T2 Coach: A Qualitative Study of an Automated Health Coach for Diabetes Self-Management

Elliot G Mitchell
Geisinger Health
Danville, Pennsylvania, USA
emitchell3@geisinger.edu

Pooja Desai
Department of Biomedical Informatics
Columbia University Irving Medical Center
New York, New York, USA
p.desai@columbia.edu

Arlene Smaldone
School of Nursing
Columbia University
New York, New York, USA
College of Dental Medicine
Columbia University
New York, New York, USA
ams130@columbia.edu

Andrea Cassells MPH
Clinical Directors Network
New York, New York, USA
acass@cdnetwork.org

Jonathan N. Tobin PhD
Clinical Directors Network (CDN)
New York, New York, USA
Center for Clinical and Translational Science
The Rockefeller University
New York, New York, USA
jntobin@cdnetwork.org

David Albers
Department of Pediatrics, Section of Informatics and Data Science
University of Colorado Anschutz Medical Center
Aurora, Colorado, USA
david.albers@cuanschutz.edu

Matthew Levine
Eric and Wendy Schmidt Center
Broad Institute of MIT and Harvard
Cambridge, Massachusetts, USA
levinema@broadinstitute.org

Lena Mamykina
Department of Biomedical Informatics
Columbia University
New York, New York, USA
om2196@cumc.columbia.edu

Abstract

Computational intelligence is increasingly common in interactive systems in many domains, including health. Health coaching with conversational agents (CA) can reach wide populations, but the level of computational intelligence needed for a positive coaching experience is unclear. We conducted a study with sixteen individuals with diabetes and prediabetes who used a CA for health coaching, T2 Coach. Qualitative interviews revealed that participants saw T2 Coach as reliable in helping them stay on track with self-management, appreciated the flexibility in choosing personally meaningful goals and engaging on their own terms, and felt it provided encouragement and even compared it favorably with human coaches. However, they also noted that coaching experience could be improved with more fluid conversations, more tailoring to their personal preferences and lifestyles, and more sensitivity to specific contexts, all of which require more computational intelligence. We discuss implications and design directions for more intelligent coaching CA in health.

CCS Concepts

• **Human-centered computing**; • **Ubiquitous and mobile computing**; • **Empirical studies in ubiquitous and mobile computing**;

Keywords

Health, diabetes, self-management, coaching, chatbots, conversational agents, mHealth, self-tracking

ACM Reference Format:
Elliot G Mitchell, Pooja Desai, Arlene Smaldone, Andrea Cassells MPH, Jonathan N. Tobin PhD, David Albers, Matthew Levine, and Lena Mamykina. 2025. T2 Coach: A Qualitative Study of an Automated Health Coach for Diabetes Self-Management. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3706598.3714404

ACM ISBN 979-8-4007-1394-1/2025/04
https://doi.org/10.1145/3706598.3714404

1 Introduction

Advances in Machine Learning (ML) and Artificial Intelligence (AI) are transforming how human beings experience and interact with computing systems in fundamental ways. New intelligent systems powered by data and AI are increasingly used not only in professional domains but in individuals' personal lives. Individuals of all walks of life rely on intelligent systems to plan their leisure

[50], receive recommendations for consumer products [49], and even seek recommendations and advice regarding their health [39].

AI-powered technologies for promoting personal health, and in particular self-management of chronic diseases such as type 2 diabetes (T2D), have the potential of a significant societal impact. The Center for Disease Control (CDC) estimates that 10.5% of the US population has diabetes and 34.5% of the adult US population has prediabetes [14]. T2D carries a high personal cost; individuals living with T2D must engage in proactive self-management and make countless daily decisions—about what to eat and how to be active—to maintain their health and long-term quality of life [10]. Health coaching has emerged as an effective approach to promoting self-management [37, 69, 83, 90]. Coaching aims to cultivate motivation and engagement, and to establish accountability in pursuing achievable health goals in a longitudinal relationship between the coach and their client [68, 70, 78, 82, 90]. Coaching with human coaches, either in-person [37] or mediated with technology [45, 80] was shown effective in supporting chronic disease management and prevention. However, both of these rely on human coaches and there are not enough coaching practitioners to support the growing population living with chronic conditions [28].

Advancements in computing technologies paved the way for fully automated digital health interventions that can complement and at times replace human health practitioners [53, 67]. In particular, Conversational Agents (CA) have shown promising results in many areas of health, including coaching [16, 41, 51]. The vast majority of CA in health proposed thus far used a simple, fully-scripted approach, in which designers outline the entire space of possible conversations in advance [53]. More recent research introduced health CA that rely on AI [75] and, in particular, on Large Language Models (LLMs) that create human-like conversational capabilities unmatched with earlier CA [44]. However, these new AI-based CA have important limitations, as they allow designers less control over their behavior and output and, thus, can provide inappropriate and inaccurate responses [44]. Others argued that health coaching has many important characteristics that are uniquely human and that are unlikely to be replicated with automated coaching systems of any kind [79]. As a result, open questions remain as to which type of CA is best suited for providing a positive coaching experience and, specifically, what degree of computational intelligence is both necessary and sufficient in digital coaching.

In this study, we set to explore these questions in the context of self-management of T2D. Specifically, we aimed to examine what coaching needs can be met with simple conversational agents that do not rely on AI and LLMs, what aspects of coaching experience can be improved with introduction of computational intelligence, and what type of computational intelligence is needed to improve digital coaching. To examine these questions, we conducted a formative study with individuals with T2D using a prototype of a digital coach for T2D, T2 Coach (Figures 1, 2, and 3). We designed T2 Coach in the spirit of technology probes [41], which focus on identifying opportunities for new design features, rather than on eliciting feedback on already completed designs. This approach is particularly compelling in case of complex technologies that may require significant effort and development time. Consistently with this, T2 Coach used a simple frame-based design, common to health CA, in which dialog flows are fully scripted, and individuals choose most of their responses from a list of available options. T2 Coach was designed using iterative user-centered design approach. We first elicited individuals' attitudes and wishes for digital coaching with T2 Coach, and then used their feedback to iteratively refine its design.

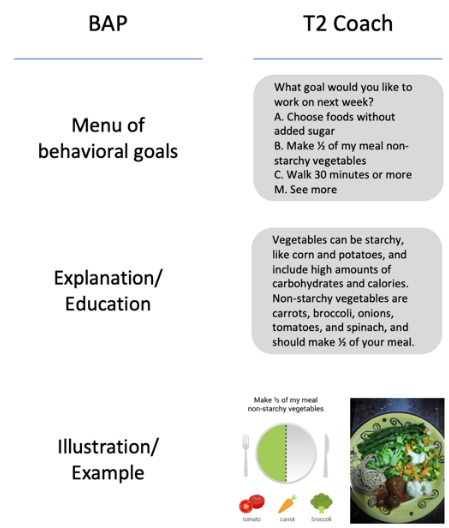

Figure 1: An example a goal setting dialog in T2 Coach following the BAP protocol

T2 Coach uses a set of daily and weekly dialogs, modeled after a clinical protocol for health coaching Brief Action Planning (BAP, [38]), to help users set self-management goals, such as management of nutrition and physical activity, pursue these goals with daily activities, and reflect on goal attainment and barriers to goal attainment over time. All the messages and responses are delivered via text messages thus minimizing technical requirements for users.

We recruited 19 individuals with T2D (n=13) and prediabetes (n=6) to use T2 Coach for 3-4 weeks using a recruitment website hosted by our university and advertisement on social media. Sixteen participants completed the study and took part in post-study qualitative interviews. The participants were predominantly female (63%), ethnically and racially diverse (26% Hispanic and 53% Black or African American), educated (with over 60% of the sample with college and graduate degrees), and overweight (average BMI=29.2).

The study showed that overall, participants had a positive experience with T2 Coach and improved their goal attainment over the course of the study, although it did not translate into self-reported changes in self-management behaviors. Furthermore, many described their experience engaging with the app as working with a coach. There were several factors that contributed to this positive experience. Its system of consistent messages and reminders created a reliable structure that helped individuals to stay on track with their self-management. Its emphasis on choice in selecting personally meaningful goals and action plans and flexibility in engagement helped to create experience that felt personal and promoted autonomy and being in control. Finally, its unassuming messages provided appropriate encouragement and contributed to a

Figure 2: The flow of the daily reflection dialog

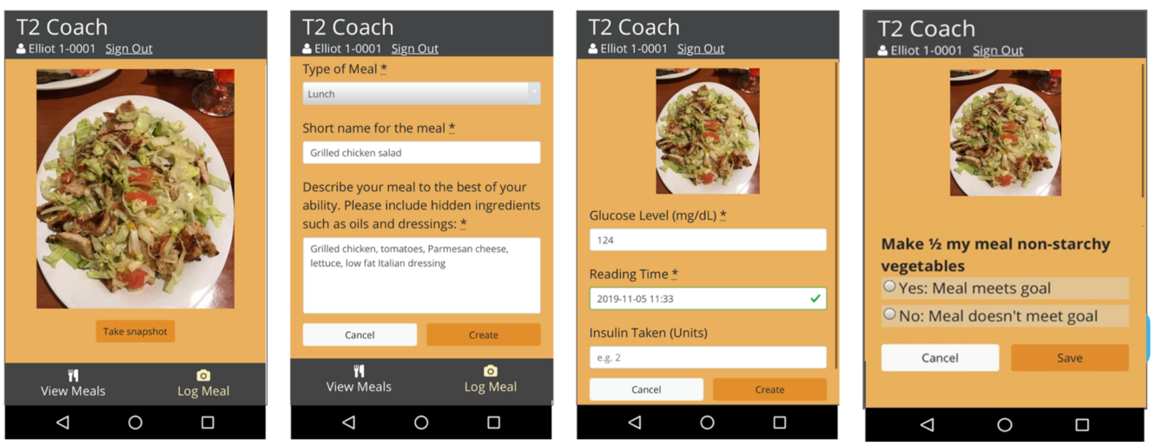

Figure 3: Screens from the progressive web application for recording meals and blood glucose readings.

blame-free experience, critical in self-management of health. These findings suggest that even simple fully scripted digital coaches can indeed create a positive coaching experience.

At the same time, the study highlighted many opportunities to introduce additional computational intelligence into the design of digital coaches. First, personal intelligence can help digital coaches stay grounded in individuals' personal preferences, lifestyles, and circumstances by integrating personal data. Second, situational intelligence can incorporate contextual information to align coaching experience with unique demands of different situations. Third, emotional intelligence can help digital coaches tailor their tone to individuals' emotional state, their success in pursuing their goals, and their need for emotional support. Finally, conversational intelligence can enable more fluid and natural conversational flows and help participants engage in richer conversations about a wide range of their experiences. These different types of intelligence require different computational solutions. For example, contemporary LLM-based chatbots are particularly well-suited for promoting conversational intelligence. Furthermore, past research in affective computing [72], personal informatics [25, 55], and context-aware computing [22] offers insights on supporting other dimensions of computational intelligence described above. However, the study highlighted the need to integrate different types of computational intelligence in a coherent coaching experience, and to balance it with user autonomy and their sense of control over their experience.

2 Related work

2.1 Technologies for personal health

The increasing availability of data related to individuals' personal histories inspired a new wave of research in Human-Computer Interaction (HCI) with the focus on facilitating individuals' engagement with personal data [25, 26, 42, 46, 55, 57, 62]. Li et al. proposed a stage-based model of personal informatics that outlined several stages of individuals' engagement with personal data [55]. Other researchers expanded on this model and discussed how individuals incorporate personal data into their daily lives [25]. Many personal informatics tools use visual representation of data captured with self-monitoring to help individuals identify patterns in their records (e.g. [2, 18, 56]). For example, Medynskiy and Mynatt's system Salud! allowed its users to specify activities with suspected impact on wellbeing and related biomarkers (for example time spent in the office and mood) and visually examined possible dependencies in these data using interactive visualizations [64]. Epstein et al. developed visualizations of variables related to specific questions posed by individual users based on their self-tracking goals [27]. Similarly, SleepExplorer by Liang et al. visualized patterns of sleep in conjunction with other data tracked by participants [58]. Somewhat in contrast to these visual approaches, Bentley et al. used simple statistical inferences to computationally identify correlations in the captured data and present their users with discoveries using language [5]. Self-experimentation framework proposed by Karkar et al. helps individuals to evaluate hypotheses in regards to the associations between specific behaviors and health outcomes [47]. However, most of these approaches focus on recognizing trends and leave it up to the users to identify mitigating strategies, which can present considerable barriers for low literacy communities [86]. Recent work in personal informatics advocated for solutions that rely on computational data analysis to identify notable trends and provide more direct support for action using predictions and recommendations [40]. However, only a few previous authors explicitly examined ways to more directly impact individuals' actions with computational inferences [20, 21].

2.2 Health coaching

Self-management of chronic conditions like T2D often necessitates modifications to daily behaviors, including diet, exercise, and sleep, as well as adherence to medication to improve glycemic control and avoid debilitating complications [10]. Behavior change is a challenging undertaking as it requires continuous motivation, skills and knowledge, and self-efficacy [73, 74]. Health coaching has been successful in promoting healthy changes to individuals' lifestyles [37]. While multiple definitions of coaching exist in literature, most prior research describes coaching as centered around achievement of personal goals: the coach and the client collaborate on identifying health goals and making concrete plans for achieving these goals. Coaching literature emphasizes a personal relationship between a coach and a client, feedback, motivation, and acquisition of skills and knowledge as part of the coaching process [69]. Furthermore, coaching places particular emphasis on promoting and supporting the client's empowerment, autonomy, and increase in their competence, rather than strict adherence to prescribed behaviors [38, 65].

While there is substantial evidence as to the benefits of human coaching interventions, they are limited in their ability to reach wide and diverse populations due to their reliance on trained human coaches. Technology-mediated coaching has broader reach; however, these interventions may still not be available to individuals in medically underserved and economically disadvantaged communities with already limited access to health practitioners and resources [24, 76, 81]. Increasing proliferation of CA and chatbots creates new opportunities for more autonomous coaching interventions that do not rely on human coaches. However, recent research argued that human skills are a critical component of successful health coaching. For example, Rutjes argued that relationship building, an ability to adapt to different contextual factors, and empathy are critical components of health coaching that are unlikely to be replicated by contemporary computing technologies [79]. Others have argued that both human and automated coaching have their respective strengths and limitations [66], but open questions remain as to the feasibility of fully automated coaching interventions in health, and whether these systems can create a positive coaching experience without involvement of a human coach.

2.3 Conversational agents in health

Conversational agents are software programs that interact with users in natural language. This includes spoken word, written or typed words, or a combination. Early CA used rules to respond based on user input, creating the illusion of human intelligence [89]. Since then, CA have been deployed to a wide array of use cases. Some are task-based, while others emulate open-ended social chit-chat.

McTear [63] described three main types of CA. Finite state or rule-based systems follow a predefined set of rules based on user input and the current state of the system. In frame or template-based systems, the agent has a template for each task it can accomplish, with a set of entities or slots that must be filled to complete the task [29, 53]. The third type, agent-based or AI agents, use ML to learn appropriate responses using large corpora of dialogue; while these agents can be powerful and more closely resemble interaction between humans, extensive corpora needed for training may not exist in many domains [34, 60, 60]. Advances in LLMs paved a way to a new generation of CA that are not bounded by constraints of specific domains; given that LLMs are trained on vast amounts of written discourse, LLM-based chatbots can converse on a broad variety of topics.

While the broader proliferation of CA is relatively recent, there is robust body of literature in HCI and related communities examining applicability of conversational agents in health. Laranjo and colleagues surveyed the use of conversational agents in healthcare [53]. The majority of applications included in their review were focused on mental health, with a sampling of other health areas like asthma and nutrition. Many agents implemented a specific clinical protocol like Cognitive Behavioral Therapy [29] or Brief Motivational Interventions [59]. None of the agents identified were AI-based; all were either rule-based or frame-based. The continued focus on rule-based and scripted agents is partly because of a low tolerance for error in the health domain. In addition, because of data security and privacy protections, health-related data sets are rarely made openly available for researcher use; therefore there is a lack of publicly available dialog corpora in health domains [60]. There are, however, emerging examples of AI-based chatbots in health, and, specifically, in diabetes self-management. For example, Gong et al. introduced Laura, an embodied mobile coach that delivered tailored diabetes education and was able to answer questions and engage in conversations with participants [35]. Similarly, Alloatti et al. introduced AIDA, a CA for providing therapeutic education for individuals with diabetes [1]. However, both of these previous studies focused more on education than on the setting and pursuing of personal goals. Many studies that examined individuals' experiences with conversational agents highlighted their benefits. Common reactions to the agent included a sense of accountability, feeling of empathy from and toward the agent, as well identifying a personality in the agent building a relationship with the agent [12, 29, 30]. Bickmore and colleagues demonstrated that conversational agents can help explain discharge paperwork to low literacy hospital patients [7] and successfully coach older adults to be more active [8]. CA can also facilitate sensitive discussion and planning around medical issues, such as in palliative care contexts [87].

At the same time, recent research identified several challenges with CA in health, particularly in relation to the new LLM-based CA. Jo et. al. studied use experiences with LLM-based chatbot used for providing emotional support to individuals living alone [44]. While overall the study highlighted many positive experiences for both individuals who received this service and healthcare providers who were monitoring these interactions, they also noted difficulties in ensuring appropriate responses from somewhat unpredictable CA and potential misalignment of these CA and needs and expectations of healthcare professionals.

3 Methods

3.1 T2 Coach

3.1.1 Iterative design. We designed T2 Coach using iterative user-centered design methods that involved focus groups with individuals with diabetes recruited from racially, ethnically, and economically diverse communities.

To this end, we conducted a series of focus groups with 23 individuals with T2D recruited from Federally Qualified Health Centers (FQHCs), and 6 diabetes care professionals at those centers. During the initial focus groups, we discussed the idea of an automated chatbot coach and collected users' perceptions of its desired behavior in different situations. After these initial focus groups, we sought an appropriate clinical standard for coaching that would be consistent with users' needs, which led us to identify BAP as coaching protocol. We then used BAP to structure the initial set of coaching dialogs and used additional focus groups to gather feedback on the perceived clarity, desirability, and utility of goals and action plans available in T2 Coach and on the appropriateness of its dialogs. Concurrent with the focus groups, we completed a two-week wizard-of-oz trial of T2 Coach to refine the conversational flows before fully developing the agent. Through feedback from these design activities, we found that, overall, participants found goal-driven coaching to be potentially useful, and they found goals and dialogs to be appropriate and easy to follow. However, participants were not always willing to continue very long dialog flows, and a large content-base was necessary to ensure diversity of goals and action plans to accommodate different preferences and lifestyles. We therefore shortened the overall length of dialogs, edited individual messages for clarity, and expanded the content base of available goals from 9 to 17 and available action plans from 3-6 to 6-9 per goal.

3.1.2 T2 Coach Features. T2 Coach included two main components, a CA that provided goal-oriented coaching and a mobile app for tracking individuals' meals and BG levels.

In the design of T2 Coach CA, we followed an established protocol, Brief Action Planning (BAP; [38]) as the basis for the scripted dialog flows. BAP defines a set of steps for health practitioners to guide an individual towards choosing a health goal and making a specific plan to achieve it. The content for goals and action plans in T2 Coach were derived from a prior knowledge base of health goals for individuals with T2D.

Consistent with BAP, T2 Coach included two primary dialogs: a longer, weekly exchange to set a health goal, (a goal-setting dialog, Figure 2) as well as a shorter, daily, follow-up exchange to check in on goal progress (daily goal reflection, Figure 3). The longer, goal-setting dialogs included the following components: 1) a menu of behavioral goals (in an order randomized for every goal setting dialog), 2) a brief explanation of the selected goal; these explanations were written by a team of registered dieticians (RD) and evaluated with users on ease of understanding and cultural appropriateness in a separate study, 3) an illustration of the goal using an infographic, 4) an example of a meal that meets the selected goal (chosen from a collection of meals recorded by participants of our prior studies), 5) a menu of action plans-specific actions that could help to meet the selected goal, 6) an option to set a goal reminder (users could choose whether to receive reminders or not, but could not change their time), and 7) a confirmation of their selections (Figure 1). These dialogs could be longer or shorter depending on how many goals and action plans individuals chose to review before finalizing their selections.

The shorter, daily goal reflection dialogs included the following components: 1) assessment of goal attainment, 2a) positive reinforcement and reconfirming daily action plan, or 2b) reflection on barriers and plan for overcoming barriers in the future (Figure 2). These last two options were included to encourage users to engage in reflection and proactive planning. In addition to these dialogs, T2 Coach sent simple non-interactive once per day messages that included reminders of selected goals accompanied by short motivational messages adapted from previous interventions to facilitate self-management of T2D [17].

The T2 Coach mobile app asked individuals to log meals by capturing photographs and textual descriptions and their blood glucose levels captured before and after meals (Figure 2). To promote in-the-moment reflection on goal attainment (and to track goal attainment overtime), those who chose nutritional goals received prompts to assess whether newly captured meals fit their selected nutritional goal at the time of logging meals. Individuals could view logs of their captured meals and corresponding BG levels in their T2 Coach log.

3.2 Participants

Participants for the study were recruited via our university's recruitment website that targets the general public as well as advertisement on social media and Craigslist. To be eligible to participate, individuals needed to be between 18 and 65 years of age, proficient in English, self-report a diagnosis of either T2D or pre-diabetes, own a functioning smartphone with a data plan, and be able to download, install, and use smartphone applications. The exclusion criteria were self-described major co-morbid illness or injury (e.g. AIDS, Cancer), and cognitive impairment. The study was approved by our university IRB; all participants signed e-consent before participating.

3.3 Study procedures

All study procedures were conducted remotely. Prior to the initial training session, participants reviewed and signed e-consent and filled out a set of baseline questionnaires. During the virtual training session conducted over the phone or Zoom, participants reviewed a pre-recorded video explaining basic principles of diabetes self-management and goal-oriented coaching (prepared by a diabetes educator on our study team), and a video explaining features of T2 Coach. After that, the participants were explained that they will have a chance to work with a digital fully-automated coach in the form of a text-messaging chatbot. The participants were told that they will have a chance to set their self-management goals once per week and choose specific actions to meet these goals every day. The participants were encouraged to use T2 Coach as much as they found convenient without setting any particular expectations for a minimal amount of engagement. Then, the participants were asked to respond to the first set of coaching dialogs, including a goal setting dialog, and use a mobile app to record a meal while a member of the research team was present to answer their questions and provide assistance. After the training, participants were asked to use T2 Coach for 3-4 weeks. If participants had not used the application within 3 days after training, researchers contacted participants to resolve any potential technical difficulties.

During the last week of the study, participants were invited for qualitative semi-structured interviews conducted over the phone or zoom about their experiences. The interview followed an interview guide that included questions about individuals overall self-management, their experience with different features of T2 Coach and their perceptions of their coaching experience with T2 Coach (the full interview guide is included in the Appendix). Here, we paid particular attention to gaps and limitations in their experience that could suggest opportunities for a more intelligent design. In addition, they were asked to fill out post-study surveys. All interviews were audio recorded and transcribed verbatim for analysis.

Study measures included several distinct outcomes. The health outcomes included diabetes self-efficacy (Diabetes Self-Efficacy, DSE, [9]), self-reported attainment of self-management goals set during the study, and self-reported self-management activities, including healthy eating, physical activity, BG monitoring, and several others (Summary of Diabetes Self-Care Activities (SDSCA),[85]). To assess participants perceptions of the quality of conversations with T2 Coach, we used Subjective Assessment of Speech Systems Interfaces (SASSI)[39]. Finally, to assess individuals' assessment of the coaching experience with T2 Coach we focused on shared decision-making, an essential component of coaching (Adapted Shared Decision-Making Questionnaire (SDM-Q-9),[52]).

3.4 Data analysis

For the quantitative data, we calculated descriptive statistics of application usage, survey measures, and goal attainment rates during different weeks of the study. Considering interactions with the chatbot, we calculated the average length of conversations, how often users replied to the daily messages they received, how long it took them to respond, as well as how often they finished the conversations they started all the way to the end. We also calculated descriptive statistics of the number of goals and action plans selected per user and overall. Considering the self-tracking mobile app, we summarized the number of meals and BG readings recorded during the study period. For post-only survey measures, we report summary scores for each scale and subscale. For the pre-post- study comparison of self-management with the SDSCA, we calculated a paired samples t-test for the relevant subscales: general diet, exercise, and blood glucose testing, applying a Bonferroni correction for multiple tests. For goal attainment, a previous study by Mitchell et al. showed that individuals' self-assessment of goal attainment over time was more optimistic than expert assessment; however, individuals and experts were consistent in their assessment of change in goal attainment over time [31]. Consistent with this, we used individuals' self-assessment in this analysis. We used univariate linear regression between days since selecting goal and average goal attainment per participant.

For qualitative data, we used thematic analysis to identify major themes in participants interviews [11]. First, four of the authors reviewed 3 transcripts together, open-coded the transcripts in several collaborative coding sessions and began identifying emerging themes. The authors discussed the labels and their meaning and used notes to track these discussions. All disagreements were reconciled in discussions. These sessions led to the initial coding scheme. After that, one of the authors coded the rest of the interviews. During this time, the authors met weekly to discuss new findings and changes to the coding scheme. New interviews were transcribed and coded as they occurred. Data saturation was reached after 15 interviews; since no new themes were identified, we stopped recruitment. To increase rigor in analysis, we used triangulation and cross-referenced findings from qualitative analysis with results of usage log analysis. We used nVivo by Lumivero for qualitative coding (NVivo 14.23.0 / 14 March 2023).

4 Results

4.1 Participant demographics

Of 19 participants recruited for this study; 16 completed the study and took part in qualitative interviews. Of the 3 who did not complete the study, one withdrew the day after enrollment, and two did not respond to the invitation for the post-study interview. Participant demographic information is included in Table 1 below. Overall, the participants were predominantly female, ethnically and racially diverse, educated (with 64% of the sample with college and graduate degrees), and on the border between overweight and obese (with BMI over 25 and close to 30).

Table 1: Participant demographics.

Demographics	Value
Sex	63% Female; 37% Male
Ethnicity	26% Hispanic
Race	53% Black or African American
	32% White
	15% Additional Race
Age	45 ± 20 years (median 47 years)
Education Level	32% Graduate degree
	32% College graduate (bachelor's degree)
	32% Some college or technical school
	4% High school graduate
Body Mass Index (BMI)	29.2 ± 15
Median Household Income	$40,000 - $59,999
Diabetes diagnosis	68% Type 2 Diabetes
	32% Prediabetes

Table 2: Overall usage statistics

Usage statistic	Weekly goal setting	Daily reflections
Response rate	73.2%	71.2%
Percent of started conversations that were completed	86.5%	89.43%
Mean number of conversational turns	9.9	4.1
Time to first response	2.3 hours	2.1 hours

4.2 Usage statistics

Overall usage statistics are presented in Table 2. Analysis of usage logs revealed overall high response rates for both weekly goal setting dialogs and daily reflection dialogs (over 70% for each). Not surprisingly, weekly goal setting dialogs were considerably longer with almost 10 conversational turns on average, while daily reflection dialogs were quite a bit shorter with just over 4 turns per conversation.

During this study, each of the 17 health goals was selected by at least one participant. Participants took some time to explore available goals before committing to one: on average they explored 5 different goals in each goal-setting dialog. They set on average 2 goals while in the study and chose 1.5 action plans for each goal with the total of 3.5 action plans tried while in the study. They recorded just over 54 meals (2.5 meals per day) and those who tracked their BG (excluding participants with prediabetes) recorded 72.5 BG readings (3.5 per day), which suggests that few captured meals included both pre- and post-meal BG levels needed for personalized goals.

4.3 Goal attainment

One of the main aims of health coaching is to help individuals set and pursue health goals. Here, we discuss our analysis of goal attainment.

Across all participants and study weeks, the average self-reported goal achievement was 80% during the study period. Participants pursued each selected goal for an average of 1-2 weeks (mean of 10 days). Fifteen participants recorded at least 1 goal self-assessment during all four weeks of the study, and those 15 participants were included in this analysis (excluding 2 participants who only had assessment data in the first week of the study).

All participants set their first goal during their first day in the study, and all continued working on their selected goal for at least the next 5 days. As one can see from Figure 4, overall, the participants' increased their level of goal attainment during those 5 days; this trend approached statistical significance (R^2 = 0.044, F(1,75) = 3.473, p = 0.066 and β =0.038, meaning that for each day since selecting a goal there was a trend of 3.8 percentage points increase in goal attainment).

After the first week, analysis of goal attainment overtime becomes less straightforward because different participants kept working on goals for different periods of times and switched goals at different timepoints in the study. Figure 5 (top) depicts their self-reported goal attainment over all 4 weeks of the study. Analysis of self-reported success achieving goals over time showed a trend of improvement during the study period, though this trend was not statistically significant. Notably, there was a drop in goal achievement in study week 3, when many participants switched to new goals.

This trend was consistent for both dietary goals and physical activity goals, with a drop in goal attainment at the time when individuals selected new goals (week 3), and gradual increase in goal attainment overtime (Figure 5, bottom).

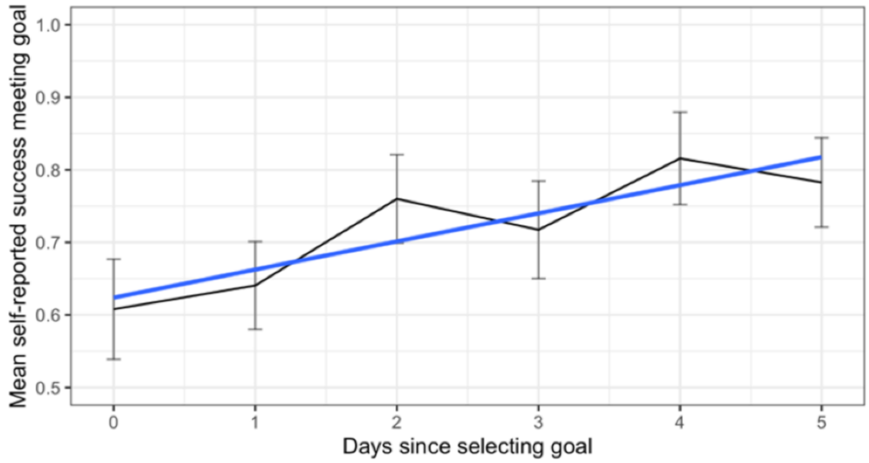

Figure 4: Average goal attainment during the first week of the study with a regression line. The mean on the Y axis ranges from 0 (none of the meals met the goal) to 1 (all of the meals met the goal).

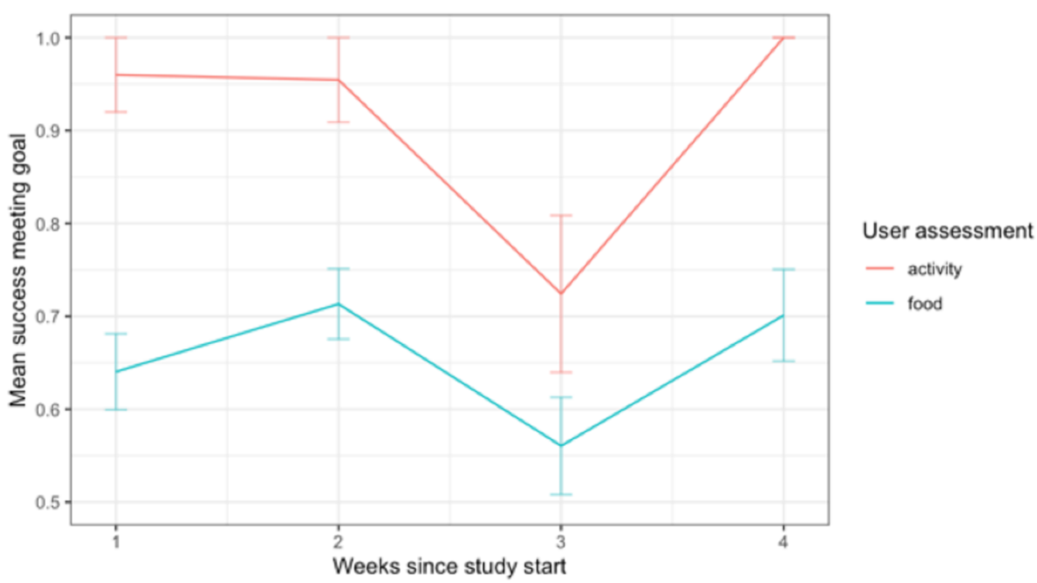

Figure 5: Average goal attainment for different study weeks with a breakdown for nutritional (food) and physical activity (activity) goals. The mean on the Y axis ranges from 0 (none of the meals met the goal) to 1 (all of the meals met the goal).

4.4 Survey measures

On the ***shared decision-making scale,*** T2 Coach received a high rating with the average median score of 5 across this 7-question scale. Participants provided higher ranking for questions that concerned their autonomy and freedom of choice (e.g., T2 Coach wanted to know exactly how I wanted to be involved in choosing a health goal), and lower ranking for questions that concerned rationale for selecting goals and their benefits (e.g., T2 Coach precisely explained the advantages and disadvantages of the health goal choices).

On the user experience ***SASSI scale***, T2 Coach received an overall median score of 5 out of 5 (overall mean 4.03 ± 0.45), where higher scores indicate a more positive assessment. Median scores for different sub-scales are included in Table 3. As one can see, participants rated T2 Coach highly on the system response accuracy, likeability, cognitive demand, and speed, and lower on annoyance and habitability.

The results of individuals' responses to the ***SDSCA*** scale and its subscales are included in Table 4. Participants reported more frequently practicing self-management behaviors related to diet, exercise, and BG testing, however these differences were not statistically significant.

Table 3: SASSI sub-scales

SASSI sub-scale	Median	Mean
System Response Accuracy	5	4.06 ± 0.4
Likeability	5	4.28 ± 0.27
Cognitive Demand	5	4.23 ± 0.26
Annoyance	4	3.48 ± 0.66
Habitability	4	3.8 ± 0.14
Speed	5	4.13 ± 0.19

4.5 Qualitative results

In this section, we first provide some details on participants' background including both their lived experiences and experiences with diabetes self-management. Then we present the results of the thematic analysis, which identified the following main themes: *1) Keeping on Track; 2) Individual differences and the need for personalization, 3) Fitting into the rhythm of daily life, 4) Appropriate encouragement, 5) Conversational fluidity, and 6) Autonomy and being in control*. Below, we describe these themes and illustrate them with quotes. We attribute quotes using participant number assigned during recruitment; as a result, these numbers exceed the actual number of participants in the study.

4.5.1 Participant background. Study participants had a wide range of backgrounds both generally and regarding diabetes self-management. Some of these differences were due to the different stages of diabetes development. For example, those with prediabetes had some experience with tracking their BG; however, most of them relied on lab values captured every 3 to 6 months, rather than daily BG levels to track their progress. Those with a diagnosis of T2D varied in their approaches to tracking BG as well. Some tracked their BG with continuous BG monitoring (CGM), while others relied on daily or even weekly checks. Most relied on either oral medications or insulin or some combination of the two. Many had well-established self-management routines, which included strict adherence to a diet and regular exercise routine. However, most talked about many challenges to self-management, spanning from environmental barriers to engaging in healthy behaviors, to lack of self-discipline, to lack of knowledge about diabetes and nutrition, among many others.

Regarding their experiences with technology, all participants had smartphones and most relied on text messaging as their preferred media for communication. In fact, most participants preferred text messages to both phone conversations and messaging apps, such as WhatsApp and Facebook messenger.

4.5.2 Theme 1: Keeping on Track. All participants came to this study with at least some experience managing their diabetes (or prediabetes) and were familiar with the many barriers to consistent self-management. Some of these barriers were internal, such as lack of motivation, struggles with cravings, and lack of knowledge as to the appropriate changes to their lifestyles. Other barriers were external and were related to lack of access to healthful foods and time and resources for cooking and preparing healthful meals. Given these challenges, participants spoke of the need to stay engaged and motivated, and to persevere with adopting and maintaining new behaviors and with overcoming barriers. In this context, participants found that T2 Coach, with its set of consistent daily messages, created a reliable persistent structure to help them "keep on track".

Participants discussed several components of this structure that they found particularly helpful. For example, the ***daily goal reminders***, which participants typically received in the morning, were found useful in setting the mood for the day and in served as orientation for daily activities:

> "The reminders in the morning. Cause that's how it starts your day. You get up and you look at it and say, oh yeah, let me make sure that's what I do for the rest of the day." (P1, 51, female, T2D)

Daily reflections and goal attainment checks helped to create a sense of accountability. Many participants used the phrase "keeping me honest" when describing their experience with T2 Coach, although the sources of this accountability differed between them. For some, it was *internal accountability,* which encouraged participants to be honest with themselves. For others, like P6, the chatbot created a sense of *external accountability* and the feeling that somebody was looking over them.

> "I knew it [T2 Coach] was going to ask me, and I didn't want to be in a position to say no I didn't." (P6, 48, female, T2D)

Weekly goal attainment summaries showed participants that they could meet their goals and encouraged them to persevere with their goal attainment:

> "Well, I would say it was encouraging. It means like, okay, fine, I was able to do it this week. And I'm sure that this is something that I could keep doing, if I stick with the goals that I made. So, it was a positive side of encouragement, saying, yes, you can do it." (P24, 52, male, T2D)

Most participants were appreciative of this consistency in messaging and did not feel burdened or overwhelmed by it. Even those that were somewhat annoyed by the frequency of messages, felt that it was necessary to achieve desired changes. For example, P4 compared T2 Coach with a parent, at times nagging, but effective in motivating good behaviors:

> "Well, I'll tell you, it felt like T2 Coach was my parents. Reminded me to eat well, which is a good thing, because if left to my own devices, I might cheat two days out of the week. But T2 Coach really kept me honest. It could say, okay, wait, you didn't meet [your goal] today, so you definitely got to meet it tomorrow and the rest of the week. So that was how I motivated myself." (P4, 58, female, T2D)

Many participants compared this consistency favorably with their previous experience with human coaches, which typically took the form of infrequent consultations with little follow-up, and was often ineffective in inspiring the desired changes. By contrast, T2 Coach was perceived as reliable, punctual, and diligent:

> "Well, the app is, you know, much more punctual and diligent. So, you know, it's on automatic and it's just up to me to respond to it in a way that takes it seriously. Whereas the primary care physician, you

Table 4: Comparison of pre- and post-study measured Summary of Diabetes Self-Care Behaviors (SDSCA), measured in days of the last week (0-7)

SASSI sub-scale	Pre-study	Post-study	Difference (p-value)
General Diet	4.4	5.2	0.8 (0.26)
Exercise	3.6	4.4	0.8 (0.26)
Blood Glucose Testing	3.0	4.7	1.7 (0.19)

> only see once a year. And when she says you need to do this or that, or you have to try these veggies instead of rice or something, when she's telling you that, you know, it's kind of like, oh yeah, okay. This is my doctor saying, you know, try to do that. Of course, I'm not that good at sticking with it" (P3, 43, male, T2D).

4.5.3 Theme 2: Individual differences and the need for personalization. While the participants appreciated the reassuring stability of the consistent daily messages, they also found that it needed to accommodate their diverse daily routines and habits. The participants varied greatly in their lifestyles, their cultural identities, their preferences for foods and in their life circumstances. Some participants had relatively stable daily schedules that could easily accommodate traditional three meals per day, while others worked night shifts and had to adjust their eating schedules to their somewhat uncommon work practices. Some were fully in charge of their own meal choices, yet others relied on their parents of partners for grocery shopping and cooking. All these differences created a rich background for their diabetes self-management practices and for their engagement with T2 Coach.

For example, participants had vastly different approaches to selection of goals and action plans. Some chose goals that were consistent with existing habits and could be incorporated into their existing routines with little effort. Others, however, chose goals that were in direct contrast to their current behaviors and required considerable effort. In many cases, these goals were aligned with individuals' aspirations for self-management, and they saw the study as an opportunity to finally make progress.

> "And I guess, like the guilt thing that I carry around, this is all I know I should eat more fruits and vegetables. It brought it into the forefront of my mind and made me actually take the step of doing it." (P20, 38 female, T2D)

In this context, participants wished that T2 Coach was more aware of their personal preferences, lifestyles, and routines and was able to accommodate those to create a more ***personalized*** experience. To some degree, this sense of personal experience was supported by the rich selection of goals and action plans offered by T2 Coach and the ability to change action plans as frequently as they wished to allow for the flexibility their desired. This flexibility created at least a perception of personalized experience, even in lieu of any actual data-driven tailoring.

> ". . . the fact that it's everything that you chose . . . definitely made it feel personal." (P18, 53, female, T2D).

However, participants felt that this experience could be greatly enriched by more extensive utilization of their data in terms of both ***tailoring*** the selection of goals and action plans to their lifestyles, and providing more tailored ***feedback*** on the activities they captured, such as their meals, steps, and BG levels.

> "You log your meal in, and I would have liked to see T2 give me maybe feedback of if I'm like eating too many calories a day or what a healthy plate should look like." (P6, 48, female, T2D)

4.5.4 Theme 3: Fitting into the rhythm of daily life. Beyond tailoring to individual lifestyles and preferences, participants often talked about unique needs of different situations and contexts. This was particularly the case for participants with unusual schedules, which required adjustments to all their daily activities. For example, P18 was a substitute teacher, whose schedule was quite consistent during the school year, but became less structured during the summer, when the study was conducted. For them, having to log meals and keep track of BG levels during summer months presented a challenge that would not have occurred had the study happen during the academic year.

> "It's that I have such a not regimented life, especially right now, because I work in a school. I'm a substitute teacher and I'm off. And so like, in September, I'm so much more regimented than I happened to be in June and July. So, I didn't do well with tracking, even though I believe in tracking, which is what attracted me to want to do this study because I want somebody to make me track." (P18, 53, female, T2D)

Even those with more regular lives felt that the app was not well adapted to handling deviations from their normal activities and routines. For example, P2 had a colonoscopy on one of the days while they were in the study and were not able to follow their regular routine in preparation for the procedure.

> "I'm in a hospital for two days, one day was colonoscopy prep. So those days were weird. So I guess it's also assuming that every day is the same, and then holidays are like normal days. [I wish] there was a way to stay, this is a weird day. What can we come up with for this day?" (P2, 50, male, T2D)

This experience was particularly frustrating given that P2 described colonoscopy as a barrier in one of their daily reflection messages. These disruptions in routines sometimes made the app with its consistently scheduled messages feel misaligned with individuals' rhythms of life.

4.5.5 *Theme 4: Appropriate encouragement.*

Throughout their interviews, participants spoke of the multiple challenges and barriers related to diabetes self-management and the importance of continuing motivation and persistence. Participants talked about the importance of their social networks and encouragement from their friends, families, and healthcare providers in helping them stay positive and motivated. In this context, participants described their interactions with T2 Coach as positive and encouraging. They found that the language used by the chatbot in both goal setting and daily reflection dialogs was able to strike the balance between being negative and guilt-inducing, and being overly positive:

> "Yeah. It felt encouraging. It didn't feel like punitive or shaming or yeah, negative really. It felt like, "all right. Yeah. Good. Try it again. You got another day tomorrow. You can do it." So, I found it just the right amount of encouraging." (P2, 50, male, T2D)

All participants were aware that T2 Coach is fully automated; yet, many perceived lack of human qualities in T2 Coach to be an advantage, rather than a limitation. For example, P2 appreciated the ability to take dialogs at their face value without having to worry about hidden meaning, common to communication among humans:

> "T2 Coach was helpful. It wasn't condescending, and it wasn't too challenging. So that was good. And since it was, I don't know, neutral, it seemed easier to take suggestions. I didn't mind, plus I know that the T2 Coach doesn't have – there's no subtext. So, if it was a human, perhaps saying somethings, I would be like, "oh, are you throwing shade? What do you mean that was tough and good – I did a good job, or I didn't do a good job? But just taking things at face value from T2 Coach made it easier." (P2, 50, male, T2D)

Furthermore, for some participants, not having a human in the loop lowed their worry about being judged and the embarrassment of revealing bad habits, both of which can have a significant negative impact on the overall coaching experience.

> "But the fact that it is not a person in a way makes it easier. Especially, at the beginning when you feel like you have so many bad habits. I would have been embarrassed to even admit how many I had, so it would be easier to do so. I mean, now I'm feeling better because I've made a lot of progress, but still, it might be easier if it's not a person." (P23, 58, female, prediabetes)

Yet other participants felt the need for more persistent and focused encouragement, particularly after a lapse in their goal attainment.

> ". . . you know, people who interact with these apps don't always get it right. And you might need like, you know, a little encouragement to – to get back with the program. . ." (P3, 43, male, T2D)

These participants perceived the consistency of the emotional tone in T2 Coach as a limitation, as it was misaligned with their emotional needs.

4.5.6 *Theme 5: Importance of conversational fluidity.*

While participants felt reassured by the simple, fully scripted nature of T2 Coach, they also felt that it lacked conversational fluidity they wished for in a coaching experience. This was exacerbated by technical challenges of using regular text messaging as a medium for conversations: network delays occasionally led to messaging arriving in a wrong order, which disrupted the flow of conversations. Furthermore, the binary choices Yes/No for the goal attainment questions felt too restrictive and lacking in nuance. Some participants, like P18, wished they could engage in a more fluid conversation and just "talk things over" because it could help them become more aware of their choices and reflect on their actions:

> "Because you know how talking just raises your consciousness about everything that's inside of your head all the time. And so, talking things out helps you to be aware of your habits." (P18, 53, female, T2D)

As a result, some participants, like P1, perceived T2 Coach as lacking flexibility and fluidity of a human conversation and as too robotic:

> "What happened was, it seems a little stagnant. Like I know I'm talking to a computer." (P1, 51, female, T2D)

4.5.7 *Theme 6: Autonomy and being in control.*

Finally, a persistent theme through all the interviews was the importance of feeling in charge of both the coaching experience with T2 Coach, and the experience of engaging in diabetes self-management. The flexibility of choosing personal goals and action plans and of engaging with T2 Coach at their own time and on their own terms boosted participants' sense of autonomy and control:

> "I thought the coach was asking me, like I was the expert. It's because it's me, who is in charge of my life. So, it is me who will answer the questions." (P21, 61, female, T2D)

Some participants compared this favorably with their previous experience working with human coaching. For example, P18's previous experience with human coaches felt too rigid because the coach's schedule and expectations did not align with their own, particularly during weekends. This ultimately led to dissatisfaction and annoyance. In contrast, participants felt they could engage with T2 Coach whenever it was convenient for them and felt at ease leaving conversations and returning to them at a later time.

5 Discussion

In this study, we investigated individuals' experiences with a fully automated fully scripted CA that provided coaching for individuals with T2D and prediabetes. We aimed to address broad questions regarding the design of digital health coaching solutions. In addition, we were interested in examining what aspects of coaching experience could be supported with simple CA that do not rely on AI, which aspect of this experience require additional computation intelligence, and what type of intelligence is needed in digital coaching solutions. The findings of this study highlight both strengths of simple fully-scripted chatbots and opportunities to improve coaching experience with additional computational intelligence. We expand on these in more detail below.

5.1 Achieving much with simplicity

One of the questions of this study was to what degree a simple fully scripted conversational agent can create a positive coaching experience. The results suggest that at least to some degree, T2 Coach was indeed able to create a genuine coaching experience even despite its simple fully scripted format and absence of human coaches. This was supported by the quantitative results. First, the increase in goal attainment overtime suggested that engagement with T2 Coach had a positive impact on individuals self-management even within the short timeline of the study and even if it did not translate into perceived changes in their self-management behaviors (captured with SDSCA). Furthermore, high scores on the shared decision-making scale indicate that T2 Coach helped to promote shared decision-making, a critical component of coaching. Finally, high scores on the SASSI scale suggest that the participants were generally satisfied with the quality of the conversations.

The qualitative results revealed several aspects of T2 Coach that contributed to these positive perceptions. First, and perhaps most importantly, the study highlighted the importance of mundane daily assistance and its role in successful health coaching (Theme 1: Keeping on track). Second, the study highlighted the importance of coaching interactions free of guilt, shame, and pressure (Theme 4: Appropriate encouragement). Participants found the tone of messages in T2 Coach to be generally positive, and interactions with it to be straightforward and lacking any hidden agenda or subtext, common to interaction between humans. Participants felt little stigma and embarrassment in disclosing behaviors that fell below their own expectations, particularly at the beginning of the study, when many of them were falling short of attaining their goals. Third, the flexibility built into the structure of setting goals and choosing action plans helped participants feel in control of both their coaching experience and their self-management (Theme 6: Autonomy and being in control). The participants chose goals that aligned with their own values and priorities and changed these goals when they were ready. Furthermore, the ability to engage with T2 Coach when it was convenient for them and on their own terms contributed to individuals' sense of autonomy and control.

These findings are consistent with reports from past studies of CA in health. Previous research on health coaching, including e-coaching, often focused on emotional and communicative aspects of coaching [79, 80]. Our study extends this perspective and highlights the importance of a stable, persistent, and consistent structure as another critical component of coaching. Creating such a structure may present a challenge to human coaches as it requires near-constant availability for consultation, reminders, and advice [44, 80]. In contrast, participants of our study found T2 Coach to be "diligent", "punctual", "persistent", and "patient". Furthermore, past research highlighted that CA can in fact inspire engagement comparable to and sometimes even superior to that with human healthcare professionals. For example, Bickmore et al showed that patients in a hospital preferred to receive discharge instructions from a conversational agent rather than a human nurse [6] and our study shows that these attitudes translate into health coaching as well. Interacting with a fully automated system often helped participants feel at ease while still experiencing the benefits of external accountability, helpful in goal attainment. Finally, previous studies demonstrated the importance of autonomy in the context of health and health management [3, 4, 36, 77]. Literature on human coaching emphasizes empowerment and autonomy as the most critical defining characteristics of coaching as a practice [19, 69]. Yet, previous studies also showed that human coaches might find it challenging to align their own expectations with their clients' priorities [66]. In contrast, participants of our study felt empowered to make their own choices

These findings suggest that even simple fully scripted CA can create a beneficial coaching experience, particularly in situations when human coaches are not available or accessible, as may be the case for many economically disadvantaged and medically underserved communities. While these communities often have higher prevalence of chronic diseases and are in a more acute need of coaching interventions, their access to healthcare professionals, including expert coaches, is limited [54]. In fact, many individuals with T2D residing in these communities never receive health coaching or even diabetes education, considered standard care [71]. Furthermore, these communities historically have complex relationships with healthcare institutions and limited trust towards healthcare professionals, which may further complicate their ability to develop close personal relationship with health coaches [48]. Consequently, new solution that incorporate human coaches are less likely to benefit these communities and may further exacerbate existing health disparities and even create new intervention-induced ones [88].

5.2 The need for different types of intelligence in intelligent coaching

While the study identified clear benefits to simple scripted chatbots, it also highlighted many opportunities to introduce additional computational intelligence.

First, our study showed that while flexibility and choice could create a perception of a personalized experience, there are clear opportunities to tailor coaching to individuals' preferences, lifestyles, and circumstances, thus suggesting the need for more ***personal intelligence*** (Theme 2: Individual differences and the need for personalization). Our study showed vast differences in participants' work schedules, eating habits, grocery shopping and cooking routines, their budgets, and availability of foods in their neighborhoods. Accounting for these differences could identify behavioral goals and actions that are accessible to different participants. Furthermore, more intelligent analysis of data captured with self-monitoring can enable personalized feedback on their goal attainment and areas for improvement. Second, participants discussed the need to balance more persistent personal preferences with demands of different situations, thus highlighting the need for ***situational intelligence*** (Theme 3: Fitting into the rhythm of daily life). Sick days, seasonal changes in schedules, and other unexpected circumstances created disruptions to the normal coaching experience and required accommodations. Third, while participants found the tone of T2 Coach messages to be comforting and reassuring, they wished for a more nuanced and tailored approach and greater ***emotional intelligence*** (Theme 4: Appropriate encouragement). This was important as participants needs for empathy and support differed, with some preferring occasional cheer-ups and others requiring more extensive engagement. Furthermore, these needs differed based on participants' goal attainment and whether they were on track and needed a simple acknowledgement of their successes, or whether they had a lapse and needed more encouragement to get back on track. Finally, while participants appreciated simplicity and consistency of the fully-scripted T2 Coach, they described it as stagnant and robotic, suggesting the need for more ***conversational intelligence*** (Theme 5: Conversational fluidity). Specifically, participants wished for the ability to engage in more fluid conversations, provide more nuanced response to questions about goal attainment, and to "talk things over" as they would with a human coach.

These findings are consistent with prior research that examined users' engagement with different types of conversational agents. For example, past research highlighted the importance of personalization in coaching experience and examined different ways to incorporate personal data to inform human coaching [67]. Furthermore, past research emphasized the importance of context in coaching and proposed ways to enhance coaching experience with data [18]. Similarly, past research emphasized the role of emotional support and encouragement in coaching [19, 78, 80]. In fact, some authors argued that empathy and emotional support, critical in coaching, are uniquely human qualities that cannot be replicated with digital coaching solutions [79]. Finally, many previous studies of frame-based chatbots, users found them too rigid and repetitive and lacking fluidity expected in daily conversations [29, 53].

These different types of intelligence call for different technological solutions. For example, LLM-based chatbots are uniquely well-suited to address the need for conversational intelligence due to their ability to maintain conversations on a variety of topics. In fact, recent studies suggested that this flexibility indeed creates a more engaging experience [44]. Past research noted some challenges that stem from these new capabilities and proposed potential solutions to address them. For example, Jo et al. noted the challenge of ensuring appropriateness of responses of LLM-based chatbots, aligning them with needs of healthcare professionals, and their inability to maintain conversational histories [44]. New solutions to address these challenges include hybrid architectures that either fine-tune LLMs on domain-specific corpora to impose a more structured approach to conversations [43], or integrate scripted dialog models with open-ended LLMs to enable both structure and flexibility [61].

However, perhaps the most important lesson learned from the study was the need to integrate different types of intelligence into a coherent user experience. A truly intelligent digital coach should be able to integrate conversational histories, assessment of a user's current health and emotional state, personal data, and situational and context-awareness. HCI community has a rich history of research on intelligent interactive computing that addresses many of these needs. For example, there is a rich body of research in affective computing that focuses on recognizing human emotions and creating interactive experiences tailored to individuals' emotional state [72]. Personal informatics provided many valuable lessons and innovative solutions in integrating personal health and wellness data for increased self-awareness through reflection [27]. Similarly, context-aware computing has been an area of active research for several decades [22]. What is needed are new approaches to integrating these different capabilities with advances in LLMs. Past work in this area mostly focused on prompt engineering and incorporating more complex information, such as personal data, into prompts [33]. Future research can help to uncover other opportunities to integrate conversational intelligence of LLMs with other types of computational intelligence.

5.3 Balancing computational intelligence and user autonomy

While our study demonstrates many opportunities to introduce different types of computational intelligence into digital coaching experience, it also highlights the need to carefully maintain a balance between increased intelligence and human autonomy and control over their experience. The simple CA used in this study provided users with flexibility to engage with it when and how they wished, choose goals that aligned with their personal priorities and preferences, and work on attaining these goals at their own pace, without pressure and blame. This flexibility helped participants feel in charge of both the coaching experience and their self-management and many participants compared it favorably with their past experience with human coaches.

The importance of individuals' autonomy has been firmly established with previous research in many research communities [4]. In health, it gave rise to the prominent patient empowerment movement [3] and Motivational Interviewing as an approach to promoting behavior change [65]. It is also at the heart of health coaching, with much of coaching literature emphasizing the need to promote patient autonomy and the role of coaches as helping individuals achieve their own goals, rather than some externally established standards [78].

Increased computational intelligence in interactive systems carries promise to create new exciting types of engagement between humans and computing. However, it can also have detrimental impact on human autonomy and control. In HCI community, these questions have been explored in the context of Human-Centered AI. Previous examinations of Human-Centered AI argued that human autonomy and control is one of the fundamental characteristics of Human-Centered AI systems [13, 15, 84]. It is important to be mindful of these questions when introducing more computational intelligence into digital coaching systems. Participants of our study often commented on negative previous experience with human coaching that felt too overbearing and prescriptive. The more human-like digital coaching systems become, the more they strive to predict users' emotions, needs, and preferences, the more opportunities they will have to diminish users' autonomy, which could in turn lead to either disengagement and under-reliance on AI, or in contrast, in inappropriate trust and overreliance on it [32]. More research is needed to identify new interactive solutions that could enable users to maintain control over their experience, for example by choosing the level of intelligence they wish for in their intelligent digital coach.

This study has several limitations. This exploratory study focused primarily on individuals' subjective experiences using T2 Coach in the context of their daily lives and on their patterns of engagement and self-reported goal achievement; given the small sample size in this study and its short duration we did not examine the impact of T2 Coach on their behaviors nor on such health outcomes as change in their blood glucose levels. Furthermore, the participants of the study self-reported their prediabetes and T2D diagnosis, but this was not confirmed as we did not have access to their medical records. Finally, while the study suggested opportunities for introducing new types of computational intelligence in digital coaching, these opportunities were inferred from individuals' experience interactive with a fully-scripted chatbot. Further research is needed to identify specific design solutions that encapsulate these different types of intelligence and to evaluate their impact on individuals' experience and health.

6 CONCLUSION

In this study, we conducted a user study of a fully automated fully scripted conversational agent T2 Coach to explore opportunities for computational intelligence in digital coaching solutions. Our study showed that even simple fully scripted digital solution can indeed create a positive coaching experience and may even have some advantaged over human coaches. These include reliability and consistency, flexibility in choices and style of engagement, and blame and guilt-free experience. At the same time, it showed many opportunities to improve coaching experience with more computational intelligence and suggested four types of computational intelligence relevant for digital coaching, including personal, situational, emotional, and conversational. Some of these types, particularly conversational intelligence, could be enabled by LLM-based CA. However, a more comprehensive intelligent digital coaching requires integration of conversational intelligence of LLM-based CA with other types of computational intelligence.

Acknowledgments

This research was funded by the National Institute of Diabetes and Digestive and Kidney Diseases award number R01DK113189, the National Science Foundation award number 2306690, and the National Library of Medicine award number T15LM007079. We thank all participants of our studies.

References

[1] Francesca Alloatti, Alessio Bosca, Luigi Di Caro, and Fabrizio Pieraccini. 2021. Diabetes and conversational agents: the AIDA project case study. Discover Artificial Intelligence 1, 1: 4. https://doi.org/10.1007/s44163-021-00005-1

[2] Ian Anderson, Julie Maitland, Scott Sherwood, Louise Barkhuus, Matthew Chalmers, Malcolm Hall, Barry Brown, and Henk Muller. 2007. Shakra: Tracking and Sharing Daily Activity Levels with Unaugmented Mobile Phones. Mobile Networks and Applications 12, 2–3: 185–199. https://doi.org/10.1007/s11036-007-0011-7

[3] Robert M. Anderson, Martha M. Funnell, Patricia M. Butler, Marilynn S. Arnold, James T. Fitzgerald, and Catherine C. Feste. 1995. Patient Empowerment: Results of a randomized controlled trial. Diabetes Care 18, 7: 943–949. https://doi.org/10.2337/diacare.18.7.943

[4] A Bandura. 2001. Social cognitive theory: an agentic perspective. Annual review of psychology 52: 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

[5] Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. 2013. Health Mashups: Presenting Statistical Patterns Between Wellbeing Data and Context in Natural Language to Promote Behavior Change. ACM Trans. Comput.-Hum. Interact. 20, 5: 30:1-30:27. https://doi.org/10.1145/2503823

[6] Timothy W. Bickmore, Suzanne E. Mitchell, Brian W. Jack, Michael K. Paasche-Orlow, Laura M. Pfeifer, and Julie ODonnell. 2010. Response to a Relational Agent by Hospital Patients with Depressive Symptoms. Interacting with computers 22, 4: 289–298. https://doi.org/10.1016/j.intcom.2009.12.001

[7] Timothy W. Bickmore, Laura M. Pfeifer, Donna Byron, Shaula Forsythe, Lori E. Henault, Brian W. Jack, Rebecca Silliman, and Michael K. Paasche-Orlow. 2010. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. Journal of Health Communication 15 Suppl 2: 197–210. https://doi.org/10.1080/10810730.2010.499991

[8] Timothy W. Bickmore, Rebecca A. Silliman, Kerrie Nelson, Debbie M. Cheng, Michael Winter, Lori Henault, and Michael K. Paasche-Orlow. 2013. A randomized controlled trial of an automated exercise coach for older adults. Journal of the American Geriatrics Society 61, 10: 1676–1683. https://doi.org/10.1111/jgs.12449

[9] Jaap van der Bijl, Ada van Poelgeest-Eeltink, and Lillie Shortridge-Baggett. 1999. The psychometric properties of the diabetes management self-efficacy scale for patients with type 2 diabetes mellitus. Journal of Advanced Nursing 30, 2: 352–359. https://doi.org/10.1046/j.1365-2648.1999.01077.x

[10] Thomas Bodenheimer, Edward H Wagner, and Kevin Grumbach. 2002. Improving primary care for patients with chronic illness. JAMA: the journal of the American Medical Association 288, 14: 1775–1779.

[11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3, 2: 77–101. https://doi.org/10.1191/1478088706qp063oa

[12] Angelo Cafaro, Hannes Högni Vilhjálmsson, and Timothy Bickmore. 2016. First Impressions in Human–Agent Virtual Encounters. ACM Trans. Comput.-Hum. Interact. 23, 4: 24:1-24:40. https://doi.org/10.1145/2940325

[13] Tara Capel and Margot Brereton. 2023. What is Human-Centered about Human-Centered AI? A Map of the Research Landscape. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), 1–23. https://doi.org/10.1145/3544548.3580959

[14] CDC. 2024. National Diabetes Statistics Report. Diabetes. Retrieved February 2, 2025 from https://www.cdc.gov/diabetes/php/data-research/index.html

[15] Stevie Chancellor, Eric P. S. Baumer, and Munmun De Choudhury. 2019. Who is the "Human" in Human-Centered Machine Learning: The Case of Predicting Mental Health from Social Media. Proceedings of the ACM on Human-Computer Interaction 3, CSCW: 147:1-147:32. https://doi.org/10.1145/3359249

[16] Céline Clavel, Steve Whittaker, Anaïs Blacodon, and Jean-Claude Martin. 2018. WEnner: A Theoretically Motivated Approach for Tailored Coaching about Physical Activity. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp '18), 1669–1675. https://doi.org/10.1145/3267305.3274190

[17] Heather J. Cole-Lewis, Arlene M. Smaldone, Patricia R. Davidson, Rita Kukafka, Jonathan N. Tobin, Andrea Cassells, Elizabeth D. Mynatt, George Hripcsak, and Lena Mamykina. 2016. Participatory approach to the development of a knowledge base for problem-solving in diabetes self-management. International Journal of Medical Informatics 85, 1: 96–103. https://doi.org/10.1016/j.ijmedinf.2015.08.003

[18] Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, Ian Smith, and James A. Landay. 2008. Activity Sensing in the Wild: A Field Trial of Ubifit Garden. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08), 1797–1806. https://doi.org/10.1145/1357054.1357335

[19] Lauren M. Denneson, Sarah S. Ono, Amira Y. Trevino, Emily Kenyon, and Steven K. Dobscha. 2020. The applicability of self-determination theory to health coaching: A qualitative analysis of patient experiences. Coaching: An International Journal of Theory, Research and Practice 13: 163–175. https://doi.org/10.1080/17521882.2019.1673457

[20] Pooja M. Desai, Matthew E. Levine, David J. Albers, and Lena Mamykina. 2018. Pictures Worth a Thousand Words: Reflections on Visualizing Personal Blood Glucose Forecasts for Individuals with Type 2 Diabetes. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), 538:1-538:13. https://doi.org/10.1145/3173574.3174112

[21] Pooja M. Desai, Elliot G. Mitchell, Maria L. Hwang, Matthew E. Levine, David J. Albers, and Lena Mamykina. 2019. Personal Health Oracle: Explorations of Personalized Predictions in Diabetes Self-Management. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), 370:1-370:13. https://doi.org/10.1145/3290605.3300600

[22] ANIND K. DEY. 2009. Context-Aware Computing. In Ubiquitous Computing Fundamentals. Chapman and Hall/CRC.

[23] Diabetes Prevention Program Research Group. 2009. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. The Lancet 374, 9702: 1677–1686. https://doi.org/10.1016/S0140-6736(09)61457-4

[24] Elizabeth G. Eakin, Elisabeth A. Winkler, David W. Dunstan, Genevieve N. Healy, Neville Owen, Alison M. Marshall, Nicholas Graves, and Marina M. Reeves. 2014. Living well with diabetes: 24-month outcomes from a randomized trial of telephone-delivered weight loss and physical activity intervention to improve glycemic control. Diabetes Care 37, 8: 2177–2185. https://doi.org/10.2337/dc13-2427

[25] Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. 2015. A Lived Informatics Model of Personal Informatics. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15), 731–742. https://doi.org/10.1145/2750858.2804250

[26] Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. 2014. Taming Data Complexity in Lifelogs: Exploring Visual Cuts of Personal Informatics Data. DIS '14 Proceedings of the 2014 conference on Designing interactive systems: 667–676. https://doi.org/10.1145/2598510.2598558

[27] Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. 2014. Taming Data Complexity in Lifelogs: Exploring Visual Cuts of Personal Informatics Data. In Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14), 667–676. https://doi.org/10.1145/2598510.2598558

[28] Alison B. Evert, Jackie L. Boucher, Marjorie Cypress, Stephanie A. Dunbar, Marion J. Franz, Elizabeth J. Mayer-Davis, Joshua J. Neumiller, Robin Nwankwo, Cassandra L. Verdi, Patti Urbanski, and William S. Yancy. 2014. Nutrition Therapy Recommendations for the Management of Adults With Diabetes. Diabetes Care 37, Supplement 1: S120–S143. https://doi.org/10.2337/dc14-S120

[29] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Mental Health 4, 2: e19. https://doi.org/10.2196/mental.7785

[30] Russell Fulmer, Angela Joerin, Breanna Gentile, Lysanne Lakerink, and Michiel Rauws. 2018. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR mental health 5, 4: e64. https://doi.org/10.2196/mental.9782

[31] Elliot G. Mitchell, Elizabeth M. Heitkemper, Marissa Burgermaster, Matthew E. Levine, Yishen Miao, Maria L. Hwang, Pooja M. Desai, Andrea Cassells, Jonathan N. Tobin, Esteban G. Tabak, David J. Albers, Arlene M. Smaldone, and Lena Mamykina. 2021. From Reflection to Action: Combining Machine Learning with Expert Knowledge for Nutrition Goal Recommendations. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21), 1–17. https://doi.org/10.1145/3411764.3445555

[32] K. Z. Gajos and L. Mamykina. 2022. Do people engage cognitively with AI? Impact of AI assistance on incidental learning. In In Proceedings of the ACM Conference on Intelligent User Interfaces.

[33] Boris A. Galitsky. 2024. LLM- Based Personalized Recommendations in Health. https://doi.org/10.20944/preprints202402.1709.v1

[34] Jianfeng Gao, Michel Galley, and Lihong Li. 2018. Neural Approaches to Conversational AI. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18), 1371–1374. https://doi.org/10.1145/3209978.3210183

[35] Enying Gong, Shaira Baptista, Anthony Russell, Paul Scuffham, Michaela Riddell, Jane Speight, Dominique Bird, Emily Williams, Mojtaba Lotfaliany, and Brian Oldenburg. 2020. My Diabetes Coach, a Mobile App-Based Interactive Conversational Agent to Support Type 2 Diabetes Self-Management: Randomized Effectiveness-Implementation Trial. Journal of Medical Internet Research 22, 11: e20322. https://doi.org/10.2196/20322

[36] Amy A. Gorin, Theodore A. Powers, Richard Koestner, Rena R. Wing, and Hollie A. Raynor. 2014. Autonomy Support, Self-Regulation, and Weight Loss. Health psychology: official journal of the Division of Health Psychology, American Psychological Association 33, 4: 332–339. https://doi.org/10.1037/a0032586

[37] The Diabetes Prevention Program (DPP) Research Group. 2002. The Diabetes Prevention Program (DPP). Diabetes Care 25, 12: 2165–2171. https://doi.org/10.2337/diacare.25.12.2165

[38] T. Gutnick, C. Davis, K. Reims, and H.L Gainforth. 2014. Brief Action Planning to Facilitate Behavior Change and Support Patient Self-Management. Journal of clinical outcomes management: JCOM 21, 1: 17–29.

[39] Claudia E. Haupt and Mason Marks. 2023. AI-Generated Medical Advice—GPT and Beyond. JAMA 329, 16: 1349–1350. https://doi.org/10.1001/jama.2023.5321

[40] Victoria Hollis, Artie Konrad, Aaron Springer, Matthew Antoun, Christopher Antoun, Rob Martin, and Steve Whittaker. 2017. What Does All This Data Mean for My Future Mood? Actionable Analytics and Targeted Reflection for Emotional Well-Being. Hum.-Comput. Interact. 32, 5–6: 208–267. https://doi.org/10.1080/07370024.2016.1277724

[41] Eva Hudlicka. 2013. Virtual training and coaching of health behavior: example from mindfulness meditation training. Patient Education and Counseling 92, 2: 160–166. https://doi.org/10.1016/j.pec.2013.05.007

[42] Maria L. Hwang and Lena Mamykina. 2017. Monster Appetite: Effects of Subversive Framing on Nutritional Choices in a Digital Game Environment. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17), 4082–4096. https://doi.org/10.1145/3025453.3026052

[43] Cheonsu Jeong. 2024. Fine-tuning and Utilization Methods of Domain-specific LLMs. Journal of Intelligence and Information Systems 30, 1: 93–120. https://doi.org/10.13088/jiis.2024.30.1.093

[44] Eunkyung Jo, Daniel A. Epstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), 1–16. https://doi.org/10.1145/3544548.3581503

[45] Bart A. Kamphorst. 2017. E-coaching systems. Personal and Ubiquitous Computing 21, 4: 625–632. https://doi.org/10.1007/s00779-017-1020-6

[46] Ravi Karkar, Jasmine Zia, Roger Vilardaga, Sonali R Mishra, James Fogarty, Sean A Munson, and Julie A Kientz. 2016. A framework for self-experimentation in personalized health. Journal of the American Medical Informatics Association 23, 3: 440–448. https://doi.org/10.1093/jamia/ocv150

[47] Ravi Karkar, Jasmine Zia, Roger Vilardaga, Sonali R. Mishra, James Fogarty, Sean A. Munson, and Julie A. Kientz. 2016. A framework for self-experimentation in personalized health. Journal of the American Medical Informatics Association 23, 3: 440–448. https://doi.org/10.1093/jamia/ocv150

[48] Bernice Roberts Kennedy, Christopher Clomus Mathis, and Angela K. Woods. 2007. African Americans and their distrust of the health care system: healthcare for diverse populations. Journal of Cultural Diversity 14, 2: 56–60.

[49] Jungkeun Kim, Marilyn Giroux, and Jacob C. Lee. 2021. When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations. Psychology & Marketing 38, 7: 1140–1155. https://doi.org/10.1002/mar.21498

[50] Chulmo Koo, Zheng Xiang, Ulrike Gretzel, and Marianna Sigala. 2021. Artificial intelligence (AI) and robotics in travel, hospitality and leisure. Electronic Markets 31, 3: 473–476. https://doi.org/10.1007/s12525-021-00494-z

[51] Lean L. Kramer, Silke Ter Stal, Bob C. Mulder, Emely de Vet, and Lex van Velsen. 2020. Developing Embodied Conversational Agents for Coaching People in a Healthy Lifestyle: Scoping Review. Journal of Medical Internet Research 22, 2: e14058. https://doi.org/10.2196/14058

[52] Levente Kriston, Isabelle Scholl, Lars Hölzel, Daniela Simon, Andreas Loh, and Martin Härter. 2010. The 9-item Shared Decision Making Questionnaire (SDM-Q-9). Development and psychometric properties in a primary care sample. Patient Education and Counseling 80, 1: 94–99. https://doi.org/10.1016/j.pec.2009.09.034

[53] Liliana Laranjo, Adam G. Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie Y. S. Lau, and Enrico Coiera. 2018. Conversational agents in healthcare: a systematic review. Journal of the American Medical Informatics Association 25, 9: 1248–1258. https://doi.org/10.1093/jamia/ocy072

[54] Karen E. Lasser, David U. Himmelstein, and Steffie Woolhandler. 2006. Access to Care, Health Status, and Health Disparities in the United States and Canada: Results of a Cross-National Population-Based Survey. American Journal of Public Health 96, 7: 1300–1307. https://doi.org/10.2105/AJPH.2004.059402

[55] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-based Model of Personal Informatics Systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10), 557–566. https://doi.org/10.1145/1753326.1753409

[56] Ian Li, Anind K. Dey, and Jodi Forlizzi. 2011. Understanding My Data, Myself: Supporting Self-reflection with Ubicomp Technologies. In Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11), 405–414. https://doi.org/10.1145/2030112.2030166

[57] Zilu Liang, Bernd Ploderer, Wanyu Liu, Yukiko Nagata, James Bailey, Lars Kulik, and Yuxuan Li. 2016. SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. Personal and Ubiquitous Computing. https://doi.org/10.1007/s00779-016-0960-6

[58] Zilu Liang, Bernd Ploderer, Wanyu Liu, Yukiko Nagata, James Bailey, Lars Kulik, and Yuxuan Li. 2016. SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. Personal and Ubiquitous Computing 20, 6: 985–1000. https://doi.org/10.1007/s00779-016-0960-6

[59] Christine Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rishe. 2013. I Can Help You Change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions. ACM Trans. Manage. Inf. Syst. 4, 4: 19:1-19:28. https://doi.org/10.1145/2544103

[60] Ryan Lowe, Nissan Pow, Iulian Serban, and Joelle Pineau. 2016. The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems. https://doi.org/10.48550/arXiv.1506.08909

[61] Wookjae Maeng and Joonhwan Lee. 2021. Designing a Chatbot for Survivors of Sexual Violence: Exploratory Study for Hybrid Approach Combining Rule-based Chatbot and ML-based Chatbot. In Proceedings of the Asian CHI Symposium 2021 (Asian CHI '21), 160–166. https://doi.org/10.1145/3429360.3468203

[62] Lena Mamykina and Elizabeth D. Mynatt. 2007. Investigating and supporting health management practices of individuals with diabetes. In Proceedings of the 1st ACM SIGMOBILE international workshop on Systems and networking support for healthcare and assisted living environments (HealthNet '07), 49–54. https://doi.org/10.1145/1248054.1248068

[63] Michael F. McTear. 2002. Spoken dialogue technology: enabling the conversational user interface. ACM Computing Surveys 34, 1: 90–169. https://doi.org/10.1145/505282.505285

[64] Y. Medynskiy and E. Mynatt. 2010. Salud!: An open infrastructure for developing and deploying health self-management applications. In 2010 4th International Conference on Pervasive Computing Technologies for Healthcare, 1–8. https://doi.org/10.4108/ICST.PERVASIVEHEALTH2010.8897

[65] William R. Miller and Stephen Rollnick. 2013. Motivational interviewing: Helping people change, 3rd edition. Guilford Press, New York, NY, US.

[66] Elliot G. Mitchell, Rosa Maimone, Andrea Cassells, Jonathan N. Tobin, Patricia Davidson, Arlene M. Smaldone, and Lena Mamykina. 2021. Automated vs. Human Health Coaching: Exploring Participant and Practitioner Experiences. Proceedings of the ACM on Human-Computer Interaction 5, CSCW1: 99:1-99:37. https://doi.org/10.1145/3449173

[67] Francisco Monteiro-Guerra, Octavio Rivera-Romero, Luis Fernandez-Luque, and Brian Caulfield. 2020. Personalization in Real-Time Physical Activity Coaching

Using Mobile Applications: A Scoping Review. IEEE Journal of Biomedical and Health Informatics 24, 6: 1738–1751. https://doi.org/10.1109/JBHI.2019.2947243
[68] Jeanette M. Olsen. 2014. Health Coaching: A Concept Analysis. Nursing Forum 49, 1: 18–29. https://doi.org/10.1111/nuf.12042
[69] Jeanette M. Olsen. 2014. Health coaching: a concept analysis. Nursing Forum 49, 1: 18–29. https://doi.org/10.1111/nuf.12042
[70] Jeanette M Olsen and Bonnie J Nesbitt. 2010. Health Coaching to Improve Healthy Lifestyle Behaviors: An Integrative Review. American Journal of Health Promotion 25, 1: e1–e12. https://doi.org/10.4278/ajhp.090313-lit-101
[71] Mark Peyrot, Richard R. Rubin, Martha M. Funnell, and Linda M. Siminerio. 2009. Access to Diabetes Self-management Education. The Diabetes Educator 35, 2: 246–263. https://doi.org/10.1177/0145721708329546
[72] Rosalind W. Picard. 2000. Affective Computing. MIT Press.
[73] James O. Prochaska and Janice M. Prochaska. 1999. Why Don't Continents Move? Why Don't People Change? Journal of Psychotherapy Integration 9, 1: 83–102. https://doi.org/10.1023/A:1023210911909
[74] James O. Prochaska and Wayne F. Velicer. 1997. The Transtheoretical Model of Health Behavior Change. American Journal of Health Promotion 12, 1: 38–48. https://doi.org/10.4278/0890-1171-12.1.38
[75] Judith J. Prochaska, Erin A. Vogel, Amy Chieng, Matthew Kendra, Michael Baiocchi, Sarah Pajarito, and Athena Robinson. 2021. A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study. Journal of Medical Internet Research 23, 3: e24850. https://doi.org/10.2196/24850
[76] Neesha Ramchandani. 2019. Virtual Coaching to Enhance Diabetes Care. Diabetes Technology & Therapeutics 21, S2: S248–S251. https://doi.org/10.1089/dia.2019.0016
[77] Meghan Reading Turchioe, Marissa Burgermaster, Elliot G. Mitchell, Pooja M. Desai, and Lena Mamykina. 2020. Adapting the stage-based model of personal informatics for low-resource communities in the context of type 2 diabetes. Journal of Biomedical Informatics 110: 103572. https://doi.org/10.1016/j.jbi.2020.103572
[78] Michael Reinboth, Joan L. Duda, and Nikos Ntoumanis. 2004. Dimensions of Coaching Behavior, Need Satisfaction, and the Psychological and Physical Welfare of Young Athletes. Motivation and Emotion 28, 3: 297–313. https://doi.org/10.1023/B:MOEM.0000040156.81924.b8
[79] Heleen Rutjes, Martijn C. Willemsen, and Wijnand A. IJsselsteijn. 2019. Beyond Behavior: The Coach's Perspective on Technology in Health Coaching. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), 670:1-670:14. https://doi.org/10.1145/3290605.3300900
[80] Kathleen Ryan, Samantha Dockray, and Conor Linehan. 2022. Understanding How eHealth Coaches Tailor Support For Weight Loss: Towards the Design of Person-Centered Coaching Systems. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22), 1–16. https://doi.org/10.1145/3491102.3501864
[81] Pouya Saeedi, Inga Petersohn, Paraskevi Salpea, Belma Malanda, Suvi Karuranga, Nigel Unwin, Stephen Colagiuri, Leonor Guariguata, Ayesha A. Motala, Katherine Ogurtsova, Jonathan E. Shaw, Dominic Bright, Rhys Williams, and IDF Diabetes Atlas Committee. 2019. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Research and Clinical Practice 157: 107843. https://doi.org/10.1016/j.diabres.2019.107843
[82] Gary A. Sforzo, Miranda P. Kaye, Irina Todorova, Sebastian Harenberg, Kyle Costello, Laura Cobus-Kuo, Aubrey Faber, Elizabeth Frates, and Margaret Moore. 2018. Compendium of the Health and Wellness Coaching Literature. American Journal of Lifestyle Medicine 12, 6: 436–447. https://doi.org/10.1177/1559827617708562
[83] Diana Sherifali, Virginia Viscardi, Johnny Wei Bai, and R. Muhammad Usman Ali. 2016. Evaluating the Effect of a Diabetes Health Coach in Individuals with Type 2 Diabetes. Canadian Journal of Diabetes 40, 84–94. https://doi.org/10.1016/j.jcjd.2015.10.006
[84] Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. International Journal of Human–Computer Interaction 36, 6: 495–504. https://doi.org/10.1080/10447318.2020.1741118
[85] D J Toobert, S E Hampson, and R E Glasgow. 2000. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 23, 7: 943–950. https://doi.org/10.2337/diacare.23.7.943
[86] Meghan Reading Turchioe, Elizabeth M. Heitkemper, Maichou Lor, Marissa Burgermaster, and Lena Mamykina. 2019. Designing for engagement with self-monitoring: A user-centered approach with low-income, Latino adults with Type 2 Diabetes. International Journal of Medical Informatics 130: 103941. https://doi.org/10.1016/j.ijmedinf.2019.08.001
[87] Dina Utami, Timothy Bickmore, Asimina Nikolopoulou, and Michael Paasche-Orlow. 2017. Talk About Death: End of Life Planning with a Virtual Agent. In Intelligent Virtual Agents (Lecture Notes in Computer Science), 441–450. https://doi.org/10.1007/978-3-319-67401-8_55
[88] Tiffany C. Veinot, Jessica S. Ancker, Heather Cole-Lewis, Elizabeth D. Mynatt, Andrea G. Parker, Katie A. Siek, and Lena Mamykina. 2019. Leveling Up: On the Potential of Upstream Health Informatics Interventions to Enhance Health Equity. Medical Care 57: S108. https://doi.org/10.1097/MLR.0000000000001032
[89] Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 9, 1: 36–45. https://doi.org/10.1145/365153.365168
[90] Ruth Q. Wolever, Leigh Ann Simmons, Gary A. Sforzo, Diana Dill, Miranda Kaye, Elizabeth M. Bechard, Mary Elaine Southard, Mary Kennedy, Justine Vosloo, and Nancy Yang. 2013. A Systematic Review of the Literature on Health and Wellness Coaching: Defining a Key Behavioral Intervention in Healthcare. Global Advances in Health and Medicine 2, 4: 38–57. https://doi.org/10.7453/gahmj.2013.042

A APPENDICES

A A.1 Appendix1: T2 Coach interview guide

Introductions

Can you tell me about your history with diabetes (or prediabetes)? When where you diagnosed? What do you do to manage it?

Overall Experience:

Can you tell me about your experience using T2 Coach for the past few months?

How often did you use it? What did you use it for? What were some of the circumstances when you used it?

Overall, how would you describe your experience?

Logging

What was it like to record your meals in T2 Coach?

Have you ever kept track of your meals before?

On paper? With a smartphone application?

What was that like for you?

What was it like to check your blood sugar before and after each meal?

How often did you check your blood sugar before the study?

Text Messaging

What was it like to send text messages with T2 Coach during the study?

Normally speaking, how often do you send text messages? What messaging applications do you use?

Text messaging? Facebook Messenger? WhatsApp? Other?

Who do you message with?

Friends? Family? Coworkers? Community boards and groups?

Goal Setting

What goal did you choose to work on?

What was the experience of choosing a goal like for you?

Was it clear what the goal was about and what you needed to do to meet it?

Do you remember seeing the illustration? Was it helpful to see the illustration?

Were you able to find goals that you wanted to try?

How did it feel to only choose one goal to focus on (at a time)?

Action Plans

What did you think of the specific plans and recommendations T2 Coach made?

For example, [List some that they chose or were presented with]

____

____

____

Did you choose this?

Why did you choose this?

Did you try to follow it?

What was that like?

Were you able to do it? Why or why not? Did you try something else? Did you try it again the next day or something new?

Was the reminder about your plan each day helpful?

Assessing/meeting goals

What was your experience receiving questions whether your meal fit your goal?

How did it feel to receive prompts asking you whether your meals met your goals?

Were you able to meet your goal to ______?

How did you keep track?

Did the messages from T2 Coach give you feedback on how you were doing with your goals?

Evoking questions

[Look at transcripts to pull examples of when T2 Coach asked barriers in meeting goal]

When you said you { did not meet your goal} T2 Coach asked "__________"

What did you think about that? What that a helpful response?

What could have made it more helpful?

In response, you said "______"; T2 Coach asked "__________"; you said "________"

Can you tell me more about that? What was going on at the time? Do you remember the specific situation?

Messages and Reminders

T2 Coach sent you a number of messages every day

Was this too much? Too little?

Helpful? Not helpful?

How many times per day would you have liked to hear from T2 Coach?

App and chatbot

In this study you sometimes interacted with text messages and sometimes with an app with a graphic interface (buttons, pictures), what was this experience like?

Was it clear what you needed to use for what?

Did it feel like they were two parts of the same app? Or did they feel different? How did you access the app (from the text messages or from an icon on your home screen?)

Would it be easier to just use one mode for all interactions? If so, what would it be, text or graphic interface?

Coaching in General

Now that you've had this experience, what do you think about coaching with text messaging?

Did it feel like you were working with a coach?

How did it feel for you?

Did you feed supported? Did you feel listened to? Did you feel frustrated?

Have you worked with a dietitian, nutritionist, diabetes educator, or coach before?

How did this compare?

What was missing? What could have made it better?