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

T2教练：糖尿病自我管理自动化健康教练的定性研究 糖尿病自我管理

埃利奥特·G·米切尔 盖辛格
医疗 美国宾夕法尼亚州丹维尔
emitchell3@geisinger.edu

普贾·德赛 生物医学信息学系
哥伦比亚大学欧文医学中心 美国
纽约州纽约市
p.desai@columbia.edu

阿琳·斯马尔东 护理
学院 哥伦比亚大学 美国
纽约州纽约市 牙科医学
院
哥伦比亚大学 美国纽约
州纽约市
ams130@columbia.edu

安德烈娅·卡斯尔斯公共
卫生硕士 临床主任网络 美国
纽约州纽约市
acass@cdnetwork.org

乔纳森·N·托宾博士 临床主任
网络(CDN) 美国纽约州纽约市 临
床与转化科学中心
科学 洛克菲勒大学 美国
纽约州纽约市
jntobin@cdnetwork.org

戴维·阿尔伯斯 儿科系信息学
与数据科学部 科罗拉多大学安舒茨
医学中心 美国科罗拉多州奥罗拉
david.albers
@cuanschutz.edu

马修·莱文 埃里克与温迪·施密
特中心 麻省理工与哈佛大学布罗德
研究所 美国马萨诸塞州剑桥市
levinema@broadinstitute.org

莉娜·马梅金娜 生物医学信息
学系 哥伦比亚大学 美国纽约州纽
约市
o
m2196@cumc.columbia.edu

摘要

计算智能在包括健康在内的诸多领域的交互系统中日益普及。通过对话代理（CA）进行健康指导可覆盖广泛人群，但实现积极辅导体验所需的计算智能水平尚不明确。我们开展了一项研究，邀请16名糖尿病及糖尿病前期患者使用对话代理T2教练进行健康指导。定性访谈显示：参与者认为T2教练能可靠地协助其坚持自我管理，赞赏其允许自主选择个人有意义的目标并按自身节奏参与的灵活性，并感到其能提供鼓励，甚至评价优于人类教练。但参与者同时指出，若需提升辅导体验，应实现更流畅的对话、更贴合个人偏好与生活方式的定制化服务，以及对特定情境更敏锐的响应——这些均需更强的计算智能支持。

我们探讨了更智能系统的设计方向及其潜在影响
健康领域的辅导对话代理。

CCS概念

• **人本计算**；• **普适与移动计算**；• **普适与移动计算的实证研究**；

关键词

健康、糖尿病、自我管理、辅导、聊天机器人、对话代理、移动健康、自我追踪

ACM参考格式：
埃利奥特·G·米切尔、普贾·德赛、阿琳·斯马尔东、安德烈娅·卡斯尔斯公共卫生硕士，乔纳森·N·托宾博士、戴维·阿尔伯斯、马修·莱文与莉娜·马米金娜。2025。T2教练：糖尿病自我管理自动化健康教练的定性研究。发表于美国计算机协会举办的CHI 2025计算机系统人因会议（CHI 2025），2025年4月26日至5月1日，日本横滨。ACM出版。

美国纽约州纽约市，17页。https://doi.org/10.1145/3706598.3714404

美国计算机协会 ISBN 979-8-4007-1394-1/2025/04
https://doi.org/10.1145/3706598.3714404

1 引言

机器学习（ML）与人工智能的进步
人工智能正在改变人类的体验与互动方式。
与计算系统在基础层面深度融合。新型
由数据和人工智能驱动的智能系统不仅越来越多地应用于
专业领域，也深入个人生活。人们
各行各业的民众依赖智能系统来规划他们的休闲活动

[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

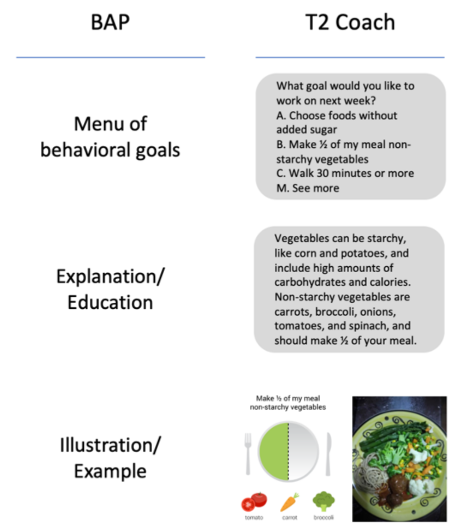

Figure 1: An example a goal setting dialog in T2 Coach following the BAP protocol

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.

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

[50], 接收消费品推荐[49], ，甚至获取关于健康的建议与指导[39]。

人工智能技术对促进个人健康，尤其是慢性病如2型糖尿病（T2D）的自我管理，具有重大社会影响的潜力。疾病控制中心（CDC）估计10.5%的美国人口中有糖尿病患者，占成年美国人口的34.5%处于糖尿病前期 [14]。2型糖尿病带来高昂的个人代价；患者必须积极参与自我管理，并做出无数日常决策——关于饮食和运动方式——以维持健康及长期生活质量 [10]。健康指导已成为促进自我管理 [37, 69, 83, 90]。辅导旨在培养动力与参与，并建立追求

可实现健康目标的长期关系，这种关系存在于教练与其客户 [68, 70, 78, 82, 90] 之间。通过人类教练进行的辅导，无论是面对面 [37] 还是借助技术 [45, 80] 媒介，已被证明能有效支持慢性病管理和预防。然而，这两种方式都依赖于人类教练且现有的辅导从业者数量不足以支持日益增长的慢性病患者群体[28]。

计算技术的进步为完全自动化的数字健康干预铺平了道路，这些干预措施可以补充甚至有时取代人类健康从业者[53, 67]。特别是，对话代理（CA）在健康领域的多个方面（包括辅导[16, 41, 51]）已展现出显著成效。迄今为止提出的绝大多数健康领域CA采用简单、完全脚本化的方法，即设计师预先规划所有可能的对话空间[53]。更近期的研究引入了依赖人工智能[75] 的健康CA，尤其是基于大型语言模型（LLM）的技术，其创造的类人对话能力是早期CA无法比拟的[44]。然而，这些新型基于人工智能的CA存在重要限制：它们赋予设计师对其行为和输出的控制权更少，因而可能产生不恰当和不准确的回应[44]。另有观点认为，健康指导具有许多独特的人类特质，任何类型的自动化辅导系统都难以复制[79]。因此，关于哪种类型的CA最适合提供积极的辅导体验，特别是数字辅导中需要何种程度且足够的计算智能，仍存在开放性问题。

在本研究中，我们旨在探索2型糖尿病自我管理背景下的这些问题。具体而言，我们试图考察哪些辅导需求可以通过不依赖人工智能和大型语言模型的简单对话代理来满足，以及引入计算智能可以改善辅导体验的哪些方面，

以及需要何种类型的计算智能来改进数字辅导。为了探讨这些问题，我们与2型糖尿病（T2D）患者合作，利用一款针对T2D的数字教练原型T2教练（图1、2和3）开展了一项形成性研究。我们设计T2教练的初衷是采用技术探针[41], 的理念，重点在于识别新设计特征的机会，而非收集对已完成设计的反馈。这种方法对于可能需要大量投入和开发时间的复杂技术尤为适用。与此一致，T2教练采用了健康对话代理（CA）中常见的简单基于框架的设计，其中对话流程完全脚本化，用户从可用选项列表中选择

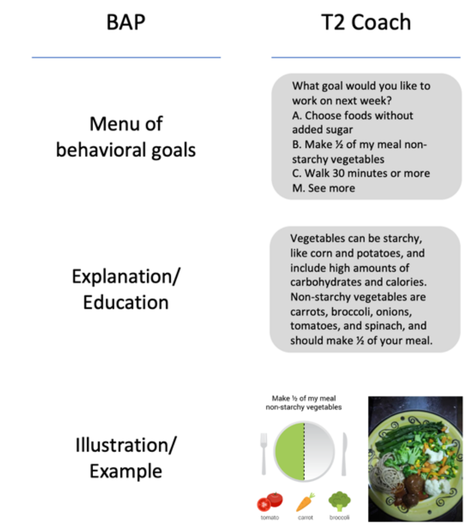

图1：T2教练中遵循BAP协议的目标设定对话示例

大部分回应。T2教练采用迭代式以用户为中心的设计方法进行设计。我们首先收集了个人对数字辅导的态度和期望，随后利用他们的反馈对T2教练进行迭代优化，最终完善其设计。

T2教练采用一系列每日和每周对话，这些对话基于健康指导的临床方案——简短行动计划（BAP, [38]），旨在帮助用户设定自我管理目标，如营养管理和体育活动，通过日常活动追求这些目标，并随着时间推移反思目标达成率及实现目标过程中的障碍。所有消息和回复均通过短信传递，从而最大程度降低用户的技术要求。

我们通过大学托管的招募网站和社交媒体广告，招募了19名2型糖尿病（n=13）和糖尿病前期（n=6）患者使用T2教练3-4周。16名参与者完成了研究并参与了研究后定性访谈。参与者主要为女性（63%）、种族和民族多样化（26%西班牙裔和53%黑人或非裔美国人）、受过教育（样本中超过60%拥有大学和研究生学位）以及超重（平均体重指数=29.2）。

研究表明，总体而言，参与者对T2教练的使用体验积极，并在研究过程中提升了目标达成率，尽管这并未转化为自我管理行为的自我报告变化。此外，许多人将使用该应用程序的体验描述为与一位教练共事。促成这一积极体验的因素包括：系统通过一致的消息和提醒构建了可靠的结构，帮助个体保持自我管理的正轨；强调选择个人有意义的目标和行动计划以及参与的灵活性，营造了个性化体验，增强了自主性与掌控感；最后，其低调的消息提供了适当鼓励，为体验增色。

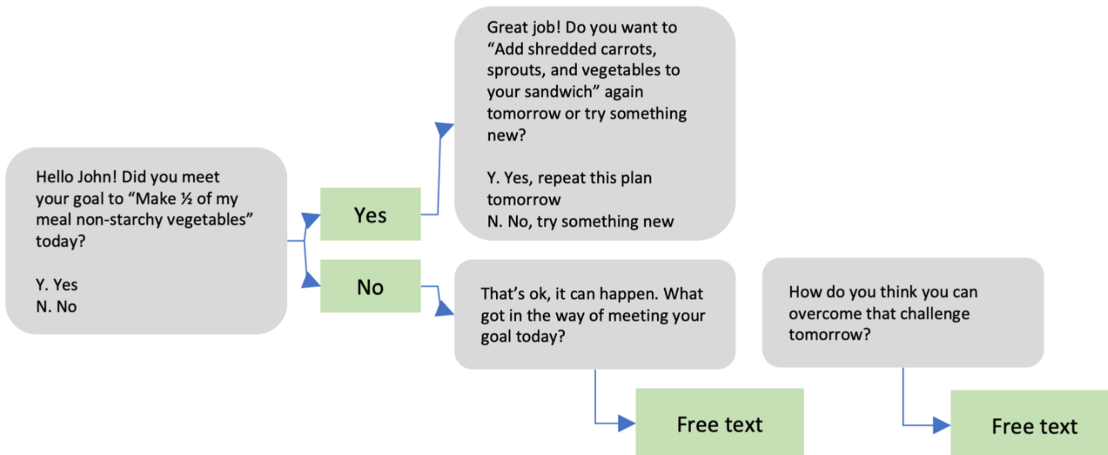

Figure 2: The flow of the daily reflection dialog

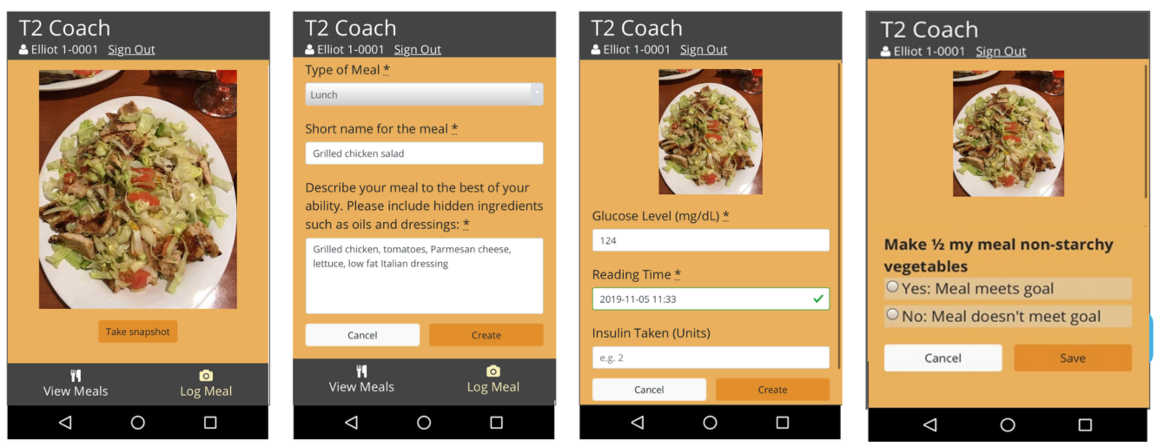

Figure 3: Screens from the progressive web application for recording meals and blood glucose readings.

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

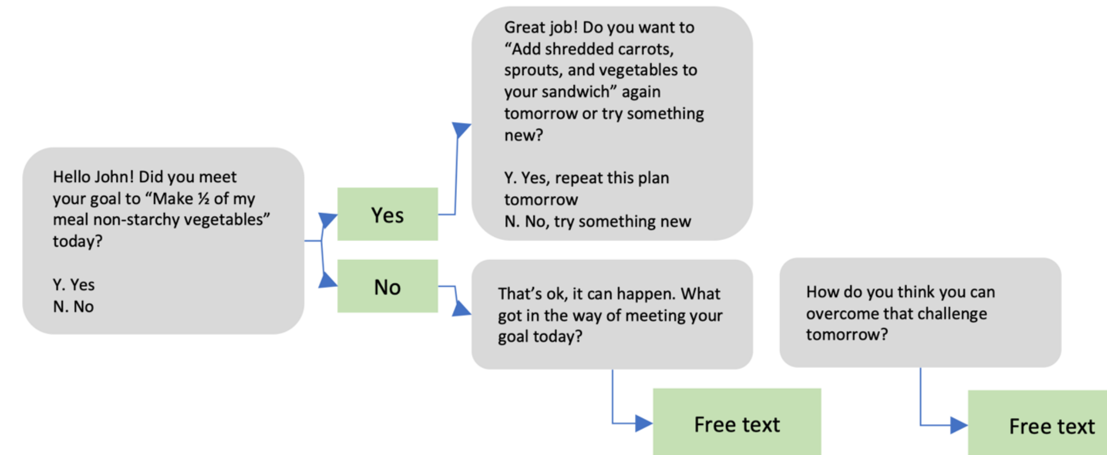

图2：每日反思对话流程

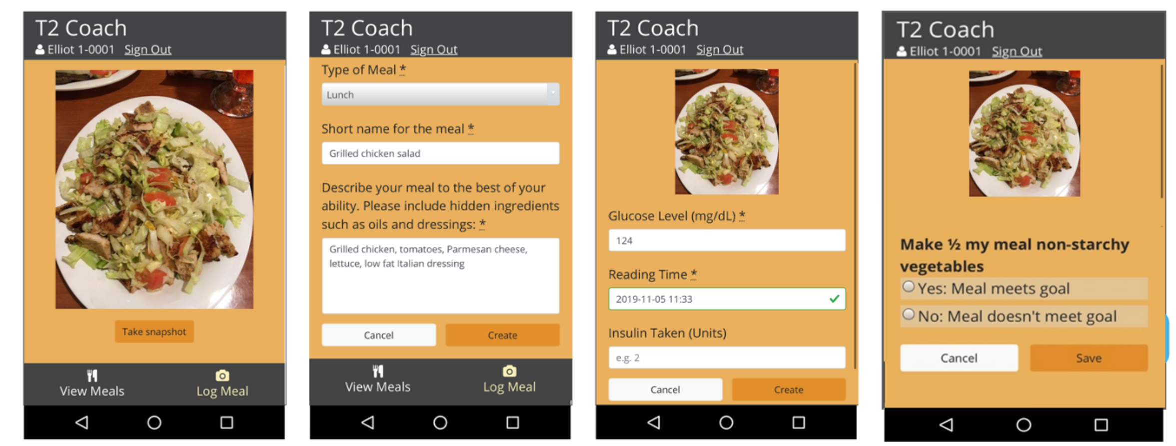

图3：记录饮食和血糖读数的渐进式网络应用界面

无责备体验，对健康自我管理至关重要。这些研究结果表明，即便是完全脚本化的简单数字教练也能确实创造积极的辅导体验。

与此同时，该研究强调了在数字教练设计中引入更多计算智能的诸多机遇。首先，个人智能可通过整合个人数据，帮助数字教练立足于个体的个人偏好、生活方式及具体情境。其次，情境智能能融合上下文信息，使教练体验与不同情境的独特需求相匹配。第三，

情商则有助于数字教练根据个体的情绪状态、追求目标的进展程度来调整沟通语调，

以及他们对情感支持的需求。最后，对话智能能够实现更流畅自然的对话流程，并帮助参与者就广泛的生活经历展开更丰富的对话。这些不同类型的情商需要不同的计算解决方案。例如，当代基于LLM的聊天机器人特别适合提升对话智能。此外，过去在情感计算[72], 个人信息学[25, 55], 和情境感知计算[22] 领域的研究为支持上述其他维度的计算智能提供了洞见。然而，该研究强调需要将不同类型的计算智能整合成连贯的辅导体验，并平衡用户自主权与其对体验的控制感。

2 相关工作

2.1 个人健康技术

与个人历史相关的数据日益增长的可获取性，激发了人机交互（HCI）领域新一轮的研究浪潮，其重点在于促进个人与个人数据的互动[25, 26, 42, 46, 55, 57, 62]。李等人提出了

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

一个分阶段的个人信息学模式，概述了若干个体参与个人数据的各个阶段[55]。其他研究人员扩展了这一模式，并探讨了个人如何将个人数据融入日常生活[25]。许多个人信息学工具采用所捕获数据的可视化表示形式自我监测帮助个体识别其记录中的模式(例如 [2, 18, 56])。例如，Medynskiy和Mynatt的系统Salud!允许用户指定疑似影响幸福感的活动及相关生物标志物(例如在办公室度过的时间与情绪)，并通过交互式可视化直观检查其中这些数据可能的关联性[64]。Epstein等人开发了针对特定问题相关变量的可视化

由个体用户根据其自我追踪目标[27]。类似地，梁等人开发的睡眠探索者可视化呈现了睡眠模式与其他参与者追踪数据[58]的关联。某种程度上,

与这些可视化方法形成对比的是，本特利等人采用简单的统计推断来通过计算识别捕获数据中的相关性，并使用语言{v1}向用户展示发现。Karkar提出的自我实验框架

语言[5]。Karkar提出的自我实验框架et al. 帮助个人评估关于特定行为与健康结果[47]之间关联的假设。然而，这些方法大多侧重于识别趋势并将制定缓解策略的任务留给用户，这可能对低文化水平社区[86]构成显著障碍。个人信息学领域的最新研究提倡开发能够依赖计算数据分析来识别显著趋势，并通过预测和推荐[40]为行动提供更直接的支持。然而，此前仅有少数研究者明确

探讨过如何更直接影响个体行为的计算推断方法[20, 21]。

2.2 健康指导

2型糖尿病等慢性病的自我管理通常需要调整日常行为，包括饮食、锻炼和睡眠，

以及提高服药依从性以改善血糖控制并避免衰弱并发症[10]。行为改变是一项具有挑战性的任务，因为它需要持续动机、技能与知识以及自我效能[73, 74]。健康指导已成功促进个体生活方式的健康改变[37]。

尽管文献中存在多种辅导定义，但多数先前研究将辅导描述为以达成个人目标为核心：教练与客户共同确定健康目标，并制定实现这些目标的具体计划。

辅导文献强调教练与客户之间的个人关系、反馈、动机以及技能与知识的获取作为辅导过程的一部分[69]。此外，

辅导特别注重促进和支持客户的赋权、自主性及其能力提升，而非严格遵循规定的行为[38, 65]。

尽管有大量证据表明人工指导干预的益处，但由于依赖训练有素的人类教练，其覆盖广泛多样人群的能力有限。技术介导的指导覆盖范围更广；然而，这些干预措施可能仍无法惠及

医疗服务不足和经济弱势的社区，这些社区本就难以获得健康从业者和资源

[24, 76, 81]。对话代理和聊天机器人的广泛普及为更多无需人工干预的自主指导创造了新机遇不依赖人类教练。然而，最近的研究认为人际技能是成功健康指导的关键组成部分。例如，鲁特耶斯提出，关系建立、适应不同情境因素的能力以及同理心，是健康指导中不太可能被现代计算技术{v1}复制的关键要素。其他人也认为人类教练和自动化辅导各有其优势和局限[66]，，但关于健康领域全自动化辅导干预的可行性，以及这些系统能否在没有人类教练参与的情况下创造积极的辅导体验仍存在开放性问题。

2.3 健康领域的对话代理

对话代理是一种能与用户以自然语言交互的软件程序。这包括口语、书面或打字文字，或两者的结合。早期对话代理通过规则来响应基于用户输入，营造出人类智能的假象[89]。此后，对话代理被广泛应用于各种用例。有些是基于任务的，而另一些则模拟开放式社交闲聊。

麦克提尔 [63] 描述了三种主要类型的对话代理。有限状态或基于规则的系统遵循一组基于用户输入的预定义规则

以及系统当前状态。在框架或基于模板的系统中，代理为每个可完成的任务配备一个模板，其中包含必须填写的实体或槽位以完成任务 [29, 53]。第三种类型，基于代理或人工智能代理，利用机器学习来利用大规模对话语料库学习适当的回应；然而这些代理可以非常强大且更接近人类之间的互动但训练所需的大量语料可能在许多领域{v1}并不存在大型语言模型的进展为新一代对话代理铺平了道路它们不受特定领域限制的约束鉴于大型语言模型是在海量数据上训练的在书面语篇中，基于LLM的聊天机器人能够就广泛话题展开对话。

尽管对话代理的广泛普及相对较新，但人机交互及相关社区已有大量文献探讨对话代理在健康领域的适用性。拉兰霍与

同事系统调研了对话代理在医疗保健中的应用[53]。其综述涵盖的应用程序多数为专注于心理健康，并涉及其他健康领域的抽样调查如哮喘和营养。许多对话代理实施了特定的临床方案，例如认知行为疗法[29] 或简短动机干预[59]。所识别的代理均未采用基于人工智能的技术；全部为基于规则或基于框架的系统。持续聚焦于基于规则和脚本化代理的部分原因在于较低健康领域中对错误的容忍度。此外，由于数据安全和隐私保护，健康相关数据集很少公开提供给研究人员使用；因此缺乏健康领域中公开可用的对话语料库[60]。然而目前正在涌现一些基于人工智能的聊天机器人在健康领域的应用案例，特别是在糖尿病自我管理方面。例如，Gong et al. 介绍了劳拉，一个具身移动教练，它提供

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

定制糖尿病教育，并能够回答问题和与参与者进行对话[35]。类似地，Alloatti等人介绍了AIDA，一个用于为糖尿病患者提供治疗教育的对话代理[1]。然而，这两项先前的研究更多地关注教育而非个人目标的设定和追求。许多研究探讨了个体与对话代理互动的体验，并强调了其益处。

对代理的常见反应包括责任感、
对代理产生及感受到的共情感受，以及识别出代理在与用户建立关系过程中展现的个性[12, 29, 30]。Bickmore及其同事证明，对话代理能帮助向低文化水平医院患者解释出院文书[7]，并成功指导老年人增加活动量[8]。对话代理还可促进围绕医疗问题的敏感讨论和计划制定，例如在姑息治疗场景中[87]。

与此同时，近期研究发现了若干挑战
特别是在健康领域使用对话代理时，尤其是新型基于LLM的对话代理。
Jo等人研究了基于LLM的聊天机器人在
为独居者提供情感支持方面的使用体验[44]。尽管
该研究总体上强调了对接受此服务的个体和
监控这些互动的医疗保健提供者双方的诸多积极体验，
但他们也指出在确保从具有一定不可预测性的对话代理
获得适当响应方面存在困难
以及这些对话代理（CA）与需求和期望之间可能存在的错位
医疗专业人员的需求和期望。

3 方法

3.1 T2教练

3.1.1 迭代设计。我们采用以用户为中心的迭代设计方法开发了T2教练，这些方法包括与来自种族、民族和经济背景多样化社区的糖尿病患者组成的焦点小组合作。

为此，我们开展了一系列焦点小组活动，参与者包括从联邦合格健康中心（FQHCs）招募的23名2型糖尿病患者以及这些中心的6名糖尿病护理专业人员。在最初的焦点小组中，我们讨论了自动化聊天机器人教练的概念，并收集了用户对其在不同情境下理想行为的看法。通过这些初步焦点小组后，我们寻求符合用户需求的临床辅导标准，从而确定了BAP作为辅导协议。随后，我们利用BAP构建了初始的辅导对话集，并通过更多焦点小组收集了关于T2教练中目标和行动计划的清晰度、吸引力及实用性的反馈，以及对其对话适当性的评价。与焦点小组同步，我们完成了为期两周的T2教练绿野仙踪试验，以在全面开发该代理前优化对话流程。通过这些设计活动的反馈，我们发现总体而言，参与者认为目标驱动的辅导可能非常有用，且目标和对话易于遵循且恰当。然而，

参与者并不总是愿意继续进行非常冗长的对话流程，因此需要庞大的内容库来确保目标和行动计划的多样性，以适应不同的偏好和生活方式。为此我们缩短了对话的整体长度，

为提升清晰度编辑了个人消息，并扩充了内容

将可用目标库从9个扩展到17个，可用行动计划从
每个目标从3-6轮缩减至6-9
轮。

3.1.2 T2教练功能。T2教练包含两个主要组件，
一个提供目标导向辅导的对话代理（CA），以及用于追踪个人餐食和血糖水平的移动应用。

在设计T2教练的对话代理时，我们遵循了既定的简短行动计划协议（BAP；[38]）作为脚本化对话流程的基础。BAP定义了一套步骤，供健康从业者引导个体选择健康目标并制定具体实现计划。T2教练中的目标和行动计划内容源自先前针对2型糖尿病患者的健康目标知识库。

与BAP一致，T2教练包含两个主要对话：较长的每周交流用于设定健康目标（即目标设定对话），

图2）以及一个更简短的每日跟进交流以检查目标进展（每日目标反思，图3）。较长的目标设定对话包含以下组成部分：1）行为目标菜单（每次目标设定对话中随机排序），2）对所选择目标的简短解释；这些解释由一组注册营养师（RD）编写，并在另一项研究中评估了用户对其易理解性和文化适宜性的评价，3）使用信息图对目标进行插图说明，

4）符合所选目标的一餐示例（从我们先前研究的参与者记录的餐食集合中选择），5）行动计划菜单——可帮助实现所选目标的具体行动，6）设置目标提醒的选项（用户可选择是否接收提醒，但无法更改提醒时间），以及7）对其选择的确认（图1）。这些对话的长度可能因个人在最终确定选择前选择查看的目标和行动计划数量而有所增减。

较短的每日目标反思对话包含以下组成部分：1) 目标达成率评估，2a) 正向强化并重新确认每日行动计划，或2b) 反思障碍及制定未来克服障碍的计划（图2）。

最后这两个选项旨在鼓励用户进行反思和主动规划。除了这些对话外，

T2教练每天会发送一次简单的非互动性消息，其中包含所选目标的提醒以及改编自先前干预措施的简短激励性文本，以促进2型糖尿病的自我管理[17]。

T2教练移动应用要求用户通过拍摄照片和文字描述来记录餐食，并记录餐前餐后的血糖水平（图2）。为促进即时反思目标达成情况（并追踪长期目标达成率），选择营养目标的用户在记录餐食时会收到提示，评估新记录的餐食是否符合其选定的营养目标。用户可在T2教练日志中查看已记录的餐食及对应血糖水平。

3.2 参与者

本研究的参与者通过我们大学面向公众的招募网站以及社交媒体和克雷格列表上的广告进行招募。要符合参与资格，

参与者年龄需介于18至65岁之间，且熟练

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 (SD-SCA),[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).

使用英语，并能自我报告确诊为2型糖尿病或糖尿病前期
拥有一部功能正常的智能手机及数据套餐，并能下载、安装和使用智能手机应用。排除标准为自述有严重共病或受伤（如艾滋病、癌症）以及认知障碍。本研究经大学机构审查委员会批准；所有参与者在参与前均签署了电子同意书。

3.3 研究流程

所有研究流程均远程进行。在初始培训课程前，参与者需审阅并签署电子同意书，并填写一组基线问卷。在通过电话或Zoom视频会议进行的虚拟培训课程中，参与者观看了一段预先录制的视频，内容涵盖糖尿病自我管理和目标导向辅导的基本原则（由研究团队中的糖尿病教育者准备），以及一段介绍T2教练功能的视频。之后，参与者被告知他们将有机会与一个完全数字化的、通过短信聊天机器人形式呈现的教练合作。参与者每周可设定一次自我管理目标，并每天选择具体行动以实现这些目标。鼓励参与者根据自身便利程度自由使用T2教练，无需设定最低参与量。接着，参与者需回应第一组辅导对话（包括目标设定对话），并在研究团队成员在场协助解答问题时，使用移动应用记录一餐。培训结束后，参与者被要求使用T2教练3-4周。若参与者在培训后3天内未使用该应用程序，研究人员会联系他们以解决潜在技术问题。

在研究最后一周，参与者受邀通过电话或Zoom视频会议参与定性半结构化访谈，分享他们的体验。访谈遵循一份访谈指南，内容涵盖个人整体自我管理情况、对T2教练各项功能的使用体验，以及对该应用程序辅导体验的感知（完整访谈指南见附录）。此处，

我们特别关注他们体验中存在的差距与限制，这些可能为更智能的设计提供改进机会。此外，参与者还需填写研究后调查问卷。

所有访谈均进行录音并逐字转录以供分析。

研究指标包含多项独立结果。健康结果涉及糖尿病自我效能（糖尿病自我效能量表，

DSE, [9]）、研究期间设定的自我管理目标自我报告达成率，以及自我报告的自我管理活动情况，

包括健康饮食、体育活动、血糖监测以及其他几项（糖尿病自我护理活动总结（SD-SCA），[85]）。为评估参与者对T2教练对话质量的感知，我们采用了语音系统主观评估量表（SASSI）[39]。最后，为评估个体对T2教练辅导体验的评价，我们聚焦于共同决策这一辅导的核心要素（改编版共同决策问卷（SDM-Q-9），[52]）。

3.4 数据分析

针对定量数据，我们计算了应用程序使用情况、调查措施以及研究不同阶段目标达成率的描述性统计。关于与聊天机器人的互动，我们统计了对话平均长度、用户回复每日消息的频率、响应耗时以及完整完成对话的比率。同时，我们还计算了每位用户及整体所选目标和行动计划数量的描述性统计。对于自我追踪移动应用程序，我们汇总了研究期间记录的餐食次数和血糖读数。至于仅后测调查测量，

我们报告了各量表及子量表的总结分数。针对使用SDSCA量表进行的自我管理前后研究比较，我们对相关子量表（常规饮食、锻炼和血糖检测）计算了配对样本t检验，并应用Bonferroni校正进行多重检验。关于目标达成率，Mitchell等人先前的研究表明，个体对目标达成率的自我评估随时间推移比专家评估更为乐观；然而，个体与专家对目标达成率随时间变化的评估具有一致性[31]。与此一致，我们在本分析中采用了个体的自我评估数据。我们使用单变量线性回归分析目标选择后天数与每位参与者平均目标达成率之间的关系。

对于定性数据，我们采用主题分析法来识别主要
参与者访谈[11]中的主题。首先，四位作者共同
审阅了3份转录本，通过多次协作编码会议进行开放式编码，并开始识别新兴

主题。作者们讨论了标签及其含义，并
使用笔记记录这些讨论内容。所有分歧均通过讨论达成一致。这些会议最终形成了初始的编码方案。

之后，由其中一位作者对剩余的访谈进行编码。在此期间，作者们每周会面讨论新发现与

编码方案的修改。新的访谈内容会在发生时
立即转录并编码。在完成15次访谈后达到数据饱和；
由于未再识别出新主题，我们停止了
受试者招募。为提高分析严谨性，我们采用了三角验证
并将定性分析的结果与使用日志分析的结果进行交叉验证。
我们使用Lumivero的nVivo软件进行定性编码（NVivo 14.23.0 / 2023年3月14日）。
编码（NVivo 14.23.0 / 2023年3月14日）。

4 结果

4.1 参与者人口统计

本研究共招募19名参与者，其中16人完成研究并参与定性访谈。未完成研究的3人中，1人在入组次日退出，2人未回应研究后访谈邀请。参与者人口统计信息如下表1所示。总体而言，

参与者主要为女性，种族和民族多样化，受过教育（样本中64%拥有大学和研究生学位），且处于超重与肥胖的临界值（体重指数超过25并接近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).

表1：参与者人口统计

人口统计	数值
Sex	63%女性；37%男性
民族	26%西班牙裔
Race	53% 黑人 or 非洲裔 美国人
	32% 白人
	15% 其他种族
Age	45 ± 20 岁（中位数 47 岁）
教育水平	32% 研究生学位
	32% 大学毕业生（学士学位）
	32% Some college or 技术 学校
	4% High 学校 毕业生
身体质量指数（BMI）	29.2 ± 15
家庭收入中位数	$40,000 - $59,999
糖尿病诊断	68% 2型糖尿病 32%
	糖尿病前期

表2: 总体使用统计

使用统计	每周目标设定	每日反思
响应率	73.2%	71.2%
Percent of started conversations that were completed	86.5%	89.43%
平均对话轮次	9.9	4.1
首次响应时间	2.3小时	2.1小时

4.2 使用统计

总体使用统计数据显示于表2。使用分析日志显示，每周目标的总体响应率均较高设置对话框和每日反思对话（每种占比均超过70%）。不出所料，每周目标设定对话相当日常对话的平均轮次接近10次，时间更长反思对话明显更简短，平均每次对话仅略超4轮交流。

在本研究中，17个健康目标中的每一个都至少被一名参与者选择过。参与者在最终确定目标前会花些时间浏览可选目标：平均每次目标设定对话中他们会探索5个不同的目标。参与者平均设定在研究期间设定了2个目标，并为每个目标选择了1.5个行动计划目标总计尝试了3.5个行动计划（在研究期间）。他们记录了略多于54餐（每日2.5餐），而那些追踪了血糖的人群（不包括糖尿病前期参与者）记录了72.5次血糖读数（每日3.5次），这表明很少有记录的餐食同时包含所需的餐前和餐后血糖水平用于个性化目标。

4.3 目标达成率

健康指导的主要目的之一是帮助个人设定并追求健康目标。在此，我们讨论对目标达成率的分析。

在所有参与者和研究周次中，研究期间自我报告的目标实现平均值为80%。参与者平均每项选定目标的持续时间为1-2周（平均10天)。15名参与者记录了至少1项目标在研究的所有四周内进行自我评估，而那15本分析共纳入参与者（排除2名仅在研究第一周有评估数据的参与者仅在第一周的研究中拥有评估数据的人）。

所有参与者均在研究首日设定了首个目标，且全部参与者至少持续跟进所选目标达5天。如图4所示，总体而言，在这5天内参与者的目标达成率有所提升；

该趋势接近统计显著性（R^2 = 0.044，F(1,75) = 3.473，p值= 0.066以及β =0.038，这意味着自那天起每过一天选择目标时，出现了3.8个百分点的增长趋势。在目标达成率方面。

第一周过后，由于不同参与者持续致力于目标的时间段各异，并在研究的不同时间点切换目标，对目标达成率的分析变得不那么直观。图5（顶部）展示了他们在整个4周研究期间自我报告的目标达成情况。随时间推移对自我报告的成功实现目标的分析显示，研究期间存在改善趋势，尽管该趋势未达到统计学意义。值得注意的是，在研究第3周出现目标达成率下降，当时许多参与者转向了新目标。

这一趋势在饮食目标和身体活动目标上均保持一致。活动目标，当时目标达成率有所下降个体选择了新目标（第3周），并逐步增加目标达成率随时间变化（图5，底部）。

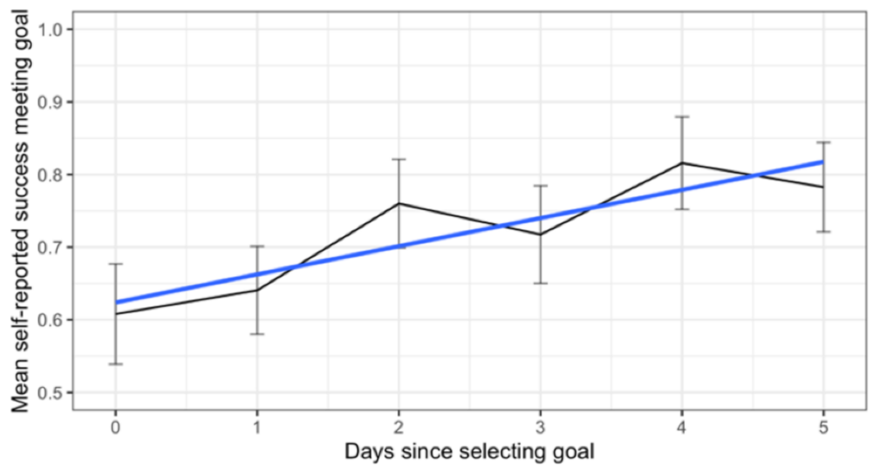

Figure 4: Average goal attainment during the first week of the study with a regression line. The mean on the Y axis ranges from 0 (none of the meals met the goal) to 1 (all of the meals met the goal).

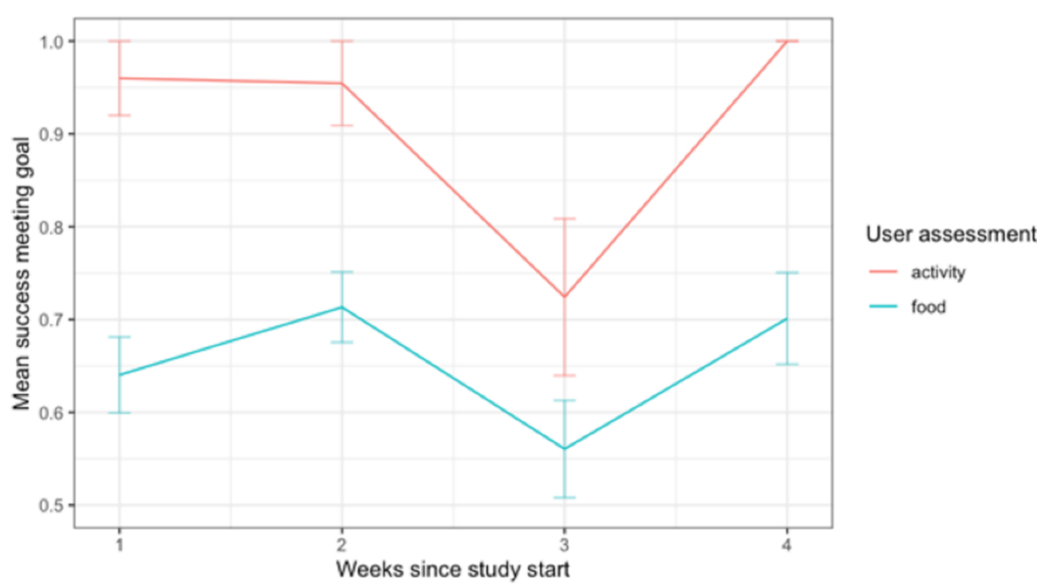

Figure 5: Average goal attainment for different study weeks with a breakdown for nutritional (food) and physical activity (activity) goals. The mean on the Y axis ranges from 0 (none of the meals met the goal) to 1 (all of the meals met the goal).

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.

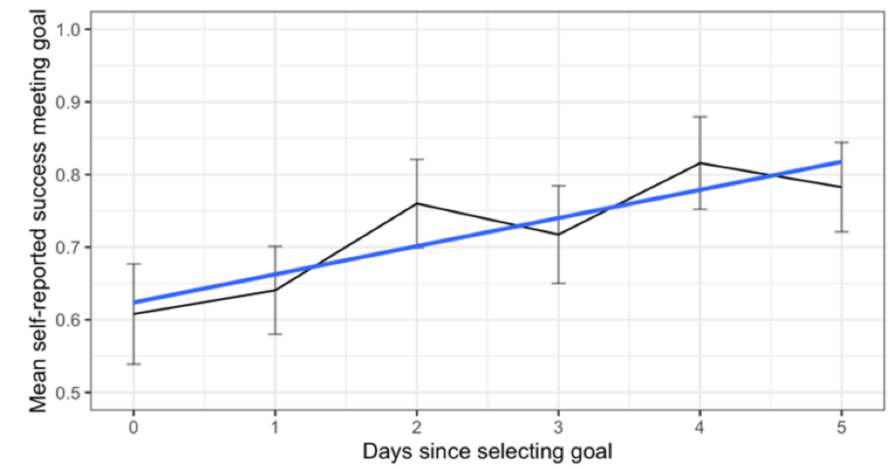

图4：研究第一周的平均目标达成率及回归线。Y轴上的平均值范围为0（所有餐食均未达成目标）至1（所有餐食均达成目标）。

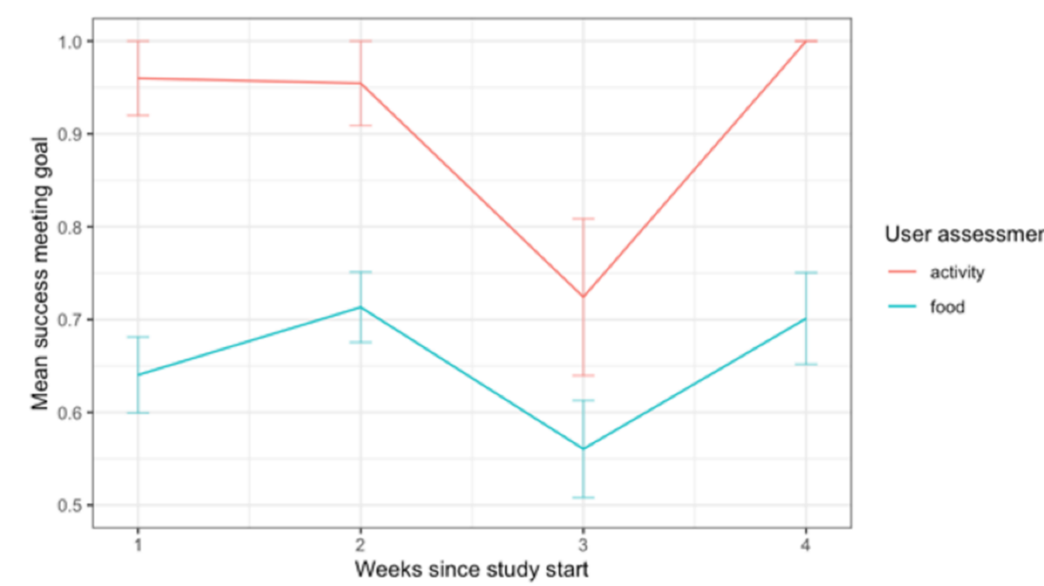

图5：不同研究周的平均目标达成率，按营养（食物）和身体活动（活动）目标细分。Y轴上的平均值范围为0（所有餐食均未达成目标）至1（所有餐食均达成目标）。

4.4 调查措施

在共享决策量表上，***T2***教练获得了高分评分，这7个问题的平均中位数分数为5
参与者对那些涉及他们的自主权和选择自由的问题给出了更高的排名（例如，T2教练想要确切知道我希望如何参与选择一个健康目标），而对那些涉及选择目标的理由及其益处（例如，T2教练详细解释了健康目标选择的优缺点）。

在用户体验SASSI量表上，T2教练获得了5分制下的整体中位数分数5分（整体平均值4.03 ± 0.45），其中分数越高表示评估越积极。中位数分数不同子量表的详细结果见表3。可以看出，参与者对T2教练在系统响应准确性、好感度、认知需求和速度方面评分较高，而在烦恼和宜居性方面评分较低。

个体对SDSCA量表的反应结果与其子量表包含在表4中。参与者报告了更频繁地实践与饮食相关的自我管理行为，锻炼和血糖测试，然而这些差异在统计学上并不显著。

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

表3：SASSI子量表

SASSI子量表	中位数	Mean
系统响应准确性	5	4.06 ± 0.4
好感度	5	4.28 ± 0.27
认知需求	5	4.23 ± 0.26
烦恼	4	3.48 ± 0.66
宜居性	4	3.8 ± 0.14
速度	5	4.13 ± 0.19

4.5 定性结果

在本节中，我们首先提供一些关于参与者背景的详细信息，包括他们的生活经历和糖尿病自我管理经验。然后我们展示主题分析的结果，该分析确定了以下主要主题：

1) 保持正轨；2) 个体差异与个性化需求，3) 适应日常生活节奏，4) 适当鼓励，5) 对话流畅性，以及6) 自主性与掌控感。下文我们将阐述这些主题并通过参与者引述加以说明。引述标注采用招募时分配的参与者编号，因此这些数字可能超过研究中实际参与者数量。

4.5.1 参与者背景。研究参与者在通用背景及糖尿病自我管理方面具有广泛多样性。这些差异部分源于糖尿病发展阶段的不同。

例如，糖尿病前期患者虽有监测血糖的经验，但多数人

依赖每3至6个月采集的实验室数值，而非日常血糖水平来追踪进展。确诊患者则2型糖尿病患者在追踪血糖的方法上也各不相同。有些患者使用连续血糖监测（CGM）来追踪血糖，而其他人则依赖每日甚至每周的检测。大多数人依赖口服药物或胰岛素或两者的某种组合。许多患者已建立了完善的自我管理常规，包括严格遵守饮食和定期锻炼计划。然而，讨论最多的是自我管理面临的诸多挑战，涵盖范围广泛从环境障碍到参与健康行为再到缺乏自律，以及对糖尿病和营养知识的匮乏等等诸多方面。

关于他们使用技术的体验，所有参与者都拥有智能手机且大多数人依赖短信作为首选通讯方式沟通媒介。事实上，大多数参与者更倾向于使用短信而非电话交谈和消息应用，例如WhatsApp和Facebook Messenger。

4.5.2 主题1：保持正轨。所有参与者参与本研究时都至少具备一些管理自身糖尿病（或糖尿病前期）的经验，并且熟悉持续自我管理面临的诸多障碍。其中部分障碍是内在的，例如缺乏动力、渴望的挣扎以及缺乏关于如何适当改变生活方式的知识。其他障碍则是外在的，与缺乏获取健康食品的途径以及烹饪和准备健康餐食所需的时间和资源有关。面对这些挑战，参与者谈到了保持参与和有动力的必要性，并坚持不懈地采纳和维持新的行为和克服障碍。在此背景下，参与者发现T2教练通过其一系列一致的每日消息，创建了一个可靠的持久结构来帮助他们"保持正轨"。

参与者们讨论了这一结构中他们认为特别有帮助的几个组成部分。例如，每日目标提醒通常在早晨发送给参与者，

这些提醒在为一天定下基调和作为日常活动的导向方面非常有用：

> "早晨的提醒。因为那是一天的开始。你起床后看到它，就会说，哦对了，让我确保这就是我接下来一天要做的事。"（P1，51岁，女性，2型糖尿病）

每日反思和目标达成检查有助于建立责任感。许多参与者用'让我保持诚实'来描述他们使用T2教练的经历，尽管这种责任感的来源因人而异。对一些人来说，这是内在责任感，鼓励参与者对自己诚实。对其他人，如P6，聊天机器人创造了一种外在责任感，让他们感觉有人在监督自己。

> "我知道[T2教练] 会问我，我不想处于一个必须回答'不，我没有'的境地。"（P6，48岁，女性，2型糖尿病）

每周目标达成总结向参与者展示了他们能够实现目标，并鼓励他们坚持不懈他们的目标达成率：

> "嗯，我会说这很鼓舞人心。这意味着，好吧，我这周做到了。而且我相信只要坚持自己设定的目标，我就能持续做到。所以，这是一种积极的鼓励，告诉我是的，你能做到。"（P24，52岁，男性，TD）

大多数参与者对这种消息推送的一致性表示赞赏，并不觉得有负担或压力。即使是那些对消息频率感到些许厌烦的人，也认为这是实现预期改变的必要手段。例如，P4将T2教练比作父母，有时唠叨，但能有效激励良好行为：

> "嗯，我告诉你，感觉T2教练就像我的父母。提醒我好好吃饭，这是好事，因为如果放任自流，我一周可能会偷懒两天。但T2教练真的让我保持诚实。它会说，好吧，等等，你今天没达成[目标]，所以明天和这周剩余时间必须完成。我就是这样激励自己的。"（P4，58岁，女性，T2D）

许多参与者将这种一致性与他们之前与人类教练的经历进行了积极对比，后者通常表现为不频繁的咨询且缺乏跟进，往往难以激发预期的改变。相比之下，

T2教练被认为可靠、准时且勤奋：

> "嗯，这个应用程序，你知道的，要准时的多也勤奋的多。所以你看，它是自动运行的，而我只需要认真对待并回应它就行。至于初级保健医生，你

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.

表4：研究前后测量的糖尿病自我护理行为摘要（SDSCA）对比，测量单位为上周天数（0-7）

SASSI子量表	研究前	研究后	差异（p值）
一般饮食	4.4	5.2	0.8 (0.26)
锻炼	3.6	4.4	0.8 (0.26)
血糖检测	3.0	4.7	1.7 (0.19)

一年才见一次。当她告诉你要做这个或那个，或者你得试试用蔬菜代替米饭之类的时候，她这么说的时候，你知道，感觉就像是，哦好吧。这是我的医生在说，你知道的，尽量照做。当然，我其实没那么擅长坚持"（P3，43岁，男性，2型糖尿病）。

4.5.3 主题2：个体差异与个性化需求。尽管参与者赞赏一致的日常消息带来的安心稳定感，但他们也认为这些消息需要适应各自不同的日常作息和习惯。参与者的生活方式、文化认同、

饮食偏好及生活状况存在显著差异。部分参与者拥有相对固定的日程，能轻松适应传统的一日三餐；而另一些人因轮值夜班，不得不根据非常规工作模式调整进食时间。有人完全自主决定餐食选择，也有人依赖父母或伴侣负责食品采购与烹饪。这些差异构成了他们糖尿病自我管理实践及使用T2教练应用程序的丰富背景。

例如，参与者在目标和行动计划的选择上存在巨大分歧。部分人选择的目标是与现有习惯一致，可以轻松融入他们的日常习惯中。然而，其他人选择的目标却与当前行为截然相反，需要付出相当大的努力。在许多情况下，这些目标与个人对自我管理的渴望一致，他们将这次研究视为最终取得进展的机会。

> "我想，就像我一直背负的内疚感，我知道自己应该多吃水果和蔬菜。这让我真正意识到这一点，并促使我付诸行动。"（P20，38岁女性，2型糖尿病）

在此背景下，参与者希望T2教练能更了解他们的个人偏好、生活方式和日常习惯，并能够适应这些因素以创造更具个性化的体验。在某种程度上，这种个人体验感得益于系统提供的丰富目标和行动计划选择。T2教练以及能够根据需要频繁更改行动计划的能力使他们能够获得所期望的灵活性。这种灵活性至少营造了一种个性化体验的感知，即便在缺乏实际数据驱动的定制的情况下。

> "……事实上它包含了你选择的一切……这无疑让它感觉非常个人化。"（P18，53岁，女性，2型糖尿病）。

然而，参与者认为这种体验可以极大地通过更广泛地利用其数据得以增强，具体表现在既能针对性地选择目标，又能定制行动计划以满足其需求生活方式，并提供更多针对活动的定制化反馈他们记录的内容，例如他们的餐食、步数和血糖水平。

> "你记录下你的餐食，我希望T2能给我一些反馈，比如我是否摄入过多卡路里，或者健康餐盘应该是什么样子。"（P6，48岁，女性，2型糖尿病）

4.5.4 主题3：适应日常生活节奏。除了适应个人生活方式和偏好外，参与者经常谈到不同情境和背景下的独特需求。这对于作息不规律的参与者尤为明显，他们需要调整所有日常活动。例如，P18是一名代课老师，其作息在学年期间相当规律，但在夏季（研究进行期间）变得松散，

对他们而言，在夏季月份记录餐食并追踪血糖水平带来了挑战——若研究在学年期间进行，这一挑战本不会出现。

> "因为我的生活特别缺乏规律性，尤其是现在，因为我在学校工作。我是一名代课老师，目前休假。所以，比如到了九月份，我的生活会比六七月时规律得多。因此，我在追踪方面做得不好，尽管我相信追踪的重要性，这也正是吸引我参与这项研究的原因——我希望有人能督促我坚持追踪。"（P18，53岁，女性，2型糖尿病）

即便是那些生活更规律的人也觉得这款应用程序无法很好地适应他们日常活动和作息的变化。例如，参与者2在某天做了结肠镜检查在研究期间的日子里，他们无法遵循自己的常规流程来为手术做准备。

> "我在医院住了两天，其中一天是结肠镜检查准备。所以那几天很奇怪。我想这也是假设每天都是一样的，然后节假日就像平常日子一样。[我希望] 有办法标记，这是特殊的一天。我们能为这一天想出什么办法？"（参与者2，50岁，男性，2型糖尿病）

这一经历尤其令人沮丧，因为参与者2在其一条日常反思消息中将结肠镜检查描述为障碍。这些日常习惯的中断有时会让这款应用程序及其定时发送的消息感觉与个人的生活节奏脱节。

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.

4.5.5 主题4：适当的鼓励。在整个访谈过程中，参与者们谈到了与糖尿病自我管理相关的多重挑战和障碍，以及持续保持动机和坚持的重要性。参与者们讨论了他们的社交网络以及来自朋友、家人和医疗保健提供者的鼓励在帮助他们保持积极和动力方面的重要性。在此背景下，参与者们描述了他们与T2教练的互动是积极且鼓舞人心的。他们发现，聊天机器人在目标设定和每日反思对话中使用的语言能够在消极、引发内疚与过度积极之间取得平衡：

> "是的。感觉很有鼓励性。它不像惩罚或羞辱，或者真的带有负面情绪。感觉像是，'好吧。是的。很好。再试一次。明天还有一天。你能做到的。'所以，我觉得这种鼓励恰到好处。"（参与者2，50岁，男性，2型糖尿病）

所有参与者都清楚T2教练是完全自动化的；然而，许多人认为T2教练缺乏人类特质反而是一种优势，而非限制。例如，参与者2欣赏能够直接理解对话的表面意思，而无需担心人类沟通中常见的隐含意义：

> "T2教练很有帮助。它既不居高临下，也不过于挑战性。这很好。而且因为它，我不知道，是中立的，所以接受建议似乎更容易。我不介意，而且我知道T2教练没有——没有潜台词。所以，如果是一个人类，也许会说一些话，我可能会想，'哦，你是在暗示什么吗？你说那很艰难也很好——我做得很好，还是我没做好？但直接从T2教练那里接受表面意思让事情变得更容易。'（参与者2，50岁，男性，2型糖尿病）

此外，对一些参与者而言，没有人类介入反而降低了他们担心被评判的焦虑以及暴露不良习惯的尴尬，这两者都会对整体辅导体验产生显著的负面影响。

> "但事实上它不是人类反而让事情变得更容易。尤其是在刚开始你觉得自己有很多坏习惯的时候。我甚至会羞于承认自己有多少坏习惯，所以这样反而更容易开口。我是说，现在我感觉好多了因为我取得了很大进展，但即便如此，如果不是人类可能还是会更容易些。"（P23，58岁，女性，糖尿病前期）

然而其他参与者则感到需要更持久且专注的鼓励，尤其是在目标达成出现失误之后。

> "……你知道的，使用这些应用程序的人并不总能做对。你可能需要一点鼓励来——来重新跟上计划进度……"（P3，43岁，男性，2型糖尿病）

这些参与者感知到了情感的一致性
在T2教练中，语气被视为一种限制，因为它与他们的初衷不符。
情感需求。

4.5.6 主题5：对话流畅性的重要性。尽管参与者对T2教练简单、完全脚本化的性质感到安心，

他们还认为，辅导体验中缺乏所期待的对话流畅性。使用常规短信作为对话媒介带来的技术挑战加剧了这一问题：网络延迟偶尔会导致消息乱序，从而打断对话的连续性。此外，目标达成问题的是/否二元选择也显得过于局限且缺乏细微差别。部分参与者（如P18）希望能进行更流畅的对话，单纯"把事情聊开"，因为这有助于他们更清晰地认识自身选择并反思行为：

> "因为你知道交谈如何能持续提升你对自己脑海中一切事物的觉察。所以，把事情说出来有助于你意识到自己的习惯。"（P18，53岁，女性，2型糖尿病）

因此，一些参与者，如P1，将T2教练视为
缺乏人类对话的灵活性与流畅性，且过于
机械：

> "实际情况是，感觉有点呆板。就像我知道自己在和电脑说话。"（P1，51岁，女性，2型糖尿病）

4.5.7 主题*6*：自主性与掌控感。最后，一个贯穿所有
访谈的持续性主题是，对T2教练的辅导体验以及
自身状况保持掌控感的重要性
参与糖尿病自我管理的体验。选择个人目标和行动计划的灵活性以及按照自己的时间和方式
与T2教练互动的自由性增强了
参与者的自主性和控制感：

> "我觉得教练是在询问我，仿佛我才是专家。因为这是我，掌控着自己生活的人。所以，回答问题的人应该是我。"（P21，61岁，女性，TD）

部分参与者将这种体验与他们之前
接受人工辅导的经历进行了积极对比。例如，P18的
之前与人类教练的经历感觉过于刻板，因为
教练的时间安排和期望与他们自身不符，
尤其是在周末。这最终导致了不满
和烦恼。相比之下，参与者觉得他们可以随时
与T2教练互动，只要对他们方便，并且可以轻松地
离开对话并在稍后时间重新继续。

5 讨论

在本研究中，我们调查了个体与一个全自动全脚本对话代理（CA）的互动体验，该代理为2型糖尿病（T2D）及糖尿病前期患者提供健康指导。我们旨在探讨数字健康辅导解决方案设计的广泛问题。此外，我们关注于探究：哪些辅导体验要素可通过不依赖人工智能（AI）的简单对话代理实现，哪些体验要素需要额外的计算智能支持，以及数字辅导解决方案中需要何种类型的智能。本研究结果既揭示了简单全脚本聊天机器人的优势，也指出了通过增强计算智能来优化辅导体验的可能性。下文将对此展开详细阐述。

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

5.1 以简约成就卓越

本研究的核心问题之一是：简单全脚本对话代理能在多大程度上创造积极的辅导体验。结果表明，即使采用完全预设的简单交互形式且无人人类教练参与，T2教练仍能在一定程度上营造真实的辅导体验。

定量结果支持了这一结论。首先，目标达成率随时间的提升表明，即使在这项研究的短暂时间范围内，且即便未转化为参与者自我管理行为的感知变化（通过SDSCA量表测得），使用T2教练仍对个人自我管理产生了积极影响。此外，共享决策量表的高分表明T2教练有助于促进共同决策——这是健康指导的关键组成部分。最后，

SASSI量表的高分表明参与者总体上对对话质量感到满意。

定性结果揭示了T2教练促成这些积极认知的多个方面。首要且最关键的是，该研究强调了日常琐事协助的重要性及其在成功健康指导中的作用（主题1:

保持正轨）。其次，该研究强调了健康指导互动中避免内疚、羞耻和压力的重要性（主题4：适当的鼓励）。参与者认为T2教练的消息语调总体积极，与其互动直截了当且不存在人类互动中常见的隐藏议程或潜台词。在披露未达自身期望的行为时（尤其在研究初期许多人未能达成目标时），参与者几乎感受不到污名化或尴尬。第三，目标设定与行动计划选择环节内置的灵活性，帮助参与者获得了对辅导体验和自我管理的掌控感（主题6：自主性与掌控感）。参与者根据自身价值观和优先级选择目标，并在准备就绪时调整这些目标。此外，能够在自己方便时按个人条件使用T2教练，也增强了使用者的自主性与控制感。

这些发现与既往关于健康领域对话代理的研究报告一致。先前关于健康指导（包括电子指导）的研究多聚焦于辅导的情感与沟通层面[79, 80]。我们的研究拓展了这一视角，强调稳定、持久且一致的结构作为辅导另一关键要素的重要性。对人类教练而言，创建此类结构可能构成挑战，因其需要近乎持续的咨询可用性、提醒服务与建议提供[44, 80]。相比之下，本研究的参与者认为T2教练具有"勤奋"、"准时"、"持久"和"耐心"的特性。此外，

过去的研究强调，对话代理(CA)实际上能激发的参与度可与医疗专业人员相媲美，有时甚至更胜一筹。例如，Bickmore等人的研究表明，医院患者更倾向于从对话代理而非人类护士处接收出院指导[6]，而我们的研究显示这种态度同样适用于健康指导领域。与全自动化系统互动常使参与者感到放松，同时仍能体验到外部责任感带来的益处，这对目标达成率颇有助益。最后，先前

研究证实了自主性在健康及健康管理情境中的重要性[3, 4, 36, 77]。关于人工辅导的文献将赋权与自主性强调为辅导实践中最关键的定义性特征[19, 69]。

然而，先前的研究也表明，人类教练可能难以将自己的期望与客户的期望保持一致

优先级[66]。相比之下，我们研究的参与者感到有能力做出自己的选择

这些发现表明，即使是完全脚本化的对话代理也能创造有益的辅导体验，尤其是在人类教练不可用或难以接触的情况下，这对于许多经济弱势和医疗服务不足的社区可能尤为常见。尽管这些社区通常慢性病患病率更高，对辅导干预的需求更为迫切，但他们获得医疗专业人员（包括专业教练）的机会有限[54]。事实上，居住在这些社区的许多2型糖尿病患者从未接受过健康指导，甚至未接受过被视为标准护理的糖尿病教育[71]。此外，这些社区历史上与医疗机构关系复杂，对医疗专业人员的信任有限，这可能进一步阻碍他们与健康教练建立密切的个人关系[48]。因此，

整合人类教练的新解决方案不太可能使这些社区受益，反而可能加剧现有的健康差异，甚至引发新的干预导致的差异[88]。

5.2 对多种智能类型的需求在智能辅导中的应用

虽然该研究明确了简单脚本聊天机器人的益处，但也揭示了引入更多计算智能的诸多可能性。

首先，我们的研究表明，尽管灵活性和选择性能营造个性化体验的感知，但显然存在根据个体偏好、生活方式

及具体情境定制辅导方案的空间，这表明需要更多个人智能（主题2：个体差异与个性化需求）。我们的研究揭示了参与者

在工作日程、饮食习惯、食品采购与烹饪惯例、预算及社区食品供应方面存在的巨大差异。

考虑这些差异可以识别出适用于不同参与者的行为目标和行动。此外，

对通过自我监测获取的数据进行更智能的分析，能够针对其目标达成率和改进领域提供个性化反馈。其次，参与者讨论了在更持久的个人偏好与不同情境需求之间取得平衡的必要性，从而凸显了情境智能的重要性（主题3：适应日常生活节奏）。病假、季节性的日程变化以及其他意外情况会打乱正常的辅导体验，需要做出相应调整。第三，尽管参与者认为T2教练消息的语气令人感到安慰和安心，但他们希望采用更细致、定制化的方法，并提升情商（主题4：适当的鼓励）。这一点很重要，因为参与者对同理心和支持的需求各不相同，

有些参与者偏好偶尔的鼓励，而另一些则需要更广泛的参与。此外，这些需求还因

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

参与者的目标达成率以及他们是否按计划进行而异——若进展顺利仅需简单认可其成功，若出现失误则需更多鼓励以重回正轨。最后，尽管参与者赞赏全脚本化T2教练的简洁性和一致性，但也认为其内容僵化机械，暗示需要更强的对话智能（主题5：对话流畅性）。具体而言，参与者希望能进行更流畅的对话，

对目标达成率相关问题提供更具细微差别的回应，
并能像与人类教练那样"深入探讨"问题。
这些发现与先前研究用户参与不同类型对话代理的结论一致。

例如，过去的研究强调了辅导体验中个性化的重要性，并探索了利用个人数据优化人工辅导的不同方法[67]。此外，既往研究强调了辅导中情境的重要性，提出了通过数据提升辅导体验的途径[18]。类似地，早期研究也强调了情感支持与鼓励在辅导中的关键作用[19, 78, 80]。事实上，有学者认为同理心和情感支持是辅导的核心要素，这些独属于人类的特质无法通过数字辅导解决方案复现[79]。最后，多项关于基于框架的聊天机器人的研究表明，用户认为其过于僵化重复，缺乏日常对话应有的流畅性[29, 53]。

这些不同类型的情报需求需要不同的技术解决方案。例如，基于LLM的聊天机器人因其能就多样话题维持对话的能力，特别适合满足对话智能的需求。

事实上，近期研究表明这种灵活性确实能创造更具吸引力的体验[44]。过去研究指出这些新能力带来的一些挑战，并提出了潜在解决方案。例如，Jo等人注意到确保基于LLM的聊天机器人回答恰当性的挑战，

使其符合医疗专业人员的需求，以及它们无法维持对话历史[44]。应对这些挑战的新方案包括混合架构——要么通过在领域特定语料库上微调LLM来强制对话结构[43]，，要么将脚本对话模型与开放式LLM结合以实现结构性与灵活性并存[61]。

然而，该研究最重要的启示或许是需要将不同类型的情报整合成连贯的用户体验。真正智能的数字教练应能整合对话历史、用户当前健康与情绪状态评估、个人数据以及情境与上下文感知。人机交互社区在智能交互计算领域有丰富的研究历史，可满足其中许多需求。例如，情感计算领域有大量研究专注于识别人类情感并创建适应用户情绪状态的交互体验[72]。个人信息学通过整合个人健康和健康数据进行反思以增强自我意识，提供了诸多宝贵经验与创新方案[27]。

同样地，情境感知计算几十年来一直是活跃的研究领域[22]。当前需要的是新方法将这些不同能力与大语言模型的进展相结合。

该领域过去的工作主要集中在提示工程和

整合更复杂的信息，例如个人数据，
转化为提示[33]。未来研究有助于发现更多机会，将大型语言模型的对话智能与其他类型的计算智能相整合。

5.3 平衡计算智能与用户自主权

虽然我们的研究展示了在数字辅导体验中引入不同类型计算智能的诸多机会，但也强调了需谨慎维持增强智能与人类自主权及体验控制权之间的平衡。本研究中使用的简单对话代理为用户提供了灵活性：可自主选择参与时机与方式，根据个人优先级和偏好设定目标，并按自身节奏实现这些目标，无需承受压力或责备。这种灵活性使参与者感受到对辅导体验和自我管理的主导权，许多参与者认为其体验优于过往与人类教练的合作经历。

个体自主权的重要性已在多个研究领域的前期研究中得到充分确立[4]。在健康领域，它催生了声势浩大的患者赋权运动[3] 以及作为促进行为改变方法的动机访谈[65]。这也是健康指导的核心所在，大量辅导文献强调需要提升患者自主权，并将教练角色定位为协助个体实现自身目标，而非遵循外部强加的标准[78]。

交互系统中计算智能的提升为创造人机之间新颖有趣的互动形式带来了希望。然而，这也可能对人类自主权和控制力产生负面影响。人机交互社区已在以人为中心的人工智能框架下探讨了这些问题。

先前对以人为中心人工智能的探讨指出，人类自主权和控制力是该类系统的基本特征之一[13, 15, 84]。在数字辅导系统中引入更多计算智能时，必须审慎考虑这些问题。我们研究的参与者常提及人工辅导的负面体验——感觉过于专横和说教。数字辅导系统越拟人化，就越倾向于预测用户情绪、需求和偏好，这反而可能削弱用户自主权，导致两种极端后果：要么降低对人工智能的参与度和依赖度，要么产生不恰当的信任和过度依赖[32]。

需要更多研究来确定新的交互式解决方案，使用户能够保持对其体验的控制，例如通过选择他们希望在智能数字教练中拥有的智能水平。

本项研究存在若干限制。这项探索性研究主要关注个体在日常生活中使用T2教练的主观体验，以及他们的参与模式和自我报告的目标实现情况；鉴于本研究样本量较小且持续时间较短，我们未考察T2教练对其行为或诸如血糖水平变化等健康结果的影响。此外，该研究的

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, Ad 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

参与者自我报告了其糖尿病前期和2型糖尿病诊断情况，但这一由于我们无法获取他们的医疗记录，这一点未能得到确认。最后，虽然该研究提出了在数字辅导中引入新型计算智能的机遇，但这些机遇是从个体与交互系统的体验中推断得出的与完全脚本化的聊天机器人相比。需要进一步研究来确定能够体现这些不同类型智能的具体设计解决方案，并评估它们对个体体验和健康的影响。

6 结论

在本研究中，我们对完全自动化、完全脚本化的对话代理T2教练进行了用户研究，以探索计算智能在数字辅导解决方案中的应用潜力。我们的研究表明，即使是简单的完全脚本化数字解决方案也能创造积极的辅导体验，甚至可能在某些方面优于人类教练。这些优势包括可靠性和一致性、选择和参与风格的灵活性，以及无责备和无内疚体验。同时，研究也揭示了通过更多计算智能提升辅导体验的诸多机会，并提出了与数字辅导相关的四种计算智能类型：个人化、情境化、情感化和对话化。其中某些类型，特别是对话智能，可以通过基于LLM的对话代理实现。然而，更全面的智能数字辅导需要将基于LLM的对话代理的对话智能与其他类型的计算智能相结合。

致谢

本研究由国家糖尿病、消化和肾脏疾病研究所资助，项目编号R01DK113189，

国家科学基金会项目编号2306690，以及国家医学图书馆项目编号T15LM007079。我们感谢所有研究参与者。

参考文献

[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, Ad 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

[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?

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 附录

A A.1 附录1：T2教练访谈指南

介绍

你能告诉我你与糖尿病（或糖尿病前期）的历史吗？你是什么时候被诊断的？你采取了哪些措施来管理它？

整体体验：

你能告诉我你使用T2教练的体验吗？过去几个月里？

你使用它的频率如何？你用它来做什么？你在哪些情况下使用过它？

总的来说，你会如何描述你的体验？

记录

在T2教练中记录你的餐食感觉如何？

你以前有记录过自己的餐食吗？

用纸笔记录？还是用智能手机应用程序？

对你来说那种方式体验如何？

在每顿餐食前后检查血糖是什么感觉？餐食？

在研究开始前，你多久检查一次血糖？

短信

在研究期间与T2教练发送短信的体验如何？这项研究？

通常情况下，你多久发送一次短信？你使用哪些消息传递应用程序？

短信？Facebook Messenger？WhatsApp？其他？

您通常与谁互发消息？

朋友？家人？同事？社区公告板和群组？

目标设定

您选择致力于实现哪个目标？

对您而言，选择目标的过程是怎样的体验？

是否清楚了解该目标的具体内容以及你需要采取哪些行动来实现它？

你是否记得看到过那张插图？看到那张插图对你有帮助吗？

你能否找到自己想要尝试的目标？

每次只选择一个目标来专注（在同一时间段内），这种感觉如何？

行动计划

你对T2教练制定的具体计划和推荐有何看法T2教练制定的？

例如，[列出他们选择或被提供的一些选项]

____
____
____

你选择这个了吗？你为什么选择这个？你尝试遵循它了吗？

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?

那是什么感觉?

你成功做到了吗？原因是什么？你尝试过其他方法吗？第二天是否重新尝试或采取了新方案?

Was the 提醒 关于 your plan each day 有帮助吗?

评估/达成目标

收到关于你的餐食是否符合目标的提问时，你的体验如何?

收到询问你的餐食是否达成目标的提示时，感受如何?

你是否达成了_______的目标?

How did you keep 追踪?

T2教练发送的消息是否为你提供了关于你如何

你在处理你的目标时做了什么?

引发 问题

[查看转录本以提取T2教练询问达成目标障碍的示例]

当你提到你{未达成目标}时，T2教练询问了 "__________"

你觉得怎么样？那是一个有帮助的回应吗？怎样才能让它更有帮助?

作为回应，你说"______"；T2教练问"__________"；你回答"________"

你能详细说说吗？当时发生了什么？你还记得具体情况吗?

消息 and 提醒

T2教练每天向您发送一定数量的消息

这是太多了还是太少了?

有帮助？没有帮助?

您希望每天收到T2教练多少次消息?

App and 聊天机器人

在这项研究中，您有时会与短信互动，有时会与带有图形界面（按钮、图片）的应用程序互动，

这种体验如何?

Was it 清晰 what you 需要 to use for 什么?

它们感觉像是同一个应用程序的两个部分吗？还是感觉不同？你是如何访问该应用程序的（通过短信还是主屏幕上的图标？）

对所有互动只使用一种模式会更容易吗？如果是这样，会是哪种模式，文本还是图形界面?

辅导 in 通用

现在你有了这样的体验，你对通过短信进行辅导有什么看法?

Did it feel like you were 工作中 with a 教练?

How did it feel for you?

你得到支持了吗？你感受到被倾听了吗？你是否感到沮丧?

Have you 曾合作 with a 营养师，营养学家，糖尿病 教育者，或教练之前?

相比之下如何?

缺少了什么？怎样才能让它变得更好?