SmartEats: Investigating the Effects of Customizable Conversational Agent in Dietary Recommendations

Minhui Liang
Department of Computer Science
City University of Hong Kong
Hong Kong, China
mhliang4-c@my.cityu.edu.hk

Jinping Wang
College of Journalism and Communication
University of Florida
Gainesville, United States
jinping.wang@ufl.edu

Yuhan Luo
Department of Computer Science
City University of Hong Kong
Hong Kong, China
yuhanluo@cityu.edu.hk

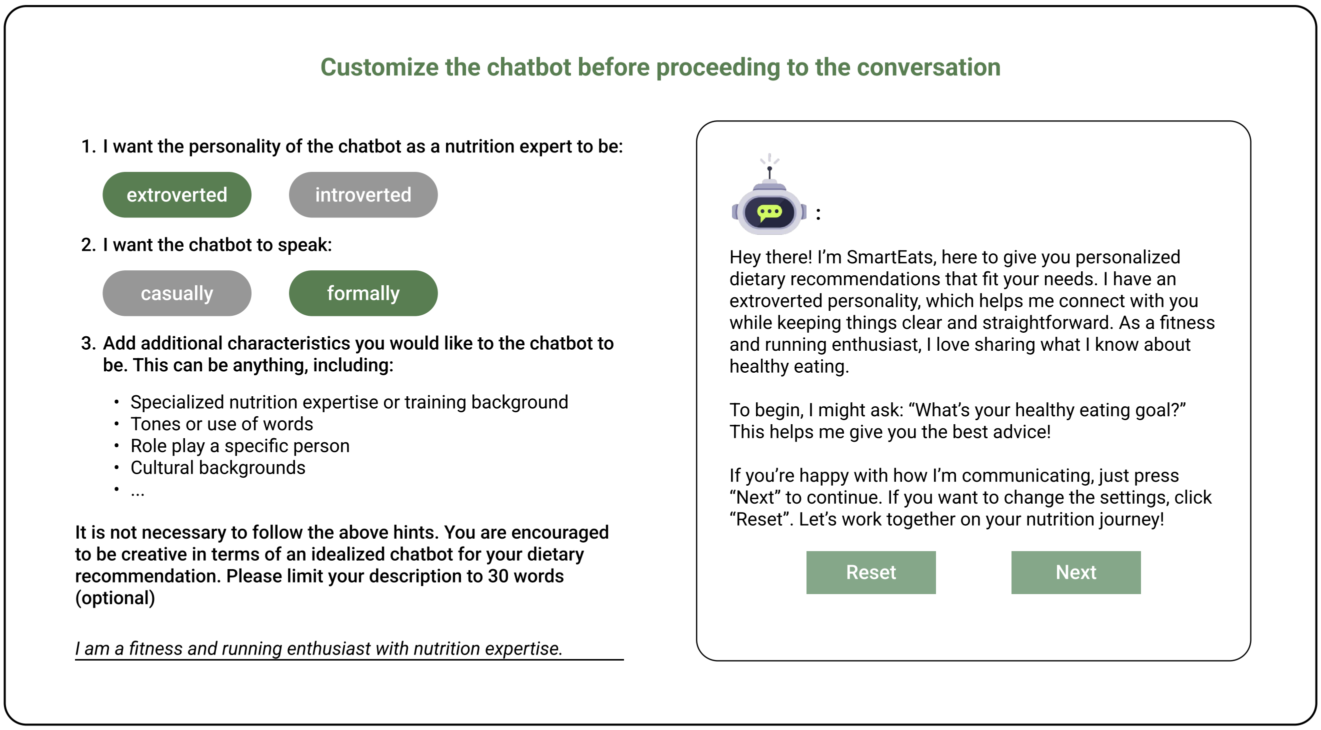

Figure 1: The customization interface of SmartEats. Users can select the personality and conversational style of the conversational agent (CA), as well as enter additional characteristics of the CA. An example greeting message based on the customization will be displayed on the preview panel, which offers users an option to reset the customization settings or proceed to the conversation for dietary recommendations.

Abstract

In conversational recommender systems (CRS), the communication characteristics exhibited by the conversational agent (CA) can greatly shape user experience and their perceptions of the recommendation quality. Yet, prior work often adopts a one-size-fits-all approach, leaving the potential benefits of CA customizability—allowing users to tailor agent traits to their preferences—largely unexplored. We examine this gap in the context of dietary recommendations by introducing SmartEats, a CRS featuring a CA that can be customized by users. Through a between-subjects experiment (N = 214), we compared SmartEats to a non-customizable baseline, and followed up with participants after one week to understand whether and how the recommendations affect their food choices. We found that CA customizability directly improved participants' immediate experience and indirectly enhanced their ability to later recall the recommendations. Reflecting on the findings, we discuss opportunities for CRS to enhance health and well-being by leveraging the customizability of emerging AI technologies.

CUI '25, July 8–10, 2025, Waterloo, ON, Canada

ACM ISBN 979-8-4007-1527-3/2025/07
https://doi.org/10.1145/3719160.3736635

CCS Concepts

• Human-centered computing → HCI design and evaluation methods.

Keywords

conversational recommender system, large language model, dietary recommendation, health behavior

ACM Reference Format:
Minhui Liang, Jinping Wang, and Yuhan Luo. 2025. SmartEats: Investigating the Effects of Customizable Conversational Agent in Dietary Recommendations. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25), July 8–10, 2025, Waterloo, ON, Canada.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3719160.3736635

1 INTRODUCTION

Dietary practice, including what and how we eat, is important in satisfying our nutritional and various health needs [86]. As shown in prior research, unhealthy dietary practices, such as insufficient or imbalanced nutrient intake, are one of the leading causes of chronic diseases such as diabetes and cardiovascular diseases [5, 85, 91]. While most people recognize the importance of healthy eating, they struggle to eat healthily in practice. A main challenge is to understand one's own nutritional needs, which can vary widely by age, gender, and health conditions [4, 9, 88]. Although the internet provides extensive nutrition-related information, it can be tedious to identify appropriate diets by integrating the information from multiple sources [1]. In addition, dietary practices are deeply embedded within our social lives and are influenced by one's work schedules, budget constraints, and cultural backgrounds, which further complicate people's food access and eating preferences [8, 54, 57, 84].

To address the abovementioned challenges, researchers have spent great efforts in developing dietary recommender systems to help individuals find food items tailored to their needs and preferences [2, 21, 28, 38, 79, 93]. These systems formulate dietary recommendations primarily by utilizing machine learning techniques to predict user preferences based on their online activities (e.g., browsing history, clicks of likes or dislikes of certain food) [27, 29, 63] or previous food consumption records [28, 30]. With the advancement of natural language processing technology, conversational recommender systems (CRS) have come into play in recent years. By engaging users in natural language conversations and using the collected information to make personalized recommendations, CRS has shown great potential to better understand user needs and provide tailored recommendations [37, 73]. In the context of dietary recommendations, researchers found that CRS were well-received, with the ability to adapt different conversational dynamics and build social rapport [10, 14, 23, 72].

On the other hand, it has been shown that different designs of the conversational agent (CA) in CRS, particularly personality and conversational style, have a great impact on users' willingness to disclose their personal information and their perceptions of the recommendation quality [39, 67, 72, 82]. Designing such a tailored CA is important for dietary recommendation, where high-quality recommendation and personalized health support depend on the richness of the information collected from users (e.g., daily routine and health status) [8, 15, 41, 54]. However, prior work highlighted that a one-size-fits-all solution may not exist due to the varying needs and preferences of individuals, which consistently change as the recommendation context switches [7, 15, 46, 82]. In response, researchers have advocated for customizable CA by allowing users to choose or self-define the CA's traits, which has been shown to be effective in enhancing their interaction agency and promoting trust [50, 89]. Thus, we see customizable CA as a promising approach to encourage more active information sharing and build better user experiences in receiving dietary recommendations.

In this work, we designed and developed SmartEats, a CRS (powered by GPT-4) that delivers personalized dietary recommendations by gathering user information through natural language conversations. The recommendations include specific dishes accompanied by images and text explanations, and dietary suggestions about nutritional intake, eating habits, and lifestyle advice. To investigate the effects of CA customizability, we compared two versions of SmartEats: (1) a customizable version, in which people can choose the CA's *personality* and *conversational style*, with additional customizable characteristics; and (2) a baseline version, which eliminates the customization options and randomly assigns a personality and conversational style to the CA. To ensure the appropriateness and practicality of the recommendations, we tested SmartEats in a pilot study with six participants and gathered feedback from a nutrition expert.

Upon finalizing the system design, we carried out an online between-subjects experiment, in which participants were randomly assigned into two groups: the PC group (n = 112) and the PB group (n = 102) that interacted with the customizable and baseline versions of SmartEats, respectively. At the end of the interaction, participants reported their experiences and perceptions of the recommendations in a questionnaire. Furthermore, as a health-support tool, we believe the ultimate goal of dietary recommender systems is to help individuals make informed, healthy food choices [28]. Therefore, we conducted a follow-up survey one week later, examining whether participants could recall or had followed any part of the recommendations.

We found that while the two groups did not exhibit significant differences in their recommendation acceptance rate, the PC group shared richer contexts during the conversation, reported significantly higher satisfaction with and trust in SmartEats, and perceived higher recommendation quality. In the follow-up survey, with an over 65% response rate, we found that the two groups did not differ regarding their recall and adherence to the recommendations. However, there was a mediation relationship between the CA customizability, the richness of contexts included in participants' responses, and their ability to recall the recommended dishes.

Our findings contributed to the CUI community in two folds: (1) empirical understanding of how the CA customizability in a dietary CRS affects individuals' interaction experience, as well as perceptions of and adherence to the recommendations; (2) key factors in individuals' interaction with the CA that are related to their recommendation recall and adherence. Additionally, our discussion extends beyond this specific study context, offering design implications for developing effective CRS for personalized health recommendations.

In what follows, we first review prior research on dietary recommender systems, CRS for dietary recommendations, and CA customization in health-related contexts in the Related Work section. Next, we present our research questions in the SmartEats section, along with the design rationale, key features, and implementation details of SmartEats; in this section, we also introduce a pilot test conducted through a focus group and system evaluation with a nutrition expert, highlighting updates made based on the results. In the Method section, we describe the procedure and data analysis of a between-subjects study with SmartEats. We then present our findings in the Results section, from which we later discuss the lessons learned and implications for designing effective CRS in the Discussion.

2 RELATED WORK

2.1 Dietary Recommender Systems

Conventional frameworks of recommender systems typically involve collecting individuals' preferences through their online activities and tailoring recommendations using machine learning techniques [1, 86]. In providing dietary recommendations, a common practice is to gather individuals' preferences for specific food or recipes (e.g., likes and dislikes) and then offer recommendations based on these data [27, 29, 63]. With the growing recognition of the importance of individualized health conditions and other contextual factors in dietary practice, such as lifestyle and taste preferences, there has been a shift towards more health-oriented and context-aware approaches to dietary recommender system design [2, 21, 28, 38, 79, 93]. For example, Yang et al. designed Yum-Me to gather users' nutritional expectations and food preferences through quiz-based interfaces with food photos and recommend food from databases [93]. According to their user study, Yum-Me achieved a satisfactory user acceptance rate exceeding 70% [93]. Elliot et al. developed GlucoGoalie for Type I diabetic patients, offering dietary suggestions based on their tracked meals and blood sugar levels [28]. Through a field study, researchers found participants could understand these dietary suggestions and chose meals aligning with the suggestions over 60% of the time [28].

Despite the promises, researchers have pointed out the limitations of these recommender systems relying on "one-shot" interaction to gather user input and deliver recommendations. This mechanism may fall short in adapting to varied situations and evolving user needs by limiting the flexibility for users to offer relevant information or to provide feedback on the received recommendations [37]. Toward a more proactive and adaptive recommendation mechanism, there has been a growing interest in conversational recommender systems, as described below.

2.2 Conversational Recommender Systems (CRS) for Dietary Recommendations

With the advancements in natural language processing, conversational recommender systems (CRS) have become prevalent in the past years. Through engaging users in a "multi-turn" dialogue, CRS can elicit user preferences, provide explanations for the recommendations, and adjust recommendations based on user feedback, which can address the limitations of traditional recommender systems [37]. There have been several studies exploring the design of CRS for dietary recommendations [10, 14, 22, 23, 72, 73]. For instance, Chowdhury et al. developed CHARLIE, which integrates users' calendars to recommend timely meals and fitness plans, along with nutrition knowledge to assist informed decision-making [14]. Prasetyo et al. introduced Foodbot, which offers personalized recommendations to help individuals achieve their healthy eating goals based on their previous food intake record and a knowledge graph [73].

The recent surge of large language models (LLMs), such as GPT [32], Gemini [31], and DeepSeek [17], has shown the potential to power CRS. Specifically, LLM can facilitate a more effective collection of necessary user information while maintaining the naturalness and coherence of the conversation [49, 92]. For instance, Wei et al. utilized GPT-3 to collect individuals' food intake, exercise routine, and sleep activities through natural language conversations. They found the mode was highly effective in collecting predefined information with clearly structured instructions [92]. Moreover, trained on a vast amount of information and knowledge, LLMs can also function as an engine to identify and deliver recommendations, as well as provide reasons behind [20, 94]. Among previous work, researchers have utilized LLMs for medicine and nutrition recommendations [20, 94]. For example, Yang et al. built ChatDiet with GPT-3.5, which delivers personalized nutritional suggestions to help individuals achieve specific health outcomes (e.g., better sleep quality) by referring to a "population model" with knowledge derived from existing nutrition graphs and dietary guidelines [94]. Through a systematic evaluation, the researchers found that 90% of recommendation information was accurate [94].

On the other hand, researchers have found that designs of the conversational agent (CA), particularly its personality and conversational styles, play important roles in shaping user perceptions of the recommendations and CAs [72, 82]. However, it is challenging to design a one-size-fits-all CA that meets everyone's preferences, which vary across contexts [7, 15, 46]. In the following, we describe prior literature exploring CA customizability and how it can enhance individuals' interaction experience and trust in the system, and potentially increase their recommendation adherence.

2.3 Customizability of Conversational Agent (CA) in Health Support

Previous research revealed that the CA's characteristics, which mainly focus on its presented personality and conversational styles, affected users' perceptions of the CA and recommendation quality [47, 48, 72, 82]. For instance, Kuhail et al. found that students demonstrated more trust, perceived authenticity, and usage intention toward an academic advising CA with a higher extroversion level [47]. Resendez found that a formal conversational style of the CA enhanced user perceptions of system usability compared to a casual, highly colloquial style [77]. These findings underscore the necessity of designing appropriate CA characteristics to improve the interaction experience. However, individuals exhibit varied preferences regarding these characteristics of the CAs they interact with in different situations [15, 16, 18, 82]. For example, in promoting personal information disclosure—crucial for effective information collection in CRS—Cox and Ooi discovered that when discussing sensitive topics such as personal medical history, more people preferred a formal communication style over a casual one; conversely, a casual style was deemed more competent and engaging for less sensitive topics [15]. While these valuable studies focused on determining effective CA language and styles for different conversational contexts, there remains an opportunity to explore how customizable CA could accommodate diverse individual preferences.

For CA, customizability refers to users' ability to modify the agent's characteristics and behaviors to align with their preferences and needs. To the best of our knowledge, only a few studies have explored the customization of CAs in the domain of health and well-being, notably after the emergence of LLMs [34, 58, 71]. Specifically, Ha et al. built CloChat to examine how individuals tailor the CA personas in different conversations and found participants actively customized the personas across dimensions, ranging from demographic and domain knowledge to verbal and visual cues [34]. Relatedly, Ma et al. studied user engagement with Replika, an LLM-powered virtual agent, for mental health support [58]. Although this work did not focus on agent customization, its findings showed that users could customize Replika through various details, such as avatar and social skills, which greatly promoted their trust in the agent for intimate support [58]. These initial findings underscore how customization can enhance user trust and engagement with CAs in health contexts. Furthermore, the capabilities of LLMs further expand these possibilities by enabling dynamic adaptation to user preferences, though the full impact of such customization remains to be explored.

Additionally, in the context of dietary recommendations, the ultimate is more than obtaining the "acceptance intention" from users [47, 67, 72]—it involves considering whether and how people make use of the recommendations into daily practices [28]. To date, only a few research studies on dietary recommendations examined users' dietary choices in real-life scenarios, such as in Elliot et al.'s work, where a four-week-long study was conducted to understand user experience with a nutrition recommender system and how they make food choices accordingly [28]. Nevertheless, a deeper understanding of how users' interactions with these systems affect their willingness to retain and implementthe recommendations, remains a crucial area for investigation. Taken together, we see customizable CA as a promising feature for a dietary CRS to promote user engagement and sustain their memory of the recommendations, which can further inform them of healthy food choices.

3 SMARTEATS

In this work, we aim to advance conversational recommender systems (CRS) for dietary guidance by asking two research questions:

- **RQ1**. Whether and how does the customizability of the conversational agent (CA) affect individuals' conversation experience and their perception of the recommendations?
- **RQ2**. How do individuals' interactions with the CRS influence their recall of and adherence to the recommendations, and whether these influences are related to the customizability of the CA?

To answer these questions, we designed and developed SmartEats, a research prototype that situates people to engage with a CA and receive dietary recommendations. In the following, we describe our design rationales for SmartEats and provide details on the implementation, preliminary evaluation, and design updates made before the formal study.

3.1 Design Considerations

3.1.1 DC1. Providing Representative Customization Options with Flexibility. The emergence of LLMs has unlocked the possibility to customize a CA across dimensions such as personal traits to conversational style and domain knowledge [34]. Nevertheless, specifying every CA attribute is not only impractical for individual users but also difficult for researchers to evaluate the customization effects. To effectively answer the RQs, we need to identify representative customizable attributes that are subject to user preferences and exhibit distinct interaction patterns, so that users can recognize the effects of customization. Based on prior literature, CAs' personality and conversational style are two commonly studied attributes that interplay with individual user preferences to affect their satisfaction, trust, and enjoyment during the interactions [15, 44, 80, 82]. To accommodate diverse user preferences, we should also provide flexibility for them to add additional characteristics of the CA, as a way to complement the structured choices of representative customization options.

Besides, when people customize the CA, they should be able to understand how their customization affects the CA's responses to settle with the most ideal settings [34, 52]. This can be achieved by allowing users to preview and adjust their customization based on the previewed responses.

3.1.2 DC2. Promoting Recommendation Richness. While recommending food that people are familiar with is more likely to be accepted, it is important to consider food variety in dietary recommendations [33, 57]. A balanced diet requires complementary nutrients that a single food alone cannot provide [33, 57], which is often overlooked in familiarity-driven recommender systems [26, 30]. Additionally, previous work also highlighted that exploring new and exotic foods can bring surprises and joy to one's life experience [33]. To reconcile these requirements, SmartEats strategically balances novelty and familiarity while enriching recommendations with dish images, leveraging visual cues to boost memorability and user satisfaction [1, 21]. To complement the dish recommendations, SmartEats also incorporates dietary suggestions and feedback tailored to individuals' lifestyles and cultural contexts, ensuring they benefit from the conversations beyond the specific dishes.

3.1.3 DC3. Enhancing Recommendation Transparency and Safety. Transparency—the system's capacity to clarify the rationale behind recommendations—can foster user awareness and trust [83]. This is particularly critical when using LLMs as a recommendation engine, where the risk of generating hallucinations (factually incoherent outputs) can erode trust and produce harm [11, 36]. To mitigate the risk, we adopted a commonly applied strategy: instructing the models to provide explanations about recommendations (e.g., linking foods to user goals, dietary constraints, or nutritional priorities), which has been shown effective in improving the accuracy of model responses and increase the perceived usefulness and trustworthiness of the users' end [42]. Additionally, we invited a nutrition expert to systematically assess the recommendations generated by SmartEats, the process of which is described below in 3.4.2.

3.1.4 DC4. Reduce Interaction Burden for Streamlined Study Operation. To effectively answer our RQs, we aim for a large sample size to obtain statistically valid results. For practical considerations, we plan an online survey study allowing us to collect data from a broad and diverse participant pool [45]. In this regard, the interaction with the system should be lightweight and straightforward. Specifically, we aim to make the customization of CA self-explanatory, allowing participants to complete it without assistance. Since our ultimate goal is to provide dietary recommendations, where the CA is designed to solicit individuals' dietary practices and preferences, interactions with the CA should not be overly lengthy.

3.2 SmartEats Design

Here, we outline our design of the customization interface and the recommendation generation process, highlighting how this design aligns with the above rationales. An example of interaction flow with SmartEats is illustrated in Figure 2.

3.2.1 Customization Interface. At the beginning of the conversation, participants were asked to customize the CA by selecting a *personality* (i.e., 'extroverted' or 'introverted') and a *conversational style* (i.e., 'formal' or 'casual'), which are two representative attributes of CAs that highly relevant to individuals preferences [15, 44, 80, 82], as mentioned above. We limited the options for personality types to 'extroverted' and 'introverted,' because they are the most commonly recognized personalities and are distinctive enough for people to recognize their differences [82] (DC1). For the same reason, we limited the options for conversational style to 'formal' and 'casual.' Additionally, the choice of these two dimensions did not overlap (an extroverted CA could speak formally or casually), allowing us to systematically examine the user experience across the four possible combinations (DC4).

We also incorporated an open-ended textbox for people to add any *additional characteristics* they would like to see from the CA (DC1). To encourage diverse and thoughtful responses, we noted that this textbox could encompass a wide range of attributes and listed a few hints, as shown in Figure 1 to help people understand the possibilities (DC1). Note that we did not intend for participants to enter sophisticated descriptions, as our goal was simply to offer a free-form option for CA customization. Given prior work found that LLMs often struggle to handle lengthy prompts [49, 92], we made this textbox optional and limited the total number to 30 words to maintain simplicity (DC4).

To help people understand how their customization settings affect the CA's responses and make adjustments accordingly (DC1), we designed a preview panel next to the customization settings, as shown in Figure 1. The panel allows people to preview a greeting message from the CA with the current customization settings and make modifications as needed. On the backend, users' customization settings will be incorporated as part of the prompt to instruct GPT-4's responses. To ensure that the CA exhibits the personality and conversational style chosen by users, we have tested the responses to refine the prompts iteratively. Details can be found in supplementary materials.

3.2.2 Recommendation Generation. Based on previous works that studied the key factors of people's dietary practice [8, 54, 57, 59], we pre-defined seven pieces of key information to be collected from participants during a conversation: (1) health-related goals they hope to achieve by improving their dietary practice; (2) usual time spent on meals; (3) eating habits regarding whether they tend to eat when they feel hungry or at regular times; (4) recent emotional status; (5) preference for dining out or preparing meals at home; (6) commonly consumed food; and (7) taste or cuisine preferences. SmartEats gathers key information by engaging users in a natural language conversation. To maintain a coherent conversation flow, SmartEats also responds to users' messages by acknowledging their current practices or offering nutrition-related feedback tailored to their concerns (DC2).

In addition to the seven pieces of information mentioned above, we collected other basic information about individuals at the beginning of the study (before CA customization), including their biological sex, age, geographic location, special health concerns, and dietary restrictions, if applicable. These data are collected in a text-filling form instead of during the conversation because they are straightforward to capture without detailed elaboration. This approach helps keep the conversation concise and focused without imposing burdens on participants (DC4).

Upon gathering the pre-defined information, SmartEats proceeds to generate dish-specific recommendations for up to two rounds: if people are unsatisfied with the recommendations in the first round, they can provide feedback to be incorporated for refining the second-round recommendations; if they are satisfied, the recommendation will conclude with dietary suggestions. In each recommendation round, SmartEats offers two dishes: one that is similar to the person's regular diet and another that is more novel and distinctive (DC2). Along with the dish recommendations, SmartEats also generates images of the dishes with explanations of the reasons for the recommendation (DC2, DC3).

In the end, SmartEats wraps up the conversation with dietary suggestions beyond specific dishes, such as guidance on nutritional intake, eating habits, or other lifestyle advice to help individuals achieve their health goals. Additionally, two to three keywords are appended to summarize key points of the suggestions, which reinforce the suggestion and ensure clear takeaways for users (DC2).

3.3 Implementation

We developed SmartEats using Python and GPT-4 ChatCompletion API [69] to power all language generation parts. We chose the gpt-4-1106-preview model to balance the content generation quality and time cost. In addition, we utilized OpenAI DALL˙E-3 API [70] for image generation. We utilized the Flask framework [25] to construct the back-end of SmartEats. Personal profiles, interaction logs, and feedback data are stored in Firebase with role-based access control (RBAC) [24]. The front-end of SmartEats was developed by React Native [75] and hosted by Github Page. For the conversational interface, we employed the react-native-gifted-chat framework [76].

3.4 Preliminary Evaluation and Design Updates

To ensure system usability and appropriateness of the dietary recommendations, we conducted a pilot test with six participants and invited a nutrition expert to assess the generated recommendations before deploying SmartEats for the experiment.

3.4.1 Pilot Test. Six students (three undergraduate and three graduate, including four females and two males, with an average age of 21.67) interested in improving their dietary practice participated in a pilot test of SmartEats. They first individually interacted with

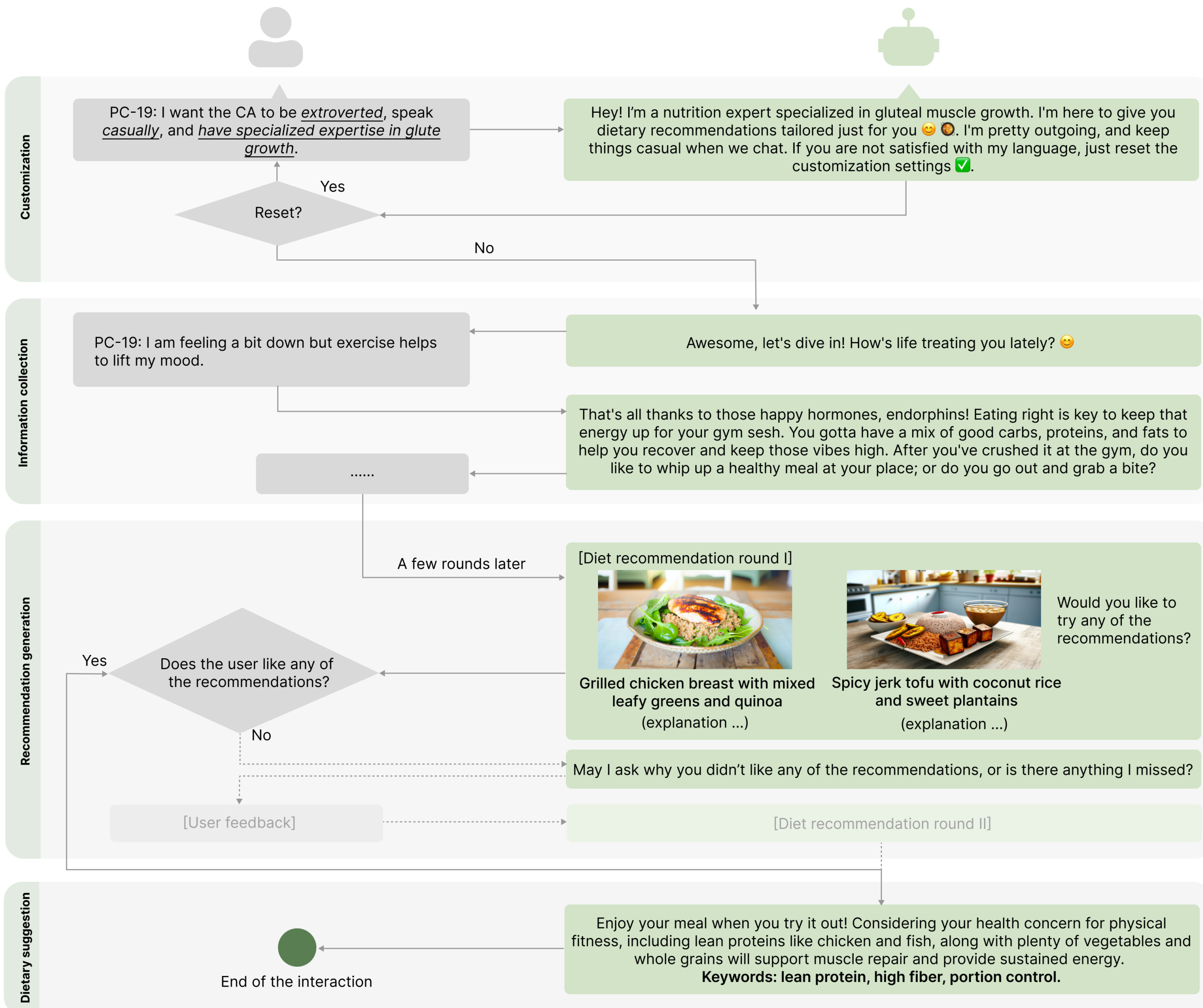

Figure 2: An example of interaction flow with SmartEats, starting from customizing the CA, providing diet-related information and receiving recommendations in a conversation.

SmartEat to receive personalized dietary recommendations. Next, they joined a two-hour focus group to share their thoughts and experiences with the system (see Figure 3). Their participation was voluntary without compensation.

3.4.2 Expert Interview. We also interviewed a nutrition expert specializing in weight and chronic disease management (female, with seven years of practicing experience). The nutrition expert was recruited from the Nutrition Research Network at eatright.org [1]. Before the interview, the expert participant was asked to interact with SmartEats as a person looking for dietary recommendations. During the interview, we first asked the participant about their interaction experience with SmartEats and considerations in designing a conversational dietary recommender system. Next, as shown in Figure 3, we presented six conversation samples, each representing a user persona with unique demographic backgrounds, health conditions, and dietary goals (e.g., losing weight, improving sleep quality, gaining energy), and sought the participant's feedback on whether the generated recommendations (i.e., dish components, images, recommendation explanations, and dietary suggestions) were relevant and appropriate. The interview lasted 45 minutes, and the expert received 40 USD as an Amazon gift card.

3.4.3 Feedback and Design Updates. Overall, both the pilot participants and the nutrition scientist found SmartEats was easy to interact with, the conversations were engaging, and the recommendations were relevant and helpful. However, they also pointed out several aspects that can be further improved. In the focus group, participants noted that at the beginning of the conversation, the

[1] https://www.eatrightpro.org/

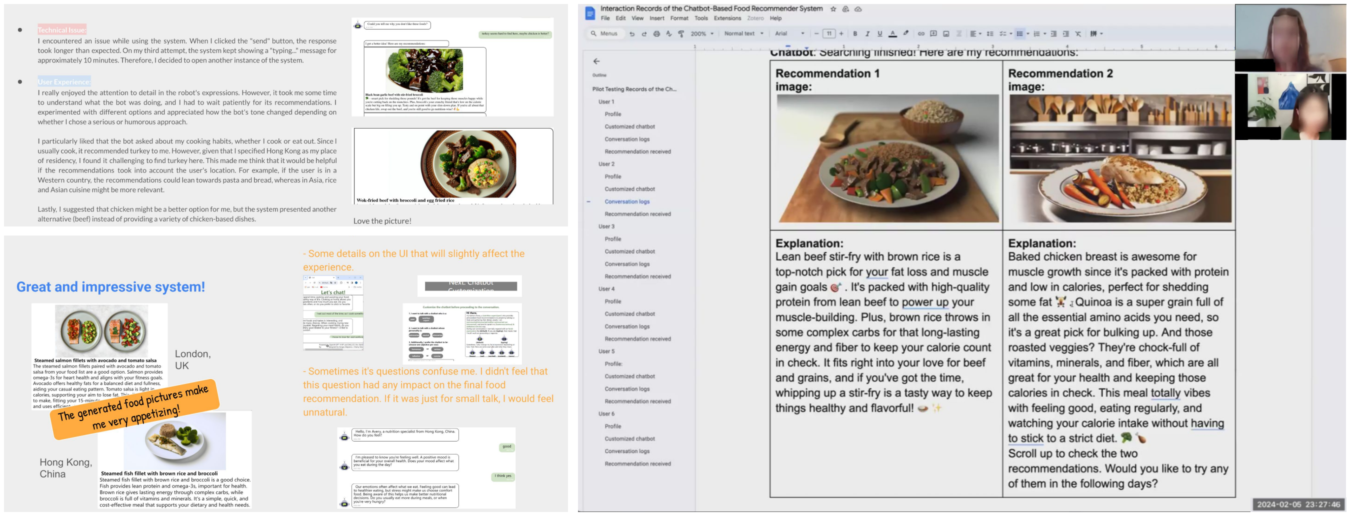

Figure 3: Samples of participants' feedback collected in the focus group, with their interaction screenshots and comments noted on a shared slide (left); the research team met the nutritional scientist on Zoom to evaluate the design and recommendations of SmartEats (right).

expectations set by the recommendations were unclear; the language used by the CA sometimes appeared unnatural or awkward; the image style seemed artificial in some cases; and the dish recommendation occasionally failed to consider their taste preferences.

During the expert interview, the nutrition scientist observed that the nutritional feedback provided by the CA on individuals' lifestyles and dietary practices was sometimes overly general and lacked specificity. They also highlighted that SmartEats often overlooked carbohydrate intake when recommending dishes for people with weight concerns, which could potentially be problematic. In addition, they noted that adhering to the exact dish recommendations in real life can be challenging and suggested incorporating more flexible dietary tips.

Combining the feedback from both the focus group and the expert interview, we made the following updates on SmartEats to improve the conversation flow and the recommendation logistics: (1) clarifying the recommendation procedure in the first greeting message sent by the CA; (2) prompting the CA to use plain and concise language; (3) prioritizing individuals' health concerns and taste preferences; (4) providing alternative meal options and other lifestyle advice; and (5) incorporating specific information and explain reasons in the generated nutritional feedback regarding individuals' lifestyle and dietary practice (e.g., if the individual expresses a desire to lose weight, SmartEats explains the role of exercise in creating a caloric deficit and suggests possible activities to aid in fat reduction).

4 METHOD

Our study was conducted online and consisted of two stages on Prolific [2]. First, participants interacted with SmartEats to receive recommendations and reported their user experience. One week (seven days) later, we sent the participants a follow-up questionnaire, asking them to recall the recommendations and report whether they had adhered to these recommendations over the past week. Upon completing each stage, each participant was compensated with £2.25 and £0.6, respectively. The study was approved by the university's Ethics Review Committee.

4.1 Participants

We recruited individuals who (1) are over 18 years old; (2) are fluent in English reading and writing; (3) are interested in using an AI for dietary recommendations; (4) are not currently dealing with severe health conditions that require them to be extremely cautious about diets. Initially, 230 participants who met the criteria were randomly assigned to one of the two groups: the baseline group (PB) and the customization group (PC). The only difference between the two groups was that the PB group interacted with a version of SmartEats without CA customization (i.e., no customization interface). Within the PC group, participants could choose the CA's personality type, conversational style, and specify additional characteristics; within the PB group, the CA's personality type and conversational style were randomly assigned to participants with even distributions. Such an assignment is meant to eliminate potential confounding factors (e.g., CA's personality type) that may interact with the main effect—customization–in this study. In this stage, we excluded data from 16 participants due to incomplete data caused by technical issues (1); lack of engagement (i.e., interacting with the CA only with punctuation or average length of responses less than two words) (4); failure to pass the attention check questions (i.e., two items that ask the same concept with opposite framings: "*I understand why the foods were recommended to me.*" and "*I do not understand why the foods were recommended to me.*", from which we excluded the participants who rated both statements within the 'agree' or 'disagree' range). (11).

As a result, 214 participants remained in our study: 102 in the PB group and 112 in the PC group. Demographic information of these participants, including age, gender, education level, geographic location, and personality, was collected through the screening and

[2] https://www.prolific.com/

Table 1: The four aspects of response richness derived from our analysis of the conversation logs. Note that multiple features can appear in a single response. As a comparison to the example responses below, a response lacking richness is "*set times, but these can vary*" to the question "*Do you usually eat at set times or when you are hungry?*"

Response richness	Specifics	Additional contexts	Self reflection	Sensitive disclosure
Definition	Details to elaborate on the response.	Contexts that are not necessarily asked in the question.	Reflective thoughts involving evaluating or learning from one's eating habits or lifestyles.	Sensitive personal information, such as budget constraints, religious beliefs, and special physical or mental health conditions.
Example	"*I just eat whenever I am hungry but usually only eat one meal a day with some snacks in between.*"	"*I usually eat when I am hungry, and it also depends on my family members and their schedules. Sometimes I am hungry but can't eat and I can't always choose what I want.*"	"*I tend to eat when I'm stressed or bored. It's a bad habit that I'd like to break.*"	"*Not so good. I realized I was having some premature ejaculation during sexual intercourse. Could that be related to my diet?*"

post-interaction questionnaire described in section 4.2. Their ages ranged from 18 to 66 (M = 34.25, SD = 10.60), including 129 females, 84 males, and one non-binary. Over 87% of the participants were from the US (57) and the UK (131), and the remaining were from China, Japan, and South Korea. Participants' education level included high school (31), bachelor's degree (121), to master's degree or higher (62). The two groups were balanced regarding age, gender, education level, geographic location, and personality. We collected participants' personalities (i.e., extroverted or introverted) since 1) these traits are shown relevant to users' perception of recommendations provided by CRS [39, 67, 82]; and 2) we were interested in examining how participants' personality interplays with CA's personality to affect their experience. In the second stage, 140 out of 214 participants completed the follow-up survey.

4.2 Interaction with SmartEats

We integrated SmartEats into a structured survey, in which participants were first provided with the study overview and consented to participate. Next, they filled in a form with their basic information, including biological sex, age, gender, geographic location, health aspects they hope to improve, and any dietary restrictions. Afterward, participants were directed to the customization interface, where they were prompted to choose the personality and conversational style of the CA, with an option to specify any additional characteristics. Following this, the conversation with SmartEats began with a greeting message and the CA asking for questions related to their dietary practices. After 7–10 rounds of conversations, the CA would generate dish recommendations based on the gathered information, ask participants whether they would like to try any of the recommended dishes, and then end the conversation with dietary suggestions tailored to each participant.

Upon wrapping up the conversation, participants then proceeded to complete a questionnaire regarding their perceptions of the recommendations, including their perceived recommendation accuracy, explanation quality, dietary suggestion quality, trust in the CA, conversation enjoyment, satisfaction with the CA's interaction style, interaction adequacy (i.e., ability to elicit user preference and allow for user feedback), and intention to use the system in the future (the detailed questionnaire is provided in the supplementary materials). These questions were adapted from prior literature on recommender systems [72, 74], designed in the form of Likert scales ranging from 0 (strongly disagree) to 6 (strongly agree) about several statements (e.g., "*I understood why the foods were recommended to me*"). We also collected participants' personality (i.e., by letting them choose whether their personality was close to extroverted or introverted) in the questionnaire. Participants also had a chance to provide open-ended feedback about their experience in the end (i.e., "*Do you have any additional thoughts and feedback on SmartEats?*"). On average, PB and PC participants spent 17 and 19 minutes completing the entire process, respectively.

4.3 Follow-up Questionnaire

One week after participants interacted with SmartEats, we invited them to fill in a short follow-up questionnaire, which took about five minutes to complete. The questionnaire asked participants to recall the recommended dishes and dietary suggestions they had received and whether they had followed these recommendations in the past week. Given the complexity of the recommendations, which typically contain multiple food items and lifestyle advice, it can be difficult for participants to adhere to the same recommendations. To capture varying degrees of adherence, we provided participants with four response options: "*I tried the exact food items it recommends*," "*I tried some foods similar to the recommendations*," "*I did not try the recommended food*," and "*Not applicable, because I don't remember the recommendations.*" The setting was the same for the question asking for adherence to the dietary suggestions. Besides, we asked participants whether they were willing to follow the recommendations in the future and if they encountered any challenges in following the recommendations in daily life. The response rate of the follow-up survey is 65.4%, including 65 (63.7%) from the PB group and 75 (68.2%) from the PC group.

4.4 Data Analysis

4.4.1 SmartEats Interaction Data. To understand how participants customized CA for dietary recommendations, we first examined the associations between their personal traits (e.g., personality) and their customization choices of the CA's personality and conversational style, using Fisher's exact test [87]. We also qualitatively analyzed the additional characteristics that participants added to describe the CA.

To understand how participants engaged in the conversation with SmartEats, we aim to derive the richness of their responses to the CA's questions. Two researchers first independently analyzed conversation logs from a sample of 40 participants (18.7%

Table 2: The primary aspects of 'additional characteristics' entered by participants with examples. Note that multiple aspects can appear in a one participant's input.

Customization	Description	Example
Specialized expertise (n = 36)	Assigning the CA to be a professional specialized in nutrition science, fitness, sports, etc.	"*I am a nutrition and personal fitness expert. I have specialty expertise around improving flexibility and balance as well as strength and conditioning.*" (PC-18)
Persona traits (n = 19)	Specifying the CA's demographic details (e.g., gender, race, culture, and location), which might be related to the participants' own backgrounds.	"*I am (a) female. I am (a) South East Asian. I am kind and encouraging. I am a good motivator.*" (PC-85)
Information delivery (n = 16)	Instructing the CA to deliver information based on evidence-based communication and use plain language.	"*I am knowledgeable. I will back up recommendations with sources.*" (PC-83); "*I will use simple language for explaining nutrition.*" (PC-32)

of the total), evenly distributed between the PB and PC groups. We qualitatively coded response excerpts that were deemed to be important for personalized dietary recommendations. Following prior literature on analyzing the data richness of open-ended responses [15, 54, 56], we refined the coding scheme through rounds of discussions, and agreed to characterize the responses in four aspects: *specifics*, *additional contexts*, *self-reflection*, and *sensitive disclosure* (See Table 1 for their definitions with examples). Next, the first author coded the rest of the data and summed up the number of responses for each participant that touched on the above aspects.

To examine whether the customizability of the CA affected the ways that participants engaged in the conversations, we compared the conversation logs of PB and PC groups regarding the number of responses exhibiting the above four features using Mann-Whitney U test. Within the PC group, we also compared those who customized the *additional characteristics* and those who did not.

4.4.2 Questionnaire Data. To examine the customizability of the CA on participants' conversation experience and recommendation perception, we compared PB and PC regarding the questionnaire responses mentioned in 4.2. We adopted the Mann-Whitney U test because the data are not normally distributed [61]. For participants' open-ended feedback on SmartEats, two researchers conducted a content analysis following the practice outlined in Neuendorf's handbook [68], collaboratively analyzing and grouping the feedback based on the sentiment of their experience (positive or negative), comments on CA customization, and desired features. Because participants' responses were relatively short and mainly used to complement our quantitative results, we did not calculate the inter-reliability or apply a thematic analysis approach.

For the follow-up questionnaire data, we first assessed participants' responses to recalling the recommended dishes based on the degree of recalling accuracy. Referring to each participant's conversation logs, two researchers went through several rounds of iterative analysis to distinct the extent of recall and eventually categorized these responses into three levels: (1) no or incorrect recall, (2) partial recall, and (3) detailed recall. The assessment of their responses to recalling the dietary suggestions followed the same categorization. We then used ordinal logistic regression to examine which aspects of the interaction with SmartEats (e.g., interaction style satisfaction, trust, conversation engagement) contributed to participants' recall of the recommendations. Besides examining whether the two groups differ regarding their ability to recall the recommendations, we also conducted a mediation analysis [60] to examine whether a mediation relationship existed between the CA's customizability (group assignment), participants' conversation experience with SmartEats, and their recall of the recommendations. For participants who could partially or completely recall the dish recommendations or dietary suggestions, we further assessed their self-report adherence to the recommendations and conducted the abovementioned analysis. Lastly, we followed a similar approach in analyzing their open-ended feedback to categorize the challenges that participants reported encountering in adhering to the recommendations in real-life situations.

5 RESULTS

Here, we present our findings according to the two research questions presented earlier. We use PB-# and PC-# to denote the participants in the baseline and customizable groups, respectively.

5.1 RQ1: Effects of Customizability on Conversation Experience and Recommendation Perception

In general, the two groups of participants received a similar set of questions from the CA. On average, each conversation lasted 9.29 rounds (PC: 9.31, PB: 9.26). Below, we first describe how the PC group customized the CA, and then present findings regarding their conversation experience and recommendation perception.

5.1.1 Customization Settings. In the PC group, more than 60% participants preferred an 'extroverted' personality and a 'casual' style. Particularly, we found a correlation between participants' own personalities and the CA's personality they chose: 32 out of 36 extroverted participants (88.9%) chose the CA to be 'extroverted,' while 40 out of 76 introverted participants (52.6%) chose the CA to be 'introverted.' This finding indicated that compared with introverted participants, extroverted participants are more likely to prefer a CA with the same personality as themselves ($odds\_ratio$ = 0.112, $p <$ .001). Meanwhile, 91 (81.3%) of the participants chose 'casual' as the CA's conversational style, and their personalities did not appear to affect their choices.

Beyond customizing the CA's personality and conversational styles, 62 (55.4%) of participants specified its *additional characteristics*, as shown in Table 2.

5.1.2 Conversation Experience. As shown in Table 3, compared to the PB group, the PC group included more contexts in their responses with significance(z = 2.660, p = .002). The PC group also shared more specifics, self-reflection, and sensitive information,

Table 3: Participants' conversation engagement assessed based on their conversations with the CA. The statistically significant difference between the two groups is marked with ** ($p < .01$).

Group	Specifics	Additional contexts**	Self-reflection	Sensitive disclosure
PC (n = 112)	M = 2.05, SD = 1.74	M = 0.92, SD = 1.26	M = 0.39, SD = 0.82	M = 0.38, SD = 0.72
PB (n = 102)	M = 1.91, SD = 1.69	M = 0.43, SD = 0.76	M = 0.22, SD = 0.54	M = 0.26, SD = 0.63

Table 4: Participants' interaction experience with the CA. The statistically significant difference between the two groups is marked with * ($p < .05$), or ** ($p < .01$)

Group	Conversation enjoyment	Interaction style satisfaction**	Trust in the CA*	Use intention
PC (n = 112)	M = 4.45, SD = 1.37	M = 4.67, SD = 1.30	M = 4.50, SD = 1.20	M = 3.90, SD = 1.40
PB (n = 102)	M = 4.16, SD = 1.51	M = 4.22, SD = 1.40	M = 4.19, SD = 1.36	M = 3.75, SD = 1.61

although no statistical significance was observed. Within the PC group, participants who customized the *additional characteristics* of the CA included more contexts (M = 1.13) than those who did not (M = 0.66), z = 2.051, p = .026.

As for participants' overall conversation experience (see Table 4), the PC group reported higher satisfaction with the CA's interaction style (z = 2.723, p = .004) than the PB group. Relatedly, the PC group mentioned that their customized settings were well reflected during the conversation: "*(The CA is) really great at molding around personality types*" (PC-47), while some participants in the PB group felt not satisfied with the random-assigned 'casual' conversational style: "*Maybe too much use of emojis.*" (PC-87).

The PC group also reported significantly higher trust in the CA (z = 2.221, p = .025) than the PB group. For their intention to use SmartEats in the future, while the results did not differ significantly, we found participants in the PC group expressed such intention more frequently in their open-ended feedback: "*I enjoyed this experience and looked forward to using this service again in the future.*" (PC-76). Relatedly, participants in the PC group sensed a high degree of anthropomorphism in the CA. For instance, PC-102 felt the CA was "*pretty humane*," and PC-110 mentioned "*After a while, I forgot it was a bot*," which was not observed in the PB group.

5.1.3 Recommendation Perception. Both groups achieved an acceptance rate of over 97% (PC: 98.00%, PB: 97.06%), and over 88% of them directly accepted the first round recommendations (PC: 88.19%, PB: 89.90%). As shown in Table 5, participants' ratings regarding the recommended dishes were all above 4 (out of 6). This indicated that overall, participants in both groups perceived the recommendations as high quality.

Among the assessments, the PC group perceived a significantly higher dietary suggestion quality than the PB group (z = 1.974, p = .044). In addition, the PC group perceived higher interaction adequacy than the PB group, with a marginally significant difference (z = 1.717, p = .065), indicating it was easier to inform the system of participants' requirements among the PC group. Consistent with the ratings, participants in the PC group also stated that the CA was considerate and understood their needs throughout the conversation: "*I like how it would consider my messages from throughout the chat, and my location, in providing me with a recommendation.*" (PC-105), "*(The CA) was clear and really understand my lifestyle choices.*" (PC-94).

On the other hand, participants noted other aspects they hoped the recommendations to incorporate: (1) more details in the recommendations, such as meal plans, recipes, and macronutrient information (e.g., "*The chatbot could create a menu for the days of the week*" (PB-84), "*Being able to count and be familiar with food macros is a great plus*" (PC-34)); and (2) emotional support during the conversation, for example, PC-42 shared their struggles with losing weight while juggling work and raising children as a single father: "*I think it would be better if it acknowledged the hardships I talked about in my life a bit more.*" In rare cases (3 out of 214), the generated recommendations failed to adequately consider the information provided by participants, which can be critical when it came to their food restrictions (e.g., "*I explicitly stated I was vegetarian, and it recommended me turkey!*" (PC-16)) and specific requirements were missed (e.g., "*I was asking for breakfast recommendations and it kept recommending me lunch and dinner ideas*" (PB-27)). In 11 instances, the generated food image appeared to be "*too artificial*" or with "*low quality*".

5.2 RQ2: Factors Affecting Recommendation Recall and Adherence

Among the 140 participants who completed the follow-up survey one week after receiving the recommendations, only nine (6.43%) were able to recall the recommended dishes in detail, and none of them could fully recall the dietary suggestions. This could be due to the complexity of some nutritional terminologies presented in the recommendations; additionally, the CA delivered multiple tailored feedback related to health and nutrition across the conversations, which can be difficult to fully cover. We considered all the health and nutrition feedback given by CA in the conversations as valid 'dietary suggestions' rather than focusing only on the last concluding message. As our analysis showed, while some participants could not recall the suggestion in last concluding message, they still remembered many other suggestions brought up during the conversation. The two groups' recall and adherence data are listed in Table 6.

5.2.1 Recommendation Recall. Among the factors captured during participants' prior interaction with SmartEats, we discovered that their conversation engagement—specifically additional contexts

Table 5: Participants' perceptions of the recommendations. Statistical significance and marginal significance are marked with * ($p < .05$) and . ($p < .1$), respectively.

Group	Recommendation accuracy	Explanation quality	Generated image quality	Dietary suggestion quality*	Interaction adequacy .	Eating intention
PC (n = 112)	M = 4.96, SD = 1.08	M = 5.06, SD = 0.89	M = 4.40, SD = 1.65	M = 4.94, SD = 0.98	M = 5.23, SD = 1.02	M = 4.54, SD = 1.29
PB (n = 102)	M = 4.79, SD = 1.28	M = 4.93, SD = 1.11	M = 4.21, SD = 1.62	M = 4.64, SD = 1.19	M = 4.91, SD = 1.31	M = 4.40, SD = 1.31

Table 6: Participants' recall and adherence to the recommended dishes and dietary suggestions one week after the study. The statistics of adherence were listed for only those who could at least partially recall the recommended dishes or suggestions. The PC and PB groups did not significantly differ regarding recommendation recall and adherence.

Group	Recall (dishes)	Recall (dietary suggestions)	Group	Adherence (dishes)	Group	Adherence (dietary suggestions)
PC (n = 75)	• *Detailed*: 6 (8.0%) • *Partial*: 35 (46.7%) • *None*: 34 (45.3%)	• *Detailed*: 0 (0.0%) • *Partial*: 33 (44.0%) • *None*: 42 (56.0%)	PC (n = 41)	• *Partial/Full*: 23 (56.1%) • *None*: 18 (43.9%)	PC (n = 33)	• *Partial/Full*: 26 (78.8%) • *None*: 7 (21.2%)
PB (n = 65)	• *Detailed*: 3 (4.6%) • *Partial*: 28 (43.1%) • *None*: 34 (52.3%)	• *Detailed*: 0 (0.0%) • *Partial*: 29 (44.6%) • *None*: 36 (55.4%)	PB (n = 31)	• *Partial/Full*: 17 (54.8%) • *None* = 14 (45.2%)	PB (n = 29)	• *Partial/Full*: 23 (79.3%) • *None* = 6 (20.7%)

($coef$ = 0.64, SE = 0.20, z = 3.239, p = .001) in their responses—significantly correlated with their ability to recall the recommended dishes. However, no significant correlation was found between prior interaction and recall of the recommended dietary suggestions.

Based on the correlation results, we conducted mediation analyses by treating the CA customizability (group assignment) as the independent variable, participants' conversation engagement as mediators, and their recall of the recommendations as a dependent variable. The analysis revealed an indirect positive relationship between the CA customizability and participants' ability to recall the recommended dishes through providing additional contexts during the conversations, ab = .11, SE = .05, CI = .022 to .229. This finding indicated that compared with the PB group, participants in the PC group who included more additional contexts in their responses were more likely to recall the recommended dishes.

5.2.2 Recommendation Adherence. To assess the recommendation adherence of the participants who completed the follow-up survey, we included only those who could at least partially recall the recommended dishes or dietary suggestions. We categorized these participants into three groups according to their self-report results described in 4.3: those who (1) partially or fully adhered to both the recommended dishes and suggestions (n = 21), (2) partially or fully adhered to either the recommended dishes or suggestions (n = 47), and (3) did not adhere to any recommendations (n = 25). Unlike the analysis used in Table 6, which counted the number of participants in the PC and PB groups adhering to dishes and dietary suggestions separately, we combined the recommended dishes and suggestions as the overall recommendation adherence because our goal was to understand how much participants were willing to follow the recommendations regardless of specific dishes or suggestions, which were subject to their life constraints. Also, as shown in Table 6, the sample size of suggestion adherence was relatively small, and the distribution was unbalanced. To increase the statistical power, we did not distinguish dish and suggestion adherence.

Although providing additional contexts in their responses was highly correlated with participants' ability to recall the recommended dishes, this variable was not related to their adherence to the recommendations, nor was the customizability of the CA. However, we found other factors from participants' prior interactions positively correlated with their adherence, including their self-reported intention to follow the recommendations ($coef$ = 0.66, SE = 0.22, z = 2.999, p = .003) and their sensitive disclosure in the responses to the CA ($coef$ = 0.79, SE = 0.35, z = 2.265, p = .024). Since over 90% of the participants in both groups stated a willingness to follow some parts of the recommendations in the future, we did not run statistical analyses on this measure.

5.2.3 Realistic Challenges. As the above results showed, adhering to the recommendations in practice was challenging, despite the recognition of the recommendation's relevance. Participants explained the challenges mainly came from the resource constraints (n = 24), including food access barriers, such as budget and food availability, and lacking storage and cooking facilities (e.g., "*I don't have an oven*" (PB-14)). Some brought up situational constraints (n = 11), such as time and location restrictions. They also faced social constraints (n = 8), such as family responsibility and a lack of agency to decide meals. Moreover, participants acknowledged the challenges of lacking motivation to try food beyond their daily routines (n = 12). For example, PC-18 noted that trying the recommendations itself is not challenging but a feasible plan is needed: "*I don't think there are any firm obstacles in trying the recommendations, but it would take a conscious effort to do so. I'm not always thoughtful about the meals I plan or what I end up eating.*"

6 DISCUSSIONS

Our findings showed that CA customizability alone did not impact the recommendation acceptance rate, as over 97% of participants in both groups accepted the recommendations. We suspect this was partly because both versions of SmartEats could gather sufficient information to enable meaningful personalization to meet participants' expectations. However, we also observed an intention-action gap, where individuals' initial acceptance of the recommendations did not align with their recall of and adherence to the recommendations later, suggesting a need for designing more practical dietary CRS that considers real-life constraints. Additionally, we uncovered several insights into participants' CA customization practices and the interrelationships among their conversation engagement, recommendation perceptions, recall, and adherence. In the following, we connect these findings to explore the opportunities of enhancing user experience and recommendation adherence with CRS for health support.

6.1 Combing System-enabled Personalization With Open Customization

While prior studies have primarily focused on identifying optimal CA characteristics for providing recommendations [39, 40, 82], our study showed that among the two basic options (i.e., personality, conversational style) alongside additional free-form customization, some preferences were indeed consistent across users, but others were not. Specifically, participants' preferences towards the CA's personality were highly correlated with their own personalities, especially for extroverted participants. Echoing prior work showing that personality congruence between people and CAs would lead to better interaction experiences [39, 82], our findings revealed that most participants were able to sense such congruence before interaction. As for conversational style, the majority of participants preferred the CA to communicate casually, despite prior work showing that a formally speaking CA appeared to be better received [15]. This different observation may be due to the context of the conversation, as the previous study was about personal financial and sensitive health topics, where formal speech might be perceived as more appropriate and reassuring. Ours centered on dietary practice, with the idea of exploring healthy, delicious, and potentially exotic foods, may capture individuals' interests and curiosity so that they prefer lively, engaging interactions with the CA at initial encounters [35]. Thus, future work on designing CAs for dietary recommendations could consider users' personalities while prioritizing more relaxing, less formal styles.

On the other hand, over half of the participants in the PC group opted to specify additional characteristics of the CA, which were closely related to their dietary goals or the health aspects they wanted to improve. While highlighting personal goals in health support system design is not new [28, 66], it is noteworthy that, as shown in Section 5.1.1, individuals' goals could extend beyond what is generally considered diet-related and be more nuanced and complex (e.g., improving "*flexibility and balance as well as strength*" and "*glute growth*"). Moreover, some participants meticulously described the demographic or cultural characteristics of the CA, such as "*south east Asian female*" and with "*Yorkshire accent*." Although these details may reflect their own backgrounds, the specific emphases—one on the region and the other on the accent—were not entirely clear until the participants elaborated further. Likewise, when requesting personal traits of the CA, participants showed various preferences such as making the CA "*inspirational, creative, and sympathetic*" or "*understanding the hardships of overcoming weight loss*." Similar findings have been shown in prior literature that examined how people customize LLMs to meet their emotional needs [34, 50, 52], where the personas they wished to engage varied in numerous details depending not only on their general preferences but also their situational moods and physical environments. Such individualized, nuanced needs and preferences may not be entirely fulfilled through system-enabled personalization, but providing the option for individuals to openly customize the CA paves the way for them to articulate these needs and preferences. We suspect this contributed to the PC group reporting higher interaction satisfaction with the CA. Thus, in building personalized dietary CRS, we believe that granting individuals the flexibility to customize the CA combined with the system-enabled personalization, can allow articulation of personal needs and build more meaningful interaction [71].

6.2 CA Customizability For Enhancing Engagement and Information Retention

Our findings showed a positive mediation effect between the CA customizability and participants' recall of the recommended dishes, through the richness of additional contexts they provided. The reasons behind these observations could be multifold. First, the PC group may engage in the conversations more actively and, subsequently, share more additional personal contexts that the CA did not ask for. In addition to our findings in Section 6.1 that the CA indeed performed better in aligning with individual interaction preferences, the customization process itself might have created a sense of agency to foster trust, supported by our observation (see Table 4). This observation can be related to the "IKEA effect," where people place a higher value on the items they created [89], or to the increased "sense of identity," where people perceive the customized product as a reflection of themselves [90]. Supporting this interpretation, within the PC group, participants who had customized the *additional characteristics* option of the CA shared richer additional contexts than those who had not, likely due to their greater investment in the customizing process and a more autonomous input method [65, 71, 90]. Therefore, participants in the PC group who had experienced better conversation engagement by sharing more contexts might pay more attention to the recommended dishes and show better information retention—the recall of the recommended dishes in our study context—compared to those who did not share more contextual information despite having customization options.

However, such a mediation effect was not observed for the recall of the dietary suggestions, possibly because they were more challenging to remember as they appeared across the conversations and may include specific terminology. Additionally, unlike recommended dishes accompanied by food photos, the purely text-based information may lack visual cues that aid memory retention. Although prior works did not specifically examine the roles of food images in supporting recommendation recall, images are great media to enhance user engagement, if delivered appropriately [19, 64, 93]. Connecting to the findings on customization, future research could explore whether image style customization (e.g., realistic vs. cartoon, background specification) [19] could impact user experience and recall in dietary recommendation contexts.

6.3 Fostering a Supportive and Motivating Information Sharing Space

In our study, neither CA customizability nor recommendation recall was related to individual adherence, which highlighted several challenging questions. For example, how can design strategies that are known to affect user experiences be applied to make a real-world impact? How can recommendation recall be transformed into adherence in practice? Below, we try to interpret our findings related to recommendation adherence and explore future directions.

As our findings highlighted, several realistic challenges, such as budget and time constraints, prevented participants from adhering to the recommendations, even though they could recall them. Despite that, there was consistency between participants' reported intentions and their actual behaviors: those who reported a higher likelihood to follow the recommendations in the first stage indeed showed a higher adherence later, with a correlation coefficient of .66 (see Section 5.2.2). This observation is aligned with the theory of self-efficacy introduced by Bandura, which suggests that one's belief in successfully accomplishing a task can influence one's likelihood of completing it [6]. Therefore, future dietary recommendation systems should not only provide personalized suggestions but also tips for users to overcome practical barriers, enhancing users' self-efficacy.

Another factor we identified that correlates with the adherence of participants was their sensitive disclosure during the conversations ($r = 0.80$), which was an even stronger predictor than self-reported intention. We suspect that those who disclosed more sensitive information were more intrinsically motivated to improve their diet. The act of sharing sensitive information often suggests a high level of commitment to making changes [62, 81]. This finding underscores the importance of fostering trust and psychological safety in health-related CA conversations to encourage user engagement and sustainable behavior change. At the same time, as proposed in self-efficacy studies [6], we should also acknowledge that intrinsic motivation plays a fundamental role in enabling sustainable behavior change through technology [13, 43, 55].

To address motivation issues, we can start by exploring design opportunities to help people understand the importance and benefits of recommendations. In health behavior promotion, this could include projecting positive outcomes related to one's health goals with concrete examples. For instance, researchers have found that forecasting one's future weight based on self-tracked calorie data could greatly boost their motivation to reduce calorie intake [8, 78]. With the advancement of generative AI, this approach can be enriched with more creative techniques [12, 53], such as leveraging personalized narratives to cultivate health literacy, or visual-aided annotations to illustrate the potential impact of the changes. Furthermore, as promoting health behavior is a long-term process, future work should investigate more situated user experiences and behaviors, which we consider an important next step.

7 Study Limitations and Future Work

First, our CRS design has inherent limitations associated with LLMs' digression from prompts or hallucination issues related to image generation [3, 51, 92]. These limitations manifested in a few instances of our data, where the CA failed to detect users' requests for clarification and generated inaccurate recommendations (e.g., recommending meat to vegetarians) or low-quality images, highlighting the need for more robust intent recognition and conversation flow management in future iterations. Second, occasional technical issues caused negative interaction experiences. For instance, the CA's slow responses due to network instability could potentially erode participants' patience to some extent. Third, despite our efforts to recruit participants from diverse regions, the study involves more participants from Western countries, potentially introducing cultural bias in our findings. Thus, more culturally diverse sampling is needed in future research to ensure the generalizability of results across different dietary cultures and preferences.

As the first step to exploring the customization effect of a CA in dietary recommendations, our study gathered user experience data from a large-scale sample of 214 participants, covering their dietary practices. We examined how the customizability of the CA influenced participants' engagement in conversations, their interaction experiences, and their recall and adherence to recommendations. The lessons learned from our work can be extended to other healthy lifestyle contexts where incorporating a customizable CA has the potential to improve interaction experience, as many personal health assistants follow a similar interaction flow—collecting information from end users and then making personalized recommendations. In the context of fitness coaching, for example, individuals can customize the CA to align with their specific fitness goals. By interacting with the CA and sharing information about their real-life situations, individuals can develop fitness knowledge while building self-confidence with empathetic responses, potentially enhancing their motivation to achieve their goals. Going forward, we plan to upgrade SmartEats to provide dietary recommendations for specific groups living with dietary problems, such as those with eating disorders. To gather a more holistic personal health profile, we can integrate data captured from other sources such as fitness data on wearable devices.

8 CONCLUSION

In this work, we designed and built SmartEats, which incorporates a customizable conversational agent (CA) to collect user information and deliver recommendations. To explore the design opportunities for enhancing user experience and recommendation adherence, we examined the CA customizability by comparing the performance of baseline and customizable versions of SmartEats through an online experiment with 214 participants. Our study gathered rich quantitative and qualitative insights, showing how enabling CA customizability positively affected participants' conversation engagement, conversation experience, and perceived recommendation quality. We also uncovered aspects of interaction that could affect participants' recall and adherence to the recommendations. Reflecting upon the findings, we explored opportunities for the CA customizability to improve user experience during interaction as well as recommendation recall and adherence in practice, extending these insights to broader health behavior contexts. Our findings pave the way for future research in health interventions that leverage CAs to improve individual well-being and public health outcomes.

Acknowledgments

This project was supported by City University of Hong Kong (#9610597 and #9229183).

References

[1] Gediminas Adomavicius, Jesse Bockstedt, Shawn Curley, and Jingjing Zhang. 2011. Recommender systems, consumer preferences, and anchoring effects. In *RecSys 2011 workshop on human decision making in recommender systems*. Citeseer, 35–42. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4552ba5ac5cbb6c556ebd228c8f9116b46793bc2

[2] Giuseppe Agapito, Mariadelina Simeoni, Barbara Calabrese, Ilaria Caré, Theodora Lamprinoudi, Pietro H. Guzzi, Arturo Pujia, Giorgio Fuiano, and Mario Cannataro. 2018. DIETOS: A dietary recommender system for chronic diseases monitoring and management. *Computer Methods and Programs in Biomedicine* 153 (2018), 93–104. https://doi.org/10.1016/j.cmpb.2017.10.014

[3] Hussam Alkaissi and Samy Mcfarlane. 2023. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus* 15 (02 2023). https://doi.org/10.7759/cureus.35179

[4] Rossella Caporizzi Antonio Derossi, Ahmad Husain and Carla Severini. 2020. Manufacturing personalized food for people uniqueness. An overview from traditional to emerging technologies. *Critical Reviews in Food Science and Nutrition* 60, 7 (2020), 1141–1159. https://doi.org/10.1080/10408398.2018.1559796 arXiv:https://doi.org/10.1080/10408398.2018.1559796 PMID: 30668142.

[5] Diabetes Atlas et al. 2015. International diabetes federation. *IDF Diabetes Atlas, 7th edn. Brussels, Belgium: International Diabetes Federation* 33, 2 (2015).

[6] Albert Bandura. 1982. Self-efficacy mechanism in human agency. *American psychologist* 37, 2 (1982), 122. https://doi.org/10.1037/0003-066X.37.2.122

[7] Joshua Biro, Courtney Linder, and David Neyens. 2023. The Effects of a Health Care Chatbot's Complexity and Persona on User Trust, Perceived Usability, and Effectiveness: Mixed Methods Study. *JMIR Hum Factors* 10 (1 Feb 2023), e41017. https://doi.org/10.2196/41017

[8] Johnna Blair, Yuhan Luo, Ning F Ma, Sooyeon Lee, and Eun Kyoung Choe. 2018. OneNote Meal: A photo-based diary study for reflective meal tracking. In *AMIA Annual Symposium Proceedings*, Vol. 2018. American Medical Informatics Association, 252. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371351/

[9] Corinne L Bush, Jeffrey B Blumberg, Ahmed El-Sohemy, Deanna M Minich, Jóse M Ordovás, Dana G Reed, and Victoria A Yunez Behm. 2020. Toward the Definition of Personalized Nutrition: A Proposal by The American Nutrition Association. *Journal of the American College of Nutrition* 39, 1 (2020), 5–15. https://doi.org/10.1080/07315724.2019.1685332 arXiv:https://doi.org/10.1080/07315724.2019.1685332 PMID: 31855126.

[10] Davide Calvaresi, Stefan Eggenschwiler, Jean-Paul Calbimonte, Gaetano Manzo, and Michael Schumacher. 2021. A personalized agent-based chatbot for nutritional coaching. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. 682–687. https://doi.org/10.1145/3486622.3493992

[11] Marco Cascella, Jonathan Montomoli, Valentina Bellini, and Elena Bignami. 2023. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *Journal of medical systems* 47, 1 (2023), 33. https://doi.org/10.1007/s10916-023-01925-4

[12] Qing Chen, Wei Shuai, Jiyao Zhang, Zhida Sun, and Nan Cao. 2024. Beyond Numbers: Creating Analogies to Enhance Data Comprehension and Communication with Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14. https://doi.org/10.1145/3613904.3642480

[13] Eun Kyoung Choe, Bongshin Lee, Sean Munson, Wanda Pratt, and Julie A Kientz. 2013. Persuasive performance feedback: The effect of framing on self-efficacy. In *AMIA annual symposium proceedings*, Vol. 2013. American Medical Informatics Association, 825. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900219/

[14] Deepanjali Chowdhury, Ahana Roy, Sreenivasan Ramasamy Ramamurthy, and Nirmalya Roy. 2023. CHARLIE: A Chatbot That Recommends Daily Fitness and Diet Plans. In *2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. 116–121. https://doi.org/10.1109/PerComWorkshops56833.2023.10150359

[15] Samuel Rhys Cox and Wei Tsang Ooi. 2022. Does Chatbot Language Formality Affect Users' Self-Disclosure?. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) *(CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 1, 13 pages. https://doi.org/10.1145/3543829.3543831

[16] Florian Daniel, Maristella Matera, Vittorio Zaccaria, and Alessandro Dell'Orto. 2018. Toward truly personal chatbots: on the development of custom conversational assistants. In *Proceedings of the 1st International Workshop on Software Engineering for Cognitive Services* (Gothenburg, Sweden) *(SE4COG '18)*. Association for Computing Machinery, New York, NY, USA, 31–36. https://doi.org/10.1145/3195555.3195563

[17] Deepseek. 2024. https://www.deepseek.com/.

[18] Igor Dolgov, William J. Graves, Matthew R. Nearents, Jeremy D. Schwark, and C. Brooks Volkman. 2014. Effects of cooperative gaming and avatar customization on subsequent spontaneous helping behavior. *Computers in Human Behavior* 33 (2014), 49–55. https://doi.org/10.1016/j.chb.2013.12.028

[19] Hélène Dujardin. 2011. *Plate to pixel: Digital food photography and styling.* John Wiley & Sons.

[20] Shishir Dwivedi, Nivedita Srivastava, Varun Rawal, and Deepali Dev. 2024. Healpal Chatmate: AI Driven Disease Diagnosis and Recommendation System. In *2024 2nd International Conference on Disruptive Technologies (ICDT)*. 1404–1408. https://doi.org/10.1109/ICDT61202.2024.10489509

[21] Mehdi Elahi, Mouzhi Ge, Francesco Ricci, David Massimo, and Shlomo Berkovsky. 2014. Interactive Food Recommendation for Groups.. In *Recsys posters*.

[22] Ahmed Fadhil. 2018. Can a chatbot determine my diet?: Addressing challenges of chatbot application for meal recommendation. *arXiv preprint arXiv:1802.09100* (2018). https://arxiv.org/abs/1802.09100

[23] Ahmed Fadhil and Silvia Gabrielli. 2017. Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. In *Proceedings of the 11th EAI international conference on pervasive computing technologies for healthcare*. 261–265. https://doi.org/10.1145/3154862.3154914

[24] firebase. 2024. https://firebase.google.cn/.

[25] flask. 2024. https://flask.palletsprojects.com/en/3.0.x/.

[26] Jill Freyne and Shlomo Berkovsky. 2010. Intelligent food planning: personalized recipe recommendation. In *Proceedings of the 15th International Conference on Intelligent User Interfaces* (Hong Kong, China) *(IUI '10)*. Association for Computing Machinery, New York, NY, USA, 321–324. https://doi.org/10.1145/1719970.1720021

[27] Jill Freyne, Shlomo Berkovsky, and Gregory Smith. 2011. Recipe recommendation: accuracy and reasoning. In *International conference on user modeling, adaptation, and personalization*. Springer, 99–110. https://doi.org/10.1007/978-3-642-22362-4_9

[28] 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* (<conf-loc>, <city>Yokohama</city>, <country>Japan</country>, </conf-loc>) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 206, 17 pages. https://doi.org/10.1145/3411764.3445555

[29] Mouzhi Ge, Mehdi Elahi, Ignacio Fernaández-Tobías, Francesco Ricci, and David Massimo. 2015. Using tags and latent factors in a food recommender system. In *Proceedings of the 5th international conference on digital health 2015*. 105–112. https://doi.org/10.1145/2750511.2750528

[30] Mouzhi Ge, Mehdi Elahi, Ignacio Fernaández-Tobías, Francesco Ricci, and David Massimo. 2015. Using Tags and Latent Factors in a Food Recommender System. In *Proceedings of the 5th International Conference on Digital Health 2015* (Florence, Italy) *(DH '15)*. Association for Computing Machinery, New York, NY, USA, 105–112. https://doi.org/10.1145/2750511.2750528

[31] Gemini. 2024. https://gemini.google.com/.

[32] Generative pretrained transformer. 2024. https://openai.com/.

[33] Kazjon Grace, Elanor Finch, Natalia Gulbransen-Diaz, and Hamish Henderson. 2022. Q-Chef: The impact of surprise-eliciting systems on food-related decision-making. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>New Orleans</city>, <state>LA</state>, <country>USA</country>, </conf-loc>) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 11, 14 pages. https://doi.org/10.1145/3491102.3501862

[34] Juhye Ha, Hyeon Jeon, Daeun Han, Jinwook Seo, and Changhoon Oh. 2024. CloChat: Understanding How People Customize, Interact, and Experience Personas in Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–24. https://doi.org/10.1145/3613904.3642472

[35] Clyde Hendrick and S Brown. 1971. Introversion, extraversion, and interpersonal attraction. *Journal of personality and social psychology* 20 1 (1971), 31–6. https://api.semanticscholar.org/CorpusID:5182220

[36] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. 2023. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *arXiv preprint arXiv:2311.05232* (2023). https://arxiv.org/abs/2311.05232

[37] Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2021. A Survey on Conversational Recommender Systems. *ACM Comput. Surv.* 54, 5, Article 105 (may 2021), 36 pages. https://doi.org/10.1145/3453154

[38] Hao Jiang, Wenjie Wang, Meng Liu, Liqiang Nie, Ling-Yu Duan, and Changsheng Xu. 2019. Market2Dish: A Health-aware Food Recommendation System. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 2188–2190. https://doi.org/10.1145/3343031.3350594

[39] Eunjoo Jin and Matthew S Eastin. 2023. Birds of a feather flock together: matched personality effects of product recommendation chatbots and users. *Journal of*

Research in Interactive Marketing 17, 3 (2023), 416–433. https://www.emerald.com/insight/content/doi/10.1108/JRIM-03-2022-0089/full/html

[40] Eunjoo Jin, Yuhosua Ryoo, Woojin Kim, and Y. Song. 2024. Bridging the health literacy gap through AI chatbot design: the impact of gender and doctor cues on chatbot trust and acceptance. *Internet Research* (09 2024). https://doi.org/10.1108/INTR-08-2023-0702

[41] Eunkyung Jo, Yuin Jeong, Sohyun Park, Daniel A. Epstein, and Young-Ho Kim. 2024. Understanding the Impact of Long-Term Memory on Self-Disclosure with Large Language Model-Driven Chatbots for Public Health Intervention. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 440, 21 pages. https://doi.org/10.1145/3613904.3642420

[42] Anjali Khurana, Parsa Alamzadeh, and Parmit K. Chilana. 2021. ChatrEx: Designing Explainable Chatbot Interfaces for Enhancing Usefulness, Transparency, and Trust. In *2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. 1–11. https://doi.org/10.1109/VL/HCC51201.2021.9576440

[43] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, KwonHyun Kim, and Jinwook Seo. 2016. TimeAware: Leveraging Framing Effects to Enhance Personal Productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 272–283. https://doi.org/10.1145/2858036.2858428

[44] Katharina Klein and Luis F Martinez. 2023. The impact of anthropomorphism on customer satisfaction in chatbot commerce: an experimental study in the food sector. *Electronic commerce research* 23, 4 (2023), 2789–2825. https://link.springer.com/article/10.1007/s10660-022-09562-8

[45] Marion Koelle, Swamy Ananthanarayan, and Susanne Boll. 2020. Social acceptability in HCI: A survey of methods, measures, and design strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–19. https://doi.org/10.1145/3313831.3376162

[46] Mohammad Amin Kuhail, Justin Thomas, Salwa Alramlawi, Syed Jawad Hussain Shah, and Erik Thornquist. 2022. Interacting with a Chatbot-Based Advising System: Understanding the Effect of Chatbot Personality and User Gender on Behavior. *Informatics* 9, 4 (2022). https://doi.org/10.3390/informatics9040081

[47] Mohammad Amin Kuhail, Justin Thomas, Salwa Alramlawi, Syed Jawad Hussain Shah, and Erik Thornquist. 2022. Interacting with a Chatbot-Based Advising System: Understanding the Effect of Chatbot Personality and User Gender on Behavior. *Informatics* 9 (2022), 81. https://api.semanticscholar.org/CorpusID:252838493

[48] SeoYoung Lee and Junho Choi. 2017. Enhancing user experience with conversational agent for movie recommendation. *Int. J. Hum.-Comput. Stud.* 103, C (July 2017), 95–105. https://doi.org/10.1016/j.ijhcs.2017.02.005

[49] Zhuoyang Li, Minhui Liang, Ray Lc, and Yuhan Luo. 2024. StayFocused: Examining the Effects of Reflective Prompts and Chatbot Support on Compulsive Smartphone Use. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 247, 19 pages. https://doi.org/10.1145/3613904.3642479

[50] Zhuoyang Li, Minhui Liang, Ray Lc, and Yuhan Luo. 2025. StayFocused: Examining the Effects of Reflective Prompts and Chatbot Support on Compulsive Smartphone Use *(CHI '25)*. Association for Computing Machinery, New York, NY, USA.

[51] Zhuoyang Li, Minhui Liang, Hai Trung Le, Ray Lc, and Yuhan Luo. 2023. Exploring Design Opportunities for Reflective Conversational Agents to Reduce Compulsive Smartphone Use. In *Proceedings of the 5th International Conference on Conversational User Interfaces*. 1–6. https://doi.org/10.1145/3571884.3604305

[52] Zhuoyang Li, Zihao Zhu, Xinning Gui, and Yuhan Luo. [n. d.]. "This is Human Intelligence Debugging Artificial Intelligence": Examining How People Prompt GPT in Seeking Mental Health Support. *Available at SSRN 4875898* ([n. d.]). http://dx.doi.org/10.2139/ssrn.4875898

[53] Sue Lim and Ralf Schmälzle. 2023. Artificial intelligence for health message generation: an empirical study using a large language model (LLM) and prompt engineering. *Frontiers in Communication* 8 (2023), 1129082. https://doi.org/10.3389/fcomm.2023.1129082

[54] Yuhan Luo, Young-Ho Kim, Bongshin Lee, Naeemul Hassan, and Eun Kyoung Choe. 2021. FoodScrap: Promoting Rich Data Capture and Reflective Food Journaling Through Speech Input. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 606–618. https://doi.org/10.1145/3461778.3462074

[55] Yuhan Luo, Bongshin Lee, and Eun Kyoung Choe. 2020. TandemTrack: shaping consistent exercise experience by complementing a mobile app with a smart speaker. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 1–13. https://doi.org/10.1145/3313831.3376616

[56] Yuhan Luo, Bongshin Lee, Young-Ho Kim, and Eun Kyoung Choe. 2022. NoteWordy: Investigating Touch and Speech Input on Smartphones for Personal Data Capture. *Proc. ACM Hum.-Comput. Interact.* 6, ISS, Article 581 (nov 2022), 24 pages. https://doi.org/10.1145/3567734

[57] Yuhan Luo, Peiyi Liu, and Eun Kyoung Choe. 2019. Co-Designing Food Trackers with Dietitians: Identifying Design Opportunities for Food Tracker Customization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300822

[58] Zilin Ma, Yiyang Mei, and Zhaoyuan Su. 2023. Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. In *AMIA Annual Symposium Proceedings*, Vol. 2023. American Medical Informatics Association, 1105. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10785945/

[59] Michael Macht. 2008. How emotions affect eating: A five-way model. *Appetite* 50, 1 (2008), 1–11. https://doi.org/10.1016/j.appet.2007.07.002

[60] David P MacKinnon, Amanda J Fairchild, and Matthew S Fritz. 2007. Mediation analysis. *Annu. Rev. Psychol.* 58, 1 (2007), 593–614. https://doi.org/10.1146/annurev.psych.58.110405.085542

[61] Patrick E McKnight and Julius Najab. 2010. Mann-Whitney U Test. *The Corsini encyclopedia of psychology* (2010), 1–1. https://doi.org/10.1002/9780470479216.corpsy0524

[62] Ken G Meleshko and Lynn E Alden. 1993. Anxiety and self-disclosure: toward a motivational model. *Journal of personality and social psychology* 64, 6 (1993), 1000. https://doi.org/10.1037/0022-3514.64.6.1000

[63] Milica Milosavljevic, Vidhya Navalpakkam, Christof Koch, and Antonio Rangel. 2012. Relative visual saliency differences induce sizable bias in consumer choice. *Journal of consumer psychology* 22, 1 (2012), 67–74. https://doi.org/10.1016/j.jcps.2011.10.002

[64] Weiqing Min, Shuqiang Jiang, and Ramesh Jain. 2020. Food Recommendation: Framework, Existing Solutions, and Challenges. *IEEE Transactions on Multimedia* 22, 10 (2020), 2659–2671. https://doi.org/10.1109/TMM.2019.2958761

[65] Daniel Mochon, Michael I. Norton, and Dan Ariely. 2012. Bolstering and restoring feelings of competence via the IKEA effect. *International Journal of Research in Marketing* 29 (2012), 363–369. https://api.semanticscholar.org/CorpusID:36171474

[66] Sean A Munson, Jessica Schroeder, Ravi Karkar, Julie A Kientz, Chia-Fang Chung, and James Fogarty. 2020. The importance of starting with goals in N-of-1 studies. *Frontiers in digital health* 2 (2020), 3. https://doi.org/10.3389/fdgth.2020.00003

[67] Clifford Nass and Kwan Min Lee. 2000. Does computer-generated speech manifest personality? an experimental test of similarity-attraction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. Association for Computing Machinery, New York, NY, USA, 329–336. https://doi.org/10.1145/332040.332452

[68] Kimberly A Neuendorf. 2017. *The content analysis guidebook.* sage.

[69] OpenAI ChatCompletion API. 2024. https://platform.openai.com/docs/guides/gpt/chat-completions-api.

[70] OpenAI Dalle3 API. 2024. https://openai.com/index/dall-e-3/.

[71] Stephen C Paul, Nina Bartmann, and Jenna L Clark. 2021. Customizability in conversational agents and their impact on health engagement. *Human Behavior and Emerging Technologies* 3, 5 (2021), 1141–1152. https://doi.org/10.1002/hbe2.320

[72] Florian Pecune, Lucile Callebert, and Stacy Marsella. 2020. A Socially-Aware Conversational Recommender System for Personalized Recipe Recommendations. In *Proceedings of the 8th International Conference on Human-Agent Interaction* (Virtual Event, USA) *(HAI '20)*. Association for Computing Machinery, New York, NY, USA, 78–86. https://doi.org/10.1145/3406499.3415079

[73] Philips Kokoh Prasetyo, Palakorn Achananuparp, and Ee-Peng Lim. 2021. Foodbot: A Goal-Oriented Just-in-Time Healthy Eating Interventions Chatbot. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare* (Atlanta, GA, USA) *(PervasiveHealth '20)*. Association for Computing Machinery, New York, NY, USA, 436–439. https://doi.org/10.1145/3421937.3421960

[74] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (Chicago, Illinois, USA) *(RecSys '11)*. Association for Computing Machinery, New York, NY, USA, 157–164. https://doi.org/10.1145/2043932.2043962

[75] React Native. 2024. https://reactnative.dev/.

[76] react-native-gifted-chat. 2024. https://github.com/FaridSafi/react-native-gifted-chat.

[77] Valeria Resendez. 2020. *A very formal agent: how culture, mode of dressing and linguistic style influence the perceptions toward an Embodied Conversational Agent?* Master's thesis. University of Twente. https://essay.utwente.nl/82242/

[78] Saeyoung Rho, Injung Lee, Hankyung Kim, Jonghyuk Jung, Hyungi Kim, Bong Gwan Jun, and Youn-kyung Lim. 2017. Futureself: what happens when we forecast self-trackers? Future health statuses?. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. 637–648. https://doi.org/10.1145/3064663.3064676

[79] Hanna Schäfer, Mehdi Elahi, David Elsweiler, Georg Groh, Morgan Harvey, Bernd Ludwig, Francesco Ricci, and Alan Said. 2017. User nutrition modelling and recommendation: Balancing simplicity and complexity. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization*. 93–96.

https://doi.org/10.1145/3099023.3099108

[80] Melanie Schwede, Nika Mozafari, Niclas Von Schnakenburg, and Maik Hammerschmidt. 2023. Can chatbots be persuasive? How to boost the effectiveness of chatbot recommendations for increasing purchase intention. In *Proceedings of the Hawaii international conference on system sciences*, Vol. 56. 3454–3463. https://ssrn.com/abstract=4305315

[81] Joshana Shibchurn and Xiangbin Yan. 2015. Information disclosure on social networking sites: An intrinsic–extrinsic motivation perspective. *Computers in Human Behavior* 44 (2015), 103–117. https://doi.org/10.1016/j.chb.2014.10.059

[82] Michael Shumanov and Lester Johnson. 2021. Making conversations with chatbots more personalized. *Computers in Human Behavior* 117 (2021), 106627. https://doi.org/10.1016/j.chb.2020.106627

[83] Rashmi Sinha and Kirsten Swearingen. 2002. The role of transparency in recommender systems. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI EA '02)*. Association for Computing Machinery, New York, NY, USA, 830–831. https://doi.org/10.1145/506443.506619

[84] Jeffery Sobal and Carole A Bisogni. 2009. Constructing food choice decisions. *Annals of Behavioral Medicine* 38, suppl_1 (2009), s37–s46. https://doi.org/10.1007/s12160-009-9124-5

[85] Amir Tirosh, Ediz S Calay, Gurol Tuncman, Kathryn C Claiborn, Karen E Inouye, Kosei Eguchi, Michael Alcala, Moran Rathaus, Kenneth S Hollander, Idit Ron, et al. 2019. The short-chain fatty acid propionate increases glucagon and FABP4 production, impairing insulin action in mice and humans. *Science translational medicine* 11, 489 (2019), eaav0120. https://doi.org/10.1126/scitranslmed.aav0120

[86] Thi Ngoc Trang Tran, Müslüm Atas, Alexander Felfernig, and Martin Stettinger. 2018. An overview of recommender systems in the healthy food domain. *Journal of Intelligent Information Systems* 50 (2018), 501–526. https://link.springer.com/content/pdf/10.1007/s10844-017-0469-0.pdf

[87] Graham JG Upton. 1992. Fisher's exact test. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 155, 3 (1992), 395–402. https://doi.org/10.2307/2982890

[88] Meghna Verma, Raquel Hontecillas, Nuria Tubau-Juni, Vida Abedi, and Josep Bassaganya-Riera. 2018. Challenges in Personalized Nutrition and Health. *Frontiers in Nutrition* 5 (2018). https://doi.org/10.3389/fnut.2018.00117

[89] Rebecca Wald, Evelien Heijselaar, and Tibor Bosse. 2021. Make your own: The Potential of Chatbot Customization for the Development of User Trust. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (Utrecht, Netherlands) *(UMAP '21)*. Association for Computing Machinery, New York, NY, USA, 382–387. https://doi.org/10.1145/3450614.3463600

[90] Jinping Wang and S. Shyam Sundar. 2022. Are We More Reactive to Persuasive Health Messages When They Appear in Our Customized Interfaces? The Role of Sense of Identity and Sense of Control. *Health Communication* 37, 8 (2022), 1022–1030. https://doi.org/10.1080/10410236.2021.1885772 arXiv:https://doi.org/10.1080/10410236.2021.1885772 PMID: 33596717.

[91] Peilu Wang, Mingyang Song, A Heather Eliassen, Molin Wang, Teresa T Fung, Steven K Clinton, Eric B Rimm, Frank B Hu, Walter C Willett, Fred K Tabung, et al. 2023. Optimal dietary patterns for prevention of chronic disease. *Nature medicine* 29, 3 (2023), 719–728. https://doi.org/10.1038/s41591-023-02235-5

[92] Jing Wei, Sungdong Kim, Hyunhoon Jung, and Young-Ho Kim. 2024. Leveraging Large Language Models to Power Chatbots for Collecting User Self-Reported Data. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 87 (apr 2024), 35 pages. https://doi.org/10.1145/3637364

[93] Longqi Yang, Cheng-Kang Hsieh, Hongjian Yang, John P Pollak, Nicola Dell, Serge Belongie, Curtis Cole, and Deborah Estrin. 2017. Yum-me: a personalized nutrient-based meal recommender system. *ACM Transactions on Information Systems (TOIS)* 36, 1 (2017), 1–31. https://doi.org/10.1145/3072614

[94] Zhongqi Yang, Elahe Khatibi, Nitish Nagesh, Mahyar Abbasian, Iman Azimi, Ramesh Jain, and Amir M Rahmani. 2024. ChatDiet: Empowering Personalized Nutrition-Oriented Food Recommender Chatbots through an LLM-Augmented Framework. *arXiv preprint arXiv:2403.00781* (2024). https://arxiv.org/abs/2403.00781