long-term dialysis patients. *Am. J. Kidney Dis.* 46, 1 (2005), 111–119.

[46] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In *proceedings of the 2018 CHI conference on human factors in computing systems*, 640.

[47] Amanda Purington, Jessie G Taft, Shruti Sannon, Natalya N Bazarova, and Samuel Hardman Taylor. 2017. Alexa is my new BFF: social roles, user satisfaction, and personification of the amazon echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2853–2859.

[48] Timothy E Quill and Howard Brody. 1996. Physician recommendations and patient autonomy: finding a balance between physician power and patient choice. *Ann. Intern. Med.* 125, 9 (1996), 763–769.

[49] Ashwin Ram, Rohit Prasad, Chandra Khatri, Anu Venkatesh, Raefer Gabriel, Qing Liu, Jeff Nunn, Behnam Hedayatnia, Ming Cheng, and Ashish Nagar. 2018. Conversational ai: The science behind the alexa prize. *arXiv Prepr. arXiv1801.03604* (2018).

[50] Steven R Rick, Aaron Paul Goldberg, and Nadir Weibel. 2019. SleepBot: encouraging sleep hygiene using an intelligent chatbot. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*, 107–108.

[51] Daniel Schulman and Timothy Bickmore. 2009. Persuading users through counseling dialogue with a conversational agent. In *Proceedings of the 4th international conference on persuasive technology*, 25.

[52] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. 2018. "Hey Alexa, what's up?": studies of in-home conversational agent usage. *Proc. Des. Interact. Syst. Conf. 2018 - DIS '18* (2018), 857–868. DOI:https://doi.org/10.1145/3196709.3196772

[53] Toshiyuki Shiwa, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. 2008. How quickly should communication robots respond? *HRI 2008 - Proc. 3rd ACM/IEEE Int. Conf. Human-Robot Interact. Living with Robot.* (2008), 153–160. DOI:https://doi.org/10.1145/1349822.1349843

[54] Anselm Strauss and Juliet Corbin. 1994. Grounded theory methodology. *Handb. Qual. Res.* 17, (1994), 273–285.

[55] Didi Surian, Liliana Laranjo, Rabia Bashir, Enrico Coiera, Adam G Dunn, Blanca Gallego, Annie Y S Lau, Farah Magrabi, Ahmet Baki Kocaballi, Huong Ly Tong, and Jessica Chen. 2018. Conversational agents in healthcare: a systematic review. *J. Am. Med. Informatics Assoc.* 25, 9 (2018), 1248–1258. DOI:https://doi.org/10.1093/jamia/ocy072

[56] The New York Times Magazine. 2018. May A.I. Help You?

[57] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S Kashavan, and John Blake Torous. 2019. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *Can. J. Psychiatry* 64, 7 (2019), 456–464.

[58] Anna Vallgårda, Morten Winther, Nina Mørch, and Edit E. Vizer. 2015. Temporal form in interaction design. *Int. J. Des.* 9, 3 (2015), 1–15.

[59] Joseph Weizenbaum. 1983. Eliza—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 26, 1 (1983), 23–28.

[60] Ryen W White. 2018. Skill discovery in virtual assistants. *Commun. ACM* 61, 11 (2018), 106–113.

[61] Cary H Wing. 2014. The evolution of group fitness: Shaping the history of fitness. *ACSMs. Health Fit. J.* 18, 6 (2014), 5–7.

[62] Maria Klara Wolters, Fiona Kelly, and Jonathan Kilgour. 2016. Designing a spoken dialogue interface to an intelligent cognitive assistant for people with dementia. *Health Informatics J.* 22, 4 (2016), 854–866.