공학박사학위논문

Study on User Experience with Mimicked Persona of Conversational AI in Daily Healthcare

헬스케어를 위한 대화형 인공지능의 모사된 페르소나 디자인 및 사용자 경험 연구

2021 년 08 월

서울대학교 대학원

협동과정 인지과학

황 유 진

헬스케어를 위한 대화형 인공지능의 모사된 페르소나 디자인 및 사용자 경험 연구

지도 교수 이 준 환

이 논문을 공학박사 학위논문으로 제출함

2021년 4월

서울대학교 대학원

인지과학 협동과정

황 유 진

황유진의 공학박사 학위논문을 인준함

2021년 6월

위 원 장 ______ 윤 명 환 ______

부위원장 ______ 이 준 환 ______

위 원 ______ 한 소 원 ______

위 원 ______ 조 비 룡 ______

위 원 ______ 박 선 영 ______

Abstract

Advance in digital healthcare technologies has been leading a revolution in healthcare. It has been showing the enormous potential to improve medical professionals' ability for accurate diagnosis, disease treatment, and the users' daily self-care. Since the recent transformation of digital healthcare aims to provide effective personalized health services, Conversational AI (CA) is being highlighted as an easy-to-use and cost-effective means to deliver personalized services.

Particularly, CA is gaining attention as a mean for personalized care by ingraining positive self-care behavior in a daily manner while previous methods for personalized care are focusing on the medical context. CA expands the boundary of personalized care by enabling one-to-one tailored conversation to deliver health education and healthcare therapies. Due to CA's opportunities as a method for personalized care, it has been implemented with various types of roles including CA for diagnosis, CA for prevention, and CA for therapy.

However, there lacks study on the personalization of healthcare CA to meet user's preferences on the CA's persona. Even though the CASA paradigm has been applied to previous studies designing and evaluating the human-likeness of CA, few healthcare CAs personalize its human-like persona except some CAs for mental health therapy.

Moreover, there exists the need to improve user experience by increasing social and emotional interaction between the user and the CA. Therefore, designing an acceptable and personalized persona of CA should be also considered to make users to be engaged in the healthcare task with the CA. In this manner, the thesis suggests an idea of applying the persona of the person who is in a close relationship with the user to the conversational CA for daily healthcare as a strategy for persona personalization. The main hypothesis is the idea of applying a close person's persona would improve user engagement. To investigate the hypothesis, the thesis explores if dynamics derived from the social relationship in the real world can be implemented to the relationship between the user and the CA with the persona of a close person.

To explore opportunities and challenges of the research idea, series of studies were conducted to (1) explore appropriate host whose persona would be implemented to healthcare CA, (2) define linguistic characteristics to consider when applying the persona of a close person to the CA, and (3)implement CA with the persona of a close person to major lifestyle domains. Based on findings, the thesis provides design guidelines for healthcare CA with the persona of the real person who is in a close relationship with the user.

Keywords: Conversational AI, healthcare, mimicked persona, user perception, user experience, persona design

Student Number: 2017-39859

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Digital health technologies continue to evolve in both numbers and capabilities. To explain the concept of digital health technologies we referred to a definition from the Worldwide Health Organization (WHO) introduced in its guidelines on digital health systems (2019). According to WHO, digital health is a broad concept that includes e-Health and emerging technologies, such as advanced computing technologies of big data, personalized genomics, and artificial intelligence [1]. Also, the additional definition was suggested by U.S. Food and Drug Administration (FDA), which is similar to WHO's in a broad aspect. According to FDA [2], the scope of digital health encompasses categories such as health informatics, mobile health, wearable devices, telemedicine, and personalized medicine.

The need for the development of digital health technologies has been emphasized by FDA since they have been leading a revolution in the domain of health care showing the enormous potential to improve professionals' ability for accurate diagnosis and disease treatment and improvement in users' daily health care. Moreover, digital health technologies take over the role of healthcare providers by automating tasks with higher accuracy (e.g. exercise reminder) [3, 4]

In detail, the domain that digital health technologies are being tackled includes the following : (1) Electronic record for healthcare decision support, (2) virtual visit based on voice or video interactions, (3) digital therapeutics improving clinical outcomes and adherence, (4) wearable monitor (e.g. activity trackers and sleep monitors), (5) mobile health apps (e.g. medical diary, diet, and lifestyle tracker), and (6) artificial intelligence and machine learning that enables automation of specific tasks and performs predictive tasks [5].

Since providing effective personalized health services is a recent transformation of digital healthcare [6, 7], Conversational AI (CA) as an approach for digital health technology in the converging area of multiple digital health domains is being highlighted as an easy-to-use and cost-effective means to deliver personalized services [8]. Generally, CA refers to technologies combining natural language processing (NLP) such as chatbots or voice assistants, with which users can make conversation with [9].

Particularly, CA is gaining attention due to its opportunity in personalized care by ingraining positive self-care behavior in a daily manner [10]. Previous methods for personalized care are mainly conducted in a medical context. However, CA expand the boundary of personalized-care by enabling one-to-one tailored conversation to deliver health education and healthcare therapies [11]. Also, the personalization of CAs is showing effectiveness in improving user comprehension [12], task efficiency [13]. Due to CA's opportunity as a method for personalized care, it has been implemented with various types of roles. According to Jovanovie et al's review study identifying the roles of healthcare CA by analyzing 225 text-based CA currently in service, major roles of healthcare chatbot can be defined as CA for diagnosis, CA for prevention, and CA for therapy [14].

However, there lacks study on the personalization of healthcare CA to meet user's preferences toward the CA as a social actor to increase user engagement. According to the computers are social actors (CASA) paradigm [15], users use similar social norms and rules when interacting with computers and people. Even though the CASA paradigm has been applied to previous studies designing and evaluating the human-likeness of CA, few healthcare CAs handle social cues and emotional cues except for some CAs for mental health therapy. Moreover, Jovanovie et al's review study also emphasizes the need for improving user experience since most commercial healthcare CAs analyzed in this study include unfamiliar characteristics for users and lack social and emotional interaction with the user [14]. Therefore, designing a familiar and acceptable persona of CA should be also considered to make users to be engaged in the healthcare task with the CA.

In general, designing a familiar persona for CA has the possibility of enhancing user engagement by forming a positive relationship between the user and the CA. In previous studies, CAs have shown positive outcomes in building relationships with the user by implementing a certain persona to the agent [16]. However, users are sensitive about whether the CA's personality matches their individualized preference or expectations [17]. According to a previous study, CA's personality could result in low engagement if it does not meet user preference [18]. Therefore, personalizing the persona should also be taken into consideration for increasing user engagement with the healthcare CAs but few studies have dealt with this issue [19]. There exist studies on adapting CA's persona to the user's personality traits. For example, there is a study on adapting CA's personality based on the user's personality traits including extroversion and agreeableness [20]. However, the study did not investigate user experience, and there were also design limitations in designing the personality of the CA to meet user's preferences based on limited traits.

In this manner, our thesis suggests an idea of applying the persona of the person who is in a close relationship with the user to the conversational CA for daily healthcare to increase user engagement. The thesis explores if dynamics derived from the social relationship in the real world can be implemented to the relationship between the user and the CA with the persona of a close person implemented to it. Since the importance of social relationships in healthcare has long been studied and proved its effectiveness in making the user involved in healthcare tasks, we expect the CA mimicking the close person would also mimic positive consequences of the close relationship [21]. According to Berkman et al's study, previous evidence suggests that social dynamics and integration promote healthier behaviors and healthier lifestyles [22].

In various types of social relationships, we focused on the strong relationship by referring to the Granovetter's study since "strength" of the tie is a combination of the amount of the time of relationship that has been persisting, the emotional intensity, the intimacy, and the reciprocal services [23]. To explore opportunities and challenges of our idea, we conducted series of studies to (1) explore appropriate persona for daily healthcare, (2) define linguistic elements that affect persona perception of a close person, and (3)implement CA with the persona of a close person to the lifestyle domains. Based on findings, we provide design guidelines for CA with the persona of the user's healthcare providers while presenting case studies of food journaling, physical activity, and stress management.

My thesis suggested, explored, and evaluated the idea of applying the persona of the user's close person to improve the user experience, particularly user engagement, with the healthcare CAs in the domain of daily health management. The thesis contributes HCI community by following aspects. First, the thesis investigated the effective persona to be implemented in CA as digital health technology. Second, the thesis provided lists of linguistic characteristics to consider when applying healthcare providers' persona to the CA. I expect these features also could be applied to expanded domains when applying the persona of the user's healthcare providers to the text-based CA. Lastly, my thesis provides user-centered design guidelines for applying the persona of a close person to CA to increase user experience with daily healthcare tasks.

Chapter 2

Literature Review

In this chapter, we introduce previous work on our domain of interest. First, we are going to introduce the roles of CA as digital healthcare technology and its opportunities. Second, we are going to introduce previous works on personalization in healthcare CA. Then, we introduce the importance of the CA persona on user engagement in example domains including healthcare. Lastly, we are going to discuss methods for designing CA's dialogue style that highly impacts the persona perception.

2.1 Roles of CA in Healthcare

According to Pereira et al's mapping study on healthcare CA, CA has great possibilities as a technical enabler with the benefits of asynchronicity, consumability, anonymity, authentication, scalability, and personalization [24]. Particularly, text-based CAs provide the opportunity of immediacy and asynchronicity at the same time. When it comes to consumability which can be defined as easy-to-use and easy to access, chatbots show higher performance compared to existing technologies in various aspects: (1) installation is easy since the chatbot is mostly built on existing platforms (e.g. Instant Messaging apps). Also, for privacy issues, the function of interacting anonymously with the system could be the main opportunity. Users could be free from feeling shame when interacting with computers. In other words, they could feel more private due to anonymous interaction in comparison with speaking to real humans [25]. At the same time, chatbots have the potential to be scalable to reach large populations in a cost-effective way [26]. Also, personalization in one's healthcare is one of the most important opportunities of CA that is being utilized in daily healthcare [11].

For these opportunities, CAs are being implemented with multiple roles in the domain of healthcare. CA in healthcare takes part in some of the chores of human healthcare providers which, as a result, reducing their physical, psychological burden [27]. Moreover, CAs have been assessed to support various kinds of tasks including counseling [28, 29], monitoring [30, 31], or medication adherence [32, 33]. According to Jovanovie et al's study identifying the roles of healthcare CA by analyzing 225 text-based CA currently in service, major roles of healthcare chatbot can be defined as CA for diagnosis, CA for prevention and CA for therapy [14].

First, we introduce diagnostic CAs tracking user's symptoms and recommending action plans. Support for diagnosis, general symptom checker, and specific symptom checker are detailed roles of diagnostic CA. In this manner, CA has the opportunity in (1) facilitating access to healthcare services (e.g.Pathology Lab Chatbot) which reduces the distance between doctors and users, (2) supporting consultations with experts and professionals (e.g.iCliniq) that connects users with doctors, and (3) giving people access to the updated symptom and disease information CA also provides To-dos and action plans based on symptom checking. One example is HealthTap, a chatbot that asks about symptoms and delivers medical information about possible reasons through dialog-based exchanges. Another example is FeverBot, assisting users to recommend medical assistance. Also, Mental Care Bot support diagnosing symptoms of mental disorders, along with ADA which analyzes relevant disease based on the reported symptoms.

Second, we introduce CAs for prevention that assists data tracking and health informatics and preventing declines in health status by promoting desirable lifestyles. CAs for prevention include a range of services with the following roles of (1) assisting access to healthcare, (2) assisting health education, and (3) healthcare coaching. CAs assisting access to healthcare help users get used to healthcare services.

Healthcare CA's main goal is reducing the efforts and increasing access to the services. Additionally, CAs for health education teach users on preventive habits for specific symptoms or diseases. DoctorBot is an example that provides information on healthcare topics. Jennifer, another example, is a text-based chatbot that aims to eliminate disinformation by answering queries about COVID-19 symptoms. There are other examples for health coaching CA, whose goal is to improve general well-being by promoting healthy lifestyles. Psychological incentives could be used to facilitate desirable behaviors such as the FitCircle that uses reputation-based incentives to promote exercising and StopBreatheThink that recommends mental practices for mental well-being. Another example is Forksy is which assists the consumption of nutritional meals tailored to user's eating habits.

Lastly, the main role of CA for therapy is to provide treatment for specific symptoms or health statuses such as pregnancy or therapeutic diet. CAs for prevention include a range of services with the following roles of (1) support for therapy, (2) health therapy, and (3) cognitive-behavioral therapy. CA supporting for therapy is designed to be used during the treatment. An example of HealthRobot has the function of personalized reminders to increase adherence to daily medication as part of the disease cure, or listing and rating effectiveness of medicines based on users' reviews for health cures. CA for health therapy also provides at-home therapy for patients. For instance, KetoBot provides information about the ketogenic diet to reduce symptoms related to diabetes. CA for cognitive behavioral therapy (CBT) includes therapies for specific mental states. For example, Woebot tracks users' moods and provides action plans for mental activities. Additionally, Wysa is designed to improve patient's mental wellness through emotional support by the CA. The common goal among these CAs is to form resilience to combat psychological disorders including anxiety, stress, and depression by practicing positive thinking including increased self-awareness and optimism).

Among these roles of CAs we are focusing on the CAs for prevention since our main target group includes users without severe disease whose main motivation for involving in daily healthcare tasks is to prevent health-related diseases and maintain personal well-being, not cure or improve outcomes of a particular disease. As seen above, CA expands the boundary of personalized care by including factors such as lifestyle choices, social context, and daily environment, and personalized health care services like health education and health therapies [11] outperforming previous methods for personalized care that are mainly conducted in the medical context. We introduce more studies about personalization in healthcare CA in the following section.

2.2 Personalization in Healthcare CA

The personalization of CA is increasingly being used in healthcare applications, personalizing the structure, content, or purpose and goal of the dialogue between users and CAs [34]. Personalization in the context of digital technology typically refers to a feature that modifies a system's interface, given information, and content to boost its personal relevance [35]. There exist a range of work about personalized conversational agents focusing on dialogue personalization [36], message personalization [37], personalized recommender systems [38], and personalized adaptive systems [39].

The personalized design of CA increases user satisfaction that could lead to improved user engagement. Improvement in digital technologies could support CA to define users to provide personalized prompts. Smartphone sensors could support collect behavioral data which can strengthen AI algorithms. GPS data can be used for tracking physical activity. Also, it may recognize a user's physical status using facial or motion recognition, or it can send customized messages to patients based on their heart rate condition using a smartphone-based or wearable-based HR monitor. All these approaches promote a personalized conversation experience.

However, there lacks study on the personalization of healthcare CA to meet user's preferences toward the CA as a social actor to increase user engagement. A weak social relationship between the user and the

CA can leads users to low engagement with the CA over time. This limitation of the previous CA design approach can largely affect outcomes of healthcare CA because if the CA cannot ensure continuous user engagement, it fails to reach the system goal for user's daily healthcare regardless of its tremendous benefits [40]. As an approach to building a strong relationship between the healthcare CA and the user, our approach focuses on the persona design of healthcare CA.

Previous researches have shown that users prefer to interact and be engaged with CAs with human-like persona [41], particularly human-like use of language [42]. Moreover, users expect agents to have human-like traits that match the user context and tasks [43, 44, 45]. For example, for house chores, users preferred calm, polite, and cooperative agent [46] and for social interaction, users preferred the chatbot showing high self-disclosure [47].

In the context of daily healthcare, there was attempts to implement human-likeness to CAs by implementing human-like dialogue styles of empathizing [48], expressing gratitude [49], making jokes [50], motivating [51], and praising [52]. To be specific, Bickmore et al insisted that CA's persona implemented with empathetic dialogue style is effective in managing mental health, and demonstrated that subtle dialogue style is effective in promoting exercise [53]. Simulation of human-likeness in CAs can provide an engaging experience [16].

However, users are sensitive about whether the CA's personality matches their individualized preference or expectations [17]. If CAs do not meet users' expectations, users fail to emotionally attach to the CA, which results in low engagement or abandonment [18]. In this manner, personalization of persona should also be considered to increase user experience with the healthcare CA. Therefore, in this thesis, we suggest an idea of applying the dialogue style of the specific person who is in a close relationship with the user to the CA to increase the engagement with the CA.

2.3 Persona Design CA

Users interact with computers and people using comparable social norms and conventions, according to the computers as social actors (CASA) concept. Despite the fact that the CASA paradigm sparked a lot of research into how to use or evaluate CAs' human-likeness, few healthcare CAs manage social and emotional cues, with the exception of a few CAs for mental health therapy. CASA is commonly used to assist human-computer interface (HCI), human-robot interaction (HRI), and media effects studies [54, 55, 56]. Due to developments in technology and increased acceptance of technologies, these subfields and researches have grown over the decades in health care [57], the domestic sphere [58] and education [54].

Even though the CASA paradigm provoked lots of studies in applying or evaluating the human-likeness of CA, healthcare CAs are limited in handling social and emotional dynamics except for some CAs for mental health therapy. Moreover, Jovanovie et al's review study also emphasizes the need to improve user experience since most commercial healthcare CAs analyzed in this study include unfamiliar characteristics for users and lack social and emotional interaction with the user [14]. Therefore, designing a familiar and acceptable persona of CA should be also considered to make users to be engaged in the healthcare task with the CA.

Designing persona of CA has opportunities in increasing user experience by implementing human-like traits to the CA [59]. There are several design elements of CA that could affect the persona perception of CA. we introduce design elements for CA design. We introduce design elements by referring to Laranjo et al's systematic review on healthcare CAs and made improvements by involving recent studies [60, 24].

In the Laranjo et al's study, authors characterized conversational agents based on the following traits: (1) the type of technology, (2) dialogue management, (3) dialogue initiative, (4) input modality, (5) output modality, and (6) task orientation [60]. Software applications supplied by mobile SMS, telephone, or multi-modal platforms are among the platforms that support the conversational agent. When it comes to dialogue management, it includes three major strategies. First is finite-state dialogue management [61] that the user goes through dialogue

Type of technology	Platform supporting the conversational agent: software application delivered via mobile device (e.g. smartphone, tablet), laptop or desktop computer, or via web browser; SMS; telephone; or multimodal platform.	
Dialogue management	Finite-state	The user is taken through a dialogue consisting of a sequence of pre-determined steps or states.
	Frame-based	The user is asked questions that enable the system to fill slots in a template in order to perform a task. The dialogue flow is not pre-determined but depends on the content of the user's input and the information that the system has to elicit.
	Agent-based	These systems enable complex communication between the system, the user and the application. There are many variants of agent-based systems, depending on what particular aspects of intelligent behavior are designed into the system. In agent-based systems, communication is viewed as the interaction between two agents, each of which is capable of reasoning about its own actions and beliefs, and sometimes also about the actions and beliefs of the other agent. The dialogue model takes the preceding context into account with the result that the dialogue evolves dynamically as a sequence of related steps that build on each other.
Dialogue initiative	User	The user leads the conversation
	System	The system leads the conversation
	Mixed	Both the user and the system can lead the conversation
Input modality	Spoken	The user uses spoken language to interact with the system
	Written	The user uses written language to interact with the system
Output modality	Written, spoken, visual (e.g. non-verbal communication like facial expressions or body movements)	
Task-oriented	Yes	The system is designed for a particular task and set up to have short conversations, in order to get the necessary information to achieve the goal (e.g. booking a consultation)
	No	The system is not directed to the short-term achievement of a specific end-goal or task (e.g. purely conversational chatbots)

Figure 2.1 Laranjo et al 's classification of the CA's design elements based on the type of the technology

with pre-determined scenarios. On the other hand, in the frame-based approach [62], to complete a task, the user must answer questions that fill slots. There exists no pre-determined but the system elicits prompts depending on the user's input and the information. When it comes to the agent-based approach [63], there is intricate communication between the agent and the user. There are numerous variants that should be built into the system based on specific agent behavior. Dialogue-based interaction is defined as the interaction between agents that allows one agent to reason about its own behavior and, on occasion, the conduct of the other agent. The dialogue evolves a sequence of relevant procedures. Dialogue initiatives can be the user or the system. Input modality can be spoken language or the written language or the combination of multiple modalities including written, spoken, and visual(non-verbal).

These Laranjo et al's classification has limitations in that they do not include the dialogue style of CA which has been known to affect the persona perception toward the chatbot [47, 64]. Also, dialog styles are also emphasized as the key design dimensions defined in Jovanovic et al's review study on healthcare CA [14]. There were successful attempts to design and implement a dialogue style to increase the user's engagement in healthcare tasks as follows. For example, an empathetic dialogue style is known for its effectiveness in managing daily mood changes and the subtle dialogue style of the chatbot is effective in promoting exercise for the elderly [16]. There also exist diverse attempts to apply likeable human-like dialogues to the chatbot by implementing dialogue styles of empathizing [48], expressing gratitude [49], making jokes [50], motivating [51], and praising [52]. Based on previous findings, Bickmore insisted that the dialogue design of chatbots for daily healthcare should be tailored for its usage and purposes [16, 53].

Also, previous studies imply that users are sensitive about whether the chatbot's dialogue style matches their individualized preference or expectations [17]. If a CA does not fulfill users' expectations or preferences, users fail to be emotionally attached to the chatbot, which results in low engagement or abandonment [18]. Because various users may prefer different conversational elements to be applied to CA, the lack of personalization in the conversations can be employed to improve both usability and user experience [34]. To emphasize, dialogue style is important in regards to user perception towards the agent and can be a critical factor for users' willingness to interact with the agent consistently [33, 65]. Therefore, CA's dialogue style should be personalized to its domain and user group to be engaging over time [66].

In this manner, our thesis suggests an idea of applying a dialogue of the person who is in a close relationship with the user to personalize the persona of healthcare CA. Through this, we expect an increase in overall user experience including user engagement. To focus on the dialogue style to design the persona of CA, we chose chatbot as the main focus of the study. Moreover, since our relevant studies (study1-3) are focusing on transferring conversational styles of a close persons to CA's dialogue style, the chatbot would be the best option among other types of CAs since user experience with the chatbot without other modalities are mostly affected by the dialogue styles.

2.4 Methods for Designing Chatbot's Dialogue Style

Mostly, Natural language processing (NLP) takes a great part in understanding the user input and generate chatbot responses. For the processing, rule-based models (ex.ELIZA [67], ALICE [68]), Information Retrieval (IR)-Based Models [69], Statistical Machine Translation Generative Models (SMT) [70], and deep learning models [71] can be applied in the process of understanding and generating natural language-based user input.

Recently, Text style transfer (TST) is an important concept in natural language generation, which controls some attributes in the generated text. There are example researches of TST that control the overall politeness, emotion, humor, and many others in the generated text. Compared to its long history, TST has recently gained significant attention due to increased performance brought by AI and deep learning models [72]. In the previous researches linguistic style is usually defined by its pragmatic aspects, including both personal (e.g., personality, gender, etc.) and interpersonal (e.g., humor, romance) aspects. Most existing literature also takes these well-defined categories of styles [73]. Existing papers include several style features for text style transfer such as formality [74], politeness [75], gender [76], humor and romance [77], biasedness, toxicity, authorship [78], simplicity [79], sentiment [80], topic, and political slant [81]. Also, there exists the study, mimicking specific person's text style such as Jhamtani et al's approaches to transfer text style from modern English to Shakespear's text style by applying an end-to-end neural model to enable copy action [82]. However, when it comes to the lay user, it is still technically challenging to train a text style transfer model due to the lack of individual's conversational data, and varied conversational style depending on who someone is talking to. Also, it requires the high cost to handle the personalized model.

Few previous approaches fully solved the challenges in high cost, low user resources, and effectiveness in personalization at the same time. To cope with previous limitations on the personalized dialogue style of conversational agents, we present the idea of applying the dialogue style of the user's healthcare provider to the conversational agent. We suggest healthcare provider-sourcing as a new approach to design a personalized dialogue style of conversational agents. By doing this, we expect a chatbot's dialogue styles to reflect complex user preferences and contexts.

In this study, instead of focusing on developing state-of-art technologies to increase the performance of the dialogue model, our goal is to focus on exploring the effectiveness of personalized dialogue style by mimicking the healthcare provider of the user. This approach requires delicate interface design. Therefore, methods used in this study include interface design methods for personalized dialogue-based interaction.

The effects of message personalization have long been demonstrated in the domain of health communication. Personalized messages could stimulate behavior change [83], increase adherence to healthcare behavior [84], and engagement with the given tasks [85]. Personalization is typically characterized in terms of digital technologies as a feature that alters information access, system's interface, and content to boost its personal relevance to users [35]. Using message elements such as message tone, message appeal, message format, or use of evidence persuasively affects receivers' involvement, receptivity, and retention in messaging when designing personalized messages for the user's healthcare [86, 87]. Before exploring our method of personalizing dialogues for healthcare agents, we introduce various approaches to designing a chatbot's dialogue-based interface.

2.4.1 Wizard of Oz Method

In the prototyping process of building conversational AI, Wizard of oz has long been used as a user study tool for exploring interface concepts or interface prototypes. Usually, the design process of dialogue style is facing major challenges coming from high cost and low available resources. To prevent cost coming from inadequate interface design, dialogue design usually includes iterative design with existing chatbot prototyping tools such as collecting conversation data through the Wizard of Oz method [88, 89]. The Wizard of Oz method has been frequently used for building prototypes of intelligent agents. A method is a rapid-prototyping tool for designing systems that is costly and requiring an implementation with new technology. In this method, a human (in our case author) plays the role of "Wizard" that operates the system and interacts with the user through a computer or usually mock-up system. Most Wizard of Oz works aim to test interface design. An example could be the speech-based intelligent systems, and also synchronous text-based agents. The method is appropriate for evaluating the function of the system but also effective in analyzing human behavior toward a particular system [90]. However, sometimes designers also can be overwhelmed due to time and costs for iterative user study and evaluation [91]. For example, the Wizard of Oz study for designing personalized reflection questions of the agent Robota [89] leveraged iterative user workshops (12 workshops in total) to generate the system's mini-dialogue flows. Therefore, in some cases such as super-personalized dialogue style matching, the human-sourced dialogue could be an effective alternative which we used for our main study 3.

2.4.2 Analyzing Dialogue Data with NLP

Other ways analyze current data sources, such as Twitter [92], existing chatbot logs [93], mail threads of DBpedia [94] and extracted data from apps [95] conversation data, to build CA dialogues. Usually, natural language processing (NLP) based analysis can be a method to analyze a large amount of conversational data. For example, sentiment analysis can be used to measure the overall emotional status delivered within the words or sentences [96]. Topic modeling with Latent Dirichlet Allocation (LDA) is a method for discovering topics and extracting semantic information from unsorted documents [97]. BERT, a deep-learning model based on a Transformer type of neural network, has become the foundation for a variety of NLP applications, including answering search queries and translating user-written words [98]. With the release of GPT-3, the newest edition of a set of language models produced by the company OpenAI, performance improved even further [98]. Analyzing dialogue data with the introduced method requires a large amount of conversational data to earn meaningful outcomes.

However, collecting a large set of individual conversational data is still challenging for the general population. More if one stores a large amount of his/her conversational data, it is also challenging to get the conversational data with the particular person (e.g. one or a family member). Since one's conversational style can differ based on the channel they are making conversation on, available conversational data for analyzing conversational style is limited in both its quantity and quality. This method can only be suitable in limited cases in our series of studies. Therefore, we excluded this method in this thesis.

2.4.3 Participatory Design

Researchers can engage with people to attain a design goal, which is termed participatory design, also known as co-design, to build a specific dialogue style adapted to a specific domain or environment. It is a collaborative learning process between consumers and designers that results in a design product that fits the needs of users [99]. Moreover, participatory design methods have been used for democratic values and an underrepresented group of people [99]. This design style fosters communication between people who use products, systems, interfaces, and environments and those who create them.

Mat-telmäki and Sleeswijk Visser [100] defined four elements for participatory design to build a conversational agent. They begin by defining the function of those who are influenced by the design (e.g., end-users and other stakeholders). Second, they suggest that in order to construct an effective chatbot, all participants must collaborate through workshop-style activities. Third, they emphasize the importance of employing specific methods and tools to enable people who aren't designers to express their thoughts and ideas and create visually tangible prototypes that contribute to the final design. Fourth, they discuss collaboration, in which participants share ideas and work together to develop solutions. To design chatbots with natural humanoid interactions, Pinhanez [101] proposes a participatory method called personality workshop, where designers, end-users, and stakeholders jointly establish the personality of the chat interface persona. Participatory design, according to Donetto et al, is driven by shared ownership, in which participants not only "have a say," but also "have the right to make decisions" during product development [102]. According to Sanders et al, while participatory design can be used throughout the design process, it is most beneficial in the early stages, such as the idea-generating phase [103]. Different sorts of tools (e.g., probes [104] and generative toolkits [103])) are used to enable users to reflect on their own experiences and produce design concepts during the design process of participatory design.

Steen presents a co-design process called collaborative inquiry and imagination, based on the philosophy of inquiry, in which designers, end-users, and stakeholders collaborate to describe challenges (explore and define), conceive solutions (ideate), then implement and assess solutions (prototype and evaluate) [105]. "Say, do, and make," Sanders advises as a strategy. She claims that by listening to what individuals say in their spoken words, she can extract common knowledge [103]. Observing people's behaviors, on the other hand, can result in the development of "observable knowledge" or "observed experience." Analyzing how people express their thoughts, feelings, and goals is critical for developing tacit information that cannot be expressed in words. Regardless of the paradigm behind those techniques, they should be implemented and applied in a flexible manner based on the design context.

2.4.4 Crowdsourcing

Crowdsourcing has been used in a variety of design fields, including gathering design samples, prototyping in real-time, and receiving design critique or feedback [91]. In addition, crowdsourcing was used to create a chatbot. There has been work to collect and provide conversational data for the social chat system via crowdsourcing. For example, For context maintenance, Fantom employs a graph-based dialog model to provide appropriate responses. As actual chat exchanges and system replies are gathered by the audience, the model continually evolves [106]. Another example of an agent that crowdsources ”trigger-action” rules to automate task management is Instructable-Crowd. Other types of work used crowdsourcing to reply to end-users in real-time while keeping contexts intact [107]. To keep the conversation going, Chorus enlisted the help of a large number of volunteers who proposed responses, voted for the best answer, and shared chat history. Chorus demonstrated that the general public could generate a wide range of comments and descriptions on a given topic, and they predicted that crowdsourcing would be a useful tool for investigating a variety of chat conversations [108] . There are also hybrid systems that operate with experts as crowds so that if a user inputs an unidentified inquiry, the system gathers answers from the crowd and responds [109]. In addition, crowdsourcing was employed to evaluate the chatbot. ChatEval [110], for example, used the audience to conduct automatic evaluations of chatbots using dialogue breakdown detecting tasks. Choi et al investigated crowd worker behaviors when evaluating a dialogue design, as well as designers' objectives and expectations when integrating the crowd in the design process [111].

Protochat builds on previous work by providing an automated system with high fidelity that allows designers to evaluate the system with crowd workers. ProtoChat helps designers iterate on conversation designs quickly by allowing them to develop discussion, quick evaluation of the intended conversation with the audience, and analyze the conversation data tested with the crowd, and change the conversation design. Two main interfaces include the designer interface and the crowd-testing interface that display these features [91].

However, we did not use crowdsourcing as our dialogue-based interface design tool, since it is not appropriate for designing a personalized dialogue style. Rather than the crowd, we recruited the people who are in a close relationship with the user to design personalized CA for daily healthcare.

Chapter 3

Goal of the Study

From the literature review, I have reviewed the roles of CA in healthcare and the opportunities of CA as personalization in individual healthcare. To synthesize, CA's major opportunities in daily healthcare include personalized care depending on users' lifestyle choices, socioeconomic context, and living environment by supporting different tasks including counseling, monitoring, or increasing adherence to the medical plans. However, there lacks study on the personalization of healthcare CA to meet user's preferences toward the CA as a social actor to increase user engagement with the CA's healthcare tasks. If CAs does not meet users' preference, users fail to emotionally attach to the CA, which result in low engagement or abandonment. This challenge is the major problem I posit in the thesis. As a solution, this thesis provides the idea of designing a personalized persona of CA to meet individual user's preferences toward CA as a social actor. There have been attempts to give healthcare CA persona by implementing particular roles but those attempts still do not include personalization of personas.

Therefore, the major goal of this thesis is to investigate the idea of applying the persona of the user's healthcare provider to the healthcare CA as means of personalization strategy to increase the overall user experience with the healthcare CA. Among the many types of CAs, the thesis series of works on text-based CA (i.e. chatbot) focus on designing the persona of CA based on the personalized dialogue style since dialogue style of CA has been known to affect the persona perception toward the CA [47, 64]. In the thesis, I used the word "healthcare provider" to imply a healthcare provider who can give the user tangible support in the context of daily healthcare. The reasons of implementing the healthcare provider's persona into the healthcare chatbot are as following: (1) to increase user experience through reflecting real-world user-healthcare provider relationship into the user-CA relationship and (2) to personalize the chatbot's persona by implementing user's acquaintance who is in the close relationship.

The thesis includes three types of study : (1) exploration on imitating whose persona will be effective in the context of daily healthcare (study 1), (2) What features must be considered to implement dialogue styles of the close person to the CA (study 2), and (3) What are opportunities challenges of applying persona of a close person to healthcare CA depending on the healthcare domains (study 3). MimicTalk, which is the prototype was also deployed into several healthcare domains including diet, physical activity, and stress management to evaluate its effectiveness.

To be clear, all experiments included in this paper were conducted in Korean and translated into American English. The thesis includes multiple HCI approaches to explore and investigate the novel idea of persona design of healthcare CA to increase user engagement. Particularly, we combined the methods of qualitative and quantitative analysis to deeply explore the user experience. For the qualitative analysis, we used interviews, Wizard of Oz, think-aloud, and thematic analysis to find user insight and collect user's unrefined data. Then, with quantitative analysis, we tried to analyze patterns of user behavior and evaluate the conversational interface and the design elements implemented in the conversational interface. For the quantitative analysis, we analyzed the survey based on 7 point Likert scale, user's data log to statistically extract meaningful results.

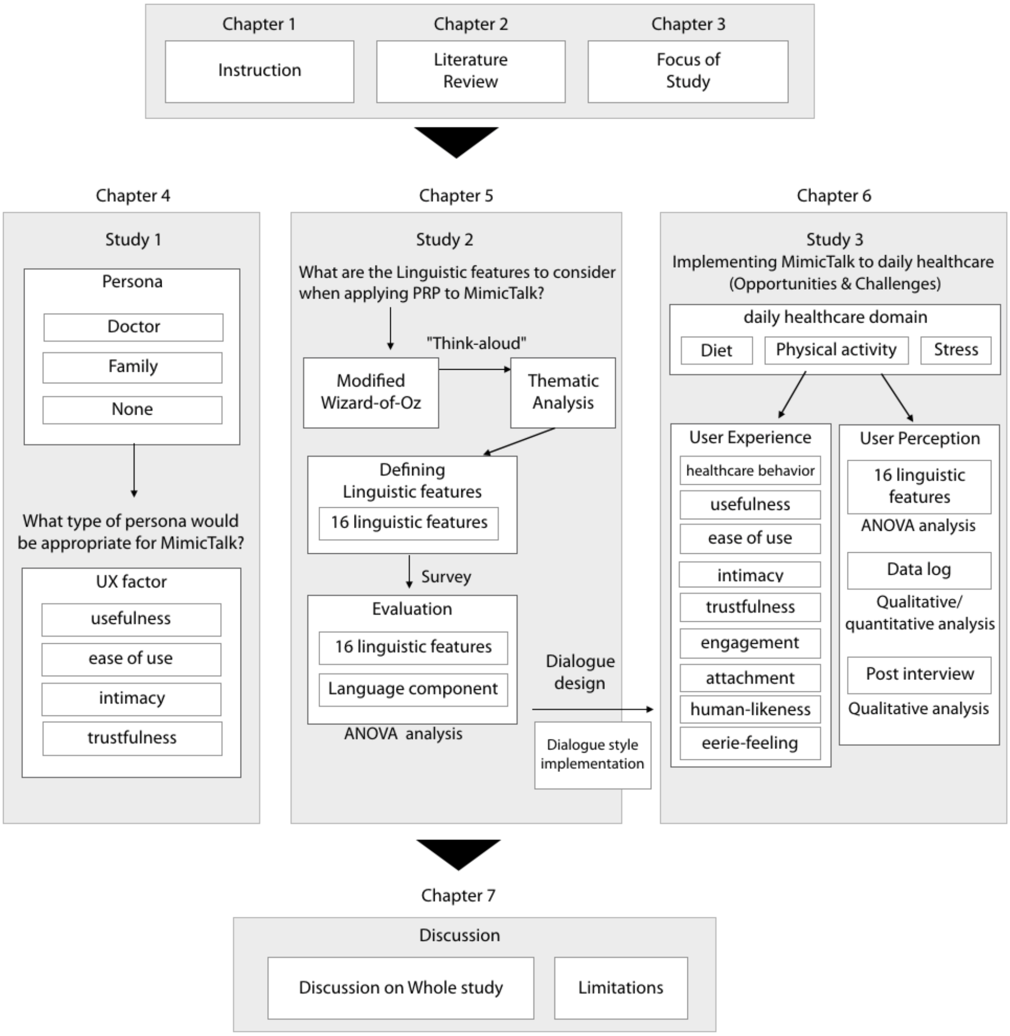


Figure 3.1 Overall structure and contents of the thesis including three main studies

Chapter 4

Study 1. Exploring Candidate Persona for CA

In this chapter, we aimed to define what kind of persona would be appropriate for the CA aiding user's daily healthcare. Since family members and healthcare experts including doctors, nutritionists, and nurses are major healthcare providers for the lay people, we explored the effectiveness of applying their personas to CA. Based on our literature review and survey results on an appropriate persona for daily healthcare, we chose candidate personas and applied them to prototype chatbot with its major function is giving intervention based on the user's daily health status. In the end, we are going to discuss our findings and share design implications.

4.1 Related Work

In this section, we introduce related work on (1) need for support in daily healthcare and (2) applying persona to text-based CA

4.1.1 Need for Support in Daily Healthcare

Managing one's habitual routine is challenging, requiring behavior interventions from external supporters. Generally, healthcare providers for daily healthcare are the expert and the family [112] with each provider's characteristics of support being different [113, 114]. For example, the main focus of expert support is to provide reliable health-related information [113], while family support is effective for making actual changes in daily life (such as diet, physical activity, medication, smoking cessation, and alcohol consumption) by leveraging its long term intimate relationship with patients [114]. Despite its effectiveness, human support is hard to persist due to the limitation of time and physical resources [113, 115, 116]. For example, the disconnection of a modern family due to the rise of one-person households and single elderly has made it hard for family support to even satisfy the basic condition - having a family [117]. Expert support also faces inevitable obstacles due to time limitations and cost issues. Other than these hurdles, there is still a potential risk of healthcare providers suffering burnout from prolonged caring [118, 119].

To tackle limitations of human support, CAs have been attracting healthcare stakeholders [3]. CAs partly take over the role of healthcare provider by automating tasks with higher accuracy (e.g. exercise reminder) [3, 4]. Due to its accuracy and effect on reducing human burden, CA is recognized as a strong alternative for human support with a lower risk of sudden termination. However, low acceptance of CA has been pointed out as a limitation of CA, since the technology becomes unsuccessful when it is not accepted by users [120]. To improve the acceptance of CA for daily healthcare, previous studies have proposed applying a persona to a CA system [121, 122]. Applying a persona to CA has been showing an opportunity to form rapport between it and users [123, 124].

4.1.2 Applying Persona to Text-based CA

What human-like factors affect the perception of an agent's persona has long been a study topic in the HCI community. These researches concentrated on broad human-like characteristics rather than those of a person in the real-world relationship with the user. Endowing a chatbot with personality is vital for this newly emerging objective of making the chatbot recognized as a human being. We must first comprehend the concept of anthropomorphism in order to comprehend how users perceive CA and why persona is so crucial in managing this process [125].

Anthropomorphism refers to a person's predisposition to regard anything non-human as a human. Anthropomorphism is the ability of humans to attribute human traits, motivations, beliefs, and sentiments to non-human objects. Anthropomorphism frequently occurs in human-computer interaction, according to previous studies, and designing a particular persona could control how users anthropomorphize machines [126].

The degree to which designers want users to anthropomorphize the system can be determined using the word "humanness." Anthropomorphism is usually utilized to develop an emotional connection between the human and the CA, but it may also be used to assess how much designers want users to anthropomorphize the agent. Previous research has revealed that a CA's level of humanness has an impact on how humans anthropomorphize it, as well as being a key aspect in trustworthiness management [127]. Humanness differs from anthropomorphism in that anthropomorphism refers to the psychological attribution of human-like features to something non-human, while humanness refers to something that appears or acts like human. While anthropomorphism is expected to result in positive consequences, different degrees of humanness can have both positive and negative effects on how humans perceive and define the agent. The uncanny valley is the result of humanity's bad repercussions[128].

However, because users' perceptions of CAs fluctuate depending on user attributes, the impact of CA on user experience varies. Users are also concerned about if the human-like CA's personality reflects their personal preferences or expectations [17]. Users become emotionally attached to CAs if they do not achieve their expectations, which leads to poor engagement or abandonment [18]. As a result, persona personalization should be considered to improve the user experience with the healthcare CA. As a result, we propose in this paper that healthcare CA be modeled after a healthcare provider's identity. Throughout the thesis, we will be focused on text-based agents because the written word transmits a lot of information and context about the host, including personality [129, 130, 131]. In this manner, dialogue style influences how users view the agent and can be a deciding factor in whether or not they want to contact the agent on a regular basis [33, 65].

4.2 Research Questions

In this study, we aim to investigate and evaluate the effectiveness of applying personas of the user's healthcare providers who are in a close relationship with the user (i.e. doctor, family member) on the acceptance of the healthcare CA. Exploration of imitating whose support will be most effective in everyday healthcare was conducted with the Wizard of Oz method. According to the survey we conducted with the 283 middle-aged users (Male: 123, Female:160) prior to study 1, the preferences toward the family member (n=101) and the doctor (n=130) was the highest followed by other healthcare provider sources (n=52). Based on findings, we decided to apply the persona of the doctor and the user's family member to the prototype of healthcare CA. Then, we compared the opportunities and challenges among prototypes. In the study, the following research questions were explored and answered.

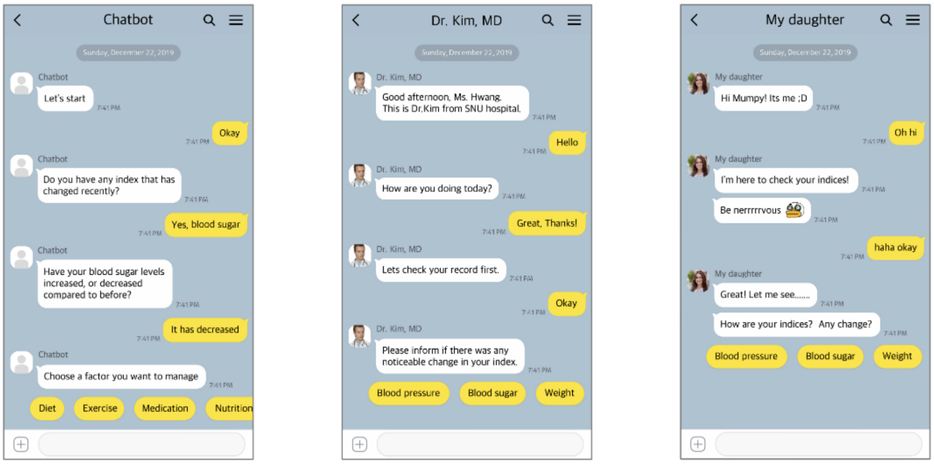
RQ1: What are the opportunities and challenges of applying the persona of the healthcare providers to the CA ?
RQ2: What kind of persona should be applied to healthcare CA for daily healthcare?

4.3 Method

We conducted a within-subject experimental study to compare opportunities and challenges among the type of personas. Two of them are persona-applied (doctor and family) chatbots and one without persona (control). To ensure that the persona of a user's family member and doctor were well implemented to the chatbot, we also conducted a manipulation with all participants to check if the personas were successfully manipulated. We used a survey and post-hoc interview to analyze the experiment results. In the end, we discuss the effects of applying the persona of the close person into the chatbot and discuss design implications for CA with persona. We recruited eleven participants (Female: 6, Male: 5) aging from 19 to 60 years old (M = 46 SD = 13.82), since these people were highly in need of continuous interventions [132]. For recruiting the participants, we used a public health bulletin board. They received 15000KRW after participating one-hour user study. Participants all showed a strong relationship with the healthcare providers who are also participated in the study to design the persona of the chatbots. We filtered out the weak relationship because we are focusing on the strong tie relationship in the study.

4.3.1 Wizard of Oz Study

During the user study, we used the Wizard of Oz method to compare and evaluate the chatbot with the persona of a family member, doctor, and the one without a persona. We used the Wizard of Oz method to investigate the effectiveness of the pre-designed personas. To design the personas of chatbots, we recruited each participants' family member and the doctor prior to the experiment. We asked them to change the given sentences, which would be implemented to a chatbot, into their own conversational style as if they are really sending messages to the participants through KakaoTalk. Through this process, we applied the collected words and sentences into the dialogue of the chatbot to apply the persona of the user's healthcare providers that are (1) family member and (2) doctor to the chatbot. To apply each persona to chatbots, we also differentiated appearance (profile image, name) and conversational style shown in Figure 4.1, given that these features are known to play a major role in persona perception [123, 124]. To customize the appearance for both chatbots, profile pictures with a recognizable face were used as the agent's profile. This profile picture



(a) Control (b) Doctor (c) Family

Figure 4.1 Example screenshots of chatbots applied with personas of user's healthcare providers (family, doctor) and a basic chatbot

was pre-collected from the user's healthcare provider prior to the user study. For the conversational style, participants' family members and the doctor were asked to transform the sentences which will be implemented to the chatbot. All chatbots were built on KakaoTalk as shown in Figure 4.1.

At the beginning of the experiment, we asked participants about their demographic information, relationship with the healthcare providers. Relationship-related questions were based on Granovetter et al's work who defined strength of tie as the combination of the amount of the time of relationship that has been persisting, the emotional intensity, the intimacy, and the reciprocal services. We also asked how much they self-care themselves and what they usually do for their daily self-care. Then, all participants interacted with the simple chatbot built for practice so that they can get used to how chatbots work and feel comfortable with the experiment environment. After, three chatbots, two of them with the persona of the user's healthcare providers (doctor and family) and one with formal and neutral persona, were assigned to participants in a randomized order. Participants were asked to complete five behavior intervention tasks per chatbot. Five behavior intervention tasks consist of lifestyle factors including (1) diet, (2) exercise, (3) medication, (4) alcohol consumption, and (5) smoking. Tasks were personalized based on the user's current status on the metabolic syndrome-related symptoms including blood sugar, blood pressure, lipid level, and weight loss/gain. For example, a chatbot proactively sends a greeting message to the user and asks the user to start an interaction. If the user agree with the interaction initiation, it asks for changes in the user's metabolic syndrome-related indicators. Based on the answers to the changes in indicators, it provides the user with 5 major lifestyle tasks including diet, exercise, medication, alcohol consumption, and smoking.

Participants were instructed to interact with the chatbots as if they are in a situation described in the action sheets. Action sheets include scenarios of changes in the health status of Metabolic syndrome. Participants were asked to interact with chatbots as if they are in the situation stated in the action sheets. After completing all fifteen tasks

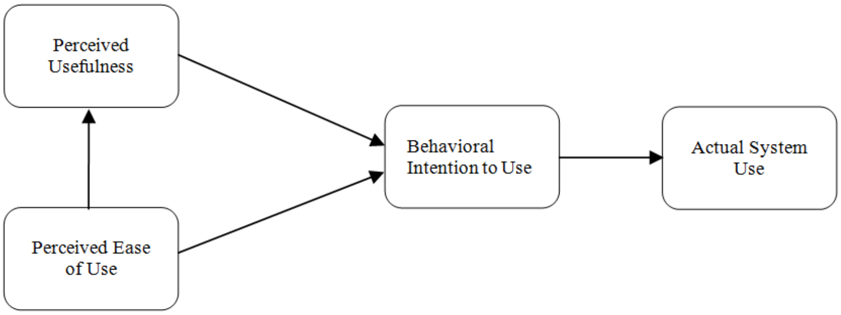


Figure 4.2 The overall constructs of technology acceptance model

(i.e. five for each type of chatbot), participants evaluated all chatbots with survey and semi-structured post-interview.

4.3.2 Survey Measurement

During the experiment, participants answered survey questions on system acceptance that consist of variables from the Technology Acceptance Model (TAM). TAM is a theoretical model to help explain and predict user behavior of information technology [133]. When users are presented with new technology, the model argues that a number of constructs influence their decision on how and when to use it. Perceived usefulness and perceived ease of use are lower constructs in TAM model that we focused on. Fred Davis' work established the definitions of perceived usefulness and perceived ease of use. In this work, perceived usefulness is defined as to what extent a person assumes that using a given system will improve his/ her performance. Perceived ease-of-use is defined as to what extent a person expects that using a system will be easy. Based on this perspective, if technology is simple to use, it is more likely to overcome barriers. It's difficult for users to have a good attitude toward anything if it's difficult to use and the interface appears confusing.

However, due to the diversity of technology, TAM model is occasionally insufficient to fully predict the acceptance level. For this reason, including appropriate variables in an original model is expected to provide a stronger outcome [134]. Thus, we included additional questionnaires measuring the perceived trustfulness and the perceived intimacy that we wanted to evaluate. These questions were selected because expert support has been known for its effectiveness in providing reliable health-related information [113] to the patient, while the family support is effective for making actual changes in daily life (such as diet, physical activity, medication, smoking cessation, and alcohol consumption) by leveraging its long term intimate relationship with patients [114].

In summary, variables measured in the survey were perceived usefulness, perceived ease of the user, behavior intention to use, perceived trustfulness, perceived intimacy, and acceptance (I asked willingness for actual use to evaluate acceptance in this study). Participants answered the questions evaluating these variables for each chatbot. Answers to all questions were collected with a 7-point Likert scale.

4.3.3 Post Interview

We also conducted a post-interview to explore user experience with three types of chatbots. We asked them how they felt the overall interaction with the chatbots that are applied with personas of healthcare providers in the real world including their perception, emotion, and their overall preferences. Also, even though we pre-collected the healthcare providers' made-up sentences to design the chatbots' dialogues, we asked participants what linguistic factors had more impact on the experience with the chatbot that resembles their actual healthcare provider. This was done to explore whether particular linguistic factors fortify users' persona perception. Also, we asked how real-world relationships affected the participants' perception toward chatbots with the persona of their healthcare providers. The participant was audio-recorded during the post-interview to allow for later transcription and analysis. Transcribed data collected was analyzed by authors through thematic analysis [135].

4.3.4 Analysis

We used the data to run a one-way repeated ANOVA to see if applying personas of users' healthcare providers affected their acceptance of the healthcare chatbot. Repeated one-way measures ANOVA (also known as a within-subjects ANOVA) is used to see if the means of three or more groups differ when the participants in each group are the same.

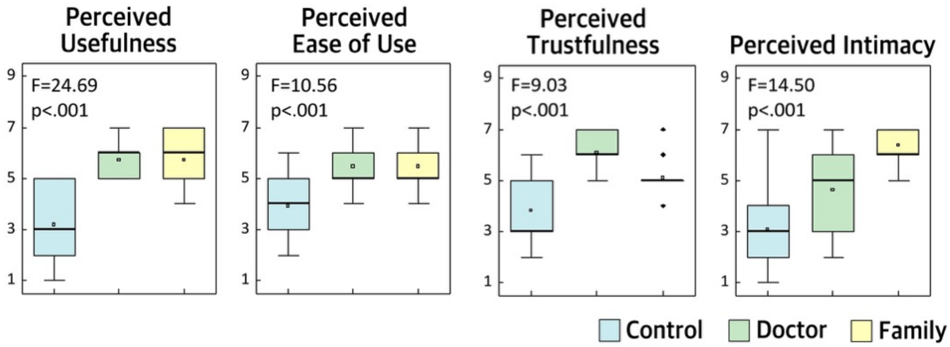


Figure 4.3 Chatbot prototypes applied without persona, with personas of the user's doctor and family member

The linear regression analysis was used to compare the predictive power of all variables on chatbot acceptance. Statistical techniques to evaluate the associations between a dependent variable and independent variables (commonly referred to as 'predictors') are known as regression analysis. Linear regression is known as the most frequently used type of regression analysis, with the line (or linear combination) that mostly fits the data according to mathematical criteria. The least-squares approach finds a line that minimizes the sum of squared discrepancies between the raw data and the line. In order to present the results of the post-interviews, we used thematic analysis'[135].

4.4 Results

Results of the study are presented in order of survey and post-interview.

4.4.1 System Acceptance

Figure 4.3 illustrates significant differences between three chatbots in all variables used in the evaluation of chatbot acceptance. Healthcare chatbots with personas of healthcare providers showed significantly higher scores in all variables compared to a control condition.

One-way repeated ANOVA analysis revealed significant difference between 3 personas (Figure 4.1.). Chatbot with the persona of the doctor showed significantly high perceived trustfulness than any other conditions($F(2{,}30) = 14.187$, $P < 0.001$) while the chatbot with the persona of family member showed significantly high perceived intimacy($F(2{,}30) = 19.421$, $P < 0.001$). However, when it comes to perceived usefulness and perceived ease of use, there was significant difference between the control chatbot and the chatbots with the personas, but there was no significant difference between the persona of the family member and the doctor. For variables from TAM model chatbot with persona showed significantly high ratings than basic chatbot without persona. Rated variables are perceived usefulness ($F(2{,}30) = 24.69$, $P < 0.001$)) and perceived ease of use ($F(2{,}30 = 10.56$, $P < 0.001$), attitude ($F(2{,}30 = 9.03$, $P < 0.001$), and behavioral intention to use ($F(2{,}30 = 14.50$,$P < 0.001$).

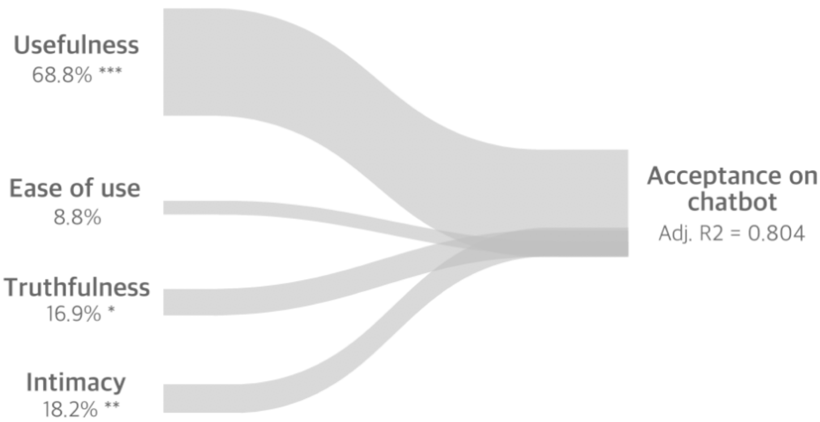


Figure 4.4 Predictive power of all variables related to acceptance of healthcare chatbots

4.4.2 Perceived Trustfulness and Perceived Intimacy

To verify the explanatory power of newly added variables, which are perceived trustfulness and perceived intimacy, we conducted regression analysis. The results demonstrated that perceived trustfulness ($Adj\ R^2 = 0.24$, $P < 0.01$) and perceived intimacy ($Adj\ R^2 = 0.54$, $P < 0.001$) significantly predicts acceptance level of healthcare chatbot, implying that both variables are valid variables for predicting acceptance level of the chatbot with the persona of healthcare providers.

4.4.3 Predictive Power of Corresponding Variables

Through the linear regression analysis, the predictive power of all variables related to users' cognitive beliefs on the acceptance of the chatbot was compared. The corresponding variables were the perceived usefulness and perceived ease of use (two from the TAM model), perceived trustfulness, and perceived intimacy (two added). Perceived usefulness showed the highest power (68.8%), and then perceived intimacy(18.2%), perceived trustfulness(16.9%), perceived ease of use in order(8.8%). The regression model for these four variables showed r squared value of 0.804 (Figure 4.4). Based on the results, we could infer that the perceived usefulness of the chatbot mostly affects acceptance on the chatbot followed by perceived intimacy, perceived trustfulness, and perceived ease of use.

4.4.4 Linguistic Factors Affecting User Perception

From the post-hoc interview, we further demonstrated that influential variables for system acceptance differ depending on the persona type. For chatbot with the persona of a user's family member, participants implied that perceived intimacy has played a major role in system acceptance. P5 laughed and expressed surprise of talking with the family-like agent: *"I feel like I am really talking to my daughter."* and P8 emphasized the feeling of intimacy from the family-like agent: *"I feel like I was talking with him. I think this made the agent friendlier.*

Using it, I even felt like I'm getting emotionally closer to my son." For a chatbot with the persona of the doctor, the interview showed the importance of perceived trustfulness for the system acceptance. For example, P10 mentioned that *"chatbot with the actual doctor's personality made it trustworthy. This may influence my decision whether I will use it or not"*

Plus, we observed that the specific linguistic factors played an important role in participants' persona perception, making the agent more like the real family member and the doctor. General patterns of such differences are endearment (e.g. *mumpy, sweetie, angel*), hedging (e.g. *em, oh, ah*), frequent typos, word choice, and emojis. For example, P7 pointed out that *"Using exclamation mark at the end of the sentence is what my sister is always doing! I feel like I am really having a conversation with her."* P4 emphasized that hedging made the chatbot more like her son by saying that *"My son always uses a word like 'OMG' before he starts to say something. When I saw this word in the chatbot, I thought it really looked similar to a conversation with my son."* In addition, five participants (P1, P2, P5, P7, P10) reported that the appearance of the chatbot consisting of the profile picture and the name played a major role in making the chatbot with the persona of the user's family member and the doctor more realistic.

4.5 Implications

Through the study, we demonstrated that applying the persona of a user's healthcare providers to the healthcare chatbot has a positive effect on the acceptance of the healthcare chatbot compared to a healthcare chatbot without a persona. Based on our results, we discuss design implications for applying the persona of user's healthcare providers on the healthcare chatbot. Through the experiment, we found that assigning a persona to a healthcare chatbot ends in higher acceptance in different ways than a healthcare chatbot without a persona. This effect can be attributed in large part to the characteristics of the original human persona. Our qualitative and quantitative data suggest that the perceived intimacy which is the successful predictor for family support is also the candidate predictor for acceptance of a family-like chatbot. In the case of the chatbot with the persona of the expert, perceived trustfulness was highly valued than other persons, which is a predictor of human expert support and acceptance of the chatbot with the persona of the expert.

Some variables have bigger predictive power than variables of TAM model when it comes to healthcare chatbot with persona. In our study, we compared four types of cognitive beliefs that are the candidate predictor of acceptance of the healthcare chatbot. Perceived usefulness and perceived ease of use were candidate predictors from TAM model, while perceived trustfulness and perceived intimacy were chosen based on previous human support [136]. Perceived trustfulness and perceived intimacy showed bigger predictive power in our regression model compared to perceived ease of use. Two major implications I learned from the study results are the following.

One implication is that critical factors for acceptance could differ according to the persona type. This tendency seems to align with the actual relationship between the user and the user's healthcare provider in the real world. Trustfulness had more influence on the chatbot with the persona of the doctor than the persona of the family member. On the other hand, intimacy had more impact on the acceptance of the chatbot with the persona of the family member than the doctor. From this, we could conclude that defining critical factors derived from a relationship with the actual healthcare provider, whose persona is implemented to the system, is important. Designers should utilize these factors appropriately when applying the persona of specific healthcare providers to healthcare chatbots. For instance, the persona of family members would be needed if a certain healthcare chatbot requires high intimacy while an expert persona would be suitable for a healthcare chatbot that requires a high trustfulness.

Another implication is that careful selection of a conversational style may contribute to designing a more acceptable healthcare chatbot. As indicated in our results, variation in the conversational style of persona made participants perceive the agents more like their actual healthcare providers. Since the ultimate goal of our agents with persona was to be accepted as an actual humans, we tried to carefully observe the effect of this facet as well. The results align with the HCI community's findings on the communication strategy of conversational agents that agents may use elements such as hedging, back-channeling, rephrasing, and emoticons to be human-like [137]. Also, a repetition of the message was found to be a critical point in making agents feel like a machine. P6, P9, and P10 mentioned that the repetition of the same contents will reduce the feeling of human likeness.

Chapter 5

Study 2. Linguistic Characteristics to Consider When Applying Close Person's Persona to a Text-based Agent

From the previous work, we earned the lessons that certain linguistic characteristics affect how users perceive the persona of the conversational AI mimicking the user's acquaintances in the real world. In this chapter, we aimed to figure out the lists of linguistic characteristics that take account when applying a real person's persona who is in the relationship with the user to the conversational AI, particularly to the text-based agent. In other words, we defined linguistic characteristics that influence user perception with a persona of a close persona (PRP). We modified the Wizard of Oz method to explore major linguistic characteristics determining the persona with PRP for use in the experiment. A separate survey was also conducted to evaluate specific features.

5.1 Related Work

We share related work on (1) linguistic characteristics that affect user's persona perception with the CA and (2) previous study on language components.

5.1.1 Linguistic Characteristics and Persona Perception

With emerging technologies making PRP more feasible and applicable, there is a greater need to investigate the features influencing user perceptions of CA's humanness and anthropomorphism based on PRP personas. The need to investigate linguistic characteristics in order to design the sophisticated dialogue flow of a text-based agent is particularly pressing. To eliminate the compound effects of other modalities such as voice, motion, and appearance, we focused on linguistic characteristics applied to the text-based agent (i.e.chatbot). The textual interface's ability to display movements and gestures is limited. To investigate how users perceive text-based agents as convincing social actors via written interactions, it is necessary to investigate linguistic elements embedded in textual interfaces. This is because a written text and dialogues convey a clue about the writer and sender. This includes the writer's personality and identity, on their own [129, 130, 131].

In the previous study, Mairesse et al. identified linguistic elements correlated with a writer's personality. The research looked at common words, textual features, punctuation, emoticons, average response time, and imitation rate [129]. They did the study by collecting individual corpus and personality ratings for each participant's conversation. Then they defined relevant features from the conversational data and build statistical models based on personality ratings.

Mairesse et al. defined the linguistic factors that influence the perception of "introvert" and "extrovert" personality in great detail. When it comes to conversational behavior, introverts prefer to listen, whereas extroverts prefer to initiate the conversation. Introverts rarely engage in back-channel behavior, whereas extroverts engage in it more frequently. When it comes to topic selection, introverts are known for their self-centered, problem-centered speech, dissatisfaction with one's current situation, strict selection, a single and monotonous topic, few semantic errors, and few self-references. Extraverts, on the other hand, are less self-centered and enjoy talking, agreeing, and complimenting others. They also tend to think out their thoughts on a variety of topics, making frequent semantic errors and self-references. Introvert users prefer a formal tone with many hedges, whereas extrovert users prefer an informal tone with few hedges. The syntax that is frequently used varies depending on the personality type. Introverts seem to use a lot of nouns, adjectives, and prepositions, and they seem to use a lot of words per sentence, as well as a lot of negations. With simple constructions and few words per sentence, extraverts use a lot of verbs, adverbs, and pronouns. They don't appear to use many negations. Introverts also use more correct, rich, diverse, and exclusive words than extroverts. However, they use fewer social and positive emotion words. Extraverts use more loose, poor, low-diversity dialogues with a lot of social and emotional words and fewer exclusive words and negative emotion words than introverts. Some of the characteristics found in Mairesse et al's study are also found in ours. This approach, on the other hand, is not a user-centered approach to text-based features that influence user perception of PRP agent personas. As a result, we investigated and redefined linguistic characteristics based on two major focuses: (1) the scope of PRP and (2) user-centered perspectives.

5.1.2 Language Component

In this study, we also considered how language components affected user perception. The components of language and their terminology align with demarcations for many of the elements of text-based communication. Language components and corresponding elements that belong to each component affect the overall delivery and understanding of the text-based outputs. Also, it delivers certain characteristics of the sender including emotion, personality, and context. Therefore, we aimed to explore those components and the corresponding elements to design the persona that consists of text-based language components.

We referred to the previous language model including three major intersecting components (i.e. content, form, and use). Content refers to language components that can be defined as semantics, the form includes major components related to morphology and syntax, and use is the component related to message context and pragmatics [138]. The syntax is the language characteristic that contains the structure of a sentence. Sentence organization including the order of clauses, network, and relationships between words and sentences, structural elements of a sentence is the major component of syntax. Syntax also includes which word combinations are acceptable or not. For example, if someone says "he went to town." the sentence including the series of words is acceptable. However, when someone says the sentence "town to went he", the sentence is not understandable and acceptable because the sentence did not follow the rules of the syntax of English. As such, there exist common rules of syntax in various languages. In English, a sentence must include a noun phrase and also a verb phrase. An example of "he went to town" contains both phrases that make sense. On the other hand, morphology more focuses on the arrangement of words. The smallest grammatical units consist of morphemes. Examples of morphemes could be any letter of the alphabet. Types of morphemes could be free or bound. When it comes to free morphemes, they can stand and be used by only themselves. Examples are words such as boy, sad, and small. These morphemes make sense when used alone. However, bound morphemes should be used with a larger word to be acceptable and make sense. Examples are prefixes and suffixes including un-, non-, -s, -ly. Lastly, semantics refers to the branch of logic and linguistics containing the meaning of words or sentences. There are various branches and subbranches derived from semantics. These include formal semantics, the logical aspects of meaning, references, such as sense, implication, and form of logic, lexical semantics, conceptual semantics, and the cognitive structure of meaning.

5.2 Research Questions

Study 2 is focusing on demonstrating linguistic factors that affect persona perception of PRP and evaluating the priority among defined factors, the study includes (1) modified Wizard-of-Oz study for defining linguistic characteristics affecting persona perception of the person in the close relationship with the user and (2) evaluating these features.

Through this study, we aim to define linguistic characteristics to consider when applying PRP to text-based CA. Moreover, we defined how a single linguistic element affects the perceived humanness of the text-based CA. We targeted text-based agents, particularly chatbots to focus on the linguistic element and exclude other compound effects coming from other modalities such as appearance, sound, and movements. At the end of the study, we present design considerations for applying PRP to the CA.

RQ1: What are linguistic characteristics to consider when applying a close person's dialogue style to the text-based CA?
RQ2: To what extent each linguistic feature affects user perception towards text-based CA?

5.3 Method

To explore linguistic factors and evaluate its influence on persona perception, we conducted modified Wizard-of-Oz study and survey. Two corresponding studies were conducted including empirical study exploring the possible linguistic elements and the survey evaluating how linguistic elements affect perceived humanness and user preference. Wizard of Oz method was used as a method for empirical study. Participants for the empirical study and the survey were seperately recruited and did not overlapped.

5.3.1 Modified Wizard of Oz Study

We recruited four teams, each with two participants, to conduct an empirical study on linguistic characteristics that influence user persona perception. More than three times a day, team members communicated with one another via a chatting platform. As shown in Table 5.1, there were four teams total in eight participants.

The study was conducted with the modified Wizard-of-Oz [139]. The traditional Wizard of Oz method has been frequently used for

Team	ID	Age	Relation	Chat Frequency
1	P1	53	Mom	>5
	P2	26	Daughter	>5
2	P3	33	Sister	>3
	P4	29	Sister	>3
3	P5	31	Friend	>3
	P6	31	Friend	>3
4	P7	36	Husband	>10
	P8	31	Wife	>10

Table 5.1 Information or participants included in Modified Wizard of Oz study

building prototypes of an intelligent agent. The method is a rapid-prototyping method for systems costly to build or requiring development with new technology. In this method, a researcher plays the role of Wizard that manipulates the system and interacts with the end-user through a real computer or mock-up system. Most Wizard of Oz tests or experiments establish the viability of a futuristic approach to interface design. An example could be a speech-based intelligent system, and also synchronous text-based agents. The method is appropriate for evaluating the function of the system but also effective in analyzing human behavior toward a particular system [90].

In this study, we introduced a user-centered approach by letting the user play the role of a wizard doing the system activity, in addition

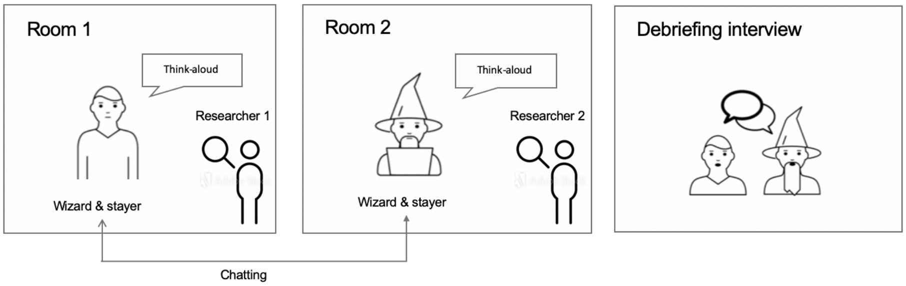


Figure 5.1 The concept of modified Wizard of Oz used in study 2

to the classic Wizard of Oz strategy of convincing the end-user that the system is working automatically. As a result, each member of the squad doubles as a wizard and a stayer. We did this by setting up two rooms for the experiment. During the experiment, we requested each team member to remain in a different room.

During the session, they were instructed that they were going to have a conversation with the chatbot with the persona of another participant even though they were talking to the other in a different room. Each participant was located in a separate room and asked to use the chat interface [140]. The think-aloud method was used in the experiment to get participants to speak out the linguistic characteristics that affect the persona perception of the other person in the team. A think-aloud method is a user research protocol used to gather data during the usability testing of system prototype or system evaluation [141]. During think-aloud, participants are thinking aloud as they are performing given tasks. Participants were asked to speak out whatever thoughts come up in their minds as they are performing the task. These include what participants are seeing, doing, thinking, or feeling. This process gave us insight into the user's cognitive processes toward the chatbot with PRP. In our study, the whole think-aloud session was voice-recorded and the recorded data were transcribed by authors. We used tlk.io to conduct modified Wizard-of-Oz.

With data from think-aloud during the modified Wizard-of-oz experiment, we conducted inductive thematic analysis [135] to identify themes based on participants' responses. For this analysis, two researchers independently coded each phrase or sentence. To validate the results we calculated inter-coder agreement among researchers which resulted in high intractability (Cohen's Kappa coefficient (κ) = 0.79.) Finally, we used Bloom et al's study to divide the themes into the categories of language components (content, form, and use). [138]. This process was conducted by experts in the linguistic domain (Cohen's Kappa coefficient (κ) = 0.90). The results found from the empirical study are in Figure 5.3.

5.3.2 Survey

We also conducted a survey with 82 participants based on the findings of the empirical study to see how a single feature affects user perception of persona in a text-based interaction. For the survey's generalizability, we recruited people through an online community that included people of all ages. The goal of the survey was to inform persona designers about the importance of each linguistic feature. On a 7-point Likert scale, participants were asked to rate each feature in the survey. 82 people took part in the survey.

5.4 Results

In this section results from the modified Wizard of Oz study and the survey will be presented. we used thematic analysis [135] to analyze the results from the modified Wizard of Oz study and used a one-way ANOVA test to statistically compare the defined linguistic characteristics and the three types of language components.

5.4.1 Linguistic Characteristics

Through the thematic analysis [135], we have defined 16 linguistic characteristics that affect the persona perception of a close person. The defined linguistic characteristics are wake-up word, emoji, response time, sentence completion, slang, punctuation, interjection, word transformation, delivery, hedging, back-channeling, abbreviation, emotion, euphemism, split sentence, and sentence structure. Examples for each linguistic feature are shown in Table 6.2.

Wake-up word

A component that explicitly requests the initial attention of a computer with a single word or single phrase is known as a wake-up word. Since our participants believed they were talking to the chatbot with the persona of PRP (i.e. co-participant), we defined the term as wake-up word. Examples of wake-up word participants used in the study were ”Hey”, ”What’s up”, etc.

Emoji

Emoji is a text-based ideogram. Emoji come in a variety of shapes and sizes, including facial expressions, objects, locations, weather, and animals. We included all kind of emojis that represents all kinds of symbols such as :-), XD, etc.

Response Time

Response time is the functional unit of the time it takes to react to a given input in technology, including conversational agents. In our study, we also defined a time and the pattern participants (all of them played a role as a wizard) took to send a response message as response time. Since response time does not have the explicit characteristic, we did not include examples for it in Table 5.2.

Sentence Completion

Sentence completion is defined as the agent’s use of communication patterns to finish a sentence. Complete sentences must usually include a capital letter at the start, follow grammatical rules, include a punctu-

Characteristics	Examples	Characteristics	Examples
Wake-up Word	”Hey”, ”Mom”	Emoji	”^^”, ”;-(”
Sentence Completion	”I’ve done”, vs ”Done”	Slang	”Y’all”, ”Cheesy”
Punctuation	”!”,”?”,”∼”	Interjection	”Wow”, ”Aha!”
Word Transformation	”Mumpy”, ”Cutie”	Hedging	”Certainly”, ”Possibly”
Back-Channeling	”Okay”, ”Uh-huh”	Abbreviation	”lol”, ”BTW”
Split Sentence	”Let’s” + ”meet”	Sentence Structure	”Let’s go to coffee shop”

Table 5.2 Defined linguistic characteristics that affect persona perception of the user’s close person

ation mark at the end of the sentence, and contain one or more major clauses. A major clause can be a standalone subject or a verbal word that expresses the entire content. For example, the sender (i.e.agent) can send ” I went to the restaurant” as ”I”, ”went”,”to”,”the”,and ”restaurant” in the chatting environment.

Slang

Slang is a type of language that includes words and phrases that are intended to be used in a casual manner. Slang is sometimes used by members of specific groups who prefer to use a standard language's specific vocabulary to establish group identity to be shared, exclude approach of outsiders, or both. There were examples such as Y'all, cheesy, and so on in our case.

Punctuation

The use of conventional signs (i.e. !,?,;,.,), sometimes spacing, and typographical words as an aid for correct reading and understanding of the written text is known as punctuation (also known as interpunction). Punctuation is necessary for written English to clarify the meaning and delivery of sentences. Punctuation was also mentioned by users in the modified Wizard of Oz experiment as one of the linguistic elements that affect the persona perception of a person in the real world

Interjection

An interjection is a word, phrase, or symbol used to express a spontaneous occurrence, feeling, or reaction. Exclamations (e.g. ouch!, ahh!, wow!), curses (e.g.damn!), greetings (e.g. hey, yo, bye), response particles (e.g. okay, huh?, mhm), hesitation markers (e.g. uh, uhm, er, hmmm....) and others are examples of interjections (e.g. please, stop, cool). The inclusion criteria of interjections occasionally overlap with profanities, fillers, and sometimes discourse markers due to its diversity. "Wow!" and "Aha!" are two other examples found in the study.

Word Transformation

For word transformation, we defined them as the results from the process of creating new words. This is different from a change in meaning in which a new meaning or meaning changes in an existing word. For example, if the user sent the message "cutie", which should be written in "cute person" in formal conversation, we called it word transformation. The word transformation is similar to word formation which is creating a new word by borrowing, derivations, compounding and blending etc.

Delivery

In the study, we defined the delivery as how clear the meaning of delivered messages. Some participants mentioned that how understandable the messages their partners usually send them determines the characteristics of the sender. Therefore this feature was in our final list of linguistic characteristics. However, since this feature is not explicitly presented through the chat-interface, we did not include the examples in Table 5.2.

Hedging

Hedging is the use of words or phrases in a sentence to reduce ambiguity or the likelihood that the sentence's meaning will be misunderstood. Hedging can be using simple words like "maybe", or "probably," in English. Hedging can also be a useful tool for expressing a stronger point of view in a polite and professional manner. During the empirical study, examples such as "Certainly" and "Possibly" were observed.

Back-channeling

Back-channeling frequently occurs during a conversation when one participant participating in the conversation is speaking and the other participant interjects the current conversation, according to linguistics. Back-channeling responses can be either verbal or nonverbal. When serving primarily social or meta-conversational goals, back-channeling responses may include phatic expressions. Rather than delivering information with high significance, goals may include grabbing the listener's attention, understanding the speaker, or having to agree. Expressions like "yeah," "uh-huh,and "right" are examples of back-channeling.

Abbreviation

A simplified or shortened word or phrase is called an abbreviation. A series of capital letters or the full version of a word can be used as an abbreviation. For instance, the abbreviation or abbr can be used to represent the word abbreviation. It could also be made up entirely with initials, or sometimes it could be a combination of initials and words that represent words with meaning in another language. For instance, in everyday conversation, e.g., i.e., or RSVP are common examples. Acronyms, oe uainf ibirila only (i.e. initialisms), and grammatical crasis are examples of abbreviations. Shortening by any of these could be an example of an abbreviation. In our experiment, words such as "lol", "BTW", "LMK", etc were found as examples of abbreviation.

Emotion

Emotions include the subjective experience of the speaker or writer, expressive behavior, emotional changes, cognitive processes, and sometimes instrumental behavior. For example, as one of the linguistic characteristics defined in this study, we defined it as the emotion perceived by the receiver via text-based messages. All participants agreed that the emotion they perceived in the text influence how they perceive the persona of the chatbot, and the PRP.

Euphemism

Euphemisms are words or phrases used to avoid saying something unpleasant, negative, or offensive. This is one of the communication strategies and we could also observe the tendency of euphemism in some participants' messages. The strategy largely impacts the receiver's perception of PRP. Therefore, we defined it as one of the linguistic factors that affect the persona perception of PRP. However, since this feature is not explicitly presented through the chat interface, we did not include the examples in Table 5.2.

Sentence Structure

Sentence structure is how users use the order of morphemes in a sentence. For example, for the sentence of "What do you want for breakfast?", one can say "you want anything?" and "for breakfast?", or can say " anything for breakfast?". The sentence structure one uses differs by everyone. Therefore, we defined it as one of the linguistic factors that affect the persona perception of PRP.

5.4.2 Priority of Linguistic Characteristics

The goal of the survey was to see how defined linguistic characteristics influence user perceptions of PRP. Participants were asked to answer questions about each linguistic feature in the survey. We ranked 16 linguistic characteristics based on the survey results (Table 5.3). Wake-up word was the highest-ranked characteristic (M=5.56, SD=1.30) from study. These included, among other things, words that participants used to address one another. Second characteristic in the ranking was emoji (M=5.40, SD=1.30). Emojis are now available in a variety of chatting applications, giving people a way to express themselves. Response time was also one of highly ranked characteristic (M=5.34,

Rank	Characteristics	Language Component	Mean	SD
1	Wake-up Word	Content	5.56	1.30
2	Emoji	Content	5.40	1.23
3	Response Time	Use	5.34	1.35
4	Sentence Completion	Form	5.19	1.40
5	Slang	Content	5.08	1.75
6	Punctuation	Content	4.92	1.51
7	Interjection	Content	4.91	1.64
8	Word Transformation	Content	4.88	1.46
9	Delivery	User	4.84	1.46
10	Hedging	Content	4.72	1.52
11	Back- channeling	Content	4.64	1.46
12	Abbreviation	Content	4.63	1.55
13	Emotion	Use	4.49	1.63
14	Euphemism	Use	4.40	1.46
15	Split Sentence	Form	4.29	1.63
16	Sentence Completion	Form	4.11	1.17

Table 5.3 Descriptive analysis of linguistic characteristics that influence persona perception

SD=1.35). And other features were evaluated in the following order : Sentence completion (M=5.19, SD=1.40), slang (M=4.40, SD=1.57), punctuation (M=4.92, SD=1.51), interjection (M=4.91, SD=1.64), word transformation (M=4.88, SD=1.46), delivery (M=4.84, SD=1.46), hedging (M=4.72, SD=1.52), back-channeling(M=4.64, SD=1.46), abbreviation (M=4.63, SD=1.55), emotion (M=4.49, SD=1.63), euphemism (M=4.40, SD=1.46), split sentence(M=4.29, SD=1.63), and sentence structure (M=4.11, SD=1.17).

5.4.3 Differences between language Component

Two linguists classified the linguistic characteristics into major language components in the empirical study (form, content, and use). The difference among language components that affect how users perceive the PRP in the text-based agent was also statistically analyzed. We used a one-way ANOVA to evaluate and compare the means of the scores (Table 5.4.). During the analysis, we found that there was a significant difference among language components defined in the study ($F(2,1320) = 8.036$, P= 0.0001). The highest mean value was found in the content, followed by use and form. There was difference between Form and content (M= -0.4417, P= 0.001), according to Tukey HSD which is post-hoc comparison.

5.5 Implications

In contrast to previous persona studies, we discovered linguistic characteristics that cause users to perceive an agent's persona as a PRP in Study 2. We assumed that some characteristics to be useful in perceiving a close person's persona but we defined additional characteristics that affect persona perception of PRP.

Our results overlap with some of the characteristics that have been shown to be important in previous studies most of which is focusing on finding linguistic characteristics to improve the human-likeness of the agent engaging in natural conversation when it comes to text-based conversational agents. However, by studying how some linguistic markers affect a user's persona perception of PRP, we discovered possibilities of mimicking a specific person's text-based chat styles.

For example, wake-up words, which is one of the highest-ranked characteristics discovered in the study may affect perceptions of text-based agents with PRP. Since they are usually used at the start of the conversation, their impact could be higher than other characteristics by affecting the user's first impression of the agent.

Wake-up word lists could be easily extracted from conversation data between the host and the user, but it should be carefully designed because it could obstruct or disturb natural conversation if it is used frequently [142]. When using PRP with text-based agents, there are 15 additional features to consider. Another major finding of study 2 is that the importance of characteristics could show different patterns depending on the type of PRP applied to the agent (i.e., host of the persona) Persona designers should always be aware of the possibilities that the importance of linguistic characteristics could differ depending on the type of PRP.

PRP has been proven to be an effective persona in view of improving user experience in close relationships. This is because social connection affects the overall quality and user experience with the conversation. By looking into linguistic characteristics, our research could help and guide designers when applying PRP to text-based agents [142].

Since we were interested in applying PRP to the CA for daily healthcare, we needed to conduct an in-the-wild study based on the findings of study 2 to investigate PRP's effectiveness in the domain of healthcare. We'll look at how defined linguistic characteristics, when used in conjunction with PRP, affect user engagement. Users may react both positively or negatively to PRP. According to the Uncanny Valley hypothesis, an agent's human-likeness sometimes can elicit eerie feelings when its human-likeness is imperfect and awkward in view of human observers [143]. Also, additional studies are required to investigate how users accept agents that talk alike healthcare providers with whom they have a close relationship with, particularly text-based agents that model personas solely using linguistic characteristics.

To explore the opportunities and challenges of MimicTalk, we implemented healthcare chatbots (i.e.MimicTalk) based on defined linguistic characteristics in multiple healthcare domains in study 3.

Chapter 6

Study3.Implementation on Lifestyle Domains

From the previous work, we have defined appropriate persona for daily healthcare and defined linguistic characteristics to consider when applying a close person's persona to the CA for daily healthcare. In this chapter, we are going to deploy MimicTalk in the wild. The domain of daily healthcare MimicTalk had been implemented include diet, physical activity, and stress. First, we are going to introduce previous works related to family as an effective persona for daily healthcare and chatbots promoting a healthy lifestyle. Then, we are going to introduce how we implemented MimicTalk to lifestyle domains. In the end, we provide design implications based on our experimental results from the data log, interview, and survey.

6.1 Related Work

We share previous work on (1) family as an effective persona for daily healthcare and (2) chatbots promoting a healthy lifestyle.

6.1.1 Family as Effective Healthcare Provider

In study 1, we confirmed that family members have great opportunities when their persona has applied to CA for daily healthcare. Therefore, we made the decision to apply the persona of the user's family member to our MimicTalk for daily healthcare in study 3. We also share additional studies to support our decision.

According to previous research, a person's family relationship has a significant impact on their overall well-being throughout their lives [144]. To be more specific, family is known to be more strongly linked to lifestyle promotion success than support from other sources [145]. Family support highly affects the receiver's self-worth, and self-esteem which can also result in higher optimism, and positive emotions, and health outcomes [146]. This is because family members frequently regulate each other's behavior and provide information and encouragement to act in a more healthy manner [21]. In comparison to formal support, a family member has been known to be an effective healthcare provider for making daily behavior changes such as diet, physical activity, medication, smoking, and alcohol consumption [114, 147]. Family support, according to the Normative-affective model, is divided into two spheres, one affective and the other instrumental, when compared to other types of human support from medical organizations [148, 149]. Marital, parent-child, grandparent and sibling relationships are all major types of family relationships [150].

Many attempts have been made in the HCI community to apply human-like traits to the agent to build emotional attachment between the user and the agent. However, despite the fact that family relationships strongly influence one's health status, applying a persona of a real family member, rather than a general human, has been rarely explored [145]. Therefore, the study focused on how effective it is to apply users' family members' personas to healthcare chatbots to increase user engagement in healthcare tasks. We also looked into whether the user's relationship with the chatbot was influenced by family dynamics.

6.1.2 Chatbots Promoting Healthy Lifestyle

Well-designed chatbots serve in certain domains to support users to achieve high efficiency with their tasks [27]. Among the various domains that chatbots are being deployed, chatbots are gaining traction in the healthcare domain by helping users achieve health-related goals through an efficient, cost-effective medium that is mobile devices and computers [151]. Also, because sensemaking and learning are accomplished through conversation, chatbots facilitate the success of lifestyle interventions that rely on them [122].

In the medical field, AI-powered chatbots can be used to triage patients and direct them to the right resources [152]. When patients are wondering what's causing their symptoms and how to treat symptoms, chatbots could be a more reliable alternative to online searches [98]. Chatbots have also been credited with improving health-related communication between patients and healthcare providers [153]. On the other hand, chatbots can also promote healthy lifestyles through daily healthcare tasks. When it comes to lifestyle promotion, there is growing recognition in public health that intervening in unhealthy lifestyles can have significant benefits in terms of reducing the risk of symptom occurrence that lead to chronic diseases including diabetes, hypertension [154]. To remain a healthy lifestyle every day, chatbots sometimes play their role as healthcare providers by provoking a healthy diet [155] and physical activity [156], or preventing drug abuse or alcohol overconsumption [157]. Chatbots achieve these outcomes through assistance, education, prevention, and training, and they can be tailored to specific populations [157]. Among the previous researches studying overview of healthcare chatbots' consequences on lifestyle promotion, Pereira et al's study analyzed the main areas being tackled by healthcare chatbot including mental health, physical wellness, nutrition, disorders [158]. This is why we selected diet, physical activity, and stress as our domain of interest.

Since the focus of the thesis is daily healthcare for preventive purposes, we implement the idea of mimicking the user's healthcare provider in the domain of lifestyle management. Managing lifestyle behaviors can prevent the prevalence of lifestyle diseases that are defined as diseases highly associated with, and often caused by lifestyle habits [159]. Non-communicable diseases can be an example that requires behavior change in daily lifestyles including type 2 diabetes, heart disease, stroke, and cancer. Unhealthy eating, lack of physical activity, overload stress, alcohol overconsumption, drug abuse, and frequent smoking are common and general causes of lifestyle diseases. In industrialized countries, these diseases appear to become more common.

Diet

For managing user's diet-related behavior, food journaling has long been used for tracking user's food intake and analyzing the nutritional value of the user's diet. In other words, food journaling has been considered the most traditional method of collecting data about the user's food intake. Food journaling is an example of using a journal to improve daily health habits [160]. Food intake also aids the treatment of obesity and chronic diseases whose outcomes could be improved through dietary changes, such as type 2 diabetes and dyslipidemia, but also helps people avoid developing those diseases [161]. The HCI community has created a variety of food journaling tools to support the user to reach a variety of goals including healthier food choices, weight loss, identifying allergies, detecting deficiencies, and detecting foods that trigger additional symptoms [162]. There exist previous researches provoking fruits and vegetable consumption [155], and adherence to longevity eating plan [163]. Also, there are commercial examples including Lark, Lysa, and some of them are implemented to chatting messengers such as Telegram or Facebook messenger (e.g.Forksy, Tasteful). However, if the chatbots cannot ensure continuous user participation, they mostly fail to reach the goal of successful nutrition management [40]. As such, food journaling is expanding its traditional use not only for personal informatics that utilizes food tracking methods for self-monitoring but also for family informatics or community informatics that enable data-mediated communication and lifestyle promotion among family health-care providers or community members [164, 165].

Despite these positive effects, decreasing engagement over time remains a major challenge. Usually, paper food journaling is a burdensome task that causes fatigue [166] and oblivion [167, 162]. To prevent the user from deviating food journaling attempts to support the user's food journaling have been made. Popular mobile apps like Noom Coach, MyFitnessPal, and others, for example, partially improved the limitations of food-intake recording by partially automating journaling. Fully automated food-journaling solutions also have been developed with limited [168, 169, 170]. However, full automation provoked experts' concerns that could reduce the mindfulness benefits coming from food journaling behavior itself [171].

As a new solution, in recent years, CAs are gaining traction as a useful tool for food journaling by proactively asks the questions [155] or gives the information [163, 172]. In the HCI community there are chatbots proposed to enable not only improving mindfulness or self-reflection based on collected data but also capturing contextual information and internal moods based on their conversation [173]. In academic approaches, there were previous attempts to journaling fruits and vegetable consumption [155], longevity eating plan [163], daily eating behaviors with feelings and contextual information [174]. Also, there are commercial examples including Lark [175], Lysa [176], and some of them are implemented to chatting messengers such as Telegram or Facebook messenger (e.g. Forksy [177], Tasteful [178]).

In this study, we explored the opportunities and challenges of applying the persona of a close person to the food journaling chatbot through the experiment.

Physical Activity

Physical activity is another important healthcare task that affects one's longevity and health. Physical activity is strongly linked to better health and a lower risk of diseases that affect one's mortality. Physical activity has numerous health benefits, including a lower risk of cardiovascular disease, ischemic stroke, type 2 diabetes, cancers, osteoporosis, depression, and injuries, in addition to its benefits on mortality [179].

Physical inactivity is also one of the leading risk factors for noncommunicable diseases [180]. The level of physical inactivity varies greatly between and within countries. For example, these risks of noncommunicable diseases are rapidly increasing in US. Nearly 80% of adults in the United States do not follow recommendations for physical ctivities [181]. To prevent the prevalence of lifestyle diseases, cost-effective and feasible physical activity interventions must be developed. According to the guidelines, some physical activity is preferable to none. They also claim that any amount of physical activity is beneficial to one's health. For many people, increasing their physical activity to the recommended levels necessitates long-term changes in their attitudes and behaviors.

Previous studies in pervasive health have demonstrated the utility of mobile or computer-based applications in assisting individuals in measuring, tracking, and reflecting on their levels of physical activity, though the success of these technologies is contingent on the individuals' motivation to participate in such activities. Existing technologies are also geared toward eliciting extrinsic motivation rather than improving users' intrinsic motivation to engage in physical activity, which could lead to longer-term behavior change [182].

However, few studies have explored the opportunities of implementing persona-based CA to the chatbot for provoking physical activity. Therefore, in the second experiment included in study 3, we investigated the opportunities and challenges of the persona of a close person for physical activity intervention.

Stress

Various non-communicable diseases are thought to be caused by stress. Stress is linked to a variety of physiological and mental disorders [183]. Unfortunately, while many people recognize that stress is a major issue, only a small percentage of people know how to effectively deal with it [183]. Recent research has demonstrated the importance of learning to cope with stress in a constructive manner in order to minimize its negative consequences [184]. When we are stressed, our bodies and physiological mechanisms go through a series of changes known as the "fight or flight" response [185]. Many external stressors confront modern humans, such as a looming paper deadline, a job interview, a critical presentation, and so on. However, there are often no practical options for "fighting" or "flighting" from these stressors. Dealing constructively with stressors in a variety of ways is a skill that everyone should learn.

HCI researchers have recently attempted to manage stress with a chatbot. For example, there was an attempt to see if Woebot could be used to help senior high school students cope with their academic workload. Participants in the study conversed with Woebot on a daily basis for two weeks. They also gave the chatbot a higher than average score for resemblance to a human and the ability to understand and empathize with the participants' feelings. Furthermore, an examination of the conversational logs revealed that the participants admired Woebot's lessons and stories, but that they had difficulty dealing with the chatbot's inappropriate responses.

6.2 Research questions

With the defined linguistic characteristics to apply PRP to the text-based agent, we applied the persona of the user's family member to the healthcare chatbot. We deployed our prototype into multiple healthcare domains including diet, physical activity, and stress management. The final study includes 3 corresponding experiments. Based on the results from study 3, we share design implications for designing the healthcare CA with PRP. The research questions for study 3 are as follows.

RQ1: How does the persona of the user's family member impact the user's behavior with healthcare tasks?
RQ2: What are the design issues to consider when applying the CA with dialogue style of the user's family member to major domains in daily healthcare?

6.3 Implementing Persona of Family Member

Chatbots are increasingly being used in everyday healthcare because they can promote health, provide education, and possibly encourage behavior change [151]. According to Juniper Research, it is predicted that we will be interacting with a veritable legion of AI-powered chatbots as part of our regular healthcare over the next five years. Particularly for daily healthcare, the HCI community has been proposing chatbot systems in various types of healthcare domains. For example, there are healthcare chatbots not only monitoring everyday changes in mood [186], dietary intake [155], physical activity [156], or medication [187] but also reminding users for daily medication [32, 33] or intervening users to promote healthy lifestyles [30, 31]. By doing so, healthcare chatbot acts as a digital healthcare provider by levying the human healthcare providers of menial work [188].

However, a weak relationship between the user and the chatbot compared leads to low participation. If the chatbot cannot ensure continuous user participation, it fails to reach the system goal for user's daily healthcare regardless of its tremendous benefits [40]. Therefore, building a strong relationship between the user and the chatbot should be considered in the design process. Among many indicators of successful relationship building, emotional attachment has been known as a critical indicator for a long-term relationship. For example, people sometimes form a strong emotional attachment to a product irrespective of its utilitarian value [189]. It has been proven that emotional attachment acts as a mediator in building a strong relationship not only in human-human relationship but also in the relationship between the user and the artifacts [190, 191].

6.3.1 Domains of Implementation

We have selected diet, physical activity and stress as the domain of implementing research ideas regarding healthcare CA with the mimicked persona of users' healthcare providers. For each domain, we selected major healthcare tasks for daily healthcare. In the diet domain, we have concentrated on the users' perception and behaviors towards the healthcare chatbot for food journaling that mainly collects the user's daily data of food intake. In the domain of physical activity, we have concentrated on the users' perception and behaviors towards the healthcare chatbot for intervening users to reach their goals with physical activity. Lastly, in the stress management domain, we focused on the users' perception and behaviors towards the healthcare chatbot for counseling the user's stress triggers and coping strategies.

To summarize, we explored and defined the users' perception and the behaviors with the healthcare CA with the functions of data collection, behavioral intervention, and counseling in the domain of diet, physical activity, and the stress management. Before introducing our study procedure, we share related work on healthcare domains of interest.

Domain	Diet	Physical activity	Stress management
Main task	Data collection (food journaling)	Behavior intervention	Counseling
Method	Dialogue design + prototyping + deployment	Dialogue design + Wizard-of-Oz	Dialogue design + Wizard-of-Oz
Days of participation	21 days	21 days	21 days
Number of participants	N= 24	N= 24	N=24
Dependent variable	Acceptance+ Engagement + User perception +impact of linguistic factors	Acceptance+ Engagement + User perception +impact of linguistic factors	Acceptance+ Engagement + User perception +impact of linguistic factors

Table 6.1 Overall introduction of experiments included in study 3

6.3.2 Measurements Used in the Study

Study 3 include three main experiments to investigate the opportunities and challenges of healthcare chatbots with the persona of a close person in the multiple domains. The first experiment focuses on the healthcare chatbot with the function of food journaling. The second experiment focuses on the healthcare chatbot with the function of physical activity intervention. Lastly, the third experiment focuses on the healthcare chatbot with the function of stress management. Through this experiment we qualitatively evaluated the overall user experience with the MimicTalk and how linguistic factors we previously defined in study 2 that affects persona perception of MimicTalk work differently

Domain	Category	Questionnaire (example)
User experience	Behavior (pre)	Before using the chatbot,I continuously did the task on the regular basis
	Behavior(post)	During the chatbot use,I continuously did the task on the regular basis
	Usefulness	Overall, I found the chatbot useful to perform the task.
	Ease of use	Overall, I am satisfied with how easy it is to use this chatbot
	Trustfulness	During the chatbot use,I trust the chatbot I used.
	Intimacy	During the chatbot use,I felt intimacy with it.
	Engagement	Overall, it was very engaging using this chatbot
	Attachment	During the chatbot use,I felt emotionally attached to the chatbot
	Human-likeness	During the chatbot use,I felt like I was interacting with human
	Eerie feeling	During the chatbot use,I felt eerie feeling toward the chatbot

Table 6.2 User experience measurements used in the study 3

on the three types of healthcare tasks.

When it comes to user experience, measurements include the questionnaires about behavior change, usefulness, ease of use, trustfulness, intimacy, engagement, attachment, human-likeness and eerie feeling. The detailed questionnaires are defined in Table 6.2.

When it comes to user perception with the MimicTalk, we measured how each linguistic factor we defined in study 3 affects persona per-

Domain	Category	Questionnaire (example)
User perception	**Wake-up word**	Wake-up word the chatbot used made me perceive the chatbot as my family member
	Response time	Response time the chatbot answered to my messages made me perceive the chatbot as my family member
	Sentence completion	How the chatbot complete the sentences made me perceive the chatbot as my family member
	Slang	Slang the chatbot used made me perceive the chatbot as my family member
	Punctuation	Punctuation the chatbot used made me perceive the chatbot as my family member
	Interjection	Interjection the chatbot used made me perceive the chatbot as my family member
	Word transformation	How the chatbot transform the specific word made me perceive the chatbot as my family member
	Delivery	How the chatbot deliver the meaning of messages made me perceive the chatbot as my family member
	Hedging	Hedging word the chatbot used made me perceive the chatbot as my family member
	Back-channeling	Back-channeling the chatbot used made me perceive the chatbot as my family member
	Abbreviation	Abbreviation the chatbot used made me perceive the chatbot as my family member
	Emotion	Emotion delivered with the chatbot's messages made me perceive the chatbot as my family member
	Euphemism	Euphemism the chatbot used made me perceive the chatbot as my family member
	Structure	Overall structure of the conversation the chatbot used made me perceive the chatbot as my family member

Table 6.3 Questionnaires measuring user perception of MimicTalk

ception of MimicTalk. The factors include the wake-up word, response time, sentence completion, slang, punctuation, interjection, word transformation, delivery, hedging, back-channeling, abbreviation, emotion, euphemism, and structure. The detailed questionnaires are defined in

Table 6.3. Throughout this study, we aimed to examine the following research questions:

RQ1. How to apply the persona of the user's family member to the CA?

RQ2. How does the persona of the user's family member impact the user's behavior with healthcare tasks?

RQ3. What are the design issues to consider when applying the CA with the dialogue style of the user's family member to the various healthcare domain?

6.4 Experiment 1: Food Journaling Chatbot

In the first experiment, we focused on food journaling for nutrition management because it is proven that nutrition is the main area that is being tackled by healthcare chatbot [158] and daily family support act as a critical determinant of success in this domain [192]. The need for applying the persona of the family member to the food journaling chatbot was discussed in the previous section.

6.4.1 Method

To investigate the effectiveness of MimicTalk, we conducted an experimental study for 21 days. 24 users without any severe diseases participated in the study. They were randomly assigned to the group using the with the conversational style of their family member (MimicTalk) and the control group using the basic chatbot (Basic bot) without the conversational style of a family member.

Designing Structure of MimicTalk for Food Journaling

Experiment 1 includes three main steps. As a first step, we conducted desk research and analyzed conversational data of families that include asking what the user ate. Through this step, we tried to define the general structure of food journaling. We decided to design the food journaling chatbot based on the defined structure since it requires continuous data collection in a structured way.

Particularly, we aimed to define the simplified dialogue structure for the food journaling chatbot. We analyzed existing food journaling chatbots and conversational data in a chatting platform. In our design process, we aimed to build a simplified food journaling chatbot over a highly advanced chatbot, because we put our priority on investigating how applying the dialogues of family member to the chatbot improve user engagement in food journaling.

Specifically, we followed the following steps to structurize chatbot: First, we analyzed existing food journaling chatbots such as NomBot [193], Forksy [177] and SLOWbot [163] to understand how and what they ask to collect the data of the user's food intake. Second, we analyzed the chat history including the conversation of asking or answering about food intakes among family members from participants (N=6) recruited for this step. We analyzed two types of data by labeling each function into themes based on the grounded theory.

Finally, with the most frequently occurred functions, we finalized the main structure of our simplified food journaling chatbot consisting of 6 functions: (1) greeting (greeting sentences to initiate the dialogue), (2) text, and (3) image-based journaling (journaling prompts asking what users ate), (4) compliment (motivating message for efforts they made while participating in food journaling), (5) cheering up (motivating message for poor participation), and (6) farewell (goodbye message to end the dialogue). The example sources we used for the first step are described in Figure 6.1.

Designing Dialogue of MimicTalk

With the defined structure, we designed and implemented MimicTalk, whose dialogues were designed by users' family members in their conversational tone to apply the persona of a close person to the MimicTalk for the second step. This is because the dialogue style of the chatbot has been known to affect the persona perception toward the chatbot [47, 64]. Therefore, we aimed to apply the dialogue style of the user's family members by involving them in the dialogue design

Figure 6.1 Screenshot of Forksy and the sample conversational data as references that were used to design a structure for food journaling chatbot

process. We aimed to design the dialogues that would be implemented to the simplified chatbot structure with 6 major functions including greeting, text-based journaling, image-based journaling, compliment, cheering up, and farewell.

At the start of the design process, we informed family members of participants about our research purpose and how the chatbot functions as a food journaling tool. Additionally, we explained the structure of the chatbot for food journaling. For this, we provided an example dialogue flow of the chatbot and introduced 6 functions including greeting, text-based journaling, image-based journaling, compliment, cheering up, and farewell. Then, we asked participants to type the messages that will be used in each function. For each function, we asked them to make 10 different sentences in the tone of their daily conversation as if they are chatting with participants who are going to participate in the main experiment,

For helping participants to design dialogues by reducing the design burden, we offered participants detailed instructions with samples done by the authors. We also shared two tips to make the process easier for the participants in the design process. First, we told them they can refer to their chatting history with their family member who is going to participate in the experimental study. Second, we provided the list of linguistic factors they can consider during their dialogue design process and told them to think about how they usually use these factors in their conversations with their parents (e.g. how many times, on what occasion, etc). The list of linguistic factors includes the wake-up word (e.g. Hey!, Mom, Hello, Surprise!!), emojis (e.g. :), ;-(, ¡3), abbreviations (e.g. lol, BTW, ASAP, LMK), punctuation (e.g. !!, ??, ˜˜˜) and more lists of factors from study 2. With the sentences collected from the participant's family members, we applied them to the MimicTalk and the MimicTalk was implemented on the Telegram. MimicTalk is easily configurable to registered users with the functions of sending pre-designed dialogues from family members. With the MimicTalk, we also implemented a basic chatbot whose dialogues were designed by authors not the family members of participants to the Telegram for the next step of the experimental study. To reduce the journaling burden of the users at the same time preventing them from forgetting their food intakes, we implemented a 24-hour recall method [194] which is a method of checking user's dietary intake in the past 24 hours once a day as the journaling strategy for all chatbots. Besides, we allowed users to set their customized time of receiving the first message from the chatbot for the day to avoid their working hours or bedtime.

As a result, two types of prototype chatbots (i.e. MimicTalk and the basic chatbot) were designed and implemented in the mobile messenger. We implemented a chatbot along with dialogues designed by users' family members into the Telegram messenger app. To precisely identify the effectiveness of MimicTalk during the evaluation study, we also implemented a basic chatbot with the dialogues constructed without the participation of the user's family members to compare it against each other. Instead of a family member's daily conversational tone, the baseline chatbot embeds a formal conversational tone by referring to a study by Casas et al. [155]. The server of each chatbot was built with Python 3.7 with Flask, deployed on AWS EC2 to record the usage logging for further analysis.

Figure 6.2 shows a sample conversation with one of MimicTalk and

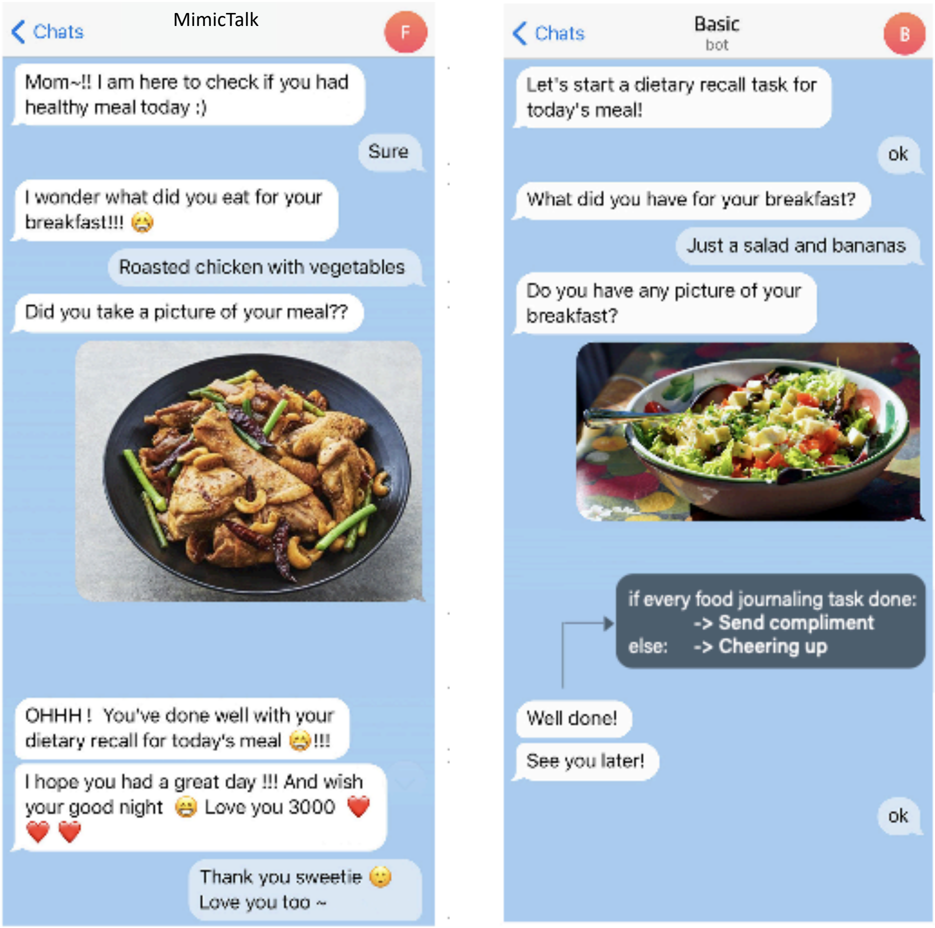


Figure 6.2 6 major functions of MimicTalk and the basic chatbot

the basic chatbot. Both chatbots have an identical dialogue structure consisting of 6 major functions, only with their dialogue styles different. At the time that each user pre-defined before using the chatbot, the chatbot sends proactive prompts with a greeting message (function 1). If a user answers the chatbot, then the chatbot sends journaling prompts that ask the food user ate for a breakfast (function 2). After the user's text-based answer to the journaling prompts, it asks the user for image upload (function 3). The user can skip the image upload if they do not have a picture of their meal. Function 2 and function 3 are repeated for lunch and dinner with the journaling prompts with different dialogue styles. After users complete their tasks, it sends the compliment message (function 4) or cheering-up message (function 5). If every food journaling task has been done, the chatbot sends compliments to users, while it sends a cheering-up message if users did not complete the task. Then, the chatbot sends the user the farewell message to end the conversation (function 6).

Evaluation

With the MimicTalk and the basic chatbot, we conducted an evaluation study for 21 days with 24 participants divided into two groups to investigate the effectiveness of MimicTalk on user engagement in food journaling. We randomly assigned participants into two groups including an experimental group using a MimicTalk, which utilizes the dialogues designed by their family members, and the control group using a basic chatbot, which has the identical dialogue structure of MimicTalk except its dialogues designed by authors based on formal dialogue style.

Table 6.4. shows the information of participants including their ID, type of chatbot each participant used, sex, age, occupation. For the par-

Participant No.	Type of chatbot used	Sex	Age	Occupation	Family participant
P1	MimicTalk	Female	55	House spouse	Son
P2	MimicTalk	Male	60	Office worker	Daughter
P3	MimicTalk	Female	57	Teacher	Son
P4	MimicTalk	Female	58	Office worker	Daughter
P5	MimicTalk	Male	60	Office worker	Daughter
P6	MimicTalk	Female	56	Accountant	Son
P7	MimicTalk	Female	53	Office worker	Daugter
P8	MimicTalk	Female	54	House spouse	Daughter
P9	MimicTalk	Female	52	House spouse	Daughter
P10	MimicTalk	Female	50	Teacher	Son
P11	MimicTalk	Male	56	Office worker	Daughter
P12	MimicTalk	Female	47	Teacher	Daughter
P13	Basic chatbot	Female	56	Official	
P14	Basic chatbot	Male	57	Teacher	
P15	Basic chatbot	Female	59	House spouse	
P16	Basic chatbot	Female	52	Teacher	
P17	Basic chatbot	Female	54	Teacher	
P18	Basic chatbot	Female	53	House spouse	
P19	Basic chatbot	Male	50	Office worker	
P20	Basic chatbot	Male	59	Office worker	
P21	Basic chatbot	Female	53	Office worker	
P22	Basic chatbot	Female	53	House spouse	
P23	Basic chatbot	Male	50	Office worker	
P24	Basic chatbot	Female	60	House spouse	

Table 6.4 Information of participants included in the evaluation of food journaling chatbot

ticipants who used MimicTalk during the experiment, the relationship with the participant who designed the dialogues of the MimicTalk is also shown in Table 6.4. Overall, participants' age ranged from 52 to 60. Participants suffering from any severe diseases (e.g. cancer, diabetes, etc) were excluded from our study because they need medical support, prior to managing eating behaviors. Participants received 150K KRW (approximately 125 USD) for 21 days of participation.

At the start day of deployment, we informed participants about our research goal and the function of our chatbots and all parent participants answered the pre-survey. Additionally, participants who used MimicTalk answered to the questionnaires about their family members including their age, how frequently they see each other, and perceived emotional attachment with their family member on a 7-point Likert scale with the questionnaire of "In recent years, I feel emotionally attached to my family member" (answering option ranged from "Strongly agree" to "Definitely not agree".) All the participants who used MimicTalk reported high emotional attachment to their family member (M= 6.67, SD = 0.51).

To manage user expectations, we conducted 30 minutes of detailed introduction sessions before starting the experiment. In the session, we shared our research goal and research questions. Then, we specifically introduced the functionality of the chatbot and its limitations coming from the study design. We not only told them they were recruited to evaluate the user experience and user perception of MimicTalk (or basic chatbot) and told them they are going to interact with the chatbot with the tasks of food journaling. We emphasized that the chatbot's functionality is limited in journaling tasks and the chatbot would be limited in responding to out-of-task conversations. Then, we showed the sample conversation of the chatbot in order to help participants be accustomed to the workflow of the chatbot. Participants were instructed to freely ask any questions regarding the chatbot and the experiment to the author. Through these processes, we made it sure for users to manage their expectations with the chatbot. After participants got used to the function of the chatbot, they were asked to freely use the chatbot for 21 days.

To analyze the user engagement with MimicTalk over time, all participants' conversational data with the chatbot were recorded for analysis. We not only measured the responses to journaling prompts but also any responses made regardless of journaling behavior to deeply explore how users were engaged with the chatbot.

Both usage frequency and intensity of usage were measured to analyze user engagement [195]. we measured the average number of days using two types of chatbots for overall usage frequency and we also measured the responses to text-journaling prompts, image-based prompts to analyze the intensity of chatbot usage. We also compared user experience and the user perception with the MimicTalk with the basic chatbot since it affects long-term user engagement with the artifact [191].

At the end of the deployment study, participants answered the post-survey and the post-hoc interview. Post survey included the questionnaire of user experience and user perception in Table 6.2 and Table 6.3. Answers to the pre/post survey were all collected with the 7-point Likert scale (answering options ranged from "Strongly agree" to "Definitely not agree".)

To explore the major findings from the conversational logs we collected, a semi-structured interview was conducted after the deployment. During the post-hoc interview, we aimed to explore the user engagement with the chatbot and how family members' dialogues affected users' attachment to the chatbot over time. To build our interview questionnaires, we referred to Short et al's study on user engagement [196]. Also, we asked participants how they perceive the personality of the chatbot that uses dialogues of their family members.

6.4.2 Results

We present our results from the evaluation study regarding user engagement in food journaling and user perception to the chatbot by comparing two types of chatbots (i.e. MimicTalk and basic chatbot) based on the data log of 24 participants collected during 21 days of deployment, pre/post survey, and interview. For the statistical analysis, we used the Wilcoxon rank-sum test.

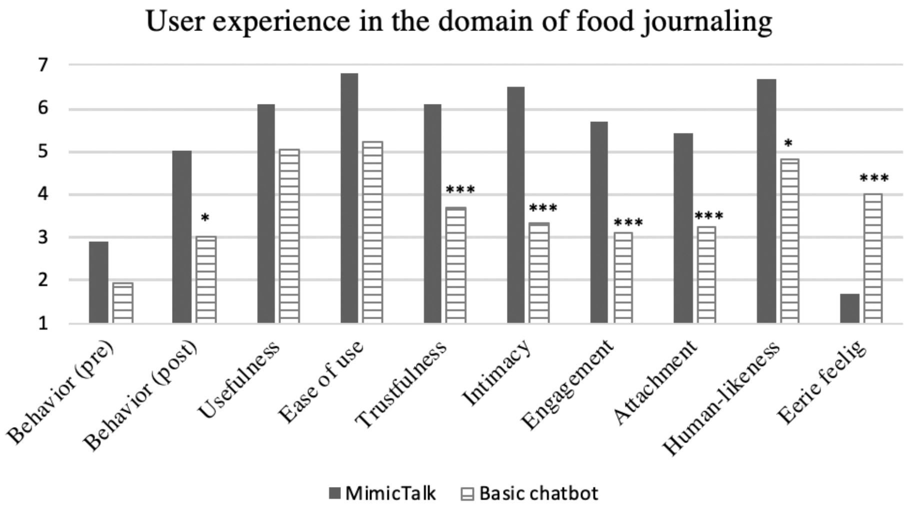


Figure 6.3 Reported user experience with MimicTalk and basic chatbot for food journaling

User Experience with healthcare Chatbots

Food journaling is a main task for both types of chatbots including MimicTalk and the basic chatbot.

We measured user experience with the MimicTalk and the basic chatbot. The mean and the standard deviation of UX factors of MimicTalk evaluated by participants are as following: user's motivation to participate in the healthcare behavior (pre (M=2.91, SD = 1.08) and post (M= 5.0, SD = 1.13)), usefulness (M= 6.08, SD = 0.67), ease of use (M= 6.83, SD = 0.67), easy of user (M 6.5, SD = 0.52), trustfulness (M= 6.08, SD = 0.67), intimacy (M = 6.5, SD = 0.52), engagement (M= 5.66, SD = 0.98), attachment (M= 5.41, SD = 1.08), human-likeness (M= 6.67, SD = 0.49), and eerie feeling (M= 1.67, SD = 1.72).

Also, the mean and the standard deviation of UX factors of basic chatbot evaluated by participants are as following: user's motivation to participate in the healthcare behavior (pre (M= 2.38, SD = 0.96) and post (M=3.92, SD = 1.03)), usefulness (M= 6.53, SD = 0.51), ease of use (M= 6.77, SD = 0.43), easy of user (M= 6.76, SD = 0.43), trustfulness (M= 4.69, SD = 0.85, intimacy (M= 4.0, SD = 2.36), engagement (M= 3.76, SD = 0.43), attachment (M= 3.53, SD = 0.87), human-likeness (M= 5.69, SD = 1.10), and eerie feeling (M= 4.46, SD =2.57).

Among these factors, there was significant difference between groups (MimicTalk and basic chatbot) in the post behavior (P < 0.05), trustfulness(P < 0.01), intimacy (P < 0.01), engagement (P < 0.01), attachment (P < 0.001), human-likeness (P < 0.05), and eerie feeling(P < 0.001).

Additionally, We analyzed the user data log to investigate how participants interacted with the MimicTalk. With the user data from the chatbot asking participants to input their meal by text (function 2. text-based journaling) and by image (function 3. image-based record), we quantitatively analyze the user's engagement with their data log, we measured two methods that were previously suggested by Perski et al: (1) frequency of use (i.e. how many days did participants use the chatbot) and (2) intensity of use (i.e., how much did they respond to the chatbot's journaling prompts) [195]. The frequency of use was analyzed based on the number of days that the participants responded to the chatbot at least once for the day. This number was automatically calculated on our server with the user's data log. The average number of days each user participated in journaling was 20.81 days (SD = *1.41*) in the group of participants who used MimicTalk during the experiment. Although it was slightly higher than that of users who used basic chatbot (M= 19.6, SD = 3.16), mean difference between two groups were not statistically significant (W= 52.5, P= 0.54).

Also, the intensity of use which implies user engagement with the chatbot was analyzed based on the data log collected during the deployment showing whether participants responded to the journaling prompts or not. The number of responses for (1) text-based and (2) image-based was collected for each day. While exploring the participants' data log, we identified that some participants make additional responses that are not directly related to the journaling behavior of the chatbot. Since analyzing such behaviors might give us novel opportunities for how users perceive chatbots, we also analyzed (3) additional responses from the participants that are irrelevant to food journaling behavior that is known to be positively associated with user engagement with the system [197].

Figure 6.4 shows the two graphs indicating the intensity of users'

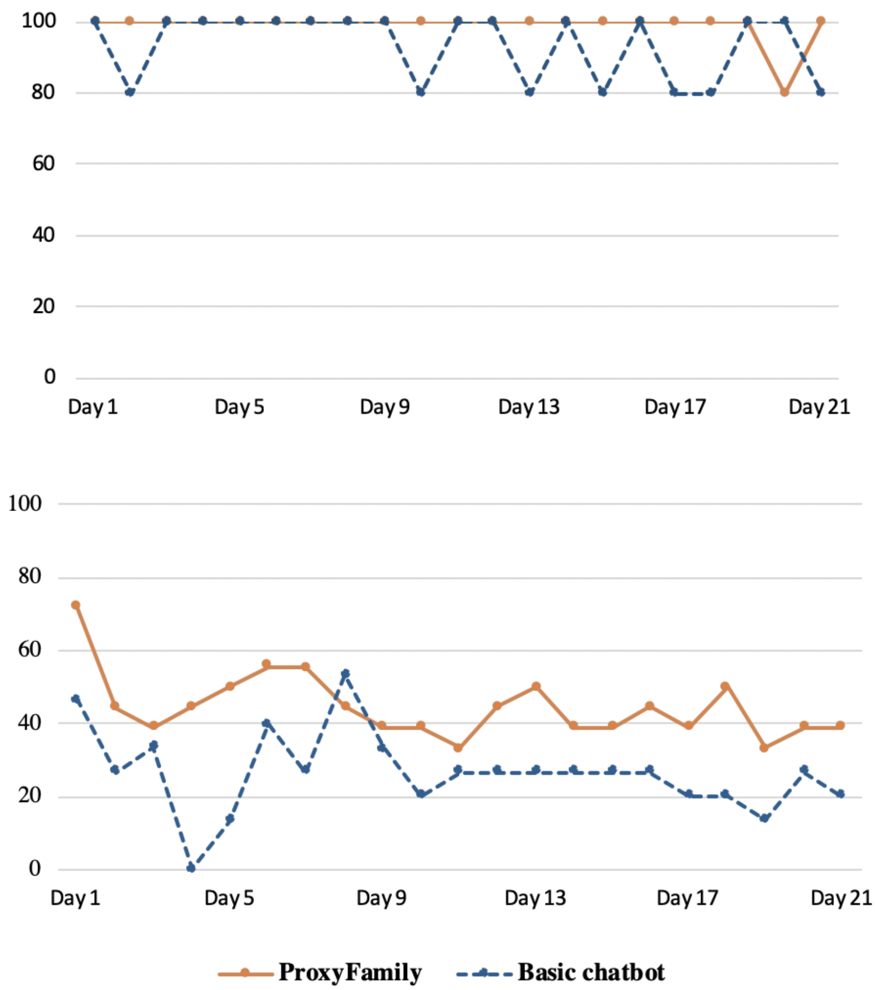


Figure 6.4 Graphs showing participation of users who used MimicTalk and the basic chatbot for food journaling

behavior responding to the food journaling chatbot during 21 days of deployment. For each day, participants had a chance to answer a total of six journaling prompts (3 types of meals(breakfast/lunch/dinner)X2 types of journaling prompts (text/image)). The solid line shows an average response rate among participants in the MimicTalk group calculated for each day, while the dotted line shows an average response rate for the basic group. The upper Graph was calculated based on the number of participants' responses in text-based journaling (function 2) for each day while the lower graph was calculated based on the number of responses in image-based journaling (function 3) for each day. The solid line (orange) shows the response rate for each day among participants in the group who used MimicTalk on average, and the dotted line (blue) shows the response rate for each day among participants in the basic group on average.

The upper graph in Figure 6.4 shows the text-based response rate in text-based journaling. In this graph, participants; text-based responses rate show no remarkable difference between two groups which resulted in no statistical significance ($W = 52.5,\ P = 0.54$). However, we found that there are differences in participants' response styles in text-based food journaling. Table 6.5 shows examples of text-based responses in two groups and we share our major findings. First, most participants who used MimicTalk responded to journaling prompts as if they are responding to the human receiver. For example, they made their responses in full sentences while participants using basic chatbot only listed the names of the foods they ate.

Second, participants who used MimicTalk frequently included contextual information with the list of food they had. They wrote down the reasons why they skipped the meal, with whom they had their meal, or how they cooked their meal, etc. This evidence led to clear quantitative differences between the two groups based on the word count calculated without a list of names of foods in their journaling responses. A remarkable number of words were reported in the group of the participants who used MimicTalk with a mean value of 59.83 ($SD = 23.73$). However, there is no single word reported to be made without lists of their food intake in the responses of the basic group for food journaling.

On the other hand, the lower graph in Figure 6.4 shows the image-based response rate in image-based food journaling. The continuous gap between the two groups indicates higher image-based responses in the MimicTalk group compared to the basic group. Moreover, the group of people who used MimicTalk shows a more steady tendency of making their responses over time compared to that of the basic group. The overall difference in image-based responses was a statistically significant difference in the average number of image uploads for 21 days between the two groups ($W = 395$, $P < 0.0001$).

Additional responses irreverent to food journaling behavior:

To explore the user's behavior of responding to the chatbot in a so-

Responses to MimicTalk
P1: "I ate pasta and salad with your aunt! I was really excited :)"
P4: "I ate fried rice cooked with green beans and lamb with your dad."
P6: "I skipped my dinner because I am still full of heavy lunch I had an hour ago :0"
Responses to Basic Chatbot
P14: "Whiitefish, rice, Kimchee, Miso soup."
P16: " I skipped (the meal)."

Table 6.5 Participants' text-based responses to the journaling prompts

cial manner, we also analyzed additional responses in participants' conversational logs excluding the user's responses to journaling prompts. Additional responses from the participants who used MimicTalk were higher than the basic group showing 19.5 words on average (*SD = 13.2*). The number of responses in the basic group was 7.6 words on average (*SD = 15.3*), which shows a statistically significant difference compared to that of users of MimicTalk (W = 27, $P < .05$).

For MimicTalk, the most frequent social responses were found after the chatbot's function of sending farewell messages (98.6%). With thematic analysis [135], we categorized these responses into 4 themes that include: (1) answering MimicTalk's farewell messages (e.g. thank you son!)(62.40%), (2) summarization of daily event (e.g. today I have been to a barbershop)(18.75%), (3) expressing feelings to their family members (e.g. Miss you so much)(12.56%), and (4) sharing updated news (e.g. I heard that it's going to be raining tomorrow)(2.09%).

On the other hand, responses from participants in the basic group only included the short responses answering MimicTalk's farewell messages (e.g. OK). Some of the participants in the basic group did not even make any responses except responses to food journaling prompts.

However, as we can see in Figure 6.5, the responses from participants are highly related to the characteristics of chatbot prompts. Therefore, there exist possibilities of increasing social responses from basic chatbots by changing the characteristics of chatbot prompts. But still, emotional interaction seems hard to apply to the basic chatbot without personality as emotional attachment between the chatbot and participants was significantly higher in the MimicTalk group.

Emotional attachment affecting user engagement: We have measured users' perceived emotional attachment to the MimicTalk at the end of deployment since it is one of the key indicators increasing long-term user engagement in using the artifact [191]. In the post-survey, we asked them the following question: "I felt emotionally attached to this chatbot after using this chatbot for 21 days (1: definitely not agree, 7: strongly agree)". As a result, perceived emotional attachment to MimicTalk on average was higher compared to the basic chatbot. The average score was 6.17 for MimicTalk (SD = 0.75), while the average score was 5 for the basic chatbot (SD = 0.75). Mean difference between two types of chatbots were statistically significant ($W = 6$, $P < 0.05$). Also, in the post-hoc interviews, most participants who used MimicTalk reported the positive impact of their emotional attachment to the chatbot on their engagement with MimicTalk through thematic analysis [135]. For example, P6 said that *"I felt like my son was taking care of me by using the chatbot that talks like my son, which made me emotionally attached to the chatbot."*. Also, P1 mentioned that *"Warm messages written by my daughter made me answer the chatbot with love."*.

Frequent emotional response, which is the key indicator of emotional attachment, found in MimicTalk:

Within the data log of users who used MimicTalk, we found frequent emotional responses in their social responses (i.e. additional responses we defined above). We deeply analyzed users' tendency of making emotional responses because it has been proven that the extent of emotional responses is positively associated with the emotional attachment to the artifacts [197].

We analyzed the emotional word frequency in the additional responses from participants made regardless of food journaling behavior. We defined the emotional word as showing feelings (e.g. happy, sad, regret) or using emotional expressions including emojis or endearment (lol, sweetie, babe, lover, :-), etc.). Two of the authors independently coded each word in participants' additional responses to the chatbot based on whether it is an emotional word or not. This analysis showed a strong inter-coder agreement between the two researchers (Cohen's Kappa coefficient (κ) = 0.89). As a result, an average of 11.67 emotional words was found among those who used MimicTalk (SD = *10.05*) in contrast to no emotional responses for the basic chatbot.

User Perception with MimicTalk

We also explored how users perceive the MimicTalk. Particularly, we focused on how linguistic factors we defined in Study 2 affected the persona perception of food journaling chatbot designed with the dialogue style of family members. To explore how the user perceives the personality of MimicTalk with the dialogues of a family member, we asked participants how they define the personality of the MimicTalk and What factors affected their perception.

	Df	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
Linguistic factors	14	171.1	12.220	10.49	10.49 <2e-16 ***
Residuals	165	192.2	1.165		

Table 6.6 Linguistic factors affecting user's perception of the chatbot with the persona of the user's family member during food journaling tasks

As a result, we found that participants using MimicTalk perceived the chatbot's personality differently even though our explanations of how the chatbot had been made were identical to all participants. Some participants defined the chatbot as their actual family member due to their dialogue styles implemented to chatbots. They insisted that the dialogues mimicking their family member made them feel like they were making real-time conversations even though they were instructed that it was not their real family member but the chatbot designed with a family member's dialogue style.

Also, We defined to what extent each linguistic factor affected the persona perception of MimicTalk as their real family member. Table 6.6 shows results from one-way ANOVA. There was significant difference among linguistic factor affecting the persona perception of MimicTalk. The mean values and the standard deviations of each linguistic factor are as following: wake-up word (M= 6.34, SD= 0.77), emoji (M= 5.83, SD= 0.38), response time (M=4.58, SD=1.83), sentence completion (M=4.25, SD=2.41), slang (M=3, SD=1.60), punctuation (M=6.34, SD=0.49), interjection (M=6.67, SD=0.49), word transformation (M=6.83, SD=0.38), delivery (M=6.17, SD=0.72), hedging (M=6.17, SD=0.72), back-channeling (M=6.18, SD=0.39), abbreviation (M=6.67, SD=0.49), emotion (M=6.17, SD=0.72), euphemism (M=3, SD=0.60), split sentence (M=3.38, SD=1.47), and sentence structure (M=4.17, SD=1.64).

To define the detailed differences among the linguistic factors, we also conducted Tukey HSD to investigate the differences among the factors. Table 6.7 shows the results from Tukey HSD. We only listed the significant factors since these are the results of our interest.

From the results, we found that the linguistic factors differently affected the persona perception of MimicTalk. Word-transformation largely affected the persona perception followed by the abbreviation, interjection, wake-up word, slang, punctuation, back-channeling, delivery, hedging, euphemism, emoji, response time, structure, and split sentence.

To explore why a family member's dialogues affected the user's perception of the chatbot's personality, we asked an additional questionnaire of "how linguistic factors affected their perception of the chatbot's personality." We analyzed the user's answers and categorized the linguistic factors into the themes that have already been defined in the previous study [198]. Some participants using MimicTalk who reported that they perceived the chatbot as their family member insisted that these factors strengthened their tendency of making responses to MimicTalk as if they were responding to their actual family member. For example, P5 mentioned that *"There is only one person in the world who calls me dad, but now the chatbot is calling me dad. That makes me feel like I am talking to my daughter."* and P1 said that *"I am sure this sentence is written by my son since there exist frequent interjections and abbreviations that he frequently uses. I think the chatbot is* *doing its work in the right way which is to deliver my son's messages.".*

Comparison	Difference	Lwr	Upr	p.adj
Euphemism-Abbreviation	-2.83	-4.35	-1.32	0.00***
Response time-Abbreviation	-2.08	-3.60	-0.57	0.00***
Sentence completion-Abbreviation	-2.42	-3.93	-0.90	0.00***
Sentence struture-Abbreviation	-2.50	-4.02	-0.98	0.00***
Euphemism-Back-channeling	-2.33	-3.85	-0.82	0.00***
Response time-Back-channeling	-1.58	-3.10	-0.07	0.03*
Sentence completion-Back-channeling	-1.92	-3.43	-0.40	0.00***
Sentence struture-Back-channeling	-2.00	-3.52	-0.48	0.00***
Euphemism-Delivery	-2.33	-3.85	-0.82	0.00***
Response time-Delivery	-1.58	-3.10	-0.07	0.03*
Sentence completion-Delivery	-1.92	-3.43	-0.40	0.00***
Sentence struture-Delivery	-2.00	-3.52	-0.48	0.00***
Euphemism-Emoji	-2.00	-3.52	-0.48	0.00***
Sentence completion-Emoji	-1.58	-3.10	-0.07	0.03*
Sentence struture-Emoji	-1.67	-3.18	-0.15	0.02*
Euphemism-Emotion	-2.17	-3.68	-0.65	0.00***
Sentence completion-Emotion	-1.75	-3.27	-0.23	0.01**
Sentence struture-Emotion	-1.83	-3.35	-0.32	0.00***
Hedging-Euphemism	2.33	0.82	3.85	0.00***
Interjection-Euphemism	2.83	1.32	4.35	0.00***
Punctuation-Euphemism	2.50	0.98	4.02	0.00***
Slang-Euphemism	2.50	0.98	4.02	0.00***
Wake-up word-Euphemism	2.50	0.98	4.02	0.00***
Word transformation-Euphemism	3.00	1.48	4.52	0.00***
Response time-Hedging	-1.58	-3.10	-0.07	0.03*
Sentence completion-Hedging	-1.92	-3.43	-0.40	0.00*
Sentence struture-Hedging	-2.00	-3.52	-0.48	0.00*
Response time-Interjection	-2.08	-3.60	-0.57	0.00***
Sentence completion-Interjection	-2.42	-3.93	-0.90	0.00***
Sentence struture-Interjection	-2.50	-4.02	-0.98	0.00***
Response time-Punctuation	-1.75	-3.27	-0.23	0.01**
Sentence completion-Punctuation	-2.08	-3.60	-0.57	0.00***
Sentence struture-Punctuation	-2.17	-3.68	-0.65	0.00***
Slang-Response time	1.75	0.23	3.27	0.01**
Wake-up word-Response time	1.75	0.23	3.27	0.01**
Word transformation-Response time	2.25	0.73	3.77	0.00***
Slang-Sentence completion	2.08	0.57	3.60	0.00***
Wake-up word-Sentence completion	2.08	0.57	3.60	0.00***
Word transformation-Sentence completion	2.58	1.07	4.10	0.00***
Slang-Sentence struture	2.17	0.65	3.68	0.00***
Wake-up word-Sentence struture	2.17	0.65	3.68	0.00***
Word transformation-Sentence struture	2.67	1.15	4.18	0.00***

Table 6.7 Significant differences among linguistic factors affecting user's perception of the chatbot's personality.

Additional Opportunities and Challenges

Family involvement in the chatbot's dialogue design increased the intensity of the user's data logging behavior by showing high participation in logging images and contextual data that gives additional information of the user's eating behavior including reasons why they skipped the meal, with whom they had their meal with, or how they cooked their meal, etc. Also, another interesting point we found is that most participants who used MimicTalk responded to the chatbot's prompts as if they were responding to humans. In other words, they were more likely to respond in full sentences and reply to messages that were not directly related to the food journaling behavior itself (e.g. frequently responded to the chatbot's farewell messages.) Moreover, within such journaling-irrelevant messages, emotional words that imply feelings and emotional expressions, such as emojis and endearments, were frequently found.

We also explored the opportunities and challenges of MimicTalk with the additional questionnaire of "how linguistic factors affected their perception of the chatbot's personality." in the post-interview. We conducted thematic analysis [135] to build up the themes from the post-interview. As a result, we got 3 major opportunities and three major challenges for applying the persona of the user's family member to the healthcare chatbot with the function of food journaling.

When it comes to MimicTalk for food journaling, benefits for data collection are the first opportunity reported from participants. As being described in 5.3.2 (user experience with healthcare chatbots), MimicTalk showed higher opportunities in collecting image-based dietary data and contextual information. In this study, we focused on three types of data (text, image, and contextual information) to collect the user's food blog. In food journaling, all of these data types, each of which indicates the intensity of the system use, are known to be beneficial to tracking one's eating behavior [173]. Family like chatbot's dialogue was found to increase the intensity of the user's data logging behavior by showing high participation in logging (1) images, which enable inferring the quantity of food, detecting ingredients, or calculating calories, and (2) contextual data that gives additional information of the user's eating behavior including reasons why they skipped the meal, with whom they had their meal with, or how they cooked their meal, etc.

The second benefit tackled by participants was its effectiveness for taking over the repetitive tasks of healthcare providers. Some participants insisted that whenever they are in negative health status, they feel worried that their family could feel burden by worrying their parents. Actually, they more frequently ask them about dietary intake and other health-related issues. Participants insisted that MimicTalk can take over the burden of their family members. At this point, we were worried that the chatbot taking over the tasks of family members could end in reduced family interactions. However, some participants said that keeping track of everyday diet log will not interfere children with asking for a meal to say hello. Also, they said that they do not bother being asked by both the chatbot and their actual family member.

Third, most participants agreed that they were engaged in food journaling with MimicTalk. Few participants who used basic chatbots also reported they felt engaged with the chatbot over time. However, these reports were relatively low compared to the user group who used MimicTalk.

There also exist challenges of MimicTalk for food journaling that can be categorized as eerie feeling coming from structured dialogue flow, psychological burden coming from the family's persona, and the privacy issue in the storage of collected data. When it comes to eerie feelings coming from structured dialogue flow, most participants who used MimicTalk and the basic chatbot reported that they noticed the chatbot's dialogue being repeated. Particularly in the user group who used MimicTalk, some participants insisted that the repeated structure of MimicTalk made them feel uncomfortable feelings by making the MimicTalk's chatbot more robot-like who is mimicking their family member.

The second challenge raised by most participants is the increased psychological burden in responding to the MimicTalk. Some participants who used MimicTalk insisted that they feel sorry or guilty when they are ignoring or forgetting to respond to the MimicTalk who resemble their family members. P4 said that ”Every time I forget to respond to the chatbot, I feel sorry since I feel like I did not respond to my real daughter.”

The third challenge raised by some participants is a privacy issue. Even though we instructed participants that the dialogue between MimicTalk and the participant would be only used for the analysis of the study, some participants expected the dialogue between MimicTalk and them would be also delivered to their real children. At the end of the experiment, P1 worried and said ”I talked a lot to the chatbot regardless of given tasks since MimicTalk made me remind of my son. And now I am worried if somebody else I do not know would see my conversation.” She also asked the authors if the conversation would be eliminated after analysis. We remind participants about our privacy policy at the start and at the end of the experiment based on the procedure of IRB approval.

6.5 Experiment 2: Physical Activity Intervention

In the domain of physical activity, CA for intervening physical activities of the users has long been explored. Particularly the approaches of gamified, text-message, social media interventions have long been implemented in the domain of physical activity.

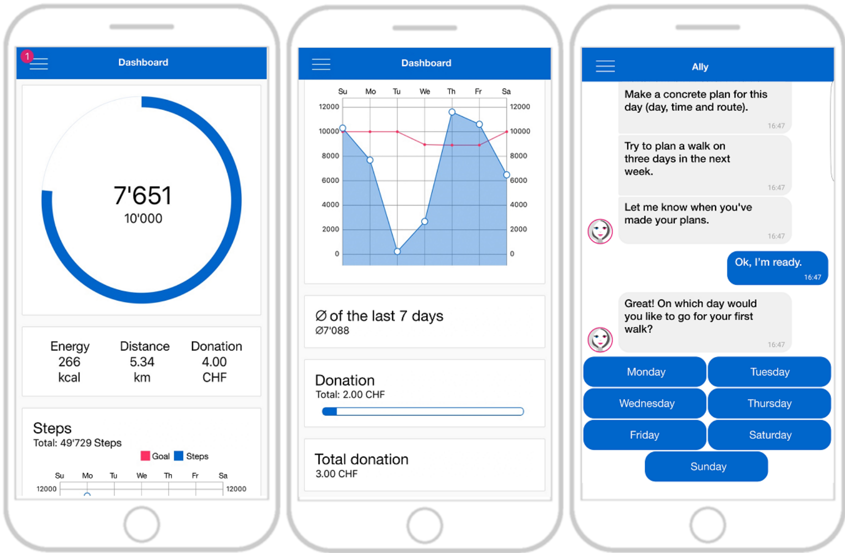


Figure 6.5 The Ally app with the chatbot for supporting user's physical activity

We present three studies that used a digital device for physical intervention as examples of gamified interventions. The first experiment was carried out with patients of an endocrinology clinic in Canada [199]. This intervention is based on a stationary bike and a cycling video game. The experimental group more participated and spent more time doing the physical activity than the control group. Two video games were used in the second example intervention to increase knowledge about energy balance [200]. The third study pitted two popular physical activity apps against a control group and was told them to use them three times per week [201]. There are also text-based messaging, social media, and email-based interventions available for physical activity.

There have also been previous studies that looked at the effects of using a chatbot for physical activity intervention. One paper discusses the effects of chatbot Ally [202] that showed effectiveness in supporting users to achieve personalized daily physical activity goals. Findings from the study revealed that Ally's daily rewards increased goal achievement increased engagement. However, over time, one-third of participants stopped using the app due to low user engagement. Another RCT study [203] which assess the chatbot for lifestyle coaching showed that after 12 weeks of intervention, this chatbot was effective in increasing physical activity among office workers. In addition, one feasibility study [204] tested Tess, a behavioral coaching chatbot, to see if it could help adolescent patients manage their weight and diabetes-related symptoms. Patients rated the chatbot as a helpful and engaging medium supporting them to reach their goals [204]. Reflection Companion, a chatbot, visualizes physical activity graphs for self-reflection. Findings from the study showed that the Reflection Companion successfully elicited self-reflection, resulting in increased motivation and frequency of physical activity.

However, only a few studies have focused on creating a healthcare chatbot persona for physical activity intervention. In the second experiment, we'll look at the benefits and drawbacks of using a user's family member's persona in a chatbot for physical activity intervention using a dialogue style. We'll also look into how linguistic factors influence user engagement with our chatbot when it comes to physical activity.

6.5.1 Method

To investigate the effectiveness of MimicTalk for physical activity intervention, we conducted the experimental study for 21 days. 24 users without any severe diseases participated in the study. They were randomly assigned to the group using the MimicTalk with the conversational style of their family member (MimicTalk) and the control group using the basic chatbot (Basic bot) without the conversational style of a family member. Table 6.8 shows information of participants in the evaluation study of experiment 2.

Wizard of Oz Study to Investigate the MimicTalk for Physical Activity Intervention.

As previous chatbots for physical intervention accompany multiple functions such as data visualization, graphs, detailed energy consumptions, etc (as in Figure 6.5 showing Ally), we could not deploy all functions of healthcare chatbot for physical activity. Therefore, we only selected limited functions to implement to our chatbot that is "setting goals for physical activity" and "checking goals and cheering up users to achieve the goals". Since goals for physical activity and how they achieve goals vary depending on users' characteristics, we did not use structured dialogue flow to design the chatbot.

Unlike Experiment 1 designing the food journaling chatbot based

Participant No.	Type of chatbot used	Sex	Age	Occupation	Family participant
P1	MimicTalk	Male	30	Student	Wife
P2	MimicTalk	Male	60	Office worker	Daughter
P3	MimicTalk	Female	33	Teacher	Husband
P4	MimicTalk	Female	30	Doctor	Husband
P5	MimicTalk	Male	59	Office worker	Daughter
P6	MimicTalk	Female	28	Office worker	Younger sister
P7	MimicTalk	Male	35	Office worker	Wife
P8	MimicTalk	Male	31	Researcher	Mother
P9	MimicTalk	Female	50	House spouse	Daughter
P10	MimicTalk	Female	45	Teacher	Son
P11	MimicTalk	Male	55	Office worker	Daughter
P12	MimicTalk	Female	47	Office worker	Son
P13	Basic chatbot	Female	30	House spouse	
P14	Basic chatbot	Female	57	House spouse	
P15	Basic chatbot	Male	33	Office worker	
P16	Basic chatbot	Female	27	Student	
P17	Basic chatbot	Female	54	Teacher	
P18	Basic chatbot	Female	32	House spouse	
P19	Basic chatbot	Male	50	Office worker	
P20	Basic chatbot	Male	35	Office worker	
P21	Basic chatbot	Female	53	Office worker	
P22	Basic chatbot	Female	28	Office worker	
P23	Basic chatbot	Male	50	Office worker	
P24	Basic chatbot	Female	50	House spouse	

Table 6.8 Information of participants included the experiment 2.

on the defined structure, we used Wizard of Oz method to investigate our idea of applying the persona of a family member to the healthcare chatbot for physical activity. However, still, we collected the family member's dialogue data to design the dialogue of the chatbot and used it during the study. Since physical activity intervention requires individualized dialogue structure, we used Wizard of Oz method to explore the effectiveness of MimicTalk. Wizard of Oz method [88, 89] has been frequently used for building prototypes of intelligent agents. The method is a rapid-prototyping method for systems costly to build or requiring development with new technology. In this method, a human plays the role of "Wizard" that simulates the system and the "Wizard" interacts with the user through a computer or mock-up system. Most Wizard of Oz tests or experiments establishes the viability of a futuristic approach to interface design. An example could be a speech-based intelligent system, and also synchronous text-based agents. The method is appropriate for evaluating the function of the system but also effective in analyzing human behavior toward a particular system [90].

Designing Dialogue of MimicTalk

To design dialogues that would be used in the Wizard of Oz study, we collected the dialogue style of the participant's family members (i.e.host whose persona would be implemented to MimicTalk) prior to the experiment. In this experiment, since we cannot fully use dialogue collected from the user's family member due to the flexible dialogue structure through Wizard of Oz, we collected raw conversational data from participants and used them for reference. Data include a conversation between the participant and the host. With the conversational data, we analyzed the conversation based on linguistic characteristics defined in study 2.

When asking participants for sharing conversational data with their family members with authors, we first asked them to at least share the conversational data with 50 turn-taking messages. While doing this, participants can eliminate or delete messages that they do not want to share with authors.

The dialogue style of each participant's family member was defined in Table 6.9. The presented sheet is the example of P7. This sheet was created for each participant and used as instruction when conducting Wizard of Oz study. The impact of each linguistic factor was defined by the authors. Examples for each factor are being described except for some factors that cannot be defined.

To compare MimicTalk with the chatbot without a family member's persona, we also made an instruction sheet for the basic chatbot too. We tried to moderate the impact of each factor to the middle level. And the dialogue style follows a formal tone as the basic chatbot used in experiment 1. As previous chatbots for physical intervention accompany multiple functions such as data visualization, graphs, detailed

Linguistic factors	Impact	Examples
Wake-up word	High	Honey ❤️, My love❤️
Response time	Low	Average response time within 10 minutes
Sentence completion	Medium	Don't know, maybe?.just kidding,you?
Slang	Low	None used
Punctuation	High	!!!!!!!!!!!!!!
Interjection		Wow, yeah~
Word transformation	Medium	Ahhhhh~,been 2 the gym, have a nice dayyyyyy~~
Delivery	Hard to define	
Hedging	Medium	Certainly not, additionally,possibly
Back-channeling	Medium	Uhmm... ha....ahh....
Abbreviation	High	OMG ! Lol😝, LMK, ASAP, BTW
Emotion	High	Love you!,I feel very happy,
Euphemism	Hard to define	
Structure	Low	I did maybe,coffee you want?

Table 6.9 Instruction sheet of dialogue style of family member

energy consumptions etc (as in Figure 6.6 showing Ally), we could not utilize the previous structure of healthcare chatbot for physical activity. Therefore, we only selected limited functions to implement to our

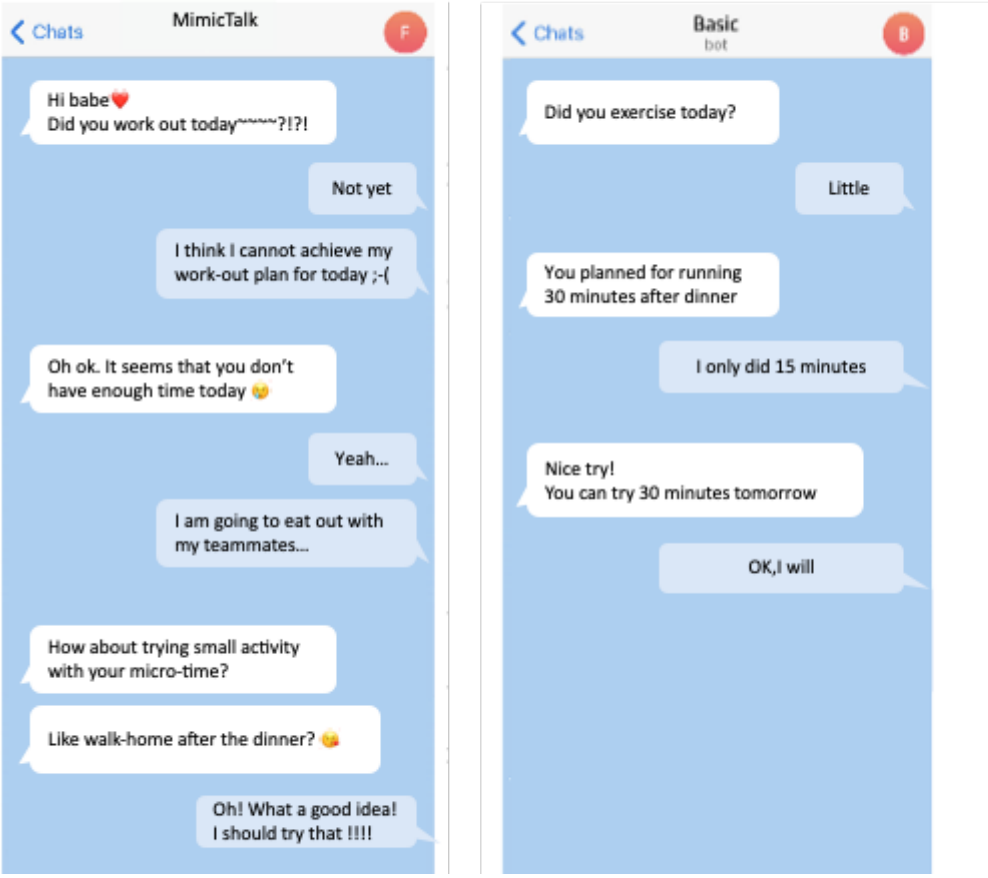


Figure 6.6 Example conversation made with MimicTalk and the basic chatbot

chatbot that are 1) setting goals for physical activity and 2) checking goals and cheering up users to achieve the goals".

Figure 6.6 shows a sample conversation with MimicTalk and the basic chatbot.

Evaluation

With the MimicTalk and the basic chatbot, we conducted a Wizard of Oz based evaluation study for 21 days with 24 participants divided into two groups to investigate the effectiveness of MimicTalk on user engagement in physical activity. We randomly assigned participants into two groups including an experimental group using a MimicTalk, which utilizes the dialogue style of their family member, and the control group using a basic chatbot, whose dialogue style was designed by authors based on formal dialogue style.

Table 6.8 shows the information of participants including their ID, type of chatbot each participant used, sex, age, occupation. For the participants who used MimicTalk during the experiment, the relationship with the participant whose dialogue style was applied to MimicTalk is also shown in Table 6.8. Overall, participants' age ranged from 27 to 60. In this study, we recruited various types of family relationships. Also, participants suffering from any severe diseases (e.g. cancer, diabetes, etc) were excluded from our study because they need medical support, prior to managing eating behaviors. They all had been building high perceived emotional attachment with their family member. Participants received 150K KRW (approximately 125 USD) for 21 days of participation. Prior to the experiment, we constructed a semi-structured dialogue flow with the dialogue style sheet (Table 6.9). The whole experiment was based on Telegram mobile messenger.

At the start day of deployment, we informed participants about our research goal and the function of our chatbots and all participants answered the pre-survey. Additionally, participants who used MimicTalk answered the questionnaires about their family members including their age, how frequently they see each other, and perceived emotional attachment with their family member on a 7-point Likert scale with the questionnaire of "In recent years, I feel emotionally attached to my family member" (answering option ranged from "Strongly agree" to "Definitely not agree".) This was to filter out the weak relationship since we are focusing on the close relationship to apply to MimicTalk. All the participants who used MimicTalk reported high emotional attachment to their family member(M= 5.67, SD = 1.31).

To manage user expectations, we conducted 30 minutes of detailed introduction sessions before starting the experiment. In the session, we shared our research goal and research questions. Then, we specifically introduced the functionality of the chatbot and its limitations coming from the study design. We not only told them they were recruited to evaluate the user experience and user perception of MimicTalk (or basic chatbot) and told them they are going to interact with the chatbot with the tasks of checking physical activity. We emphasized that the chatbot's functionality is limited in intervention tasks and the chatbot would be limited in responding to out-of-task conversations. Then, we showed the sample conversation of the chatbot in order to help participants be accustomed to the workflow of the chatbot. Participants were instructed to freely ask any questions regarding the chatbot and the experiment to the author. Through these processes, we made it sure for users to manage their expectations with the chatbot. After participants got used to the function of the chatbot, they were asked to freely use the chatbot for 21 days.

During the experiment, all participants' conversational data with the chatbot were recorded for analysis to analyze the user engagement with MimicTalk over time. We not only measured the responses to physical activity intervention but also any responses made regardless of healthcare task.

At the end of the deployment study, participants answered the post-survey and the post-hoc interview. Post survey included the questionnaire of user experience and user perception in Table 6.2 and Table 6.3. Answers to the pre/post survey were all collected with the 7-point Likert scale (answering options ranged from "Strongly agree" to "Definitely not agree".)

To explore the major findings from the conversational logs we collected, a semi-structured interview was conducted after the deployment. During the post-hoc interview, we aimed to explore the user engagement with the chatbot and how family members' dialogue style affected users' engagement and perception of the chatbot over time. We used the same survey questionnaires among experiments 1,2 and 3 to compare the results among these experiments. Surveys include user experience questionnaires and questionnaires about user perception.

6.5.2 Results

We present our results from the Wizard of Oz-based evaluation study regarding user engagement with the chatbot by comparing two types of chatbots (i.e. MimicTalk and basic chatbot) based on the data log of 24 participants collected during 21 days of experimental study, pre/post survey and interview. For the statistical analysis, we used the one-way ANOVA test.

User Experience with MimicTalk for Physical Activity

We first present the overall user experience with MimicTalk for physical activity collected through the post-experiment survey. Survey questionnaires include behavior change, usefulness, ease of use, trustfulness, intimacy, engagement, attachment, human likeness, and eerie feeling.

We measured user experience with the MimicTalk and the basic chatbot. The mean and the standard deviation of UX factors of MimicTalk evaluated by participants are as following: user's motivation to participate in the healthcare behavior (pre (M= 2.87, SD = 1.95) and post (M= 5.25, SD = 0.89)), usefulness (M= 6.5, SD = 0.76), ease of use (M= 6.62, SD = 0.52), trustfulness (M= 5.25, SD = 1.03), intimacy (M= 6.5, SD = 0.54), engagement (M= 6.4, SD = 0.52), attachment (M= 6, SD = 0.54), human-likeness (M= 6.38, SD = 0.74),

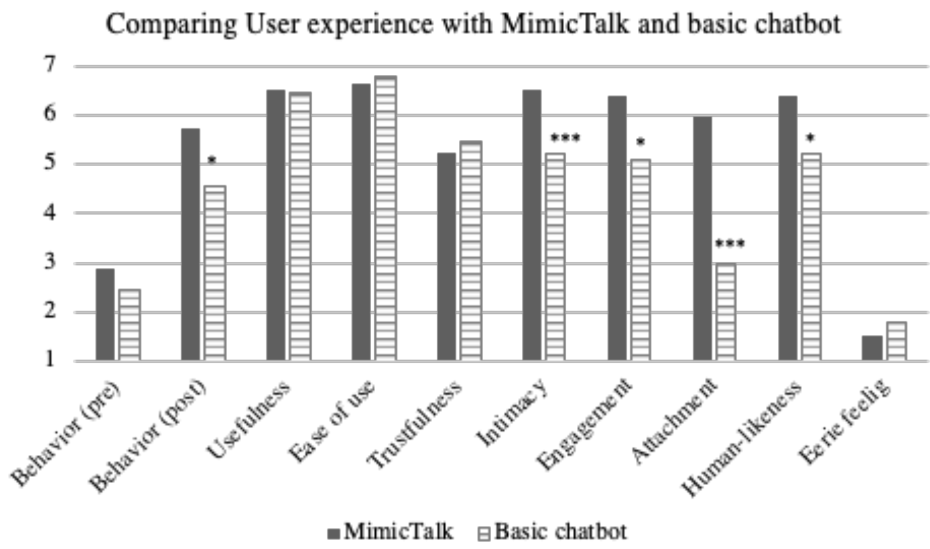


Figure 6.7 Comparing user experience with MimicTalk and basic chatbot for physical activity

and eerie feeling (M= 1.5, SD =1.07).

Also, the mean and the standard deviation of UX factors of basic chatbot evaluated by participants are as following: user's motivation to participate in the healthcare behavior (pre (M= 2.44, SD = 1.14) and post (M= 4.56, SD = 1.34)), usefulness (M= 6.45, SD = 0.53), ease of use (M= 6.78, SD = 0.44), trustfulness (M= 5.45, SD = 1.34, intimacy (M= 5.23, SD = 1.09), engagement (M= 5.11, SD = 1.67), attachment (M= 3, SD = 1.22), human-likeness (M= 5.22, SD = 1.20), and eerie feeling (M= 1.17, SD =0.66).

Among these factors, there was significant difference between groups (MimicTalk and basic chatbot) in the post behavior(p < 0.05), intimacy ($p < 0.05$), engagement ($p < 0.05$), attachment ($p < 0.0001$), and human-likeness ($p < 0.05$).

Physical activity-related topic participants made with MimicTalk:

To explore the effectiveness of MimicTalk in physical activity intervention, we first compared the conversation topic between MimicTalk and the basic chatbot. A major prompt made by both chatbots are asking about physical activity goal and their achievement, Most of the conversation was focused on those topics. However, detailed information and their motivation reflected on the conversation, reflection on failure, and reflection on physical changes differed. The examples of the conversational logs showing these differences are being described in Figure 6.9.

For detailed information, participants who used MimicTalk had the tendency to explain their physical activity-related behaviors. Most explanations made with MimickTalk with the first topic are about the type of exercise participants did. Some participants explained whether they did anaerobic exercise or aerobic exercise, Most cases explained the categories of sports such as climbing, tennis, golf, and swimming. In some cases, contextual data was also collected including whom they did an exercise with, when, where, how much they did the exercise.

For keeping motivation, participants who used MimicTalk tend to

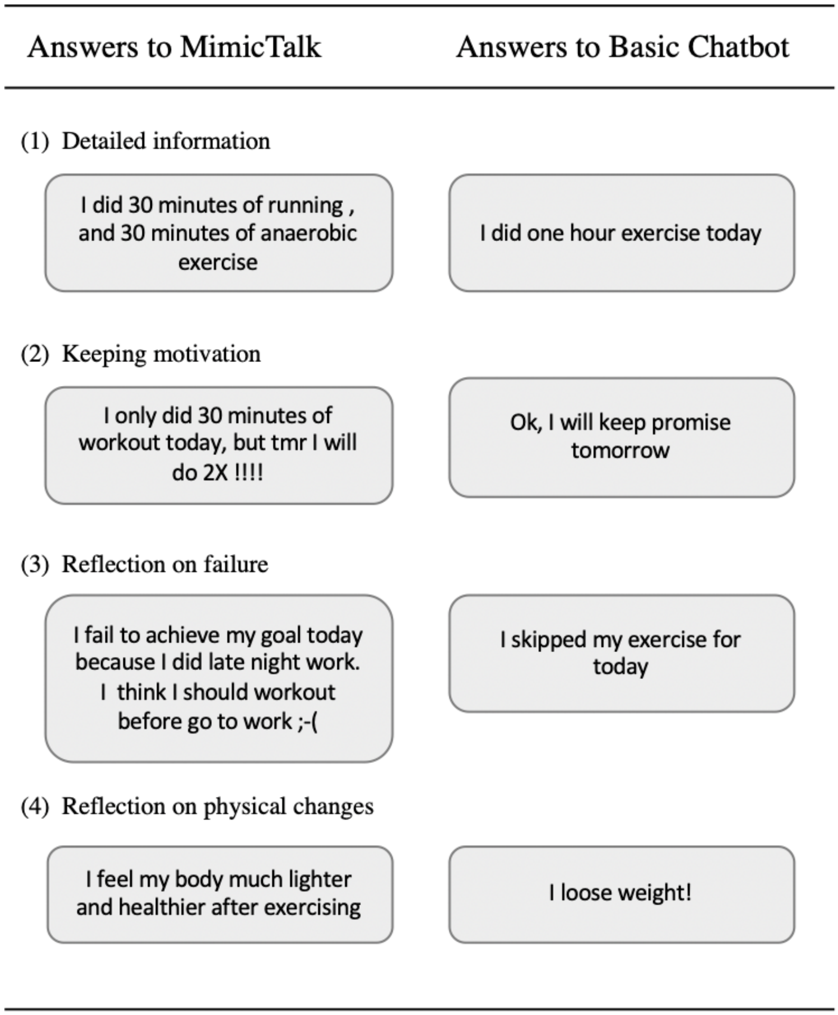


Figure 6.8 Example messages showing the difference between the quality of answers to the MimicTalk and the basic chatbot.

share their motivation with the chatbot compared to the basic chatbot. Since we did not measured the user motivation, we can not confirm that participants who used MimicTalk had higher motivation than participants who used the basic chatbot. But still, we can say that participants share their motivational behavior more often to the MimicTalk compared to the basic chatbot.

For reflection on failure, participants who used MimicTalk more frequently share their thoughts toward their failure of keeping physical activity goals while participants who used the basic chatbot tend to simply share the results of success or failure. While reflecting on their failure with the MimicTalk, participants frequently share their feelings and reasons for failure with their excuses.

For reflection on physical changes, participants who used the MimicTalk shared their detailed information of body change compared to participants who used the basic chatbot. Some participants even shared the detailed values of their body index including weight, waist circumference, fat, and muscle percentages.

User Perception with MimicTalk

We also explored how users perceive the MimicTalk. Particularly, we focused on how linguistic factors we defined in Study 2 affected the persona perception of the chatbot for physical activity intervention. To explore how the user perceives the personality of MimicTalk with the dialogues of a family member, we asked participants how they define the personality of the MimicTalk and What factors affected their perception. These questionnaires are identical to the questionnaires used in experiments 1 and 3.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Linguistic factors	14	108.3	7.736	4.583	5.59e-07 ***
Residuals	165	278.5	1.688		

Table 6.10 Linguistic factors affecting user's perception of the chatbot's persona in tasks of physical activity intervention

Also, We defined to what extent each linguistic factor affected the persona perception of MimicTalk as their real family member. Table 6.10 shows results from one-way ANOVA. There was significant difference among linguistic factor affecting the persona perception of MimicTalk. The mean values and the standard deviations of each linguistic factor are as following: wake-up word (M= 6.34, SD= 0.89), emoji (M= 5.75, SD= 1.14), response time (M=5.17, SD=1.33), sentence completion (M=5.92, SD=0.99), slang (M=3.58, SD=1.71), punctuation (M=5.92, SD=0.49), interjection (M=5.58, SD=1.62), word transformation (M=6.34, SD=0.49), delivery (M=6.08, SD=0.90), hedging (M=5.25, SD=1.54), back-channeling (M=4.91, SD=1.83), abbreviation (M=6.41, SD=0.99), emotion (M=4.33, SD=1.67), euphemism (M=3.58, SD=1.44), split sentence (M=3.38, SD=1.47), and sentence structure (M=6.16, SD=0.93).

To define explore the detailed differences among the linguistic factors, we also conducted Tukey HSD to investigate the differences among

Comparison	Difference	Lwr	Upr	p.adj	Signifiance
Emotion-Abbreviation	-2.08	-3.90	-0.26	0.01	**
Euphemism-Emoji	-2.17	-3.99	-0.34	0.01	**
Wake-up word-Emotion	1.99	0.17	3.82	0.02	*
Word transformation-Emotion	2.00	0.17	3.82	0.02	*
Sentence completion-Euphemism	2.33	0.51	4.15	0.00	***
Sentence struture-Euphemism	2.58	0.76	4.40	0.00	***

Table 6.11 Differences among linguistic factors affecting user's perception of the chatbot's persona in experiment 2

the factors. Table 6.11 shows the results from Tukey HSD. We only listed the significant factors since these are the results of our interest. As in table 6.11, there were significant differences between emotion and abbreviation, euphemism and emoji, wake-up word and emotion, word transformation and emotion, sentence completion and euphemism, and sentence structure and euphemism.

From the results, we found that the linguistic factors also differently affected the persona perception of MimicTalk in the domain of physical activity. Abbreviation largely affected the persona perception followed by back-channeling, punctuation, emotion, interjection, wake-up word, hedging, word transformation, slang, sentence structure, emoji, sentence completion, response time, delivery, and euphemism.

Additional Opportunities and Challenges

As we did for experiment 1, we also explored the opportunities and challenges of MimicTalk for physical activity intervention with the additional questionnaire of "how linguistic factors affected their perception of the chatbot's personality." and "how the persona of MimicTalk affected their physical activity behaviors." in the post-interview. We conducted thematic analysis [135] to build up the themes from the post-interview. As a result, we got 3 major opportunities and three major challenges for applying the persona of the user's family member to the healthcare chatbot with the function of food journaling.

When it comes to MimicTalk for physical activity intervention, benefits for customized framing effects are the first opportunity reported from participants. Some participants who used MimicTalk said that sharing their failure on physical activity goals made them feel guilty and feeling shame to their family partners since they feel like talking to their family members while talking to the MimicTalk. On the other hand, some participants insisted that the MiimicTalk amplifies their responsibility to achieve the physical activity-related goals. Some participants even said that they feel sorry for their family members since they felt like breaking the promise with their family members.

The second benefit tackled by participants was its effectiveness in increasing the acceptance of information. In some cases, participants asked the chatbot about the activity recommendation or effectiveness of a particular exercise. In this case, we referred to academic journals or well-known articles to give them information. For exercise recommendation, participants who used MimicTalk were willing to accept the exercise suggestion. On the other hand, participants who used the basic chatbot also agreed with the chatbot's recommendation or opinion, but we found that they did not actually take action.

Third, most participants who used MimicTalk agreed that MimicTalk helped them keep motivated. Most participants agreed that the first and the second benefits also influenced their motivation. Additionally, The feeling of being seen by family members was reported to keep them motivated.

Moreover, when it comes to the challenges of MimicTalk, low trust in information that requires expertise was reported by most participants. Since most participants had the experience of watching physical activity content or being trained by experts, they insisted that their family member is not an expert on physical activity. Some participants said they are willing to accept their suggestions but they do not fully trust the recommendation by MimicTalk. When it comes to trust in expertise information, most participants said that applying their family members is not the best option. Most participants who used MimicTalk recommended us to apply the persona of a personal trainer or well-known expert.

The second challenge addressed by participants was decreased attachment when the relationship is bad. Some participants who use MimicTalk insisted that they really liked interacting with the MimicTalk when they were in a good relationship. However, they do not want to use it when they are in trouble. Fortunately, none of our participants had a bad relationship with the family members, but when choosing the family member as the persona of MImicTalk for physical activity intervention, it should be considered.

Lastly, collision with the actual host was mentioned. While using MimicTalk, some participants experienced the case that MimicTalk sent them messages while they are with their family member whose persona was applied to the MimicTalk. In this case, some participants felt like they were cheating with someone else. This was limited to the husband-wife relationship. However, since a collision with the actual host is very likely to happen if a family member and the user live together. In this case, it may be a good idea to avoid overlapping times or target people living apart.

6.6 Experiment 3: Chatbot for Coping Stress

Stress is considered to be the cause of various non-communicable diseases. Mostly, mental disorders are highly associated with stress [183]. When it comes to stress management, chatbot with the role of counseling has been mostly developed to relieve the stress of the user in a daily context. Therefore we implemented the healthcare task of stress counseling to the domain of stress management.

Wysa, for example, is stress, depression, and anxiety therapy chatbot that coaches stress management coping strategies. Wysa is an AI-powered chatbot that is designed and developed to help users manage their stress levels, unstable emotions, and negative. Evidence-based cognitive behavioral therapy (CBT) is dominantly used in the interaction with the Wysa. Sometimes it also recommends Dialectical Behavior Therapy, introduction to meditation, breathing for reducing stress, and yoga.

Woebot is another example of a chatbot that monitors users' moods and provides appropriate coping strategies through therapeutic conversations. During the conversation, users are encouraged to express their thoughts and emotions. The cognitive Behavioral Therapy platform (CBT) is also the dominant strategy of Woebot. Other than stress, it covers a wide range of mental health issues. The Woebot inquires about users' feelings and what's going on in their lives while discussing mental health and psychological well-being. To support users to cope with stress in a daily manner, it sends personalized videos, URLs, content based on the user's current mood.

Moodkit is a mobile application that helps users cope with stressors and alleviates the symptoms of mental illness. Users can use Moodkit to identify and change negative thought patterns through activities. It also has tools for rating and charting moods over time, as well as a text-based journaling tool.

Joy is another example, as it is based on two platforms: Facebook Messenger and Slack. It used a variety of channels to reach out to more users who were already using these apps. Joy, like other stress-relieving and mental-health apps, sends a message to the user once a day to see how they're doing and what they've planned for the day. Joy interprets the user's current mood contents and provides personalized ways and solutions to promote positive emotions by detecting mood-related words or phrases in the conversation.

Chatbots for stress management, like other chatbots for lifestyle management, face challenges with long-term adherence and effectiveness. Personalized care via chatbot is also one of the major challenges, as defining the user for personalized care is difficult, and evaluating user behaviors is also difficult. We also implemented my idea of using the persona of someone who is in a real-world relationship with the user in the domain of stress management.

6.6.1 Method

To investigate the effectiveness of MimicTalk for stress management, we conducted an experimental study for 21 days. 24 users without any severe diseases participated in the study. They were randomly assigned to the group using the conversational style of their family member (MimicTalk) and the control group using the basic chatbot (Basic bot) without the conversational style of a family member. Table 6.12 shows

Participant No.	Type of chatbot used	Sex	Age	Occupation	Family participant
P1	MimicTalk	Male	33	Office worker	Wife
P2	MimicTalk	Male	33	Office worker	Wife
P3	MimicTalk	Male	32	Actor	Younger sister
P4	MimicTalk	Male	32	Researcher	Older sister
P5	MimicTalk	Female	33	Researcher	Younger sister
P6	MimicTalk	Female	56	Office worker	Husband
P7	MimicTalk	Female	35	Office worker	Husband
P8	MimicTalk	Female	33	House spouse	Older sister
P9	MimicTalk	Female	48	Office worker	Husband
P10	MimicTalk	Female	50	Teacher	Daughter
P11	MimicTalk	Male	56	Office worker	Daughter
P12	MimicTalk	Female	25	Student	Younger brother
P13	Basic chatbot	Male	55	Office worker	
P14	Basic chatbot	Male	40	Teacher	
P15	Basic chatbot	Female	33	House spouse	
P16	Basic chatbot	Male	32	Doctor	
P17	Basic chatbot	Female	33	Teacher	
P18	Basic chatbot	Female	52	House spouse	
P19	Basic chatbot	Female	51	House spouse	
P20	Basic chatbot	Male	33	Office worker	
P21	Basic chatbot	Male	36	Researcher	
P22	Basic chatbot	Female	53	House spouse	
P23	Basic chatbot	Male	30	Office worker	
P24	Basic chatbot	Female	33	Office worker	

Table 6.12 Information of participants included in the experiment 3

information of participants in the evaluation study of expriment 3. The overall study design is almost the same as experiment 2.

Wizard of Oz Study to Investigate the MimicTalk for Stress Management.

Unlike other domains such as diet management and physical activity intervention, most previous systems for stress management are focusing on counseling. As in our literature review, chatbots are known to be the effective synchronous counselors for stress management. Among domains we chose for implementing MimicTalk, stress counseling requires the longest conversation and various strategies for personalized dialogue flow. Therefore, in this experiment, we did not define specific functions for MimicTalk for stress counseling. We expected that we could demonstrate the particular topics that could be tackled by MimicTalk for stress counseling. To do this we also used Wizard of Oz method to investigate our idea of applying the persona of a family member to the healthcare chatbot for stress management. However, still, we collected the family member's dialogue data to extract the conversational features that would be applied to MimicTalk.

Designing Dialogue of MimicTalk

To design dialogues that would be used in the Wizard of Oz based evaluation, we collected the dialogue style of participant's family members prior to the experiment. However, in this experiment, since we cannot fully use dialogue collected from the user's family member due to open, non-structured dialogue interaction through Wizard of Oz, we collected raw conversational data from participants. Collected data were conversations between participants and the host whose persona would be implemented to MimicTalk). With the conversational data, we analyzed the conversational data based on linguistic characteristics defined in study 2.

When asking participants for sharing conversational data with their family members with authors, we first asked them to at least share the conversational data with 50 turn-taking messages. While doing this, participants can eliminate or delete messages that they do not want to share with authors. Dialogue style of each participant's family member was pre-defined as done in experiment 2. This sheet was also created for each participants and used as instruction for experiment 3 when conducting Wizard of Oz study.

To compare MimicTalk with the chatbot without a family member's persona, we also designed a basic chatbot for physical activity intervention like prior experiments. We tried to moderate the impact of each linguistic factor as the medium for the basic chatbot. And the dialogue style follows a formal tone as the basic chatbot used in experiments 1 and 2. MimicTalk for stress management proactively asks the question of if the user had any stressed event or stressors to deal with such as "Is there anything that stressed you?" , " Any stressors for today?". Other than greeting messages, following dialogues from MimicTalk is depending on the user's answers. Figure 6.9 shows a sample

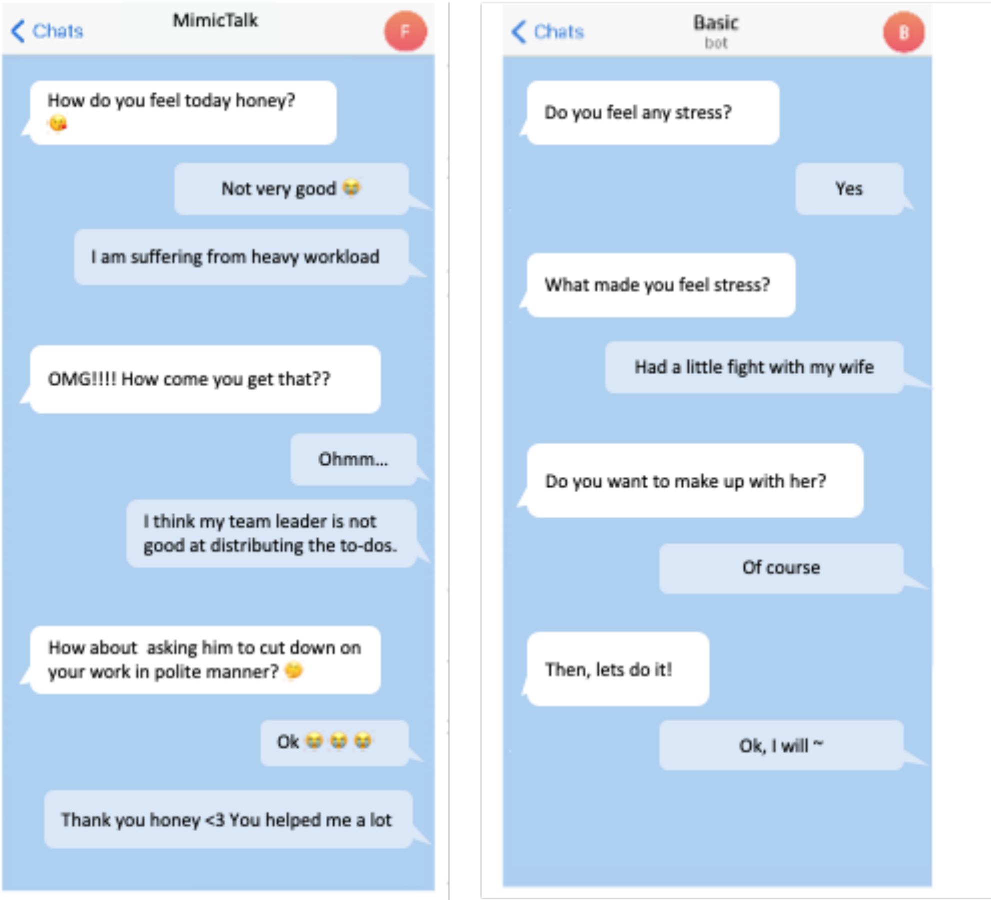


Figure 6.9 Example conversation made with MimicTalk and the basic chatbot

conversation with MimicTalk and the basic chatbot for stress counseling.

Evaluation

With the MimicTalk and the basic chatbot, we conducted a Wizard of Oz based evaluation study for 21 days with 24 participants divided into two groups to investigate the effectiveness of MimicTalk on user engagement in stress management. We randomly assigned participants into two groups including an experimental group using a MimicTalk, which utilizes the dialogue style of their family member, and the control group using a basic chatbot, whose dialogue style was designed by authors based on formal dialogue style. During the experiment authors synchronously made a conversation with participants at a particular time. We instructed participants that they can only have conversational with MimicTalk during the specific time span since we could not synchronously deal with conversations with 24 participants every day. The whole experiment was based on Telegram mobile messenger.

At the start day of deployment, we informed participants about our research goal and the role of our chatbots and all parent participants answered the pre-survey. Additionally, participants who used MimicTalk answered to the questionnaires about their family members including their age, how frequently they see each other, and perceived emotional attachment with their family member on a 7-point Likert scale with the questionnaire of "In recent years, I feel emotionally attached to my family member" (answering option ranged from "Strongly agree" to "Definitely not agree".) The emotional attachment was evaluated prior to the experiment since our inclusion criteria for selecting an appropriate persona is whose host is in a close relationship with the user. All the participants who used MimicTalk reported high emotional attachment to their family member ($M = 6.34$, $SD = 1.71$).

To manage user expectations, we conducted 30 minutes of detailed introduction sessions before starting the experiment. In the session, we shared our research goal and research questions. Then, we specifically introduced the functionality of the chatbot and its limitations coming from the study design. We not only told them they were recruited to evaluate the user experience and user perception of MimicTalk (or basic chatbot) and told them they are going to interact with the chatbot with the tasks of stress counseling. We emphasized that the chatbot's functionality is limited in counseling tasks and the chatbot would be limited in responding to out-of-task conversations. Then, we showed the sample conversation of the chatbot in order to help participants be accustomed to the workflow of the chatbot. Participants were instructed to freely ask any questions regarding the chatbot and the experiment to the author. Through these processes, we made sure for users to manage their expectations with the chatbot. After participants got used to the function of the chatbot, they were asked to freely use the chatbot for 21 days.

During the experiment, to analyze the user engagement with MimicTalk over time, all participants' conversational data with the chatbot were recorded for analysis. We not only measured the responses to physical activity intervention but also any responses made regardless of the task. What topics participants made with MimicTalk for stress management was also our research of interest in experiment 3.

At the end of the deployment study, participants answered the post-survey and the post-hoc interview. Post survey included the questionnaire of user experience and user perception in Table 6.2 and Table 6.3. Answers to the pre/post survey were all collected with the 7-point Likert scale (answering options ranged from "Strongly agree" to "Definitely not agree".)

To deeply explore the major findings from the conversational logs we collected, a semi-structured interview was conducted after the deployment. During the post-hoc interview, we aimed to explore the user engagement with the chatbot and how family members' dialogue style affected users' engagement and perception of the chatbot over time. We used the same survey questionnaires among experiments 1,2 and 3 to compare the results among these experiments. Surveys include user experience questionnaires and questionnaires about user perception.

6.6.2 Results

We present our results from the Wizard of Oz based evaluation study regarding user engagement in stress management by comparing two types of chatbots (i.e. MimicTalk and basic chatbot) based on the data log of 24 participants collected during 21 days of experimental study, pre/post survey and interview. For the statistical analysis, we used the one-way ANOVA test.

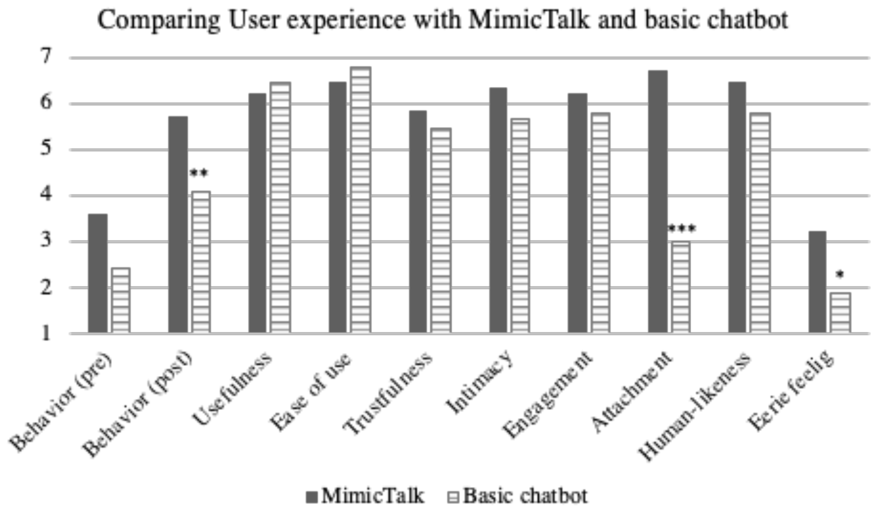


Figure 6.10 Comparing user experience with MimicTalk and basic chatbot for stress counseling

User Experience with MimicTalk for Stress Management

We first present the overall user experience with MimicTalk for stress counseling collected through the post-experiment survey. Survey questionnaires include behavior change, usefulness, ease of use, trustfulness, intimacy, engagement, attachment, human likeness, and eerie feeling.

We measured user experience with the MimicTalk and the basic chatbot. The mean and the standard deviation of UX factors of MimicTalk evaluated by participants are as following: user's motivation to participate in the healthcare behavior (pre (M= 3.62, SD = 1.84) and post (M = 5.75, SD = 0.89)), usefulness (M = 6.25, SD = 0.46), ease of use (M = 6.5, SD = 0.53), trustfulness (M = 5.87, SD = 0.99), intimacy (M = 6.37, SD = 0.74), engagement (M = 6.25, SD = 0.46), attachment (M = 6.75, SD = 0.46), human-likeness (M = 6.5, SD = 0.76), and eerie feeling (M = 3.25, SD =1.28).

Also, the mean and the standard deviation of UX factors of basic chatbot evaluated by participants are as following: user's motivation to participate in the healthcare behavior (pre (M = 2.44, SD = 1.13) and post (M = 4.11, SD = 1.05)), usefulness (M = 6.44, SD = 0.53), ease of use (M = 6.78, SD = 0.44), trustfulness (M = 5.45, SD = 1.34, intimacy (M = 5.66, SD = 1.58), engagement (M = 5.78, SD = 1.48), attachment (M = 3, SD = 1.22), human-likeness (M = 5.77, SD = 1.09), and eerie feeling (M = 1.89, SD =0.92).

Among these factors, there was a significant difference between groups (MimicTalk and basic chatbot) in the post behavior ($P < 0.01$), attachment ($P < 0.0001$), and eerie feeling ($P < 0.05$).

Stress-related topic participants made with MimicTalk:

In the domain of stress management, there was various topic made with participants. We conducted the thematic analysis to analyze the topics and categorized them into themes.

To explore the effectiveness of MimicTalk for stress counseling, we first compared the conversation topic between MimicTalk and the basic chatbot. Even though we did not define the specific topic in experiment 3, we still could categorize them into certain topics. This analy-

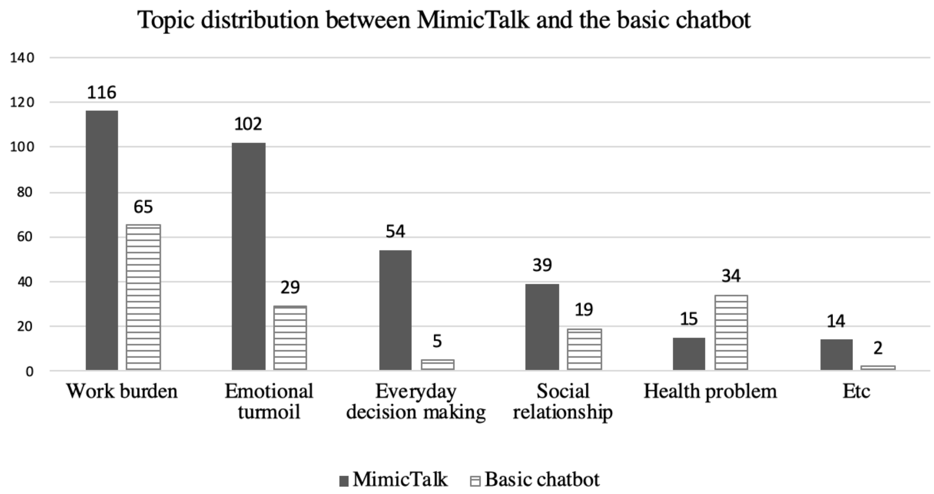


Figure 6.11 Example messages from users showing the difference between the MimicTalk and the basic chatbot.

sis showed a strong inter-coder agreement between the two researchers (Cohen's Kappa coefficient (κ) = 0.89.)

As in Figure 6.11, conversational themes include work burden, emotional turmoil, everyday decision making, social relationship, health problems, and others. To compare the difference between two groups of participants, we also compared the values for each group. Overall, conversation topics made with the MimicTalk include 340 sets and conversation topics made with the basic chatbot include 154 sets. Since there exists a big difference in the number of conversational data, we also calculated the overall portion of a particular conversation topic. For the MimicTalk the portion of each topic is as following: work burden (34.11%), emotional turmoil (30%), everyday decision making (15.88%), social relationship (11.47%), health problems (4.41%), and the others (4.11%). For the basic chatbot, the portion of each topic is as following: work burden (42.20%), emotional turmoil (18.83%), everyday decision making (3.23%), social relationship (12.33%), health problems (22.07%), and the others (1.29%).

User Perception with MimicTalk

Like prior experiments, we also explored how users perceive the MimicTalk. We focused on how linguistic factors we defined in Study 2 affected the persona perception of the chatbot for physical activity intervention. To explore how the user perceives the personality of MimicTalk with the dialogues of a family member, we asked participants how they define the personality of the MimicTalk and What factors affected their perception. These questionnaires are identical to the questionnaires used in experiments 1 and 2.

	Df	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
Linguistic factors	14	109.81	7.844	15.01	<2e-16 ***
Residuals	165	86.25	0.523		

Table 6.13 Linguistic factors affecting user's perception of the MimicTalk's persona during stress management tasks

Also, we defined to what extent each linguistic factor affected the persona perception of MimicTalk as their real family member. Table 6.13 shows results from one-way ANOVA. There was significant difference among linguistic factor affecting the persona perception of MimicTalk. The mean values and the standard deviations of each linguistic factor are as following: wake-up word (M= 6.5, SD= 0.67), emoji (M= 6.17, SD= 0.72), response time (M=5.66, SD=1.61), sentence completion (M=5.92, SD=0.90), slang (M=3.17, SD=1.77), punctuation (M=6.67, SD=0.65), interjection (M=6.58, SD=0.51), word transformation (M=6.41, SD=0.51), delivery (M=5.58, SD=1.16), hedging (M=6.5, SD=0.52), back-channeling (M=6.75, SD=0.45), abbreviation (M=6.83, SD=0.39), emotion (M=6.67, SD=0.49), euphemism (M=3.17, SD=1.40), and sentence structure (M=6.25, SD=0.45).

To define the detailed differences among the linguistic factors, we also conducted Tukey HSD to investigate the differences among the factors. Table 6.14 shows the results from Tukey HSD. We only listed the significant factors since these are the results of our interest. As in table 6.14, there were significant differences between emotion and abbreviation, euphemism and delivery, euphemism and emoji, wake-up word and emotion, word transformation and emotion, interjection and euphemism, sentence completion and euphemism, sentence structure and euphemism.

From the results, we found that the linguistic factors also differently affected the persona perception of MimicTalk in the domain of physical

Comparison	Difference	Lwr	Upr	p.adj	Significance
Emotion-Abbreviation	-2.08	-3.90	-0.26	0.01	**
Euphemism-Delivery	-2.50	-4.32	-0.67	0.00	***
Euphemism-Emoji	-2.16	-3.99	-0.34	0.01	**
Wake-up word-Emotion	1.99	0.17	3.82	0.02	*
Word transformation-Emotion	2.00	0.17	3.82	0.02	*
Interjection-Euphemism	2.00	0.17	3.82	0.02	*
Sentence completion-Euphemism	2.33	0.51	4.15	0.00	***
Sentence struture-Euphemism	2.58	0.76	4.40	0.00	***

Table 6.14 Differences among linguistic factors affecting user's perception of the chatbot's persona in experiment 3

activity. Abbreviation largely affected the persona perception followed by wake-up word, word transformation, sentence structure, delivery, sentence completion, punctuation, emoji, interjection, slang, hedging, response time, back-channeling, euphemism.

Additional Opportunities and Challenges

In this subsection, we are going to present the opportunities and challenges of MimicTalk for stress counseling based on the qualitative and quantitative findings. These findings are answers to the additional questionnaire of "how linguistic factors affected their perception of the chatbot's personality." and "how the persona of family members affected their physical activity behaviors." in the post-interview. We conducted thematic analysis [135] to build up the themes from the post-interview. As a result, we got 3 major opportunities and three major challenges for applying the persona of the user's family member to the healthcare chatbot with the function of stress management.

When it comes to MimicTalk for stress management, participants made the longest conversation compared to other domains. It might be due to its major function was counseling for daily stressors. Since the longest conversation was made with the MimicTalk for stress management, we found higher number of opportunities and challenges from the user interaction. Totally, we defined four major opportunities and four challenges from the interviews.

The first opportunity tackled by participants was opportunities for discussing coping strategies. In the interview, most participants said that they rarely deeply think about coping strategies for their stressors. P2 said that *"whenever I feel stress from my work, I just tried to forget about it but while talking to MimicTalk, I got the chance to think about personalized coping strategies."* Like this, participants who used MimicTalk came up with various coping strategies while making conversation with MimicTalk including making new hobbies, hanging out with loved ones, exercising, planning for a holiday, eating a favorite food, listening to favorite music, thinking about the realistic solution,s etc.

Second, proactively asking about current stressors is another opportunity of MimickTalk for stress management. Most participants who used both MimicTalk and the basic chatbot said that they had rarely been asked by someone about recent stressors. However, during using both chatbots, participants reported the feeling of being cared (MimicTalk:12/12. basic chatbot:3/12). However, for participants who used a basic chatbot, most participants reported that they felt the feeling of being cared for the first time, but they felt a rare emotional attachment to the basic chatbot. On the other hand, some participants who used MimicTalk said that the persona of their family members strengthened the presence of the receiver.

Third, MimicTalk also had the opportunity of provoking self-reflection. Most participants who used MimicTalk reported that they self-reflected themselves while talking with the MimicTalk. Some said that they could go through their attitude toward stressors, their corresponding behaviors, and their habits. Some of them insisted that they want to change their particular behavior of dealing with stressors. However, participants who used the basic chatbot also agreed that the chatbot has the opportunity of self-reflection at some part, but they were not willing to use the chatbot for deep talk. some of them even say that it's time-wasting talking to the non-human.

Fourth, MimicTalk has benefits in forming rapport. Most participants said that they are not familiar with talking about their stressors with strangers. However, participants who used MimicTalk insisted that the persona of their family members made it easy for them to form a rapport with the chatbot. On the other hand, participants who used the basic chatbot said that they did not feel that they form a rapport with the rapport, but still, they felt that the chatbot seem to be experts in this domain which is the reason they share their thoughts with the chatbot.

Then, we present the challenges of MimicTalk for stress counseling. Unlike other domains, participants rarely reported lower trust in information that requires expertise, but still, some participants said that they prefer MimicTalk for sympathizing and listening to them, but a professional persona would be better to counsel them for coping strategies requiring expertise.

First, since stress counseling requires relatively long conversation compared to other domains, MimicTalk for stress counseling increases the possibility of eeire feeling (M=3.25). In a short conversation with the MimickTalk for food journaling and physical activity, a few participants who used MimicTalk reported eerie feelings (M=1.67). Some participants reported that the conversational style of MimicTalk is very similar to their family members, but MimicTalk lacks contextual information between family members.

The second challenge could be said to be the other side of forming rapport which is the privacy issue of personal history. Due to the familiarity with the persona of family members implemented to MimicTalk, participants who used MimicTalk tend to make more conversation like in Figure 6.12. This tendency makes participants talk more about detailed information about their problems or private history. In our case, we went through the IRB approval process to keep the data in a safe place, but participants were somewhat worried of applying the persona of a close person to get an individual's private data could be exploited.

The third challenge addressed by participants was decreased attachment when the relationship is bad like other domains. Some participants who use MimicTalk insisted that they really liked interacting with the MimicTalk when they were in a good relationship. However, they do not want to use it when they are in trouble. Fortunately, none of our participants had a bad relationship with the family members, but when choosing the family member as the persona of the chatbot for stress management, it should be considered.

Lastly, collision with the actual host was also mentioned in this domain. While using MimicTalk, some participants experienced the case that MimicTalk sent them messages while they are with their family member whose persona was applied to the MimicTalk. In this case, some participants felt awkward. Since collision with the actual host is very likely to happen if family member and the user live together. In this case, it may be a good idea to avoid overlapping times or target people living apart.

Additionally, we are going to compare user experience, user perception toward MimicTalk, and opportunities and challenges of MimicTalk among three types of domains in the next section.

6.7 Implications from Domain Experiments

We have explored diet, physical activity, and stress as the domain of implementing research ideas regarding healthcare CA with the mimicked persona of users' healthcare providers. In the diet domain, we have concentrated on the users' perception and behaviors towards the healthcare chatbot for food journaling that mainly collects the user's daily data of food intake. In the domain of physical activity, we have concentrated on the users' perception and behaviors towards the healthcare chatbot for intervening users to reach their goals with physical activity. Lastly, in the stress management domain, we focused on the users' perception and behaviors towards the healthcare chatbot for counseling the user's stress triggers and coping strategies.

To summarize, we explored and defined the user experience, users' perception, opportunities, and challenges with the healthcare CA with the functions of data collection, behavioral intervention, and counseling in the domain of diet, physical activity, and stress management. In this section, we are now going to compare the user experience, user perception, opportunities, and challenges among three healthcare domains to extract design implications for applying the persona of the user's healthcare provider to the healthcare CA.

6.7.1 Comparing User Experience

First, we are going to compare the user experience between the MimicTalk and the basic chatbot among three healthcare domains including diet, physical activity, and stress management. Even though we presented each graph in the section of each experiment, we put together all results to be easily distinguishable. The symbol of * refers to the significant differences between the MimicTalk and the basic chatbot. * refers to the p-value below 0.05, ** refers to the p-value below 0.01, and *** refers to the p-value below 0.001. We are not going to repeatedly go through all the mean values and the standard deviations since they are being presented in the prior sections. We are going to go through the elements of user experience one by one starting from the left to the right.

To analyze how participants conducted healthcare tasks with and without the MimicTalk and the basic chatbot we analyzed the pre and post-behavior in conducting particular tasks in three healthcare domains including food journaling, physical activity, and stress management. Behavior (pre) is the score to the statement of "before using the chatbot, I continuously did the task on the regular basis." As in Figure 6.12, participants who used MimicTalk rated their pre-behavior higher than those who used the basic chatbot. However, these differences were not significant. Behavior (post) is the score to the statement

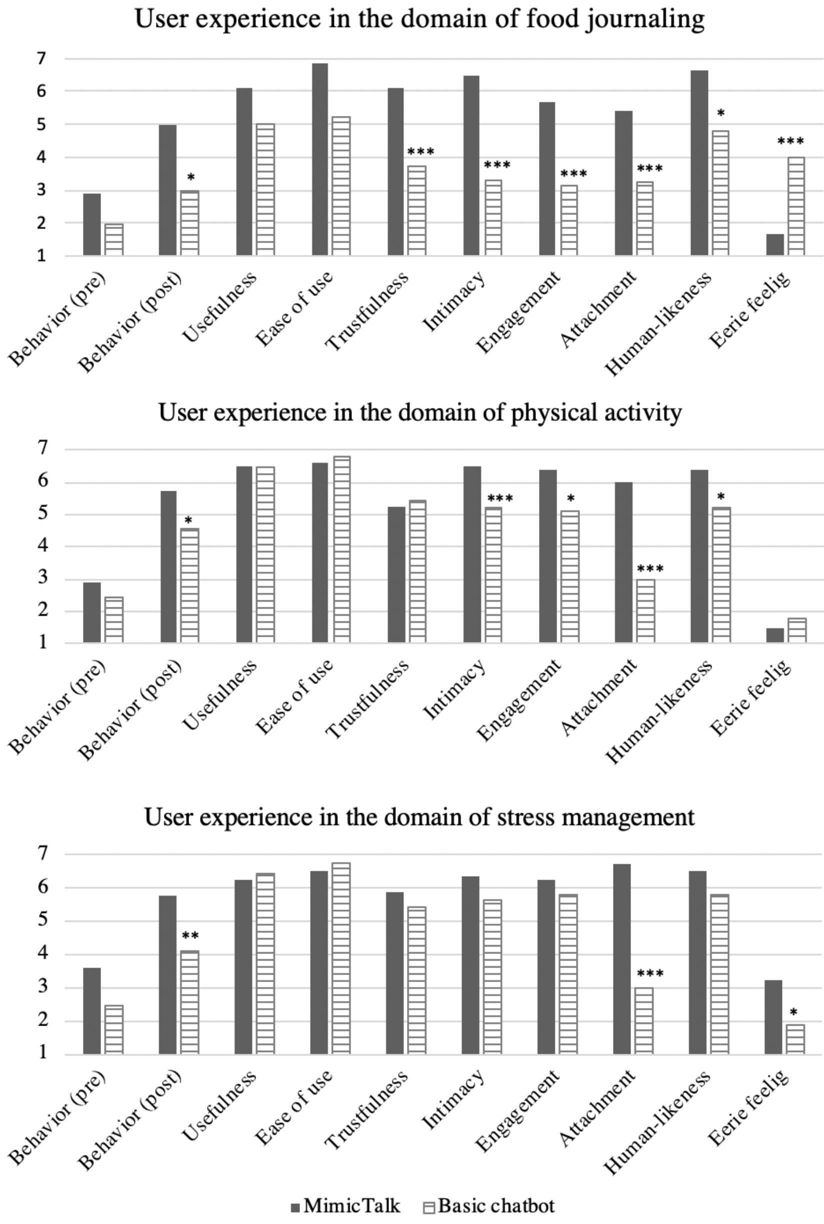


Figure 6.12 Overall user experience with the MimicTalk and the basic chatbot among healthcare domains.

of "during the chatbot use, I continuously did the task on the regular basis." As in Table 6.12, participants who used MimicTalk showed significantly higher ratings in their post behavior ratings.

To measure the system acceptance of both chatbots, we analyzed the usefulness, ease of use, trustfulness, the intimacy of both chatbots. these elements were elements we defined in study 1 that affected the overall acceptance of healthcare with the persona of a close person. To be specific, usefulness is the rating for the statement of "overall, I found the chatbot useful to perform the task." For this element, there was no significant differences between the two groups who used the MimicTalk and the basic chatbot. Ease of use is the rating for the statement of "overall, I am satisfied with how easy it is to use this chatbot." There also was no significant difference between the two chatbots.

But for trustfulness which is the rating for the statement of "During the chatbot use, I trust the chatbot I used", the basic chatbot showed a significantly higher score in the food journaling task. These results did not align with our results from study 1 where the basic chatbot showed a significantly lower score in trustfulness compared to the chatbot with the persona of the family member. We inferred the reason for this result from the post-interview. Some participants who used the basic chatbot for food journaling reported that they felt that the chatbot is made by an expert since they participated through the university's community board.

Intimacy is is the rating for the statement of "During the chatbot use, I felt an intimacy with it." Overall, participants who used MimicTalk showed significantly higher intimacy with the healthcare chatbot. For stress management, this tendency was not significant, but still, MimicTalk showed a slightly higher mean value than the basic chatbot. These results of effectiveness in acceptance did not align with our results from study 1 where all results showed the significantly high differences between the basic chatbot and the chatbot with the persona of a close person. The reason could be the difference between the study design. In study 1, we conducted a within-subject design that all participants used all types of chatbots in one place and used them for only short minutes. We suggest that results from Study 1 can be referred to as the effect of a one-time chatbot, and study 3 as the effect of long-term tasks.

We also analyzed additional elements that could affect the overall user experience with the healthcare chatbots including user engagement, emotional attachment, human-likeness, and eerie feeling. Engagement is the rating for the statement of "Overall, it was very engaging using this chatbot." Overall, participants who used MimicTalk showed significantly higher engagement with the healthcare chatbot. For stress management, this tendency was not significant, but still, MimicTalk showed a slightly higher mean value than the basic chatbot. Attachment is the rating for the statement of "during the chatbot use, I felt emotionally attached to the chatbot." For attachment, participants who used MimicTalk showed significantly higher attachment with the healthcare chatbot in all three healthcare domains. Human-likeness is the rating for the statement of "During the chatbot use, I felt like I was interacting with a human." Participants who used MimicTalk rated significantly higher human likeness with the healthcare chatbot. For stress management, this tendency was not significant, but still, MimicTalk showed a slightly higher mean value than the basic chatbot. The eerie feeling is the rating for the statement of "during the chatbot use, I felt eerie feeling toward the chatbot." In the domain of food journaling and physical activity, the mean value in the basic chatbot is higher for the eerie feeling which means that participants experience more eerie moments with the basic chatbot. However, for the stress management chatbots participants scored higher eerie feeling for the MimicTalk. We measured human-likeness and eerie feelings to quantitatively measure the uncanny valley experiences which we will be discussed in the discussion section.

6.7.2 Comparing User Perception

When it comes to user perception with the MimicTalk, we measured how each linguistic factor we defined in study 3 affects persona perception of MimicTalk. The factors include wake-up word, response time, sentence completion, slang, punctuation, interjection, word transformation, delivery, hedging, back-channeling, abbreviation, emotion, eu-

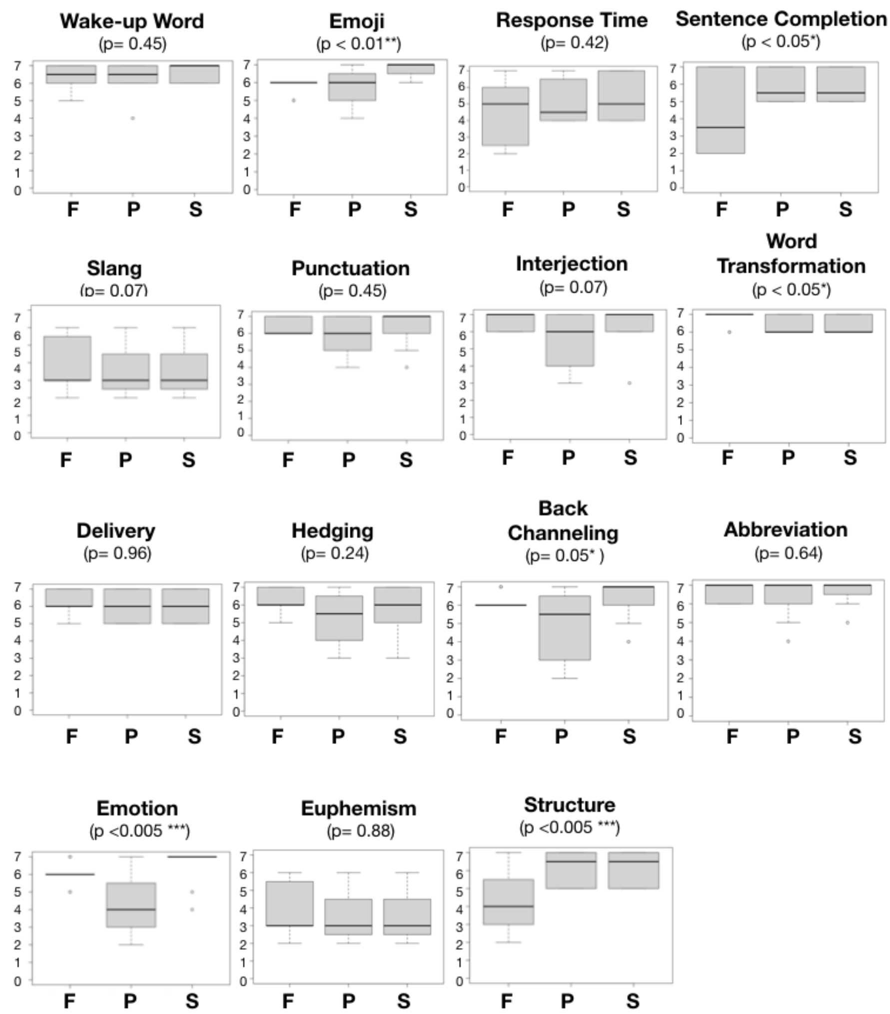


Figure 6.13 Comparison of linguistic factors affecting persona perception of family host among domains

phemism, and sentence structure. Mean values and standard deviations are all presented in the result of each experiment. Therefore, in this section, we would like to discuss the differences between domains based on Figure 6.13. Moreover, Table 6.15 and Table 6.16 show the statistical significance of scores among domains.

Wake-up word is a component that explicitly requests the initial attention of a computer with a single word or single phrase. Since our participants believed they were talking to the chatbot with the persona of PRP (i.e. co-participant), we defined the term as wake-up word. Example of wake-up word participants used in the study were "Hey", "What's up", etc. As in Figure 6.13, wakeup word highly influences the persona perception of MimicTalk as the participant's family member. In study 2, wake-up word was also considered as top elements that affects persona perception of MimicTalk. Based on our results, we conclude that wake-up word affects persona perception of Mimicked persona regardless of domain differences.

Emoji is a text-based ideogram. Emoji come in a variety of shapes and sizes, including facial expressions, objects, locations, weather, and animals. We included all kind of emojis that represents all kinds of symbols such as :-), XD, etc. The mean value of emoji element was highly ranked among all three domains with the slight low score in physical activity.

response time is the functional unit of the time it takes to react to a given input in technology, including conversational agents. In our study, we also defined a time and the pattern participants (all of them played a role as a wizard) took to send a response message as response time. However, we could not exactly mimic the response time patterns among participants. Instead, we synchronously sent the messages to participants. Due to our designing limitations, some participants reported that they felt differences in the conversation with their real family members.

Sentence completion is defined as the agent's use of communication patterns to finish a sentence. Complete sentences must usually include a capital letter at the start, follow grammatical rules, include a punctuation mark at the end of the sentence, and contain one or more major clauses. A major clause can be a standalone subject or a verbal word that expresses the entire content. It was also highly affected the persona perception of MickTalk. However, in food journaling task, it showed the lowest score among other features. In the interview, we found the possible reasons that the structured dialogue flow of the chatbot made participants rarely notice the difference in sentence completion during the experiment.

Slang is a type of language that includes words and phrases that are intended to be used in a casual manner. Slang is sometimes used by members of specific groups who prefer to use a standard language's specific vocabulary to establish group identity to be shared, exclude approach of outsiders, or both. There were examples such as Y'all, cheesy, and so on in our case. Interestingly, participants highly rated slang in the domain of diet and stress management, but in all three domains, it was rated as one of the lowest elements. For this, we asked some participants who rated the raw score. P2 who used MimicTalk for physical activity intervention said that *"my daughter rarely talk slang to me, therefore it did not impact me a lot."*, This tendency was also observed in parent-child relationships mostly in the father-daughter relationships.

Punctuation The use of conventional signs (i.e. !,?,;,.,), sometimes spacing, and typographical words as an aid for correct reading and understanding of the written text is known as punctuation (also known as interpunction). Punctuation is necessary for written English to clarify the meaning and delivery of sentences. Punctuation was also repeatedly mentioned by users as one of the linguistic characteristics that affect persona perception of MimnicTalk. Punctuation was rated highest among all three domains. Based on the interview and user data log, we found that participants highly perceived the persona of their family member with the punctuation marks. For example, P8 from experiment 3 insisted that " my older sister use too many !!!!!! too often so did the MimicTalk."

Interjection is a word, phrase, or symbol used to express a spontaneous occurrence, feeling, or reaction. Exclamations (e.g. ouch!, ahh!, wow!), curses (e.g.damn!), greetings (e.g. hey, yo, bye), response particles (e.g. okay, huh?, mhm), hesitation markers (e.g. uh, uhm, er, hmmm....) and others are examples of interjections (e.g. please, stop, cool). The inclusion criteria of interjections occasionally overlap with profanities, fillers, and sometimes discourse markers due to its diversity. "Wow!" and "Aha!" are two other examples found in the study. It showed higher ratings in the domain of stress management compared to other domains. In the user data log, we observed a more frequent appearance of interjection in the stress management conversation. It seems to be more frequent in relatively long conversations. Due to its frequent occurrence in the domain of stress management, the highest score among domains was observed.

Word transformation For word transformation, we defined it as the result of the process of creating new words. This is different from a change in meaning in which a new meaning or meaning changes in an existing word. For example, if the user sent the message "cutie", which should be written in "cute person" informal conversation, we called it word transformation. The word transformation is similar to word-formation which is creating a new word by borrowing, derivations, compounding, and blending, etc. It also highly affected the persona perception of MimicTalk. The score was significantly high in the domain of food journaling. We asked participants who highly scored the word transformation about their experience. We could learn that the feature impacted participants' perception a lot when used. However, we could found that some hosts did not transform words when they are making text-based conversations. Therefore, we concluded that the significance of food journaling is due to a higher number of occurrences.

Delivery is how clear the meaning of delivered messages. Some participants mentioned that how understandable the messages their partners usually send them determines the characteristics of the sender. It also highly affected the persona perception of MimicTalk. We could not exactly measure the degree of delivery, but at most times absence of awkward delivery made MimicTalk to be perceived as an actual family member's persona.

Hedging is the use of words or phrases in a sentence to reduce ambiguity or the likelihood that the sentence's meaning will be misunderstood. Hedging can be using simple words like "maybe", or "probably," in English. Hedging can also be a useful tool for expressing a stronger point of view in a polite and professional manner. The impact of hedging is relatively low in the domain of food journaling and physical activity. It seems to be more frequent in relatively long, and unstructured conversations since participants reported that the more they talk to the chatbot, the more they can perceive each linguistic characteristic such as hedging, back-channeling, delivery, etc.

Back-channeling frequently occurs during a conversation when one participant participating in the conversation is speaking and the other participant interjects the current conversation, according to linguistics. Back-channeling responses can be either verbal or nonverbal. When serving primarily social or meta-conversational goals, backchanneling responses may include phatic expressions. In the study 3, the impact of backchanneling is relatively low in the domain of food journaling and physical activity. It seems to be more frequent in relatively long, and unstructured conversations since participants reported that the more they talk to the chatbot, the more they can perceive each linguistic characteristic such as back-channeling.

Abbreviation is a simplified or shortened word or phrase. A series of capital letters or the full version of a word can be used as an abbreviation. For instance, the abbreviation or abbr can be used to represent the word abbreviation. It could also be made up entirely with initials, or sometimes it could be a combination of initials and words that represent words with meaning in another language. The impact of abbreviation is relatively high in physical activity and stress management compared to food journaling. We found the possible reason from the interview. According to some participants who used MimicTalk for food journaling, they found few abbreviations in the structured dialogue flow of food journaling.

Emotion include the subjective experience of the speaker or writer, expressive behavior, emotional changes, cognitive processes, and sometimes instrumental behavior. As one of the linguistic characteristics defined in this study, we defined it as the emotion perceived by the receiver via text-based messages. Emotional elements in physical activity were rated low compared to other domains. We found some evidence from the interview and dialogue data that participants' overall tone of the conversation is a lot more neutral or objective compared to other domains.

Euphemism are words or phrases used to avoid saying something unpleasant, negative, or offensive. This is one of the communication strategies and we could also observe the tendency of euphemism in some participants' messages. We could not exactly measure the degree of euphemism from the conversational data, but most participants answered that at most times absence of the euphemism element made MimicTalk to be perceived as an actual family member's persona.

Sentence structure is how users use the order of morphemes in a sentence. For example, for the sentence of "What do you want for breakfast?", one can say "you want anything?" and "for breakfast?", or can say " anything for breakfast?". The sentence structure one uses differs by everyone. Therefore, we defined it as one of the linguistic factors that affect persona perception of MimicTalk. This element was high in the domain of physical activity and stress management but not in the domain of food journaling. In the interview, some participants who participated in experiment 1 (food journaling task) reported that the structure of the dialogue flow made them rarely perceived that sentence structure since the structure seemed to be repeated.

Overall results from our study 3 are different from the ratings of the impact of linguistic factors affecting persona perception of MimicTalk in study 2 in some aspects. We assume that differences come from domain differences. Actually, in study2, the impact ratings were based on the social conversation. Therefore, based on the results from study 3, we conclude that these factors differently affect the persona perception according to the type of domains. However, generally, most factors among all three domains were rated over median which means these factors should be considered as important factors when applying MimicTalk to healthcare domains.

6.7.3 Implications from Study 3

We have gone through the opportunities and challenges of MimicTalk in each healthcare domain throughout three relevant experiments. In this section, we are going to compare the similar or different consequences of MimicTalk among healthcare domains.

For the opportunity of MimicTalk, there exist both similarities and differences among domains. The first opportunity of MimicTalk addressed in all types of domains is Keeping users engaged with the healthcare tasks. In the domain of food journaling, we could find that participants who used MimicTalk were more engaged in the healthcare tasks in the user's answers to the MimicTalk. Participants who used MimicTalk tend to report their contextual information including how they cooked their meal, how they felt about their meal experience, whom they ate their meal with, and why they skipped the meal. They also tend to make social responses to the chatbot. In the domain of physical activity, participants who used MimicTalk were more engaged with the tasks of sharing their amount of exercise time, the intensity of exercise, and their personal goals. In the domain of stress management, participants who used MimicTalk were more engaged with the tasks of sharing their recent stressors and their feelings and private experiences toward such stressors.

Second, MimicTalk strengthened the feeling of being cared for regardless of the type of healthcare domain. Mostly, the persona of MimicTalk (i.e.participant's family member) was mentioned as the primary cause for their feelings. Most participants among all types of domain agreed that just sending prompts that mimick the dialogue style of their family members made them feel like they are being cared for. Third, MimicTalk was found to take over the repetitive task including asking what the users ate for the day while giving them a feeling of intimacy due to the family member's persona which made participants keep engaged in the journaling tasks. Also, strengthening self-reflective behavior was also discussed among participants who used MimicTalk. Self-reflective behavior was also reported by some participants who used the basic chatbot. However, some participants who reported self-reflective behavior in the MimicTalk group insisted that the persona of their family member strengthened their reflective behavior.

While domains included in this study share similar opportunities of

MimicTalk, the following aspects are opportunities of MimicTalk that are specific to the type of domain. In the domain of food journaling with the function of collecting daily meal data from the user, we found the opportunities of collecting image-based data. Actually, food journaling requires the data including not only lists of food intake, but also the amount of food intake. Lists of food intake can be collected with natural language-based answers. However, when it comes to the amount of food data, it is hard to be collected with the natural language-based answers and need another type of information such as image, video, etc. In this manner, we found that MimicTalk for food journaling has an opportunity in collecting image-based data. It could be a burden to take pictures every time participants have each meal.

When it comes to the domain of physical activity, participants who used the MimicTalk frequently reported that MimicTalk motivates participants for the physical activity. While participants who used the basic chatbot described the chatbot as daily activity checker rather than a motivator. Also, customized framing effects and information acceptance of MimicTalk are emphasized by some participants in experiment 2.

In the domain of stress management, opportunities of discussing coping challenges with the chatbot were discussed by most participants. This process of discussion provoked the user's self-reflective behaviors in some participants. Proactive daily messages from chatbot increased the feeling of being cared for in some participants. Also, MimicTalk had the opportunity of forming rapport in the counseling task. There were also several reports about the rapport among the other types of domains. However, the opportunity of forming rapport was more frequently reported compared to other types of healthcare domain since the longest length of conversation was observed in the domain of stress management. We found that MimicTalk reflects the family dynamics with the dialogue styles, one of which is intimate relationships. The intimate relationship between the participants and their family members affected forming quick rapport between the chatbot and the user.

We also discuss challenges of MimicTalk that are common among types of domains and that differ among types of domains. When it comes to challenges of MimicTalk that are common among the type of domains, collision with actual hosts were considered as a major challenge of MimicTalk. While using MimicTalk, participants reported they experienced several overlapping interactions between MimicTalk with the persona of their family member and their real family member. For example, some participants said that a message was sent from the chatbot while eating together, and some participants said that a message came from the chatbot while they are on the phone with their real family member. We call this collision with the actual persona in this paper. When the persona of MimicTalk collides with the actual person, participants reported uncomfortable eerie feelings with the MimicTalk. This collision was reported in all types of healthcare domains but the intensity of uncomfortable feeling differs depending on the type of the domain. We assume that the user experience of collision with a family member affected eerie feeling with the chatbot. Moreover, the longer the interaction and the less separate the function from the social talk, the stronger the eerie feeling seems to be. Therefore, as in Figure 6.12, a higher eerie feeling was reported in stress management, where the conversation length was relatively long compared to other types of domains.

Moreover, decreased attachment when the relationship between family members is bad is another common challenge MimicTalk face in all types of domains. We recruited a group of families who are in an intimate relationship in all three experiments so we expected that persona would only have a positive effect on interactions with chatbots. However, no matter how good the relationship is, when the intimacy of the relationship decreased due to an unexpected negative event between family members, it was reflected in the relationship with the MimicTalk as well. In some cases, there have also been positive aspects that have been reported by participants. For example, some participants reported that MimicTalk which has a persona of a family member made their relationship better when he or she had a fight with each other. One participant said that *"we were not talking to each other because we* *had a fight, but my heart weakened when MimicTalk that talks alike my husband asked me hello."*

We also share the points to be considered differently for each domain. The first challenge reported particularly in food journaling was fatigue from structured dialogue flow. This is actually the difference caused by the nature of food journaling itself. Some participants said that there is an awkwardness that comes from the structure itself. Some of them said that MimicTalk is too interesting at first, but the repetitive structure makes it less interesting as time goes by. This reaction was not found in the semi-structured MimicTalk for physical activity intervention. Also, a problem related to trust was found in the physical activity domain. When it comes to physical activity, there were many mentions of expertise, but there was an opinion that in the case of areas requiring expertise, the persona of the family could rather be a factor that lowers trust. In marital relationships, this reaction was more common in men than in women, and in parent-child relationships more reported in parents than in children. However, most of these people said that if the chatbot is functionally specialized and the persona of the family is applied with it, they would more likely to use MimicTalk if it is clearly recognized by the user. When it comes to stress management, the benefits of MimicTalk in stress management are closely related to the challenges of MimicTalk. As mentioned above, MimicTalk had the opportunity of forming rapport in long conversations. Because of this advantage, some participants feared that it would be counterproductive to say too much personal information. In the case of healthcare tasks based on long-time conversations, privacy about personal data protection should be guaranteed.

Based on reported opportunities challenges of MimicTalk addressed by participants, we are going to provide design guidelines in the discussion section to improve the user experience with the MimicTalk through proper design strategies.

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
Wakeup word	(Intercept)	1	1482.25	1482.25	2680.23	<2e-16 ***
	Domain	2	0.50	0.25	0.45	0.64
	Residuals	33	18.25	0.55		
Emoji	(Intercept)	1	1320.11	1320.11	2313.1150	< 2e-16 ***
	Domain	2	5.06	2.53	4.4292	0.01977 *
	Residuals	33	18.83	0.57		
Response time	(Intercept)	1	1078.03	1078.03	580.815	< 2.2e-16 ***
	Domain	2	27.72	13.86	7.468	0.002111 **
	Residuals	33	61.25	1.86		
Sentence completion	(Intercept)	1	1067.11	1067.11	415.1041	< 2e-16 ***
	Domain	2	26.06	13.03	5.0678	0.01204 *
	Residuals	33	84.83	2.57		
Slang	(Intercept)	1	1369.00	1369.00	1260.7535	< 2e-16 ***
	Domain	2	17.17	8.58	7.9047	0.001567 **
	Residuals	33	35.83	1.09		
Punctuation	(Intercept)	1	1431.36	1431.36	2122.9176	< 2e-16 ***
	Domain	2	3.39	1.69	2.5131	0.09641
	Residuals	33	22.25	0.67		
Interjection	(Intercept)	1	1444.00	1444.00	1408.4335	< 2e-16 ***
	Domain	2	10.17	5.08	4.9581	0.0131 *
	Residuals	33	33.83	1.03		
Word transformation	(Intercept)	1	1534.03	1534.03	6982.4713	< 2e-16 ***
	Domain	2	1.72	0.86	3.9195	0.0297 *
	Residuals	33	7.25	0.22		

Table 6.15 Significant differences among linguistic characteristics affecting persona perception in daily healthcare domains(1)

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
Delivery	(Intercept)	1	1369.00	1369.00	1978.5547	<2e-16 ***
	Domain	2	0.17	0.08	0.1204	0.8869
	Residuals	33	22.83	0.69		
Hedging	(Intercept)	1	1296.00	1296.00	1161.1222	< 2e-16 ***
	Domain	2	11.17	5.58	5.0023	0.01266 *
	Residuals	33	36.83	1.12		
Back channeling	(Intercept)	1	1248.44	1248.44	947.0958	< 2.2e-16 ***
	Domain	2	18.06	9.03	6.8487	0.003252 **
	Residuals	33	43.50	1.32		
Abbreviation	(Intercept)	1	1560.25	1560.25	821.2740	< 2e-16 ***
	Domain	2	0.50	0.25	0.4521	0.6402
	Residuals	33	18.25	0.55		
Emotion	(Intercept)	1	1201.78	1201.78	1144.000	< 2e-16 ***
	Domain	2	43.56	21.78	20.731	1.474e-06 ***
	Residuals	33	34.67	1.05		
Euphenism	(Intercept)	1	484.0	484.00	229.8129	< 2e-16 ***
	Domain	2	0.5	0.25	0.1187	0.8884
	Residuals	33	69.5	2.11		
Sentence structure	(Intercept)	1	1089.00	1089.00	733.408	< 2e-16 ***
	Domain	2	32	16.00	10.775	0.0002 ***
	Residuals	33	49	1.48		

Table 6.16 Significant differences among linguistic characteristics affecting persona perception in daily healthcare domains(2)

Chapter 7

Discussion

Throughout the study, we have demonstrated that family member's dialogue style implemented to MimicTalk increased the user's persona perception of family persona, and its consequences improved overall user experience while conducting healthcare tasks. With the results and implications we identified from previous studies, we discuss the following contents to guide designers and researchers. First, we suggest design guidelines when applying the persona of the user's close person to the healthcare CA. We provide guidelines based on conversational traits, host traits, and healthcare domain traits. Second, we are going to discuss ethical considerations when applying the idea of MimicTalk to expanded domains. Lastly, we discuss the limitation of the series of studies included in the thesis to call for further researches to improve unexpected consequences.

7.1 Design Guidelines

Figure 7.1. shows an overview of design guidelines when applying the persona of the user's close person to the healthcare CA. These guidelines are only based on our results so there must be some more considerations. We discuss the limitations of the study to call for further study in the HCI community. We categorized design guidelines based on the three types of traits as follows: (1) Guidelines based on the conversational traits, (2) guidelines based on the host traits, and (3) guidelines based on the domain traits. Categorization was conducted based on the thematic analysis by three researchers including the author of the thesis.

Guidelines Based on Conversational Traits

Through the implementation of MimicTalk in study 3, we have learned that the conversational structure, conversation topic, the function of conversation can influence outcomes of MimicTalk in the healthcare domains. While discussing, we used the word host to refer to the person whose persona is implemented to the MimicTalk.

In our study, conversation structure refers to three levels of structure including structured, semi-structured, and unstructured. Due to the domain characteristics, we made a design decision that food journaling chatbot to have structured dialogue flow, physical activity to have semi-structured dialogue flow, and stress management for unstructured dialogue flow. The structure of the conversation is highly related to the function of conversation. However, we separately discuss these two elements since each of them is highly important. In the structured conversation (food journaling in our case), more variety of dialogue should be implemented to the CA since users were more likely to feel bored or fatigue from the structured conversation. This impacted the user experience with MimicTalk since some participants reported that structured dialogue flow made MimicTalk more robot-like. Even in a structured conversation, the social talk should be randomly implemented to prevent awareness of interacting with the persona of a close person. However, in the semi-structured and the unstructured conversational structure, the form of language component including sentence structure and sentence completion more impacted the persona perception of the MimicTalk. Using emojis and emotions reflected in the conversation more largely impacted the persona perception of MimicTalk in the unstructured conversation. Designers should keep in mind that as the freedom of CA responses increases, dialogues should be designed in a more detailed way.

When it comes to conversation topics, the familiarity of the topic influences the persona perception and user experience of MimicTalk. In this study, the topic is likely to refer to the healthcare tasks of MimicTalk. In some cases where the host of persona and the user rarely talks about the healthcare task of the MimicTalk, the user felt awkward in the conversation with the MimicTalk. Therefore, the familiarity of having a conversation with the host of persona for conversation topic should be determined before choosing an appropriate candidate. When choosing the persona of a close person for healthcare CA, designers could choose someone the user is frequently having the particular conversation (which is to be implemented to CA) to prevent feeling out of context. If the host who is unfamiliar with the conversation topic was chosen as a persona to be implemented to the MimicTalk, then, designing an onboarding scenario to make the user naturally adapt to the conversation with the MimicTalk is necessary.

The function of conversation refers to the main function of the MimicTalk to reach its system goals. For example, the function of MimicTalk for food journaling was data-collection, the function of MimicTalk for physical activity was the behavioral intervention, and the function of MimicTalk for stress management was interactive conversation to cope with the stressors. The function of conversation is closely related to the conversation structure since the structure should be decided based on the major function of MimicTalk. When applying the persona of a close person for counseling, check the privacy issue. Ethical concerns from the user's perception of the chatbot as an actual family member that could lead to the user's behavior of sharing sensitive information, have been raised. This concern emphasizes the need for clearly noticing the user knows about the identity of the chatbot which is not a human. Most participants worried that they were more likely to talk to MimicTalk about their private information in the conversation of counseling. Some participants worried that they were not sure where their conversation data would be stored and utilized. Sharing the data storage policy could be a way to prevent this issue. In the interactive conversation, the user's answers seem to be much longer than other functions and the mismatch of conversation style between the host and the MimicTalk was found to be more distinguishable compared to other domains. Particularly for the linguistic factors that seem to be more noticeable in interactive conversation should be implemented to MimicTalk to minimize the mismatch.

Guidelines Based on Host Traits

We empathize user-host relationship, the linguistic identity of the host, and contact channels as host traits that can influence outcomes of MimicTalk. When it comes to user-host relationship, designers have to consider the unique characteristics and the quality of the family relationship between the user and his/her family member when expanding their candidate, who is going to be included in the dialogue design, to diverse types of family members such as siblings, husbands, and wives. Designers should use a host with high intimacy for the persona of MimicTalk since intimacy with the actual host greatly influences the intimacy with MimicTalk. The relationship between the user and the host was found to be reflected in the relationship between the user and the MimicTalk. Moreover, designers should avoid persona with whom the user continuously keeps in face-to-face contact (e.g.living together). Some participants reported eerie feelings when MimicTalk talks to them while they are physical with the host of persona. Also, designers should apply a persona of the host who has spoken a lot with the user about the particular field where CA will be applied to. If they rarely talk about the task in the actual world, the MimicTak could be perceived as more non-human like in our case.

The linguistic identity of the host also greatly impacts persona perception of mimicked persona. During dialogue design, we could notice that some hosts have strong linguistic identities while others have weak or moderate linguistic identities. For example, one host used lots of emojis in the conversation, and another used lots of slang or abbreviations, and in some cases being highly emotion-centric. In the post-interview, we asked all participants about how linguistic characteristics affected their persona perception of MimicTalk, and those participants whose family member(host) has strong linguistic identity reported more linguistic characteristics that impacted their persona perception compared to those who do not. As learned from our study, using the persona of the host with the strong linguistic characteristic is an effective way to increase the persona perception. On the other hand, when using the persona of the host with the weak linguistic characteristic, designers can add other modalities to increase persona perception including profile image and voice.

Contact channels users use to reach their family members (host) also affected the outcome of persona implementation. Designers should avoid the persona of the person with whom the user continuously keeps in face-to-face contact (e.g.living together). This increases the eerie feeling in our case. In all cases, designers should be careful of confusion that the user can forget whether they have talked to CA with the persona of the family member or the actual family member for particular agenda. Our results showed that there exist some cases that users frequently forgot whom they talked about particular agenda. So separating and specifying the role between the host and the CA with the persona of the host would be also effective in this case. For example, when it comes to food journaling, MimicTalk can keep track of user's meal menu, while actual family member gives user tangible support such as making a healthy meal or gives the user useful information, etc.

Guidelines Based on Domain Traits

We also provide guidelines based on the healthcare domain traits. These traits include the priority of healthcare tasks, healthcare habits, and frequency of being cared for by others in a particular healthcare domain. When it comes to the priority of healthcare tasks if a particular task is the user's region of major interest affects the outcomes of MimicTalk. If the priority is high, increasing the interaction period to increase user engagement could be effective. For example, we suggested implementing social talk to the MimicTalk in the prior section in this chapter. However, if the priority of a particular healthcare task is relatively low, designers should avoid prolonged interaction. In this case, the concerns of the user feeling guilty not answering to MimicTalk could be raised while users feel fatigued or burden having interaction with MimicTalk. If the priority is high, guilty framing can be utilized as the strategy of the healthcare CA.

The healthcare habits of the user are another consideration when implementing MimicTalk. If the user is already an engaged in healthcare tasks, persona of the close person can be used as a tool to enhance the user experience. If not, designers should attempt to use the consequences of applying the persona of a close person to achieve the user's behavior change (e.g. guilty, shame, sorry, love, etc). However, when applying these factors, investigating users to define expected outcomes should be preceded before implementing MimicTalk since emotional consequences are different among users.

The frequency of being careful about should be considered in the design process. If the user is already receiving enough caregiving, designers should define the required task in detail and select a persona of a close person suitable for that specific task. If the frequency of caregiving is low and caregiving with emotional support is required, designers should apply the persona of the host that the user feels an emotional attachment.

7.2 Ethical Considerations

In the series of studies, we have found few pieces of evidence that require ethical considerations. Evidence from studies was not major results so it wasn't mentioned much in the results section. However, since ethical concerns accompany any technological innovation, we are going to discuss all possible concerns to prevent the worst cases in its practical application. We referenced Ruane et al's study in this section, which considers the social context in identifying ethical challenges of conversational AI design [205]. According to Ruane et al, ethical concerns derived from conversational AI could be different depending on the target user, the application domain, and the task and the goal of the agent. For instance, a chatbot used by employees of the organization will have different considerations compared to a public-facing CA. For the responsible design of conversational AI, a profound understanding of the user including user characteristics and interests, social contexts, should be prioritized. No one-fits-all principle exists to be applied to all kinds of CAs. For these reasons, previous studies suggest abstract principles including multiple approaches to develop ethical and responsible Conversational AI [205]. Since we posit our series of studies in the domain of daily healthcare for preventive purposes, we share our ethical concerns raised during our work with designers and stakeholders.

We have defined ethical concerns of MimicTalk based on the categories of (1) trust and transparency, (2) privacy, (3) agent persona, and anthropomorphism.

Trust and transparency: Transparency about an agent's status as an autonomous (non-human) agent, as well as the boundaries of its capabilities, is necessary to allow users to make informed decisions, which leads to user's overall trust toward the system. Understanding a user's expectations of an agent is essential for ensuring the user's confidence. Reasonable expectation management is a need and it should be evaluated. For example, if a user expects an anonymous conversation, then identifiable information about the user and plain text logs should not be stored by or visible to the development team.

In our case, we informed participants of MimicTalk's functionality in detail. We specifically introduced the functionality of the chatbot and its limitations coming from the study design. We not only told them they were recruited to evaluate the user experience and user perception of MimicTalk (or basic chatbot) and told them they are going to interact with the chatbot with the tasks of food journaling. We emphasized that the chatbot's functionality is limited in journaling tasks and the chatbot would be limited in responding to out-of-task conversations. Also, we informed participants that conversational data collected during the experiments would be only used for data analysis and would be eliminated after the paper is published based on the study's procedure approved by IRB. Then, we showed the sample conversation of the chatbot in order to help participants be accustomed to the workflow of the chatbot. Participants were instructed to freely ask any questions regarding the chatbot and the experiment to the author. Through these processes, we made sure for users to manage their expectations with the chatbot.

However, even though we tried to fully manage user expectations based on related work, a few users raised the transparency issue that affects their trust on MimicTalk. The issue was about how the dialogue styles of MimicTalk have been made. One participant who used fully automated MimicTalk for food journaling insisted that "It seems to affect my trust toward chatbot as MimicTalk informs me whether it was entered manually from my family member or automated with something/someone else." Therefore, when utilizing MimicTalk, an explanation for not only MimicTalk works but also how it has been designed should be implemented to the CA.

Privacy: The interaction of humans and CAs result in various ethical questions about privacy including data type, data access, and the period of data storage, and the purpose of data use. Collecting user data is one of the major privacy concerns so some countries protect user data though ethical guidelines such as GDPR in Europe [206]. Even though previous guidelines exist, since ethical issues vary depending on the domain of deployment and user's vulnerability, additional studies are needed to investigate the domain-specific issues [205].

In our case, we found two kinds of privacy issues in (1)social relationship with the chatbot and (2) data privacy issue. When it comes to social relationships, concerns arise that the social features of the agent can encourage self-disclosure of information [207]. Sometimes, the dialogue of an agent designed with social purpose impacts users to self-disclose without discernment. Self-disclosure could be beneficial in some cases by encouraging users to gather data, at the same time, improving user experience [208]. However, sometimes, it can affect user trust or decrease user experience when the functionality of CA is not properly explained prior to its use. In study 3, we have confirmed that interaction with the MimicTalk increased rapport formation, contextual information, and additional responses made regardless of healthcare tasks. With these evidence, we can conclude that the MimicTalk increased the self-disclosure of participants. In this manner, MimicTalk should be implemented with the explainability in privacy issue. Making users notice that MimicTalk is non-human, and the policy on safe data storage should be always approachable for users in a transparent way.

Agent persona and anthropomorphism: A large part of design decisions on CA is related to persona and personality. Designing an agent persona includes applying age, gender, race, and cultural affiliation to the agent. As the level of embodiment increases, these indications become more explicit. It is essential to investigate what types of relationships users prefer to build with the agent. After, to determine if the particular persona is encouraging, defining persona traits that would cause a negative social reaction is also important since inappropriate agent persona can also result in the user building harmful stereotypes of agents [205]. In our case, we have mainly applied the persona of a close person to CA, which made participants interact with CA as if they are interacting with their family members through a text interface. We confirmed that applying the persona of the user's close person has an opportunity in managing outcomes of a particular behavior which is, in our case, healthcare tasks.

One of our primary outcomes coming from applying the persona of a close person was the high tendency of anthropomorphism. Participants who used MimicTalk tend more to respond the human-like triggers than participants who used the basic chatbot. Anthropomorphism could be strengthened when users interact with a system through conversation and when the CA has been designed with personality and embodied with a human-like avatar [209]. The level of anthropomorphism can also be determined by agent persona expressions including gender, age, race, cultural affiliation. In our study, the level of anthropomorphism tends to be stronger if the CA's persona is implemented with the daughter than a son, children than parents. This tendency also varies depending on the family dynamics.

The problem is that there exists a chance of adverse effects coming from highly human-like CAs for daily use. We recruited participants who show emotional attachment with the host above the median, and it made participants emotionally attached to MimicTalk by real-world dynamics. This tendency was found to be positive in our study by increasing engagement with healthcare tasks. We also asked participants about adverse effects coming from family member's persona, but rare concerns were found. However, one participant who participated in experiment 1 insisted that "I know the chatbot's functions are limited in healthcare tasks, but when the chatbot doesn't answer my social conversation, I feel disappointed and sad. Maybe it's because I miss my daughter so much." This interview data implied to us that in some cases that users show extreme emotional attachment to both their family members and the MimicTalk, MimicTalk can be the emotional trigger that could cause unexpected consequences such as triggering negative feelings, triggering traumatic experiences related to emotions. Designers should consider minor cases, for example, while using the chatbot the host may physically move away, get injured, or die. The changes in user-host dynamics should be reflected in the design of MimicTalk. Further researches that focus on adverse effects coming from MimicTalk's triggers are also needed.

7.3 Limitations

Throughout the thesis, we have explored and evaluated the idea of applying the persona of the user's close person (i.e.MimicTalk) to CA to improve the user experience, particularly user engagement, with the healthcare tasks in the domains of daily healthcare. We have defined the opportunities and challenges of MimicTalk based on our results. However, since the thesis could not investigate all cases of relationships, stakeholders or healthcare tasks, we recommend readers to use our research results after recognizing the following limitations.

First, we share limitations coming from highly user-centered approaches in the thesis. The Series of studies included in the thesis focuses on the user and uses user-centric methods for designing and evaluating MimicTalk. It was a reasonable choice because the major goal of the thesis was to improve the user experience with daily healthcare CA. In particular, we targeted the self-care of the users with preventive healthcare, not patients. However, if the goal of healthcare CA is out of one of the two contexts that are (1) individual self-care (2) preventive purpose, additional research on existing stakeholders is essential. For example, if the MimicTalk should be implemented for the purpose of curing particular diseases such as cancer, the needs of stakeholders including doctors, nutritionists, family members should all be taken into consideration when designing and implementing MimicTalk. In this case, the context of sending and receiving support might be highly complicated, so additional research is necessary before implementing MimicTalk.

Second, we share limitations that occur from experimental design. In order to focus on the effectiveness of MimicTalk, we did not relay the conversation between MimicTalk and the user to the actual host. Our results showed that there exist some cases that users frequently forgot whom they talked about particular agenda. In the discussion, we suggested separating and specifying the role between the host and the CA with the persona of the host would be also effective in this case. For example, when it comes to food journaling, MimicTalk can keep track of user's meal menu, while actual family member gives user tangible support such as making a healthy meal or gives the user useful information etc. However, some participants suggested the conversation with the MimicTalk be passed on to the host as well. For this issue, we are going to conduct additional research on information sharing among the user, MimicTalk, and the host.

Third. we share limitations on generalizability of the study. We intended to generalize the study to various fields of application by selecting the most representative daily healthcare domains and recruiting types of participants and hosts. However, as mentioned throughout the paper, if the application field of MimicTalk goes beyond preventive daily healthcare, additional study is needed to consider unexpected consequences of MimicTalk before implementing it to a large population. Also, emotional and social characteristics may vary depending on the history of the family even if the type of relationship is the same. Therefore, designers should always consider user characteristics and user context before implementing MimicTalk.

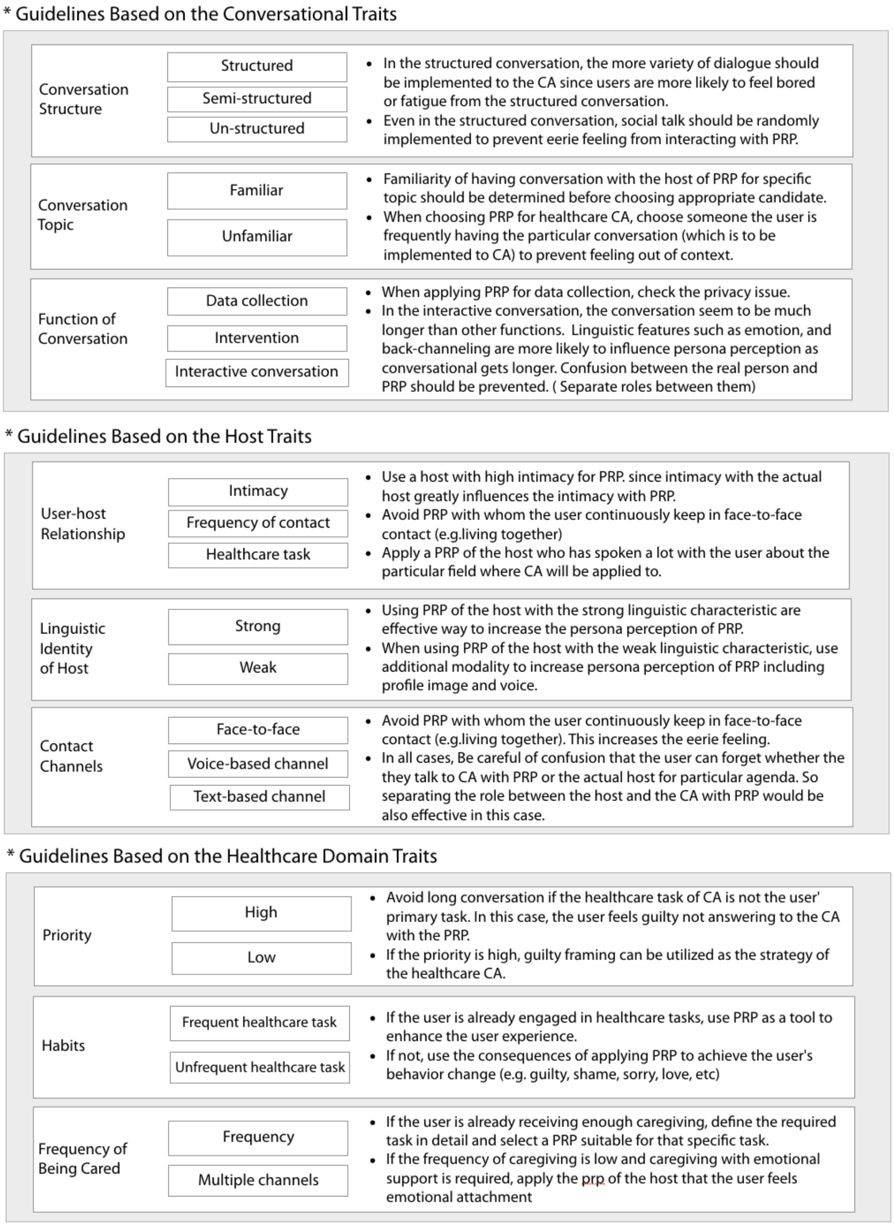


Figure 7.1 Overview of design guidelines when applying persona of close person to the healthcare CA

Chapter 8

Conclusion

My thesis suggested, explored, and evaluated the idea of applying the persona of the user's close person in the real world to improve the user experience with the healthcare CAs in the domain of daily health management. Particularly the thesis is focusing on (1) individual self-care with (2) preventive purpose. I made three major research contributions through the thesis. First, I investigated the effective persona of a close person to be implemented in CA for daily healthcare in study 1. Second, I provided lists of linguistic characteristics to consider when applying a real person's persona who is in the relationship with the user to CA in study 2. We expect these features also could be applied to expanded research areas such as Text-Style-Transfer, and automated text generation. Lastly, I investigated the effectiveness of healthcare CA with the persona of the user's family member in study 3 by applying it to major daily healthcare domains including diet, physical activity, and stress management.

Moreover, based on findings, I provide design guidelines based on conversational traits, host traits, and healthcare domain traits. Each of these traits includes elements that should be considered when designing healthcare CA with the persona of the user's close person. Since provided guidelines reflect study results in this thesis, there exists a need for further studies in this area to explore other factors that can influence the outcomes of healthcare CA with the persona of a person who is in a close relationship with the user. Additionally, ethical considerations defined in the thesis should also be fully considered.

I expect outcomes from the thesis could be applied to various domains and fields that require the application of personalized CA. When expanding boundaries of applying a real person's persona to the CA, domain of interest, target user group, and the type of host should all be taken into consideration and an additional evaluation process is always necessary to prevent adverse effects coming from following issues that are also discussed in the thesis: (1) trust and transparency, (2) privacy, and (3) anthropomorphism. By taking all these things into considerations, I expect that more relevant researches would be done so that more people would be healthier and happier with the technology.

Bibliography

[1] W. H. Organization *et al.*, *WHO guideline: recommendations on digital interventions for health system strengthening.* World Health Organization, 2019.

[2] U. Food, D. Administration, *et al.*, "Digital health innovation action plan: Fda center for devices and radiological health; 2017."

[3] S. M. Schueller, K. N. Tomasino, and D. C. Mohr, "Integrating human support into behavioral intervention technologies: The efficiency model of support," *Clinical Psychology: Science and Practice*, vol. 24, no. 1, pp. 27–45, 2017.

[4] K. Patrick, E. B. Hekler, D. Estrin, D. C. Mohr, H. Riper, D. Crane, J. Godino, and W. T. Riley, "The pace of technologic change: implications for digital health behavior intervention research," 2016.

[5] D. H. Solomon and R. S. Rudin, “Digital health technologies: opportunities and challenges in rheumatology,” *Nature Reviews Rheumatology*, vol. 16, no. 9, pp. 525–535, 2020.

[6] S. M. Noar, N. Grant Harrington, S. K. Van Stee, and R. Shemanski Aldrich, “Tailored health communication to change lifestyle behaviors,” *American Journal of Lifestyle Medicine*, vol. 5, no. 2, pp. 112–122, 2011.

[7] R. P. Hawkins, M. Kreuter, K. Resnicow, M. Fishbein, and A. Dijkstra, “Understanding tailoring in communicating about health,” *Health education research*, vol. 23, no. 3, pp. 454–466, 2008.

[8] J. L. Z. Montenegro, C. A. da Costa, and R. da Rosa Righi, “Survey of conversational agents in health,” *Expert Systems with Applications*, vol. 129, pp. 56–67, 2019.

[9] A. Ram, R. Prasad, C. Khatri, A. Venkatesh, R. Gabriel, Q. Liu, J. Nunn, B. Hedayatnia, M. Cheng, A. Nagar, *et al.*, “Conversational ai: The science behind the alexa prize,” *arXiv preprint arXiv:1801.03604*, 2018.

[10] M. Jovanovic, M. Baez, and F. Casati, “Chatbots as conversational healthcare services,” *IEEE Internet Computing*, pp. 1–1, 2020.

[11] K. Barnard, C. Lloyd, P. Dyson, M. Davies, S. O'Neil, K. Naresh, J. Lawton, R. Ziegler, and R. Holt, "Kaleidoscope model of diabetes care: time for a rethink?," *Diabetic medicine*, vol. 31, no. 5, pp. 522–530, 2014.

[12] H. H. Clark and D. Wilkes-Gibbs, "Referring as a collaborative process," *Cognition*, vol. 22, no. 1, pp. 1–39, 1986.

[13] A. B. Kocaballi, S. Berkovsky, J. C. Quiroz, L. Laranjo, H. L. Tong, D. Rezazadegan, A. Briatore, and E. Coiera, "The personalization of conversational agents in health care: systematic review," *Journal of medical Internet research*, vol. 21, no. 11, p. e15360, 2019.

[14] M. Jovanovic, M. Baez, and F. Casati, "Chatbots as conversational healthcare services," *IEEE Internet Computing*, 2020.

[15] C. Nass, J. Steuer, and E. R. Tauber, "Computers are social actors," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 72–78, 1994.

[16] T. W. Bickmore, *Relational agents: Effecting change through human-computer relationships.* PhD thesis, Massachusetts Institute of Technology, 2003.

[17] D. McDuff and M. Czerwinski, "Designing emotionally sentient agents," *Communications of the ACM*, vol. 61, no. 12, pp. 74–83, 2018.

[18] N. Svenningsson and M. Faraon, "Artificial intelligence in conversational agents: A study of factors related to perceived humanness in chatbots," in *Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference*, AICCC 2019, (New York, NY, USA), p. 151–161, Association for Computing Machinery, 2019.

[19] B. Vandenberghe, "Bot personas as off-the-shelf users," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, (New York, NY, USA), pp. 782–789, ACM, 2017.

[20] A. Yorita, S. Egerton, J. Oakman, C. Chan, and N. Kubota, "Self-adapting chatbot personalities for better peer support," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 4094–4100, 2019.

[21] S. Cohen, "Social relationships and health.," *American psychologist*, vol. 59, no. 8, p. 676, 2004.

[22] L. F. Berkman, "The assessment of social networks and social support in the elderly," *Journal of the American Geriatrics Society*, vol. 31, no. 12, pp. 743–749, 1983.

[23] M. S. Granovetter, “The strength of weak ties,” *American journal of sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[24] J. Pereira and Ó. Díaz, “Using health chatbots for behavior change: A mapping study,” *Journal of medical systems*, vol. 43, no. 5, p. 135, 2019.

[25] B. J. Fogg, “Persuasive technology: Using computers to change what we think and do,” *Ubiquity*, vol. 2002, Dec. 2002.

[26] T. Kowatsch, D. Volland, I. Shih, D. Rüegger, F. Künzler, F. Barata, A. Filler, D. Büchter, B. Brogle, K. Heldt, *et al.*, “Design and evaluation of a mobile chat app for the open source behavioral health intervention platform mobilecoach,” in *International Conference on Design Science Research in Information System and Technology*, pp. 485–489, Springer, 2017.

[27] P. Brinkman, “Virtual health agents for behavior change: Research perspectives and directions,”

[28] K. Oh, D. Lee, B. Ko, and H. Choi, “A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation,” in *2017 18th IEEE International Conference on Mobile Data Management (MDM)*, pp. 371–375, 2017.

[29] A. van Heerden, X. Ntinga, and K. Vilakazi, "The potential of conversational agents to provide a rapid hiv counseling and testing services," in *2017 International Conference on the Frontiers and Advances in Data Science (FADS)*, pp. 80–85, 2017.

[30] A. Cheng, V. Raghavaraju, J. Kanugo, Y. P. Handrianto, and Y. Shang, "Development and evaluation of a healthy coping voice interface application using the google home for elderly patients with type 2 diabetes," in *2018 15th IEEE Annual Consumer Communications Networking Conference (CCNC)*, pp. 1–5, 2018.

[31] P. Achananuparp, E.-P. Lim, and V. Abhishek, "Does journaling encourage healthier choices? analyzing healthy eating behaviors of food journalers," in *Proceedings of the 2018 International Conference on Digital Health*, DH '18, (New York, NY, USA), p. 35–44, Association for Computing Machinery, 2018.

[32] T. W. Bickmore, K. Puskar, E. A. Schlenk, L. M. Pfeifer, and S. M. Sereika, "Maintaining reality: Relational agents for antipsychotic medication adherence," *Interacting with Computers*, vol. 22, no. 4, pp. 276–288, 2010.

[33] A. Fadhil, "A conversational interface to improve medication adherence: Towards ai support in patient's treatment," 2018.

[34] A. B. Kocaballi, S. Berkovsky, J. C. Quiroz, L. Laranjo, H. L. Tong, D. Rezazadegan, A. Briatore, and E. Coiera, "The per-

sonalization of conversational agents in health care: Systematic review," *J Med Internet Res*, vol. 21, p. e15360, Nov 2019.

[35] T. Bickmore and T. Giorgino, "Health dialog systems for patients and consumers," *Journal of Biomedical Informatics*, vol. 39, no. 5, pp. 556–571, 2006. Dialog Systems for Health Communications.

[36] A. N. Pargellis, H.-K. J. Kuo, and C.-H. Lee, "An automatic dialogue generation platform for personalized dialogue applications," *Speech Communication*, vol. 42, no. 3, pp. 329–351, 2004.

[37] Y. Kim, J. Bang, J. Choi, S. Ryu, S. Koo, and G. G. Lee, "Acquisition and use of long-term memory for personalized dialog systems," in *Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction* (R. Böck, F. Bonin, N. Campbell, and R. Poppe, eds.), (Cham), pp. 78–87, Springer International Publishing, 2015.

[38] C. A. Thompson, M. H. Goker, and P. Langley, "A personalized system for conversational recommendations," *Journal of Artificial Intelligence Research*, vol. 21, pp. 393–428, 2004.

[39] D. J. Litman and S. Pan, "Designing and evaluating an adaptive spoken dialogue system," *User Modeling and User-Adapted Interaction*, vol. 12, no. 2, pp. 111–137, 2002.

[40] M. K. Hong, L. Wilcox, D. Machado, T. A. Olson, and S. F. Simoneaux, "Care partnerships: Toward technology to support teens' participation in their health care," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, (New York, NY, USA), p. 5337–5349, Association for Computing Machinery, 2016.

[41] N. A. Ahmad, M. H. Che, A. Zainal, M. F. Abd Rauf, and Z. Adnan, "Review of chatbots design techniques," *International Journal of Computer Applications*, vol. 181, no. 8, pp. 7–10, 2018.

[42] M. Jain, P. Kumar, R. Kota, and S. N. Patel, "Evaluating and informing the design of chatbots," in *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, (New York, NY, USA), p. 895–906, Association for Computing Machinery, 2018.

[43] J. Zamora, "I'm sorry, dave, i'm afraid i can't do that: Chatbot perception and expectations," in *Proceedings of the 5th International Conference on Human Agent Interaction*, HAI '17, (New York, NY, USA), p. 253–260, Association for Computing Machinery, 2017.

[44] T. W. Bickmore, H. Trinh, S. Olafsson, T. K. O'Leary, R. Asadi, N. M. Rickles, and R. Cruz, "Patient and consumer safety risks when using conversational assistants for medical information: An

observational study of siri, alexa, and google assistant," *J Med Internet Res*, vol. 20, p. e11510, Sept. 2018.

[45] E. Luger and A. Sellen, ""like having a really bad pa": The gulf between user expectation and experience of conversational agents," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, (New York, NY, USA), p. 5286–5297, Association for Computing Machinery, 2016.

[46] S. Boer, "Robot vacuum cleaner with personality traits," July 2017.

[47] Y.-C. Lee, N. Yamashita, Y. Huang, and W. Fu, ""i hear you, i feel you": Encouraging deep self-disclosure through a chatbot," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, (New York, NY, USA), p. 1–12, Association for Computing Machinery, 2020.

[48] B. Inkster, S. Sarda, and V. Subramanian, "An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: Real-world data evaluation mixed-methods study," *JMIR Mhealth Uhealth*, vol. 6, p. e12106, Nov 2018.

[49] K. H. Ly, A.-M. Ly, and G. Andersson, "A fully automated conversational agent for promoting mental well-being: A pilot rct using mixed methods," *Internet Interventions*, vol. 10, pp. 39 – 46, 2017.

[50] D. Cruz-Sandoval and J. Favela, "Semi-autonomous conversational robot to deal with problematic behaviors from people with dementia," in *Ubiquitous Computing and Ambient Intelligence* (S. F. Ochoa, P. Singh, and J. Bravo, eds.), (Cham), pp. 677–688, Springer International Publishing, 2017.

[51] C. Lisetti, R. Amini, and U. Yasavur, "Now all together: Overview of virtual health assistants emulating face-to-face health interview experience," *KI - Künstliche Intelligenz*, vol. 29, no. 2, 2015.

[52] R. J. Beun, W.-P. Brinkman, S. Fitrianie, F. Griffioen-Both, C. Horsch, J. Lancee, and S. Spruit, "Improving adherence in automated e-coaching," in *Persuasive Technology* (A. Meschtscherjakov, B. De Ruyter, V. Fuchsberger, M. Murer, and M. Tscheligi, eds.), (Cham), pp. 276–287, Springer International Publishing, 2016.

[53] E. Bendig, B. Erb, L. Schulze-Thuesing, and H. Baumeister, "The next generation: chatbots in clinical psychology and psychotherapy to foster mental health–a scoping review," *Verhaltenstherapie*, pp. 1–13.

[54] L. Takayama, "Telepresence and apparent agency in human–robot interaction," *The handbook of the psychology of communication technology*, vol. 32, p. 160, 2015.

[55] A. L. Baylor, "The design of motivational agents and avatars," *Educational Technology Research and Development*, vol. 59, no. 2, pp. 291–300, 2011.

[56] A. Gambino, J. Fox, and R. A. Ratan, "Building a stronger casa: extending the computers are social actors paradigm," *Human-Machine Communication*, vol. 1, no. 1, p. 5, 2020.

[57] P. Kenny, T. Parsons, J. Gratch, and A. Rizzo, "Virtual humans for assisted health care," in *Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments*, pp. 1–4, 2008.

[58] L. Fortunati, "Robotization and the domestic sphere," *New Media & Society*, vol. 20, no. 8, pp. 2673–2690, 2018.

[59] P. Ruijten and R. Cuijpers, "Dynamic perceptions of human-likeness while interacting with a social robot," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17, (New York, NY, USA), p. 273–274, Association for Computing Machinery, 2017.

[60] L. Laranjo, A. G. Dunn, H. L. Tong, A. B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A. Y. S. Lau, and E. Coiera, "Conversational agents in healthcare: a systematic review," *Journal of the American Medical Informatics Association*, vol. 25, pp. 1248–1258, 07 2018.

[61] H. Tanaka, H. Negoro, H. Iwasaka, and S. Nakamura, "Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders," *PloS one*, vol. 12, no. 8, p. e0182151, 2017.

[62] D. Ireland, C. Atay, J. J. Liddle, D. Bradford, H. Lee, O. Rushin, T. Mullins, D. Angus, J. Wiles, S. McBride, *et al.*, "Hello harlie: enabling speech monitoring through chat-bot conversations," *Studies in health technology and informatics*, vol. 227, pp. 55–60, 2016.

[63] J. O. De Nieva, J. A. Joaquin, C. B. Tan, R. K. Marc Te, and E. Ong, "Investigating students' use of a mental health chatbot to alleviate academic stress," in *6th International ACM In-Cooperation HCI and UX Conference*, CHIuXiD '20, (New York, NY, USA), p. 1–10, Association for Computing Machinery, 2020.

[64] S. Kim, J. Lee, and G. Gweon, "Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, (New York, NY, USA), p. 1–12, Association for Computing Machinery, 2019.

[65] Z. Callejas, R. López-Cózar, N. Ábalos, and D. Griol, "Affective conversational agents: the role of personality and emotion in spo-

ken interactions," in *Conversational agents and natural language interaction: Techniques and effective practices*, pp. 203–222, IGI Global, 2011.

[66] T. L. Smestad and F. Volden, "Chatbot personalities matters," in *Internet Science* (S. S. Bodrunova, O. Koltsova, A. Følstad, H. Halpin, P. Kolozaridi, L. Yuldashev, A. Smoliarova, and H. Niedermayer, eds.), (Cham), pp. 170–181, Springer International Publishing, 2019.

[67] A. S. Lokman and J. M. Zain, "One-match and all-match categories for keywords matching in chatbot,"

[68] B. AbuShawar and E. Atwell, "Alice chatbot: Trials and outputs," *Computación y Sistemas*, vol. 19, no. 4, pp. 625–632, 2015.

[69] Z. Yan, N. Duan, J. Bao, P. Chen, M. Zhou, Z. Li, and J. Zhou, "DocChat: An information retrieval approach for chatbot engines using unstructured documents," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, (Berlin, Germany), pp. 516–525, Association for Computational Linguistics, Aug. 2016.

[70] A. Ritter, C. Cherry, and W. B. Dolan, "Data-driven response generation in social media," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11,

(USA), p. 583–593, Association for Computational Linguistics, 2011.

[71] W. Wu and R. Yan, “Deep chit-chat: Deep learning for chatbots,” in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR’19, (New York, NY, USA), p. 1413–1414, Association for Computing Machinery, 2019.

[72] D. Jin, Z. Jin, and R. Mihalcea, “Deep learning for text attribute transfer: A survey,” *arXiv preprint arXiv:2011.00416*, 2020.

[73] E. Hovy, “Generating natural language under pragmatic constraints,” *Journal of Pragmatics*, vol. 11, no. 6, pp. 689–719, 1987.

[74] S. Rao and J. Tetreault, “Dear sir or madam, may i introduce the gyafc dataset: Corpus, benchmarks and metrics for formality style transfer,” *arXiv preprint arXiv:1803.06535*, 2018.

[75] A. Madaan, A. Setlur, T. Parekh, B. Poczos, G. Neubig, Y. Yang, R. Salakhutdinov, A. W. Black, and S. Prabhumoye, “Politeness transfer: A tag and generate approach,” *arXiv preprint arXiv:2004.14257*, 2020.

[76] S. Prabhumoye, Y. Tsvetkov, R. Salakhutdinov, and A. W. Black, "Style transfer through back-translation," *arXiv preprint arXiv:1804.09000*, 2018.

[77] C. Gan, Z. Gan, X. He, J. Gao, and L. Deng, "Stylenet: Generating attractive visual captions with styles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3137–3146, 2017.

[78] C. N. d. Santos, I. Melnyk, and I. Padhi, "Fighting offensive language on social media with unsupervised text style transfer," *arXiv preprint arXiv:1805.07685*, 2018.

[79] L. Van den Bercken, R.-J. Sips, and C. Lofi, "Evaluating neural text simplification in the medical domain," in *The World Wide Web Conference*, pp. 3286–3292, 2019.

[80] T. Shen, T. Lei, R. Barzilay, and T. Jaakkola, "Style transfer from non-parallel text by cross-alignment," *arXiv preprint arXiv:1705.09655*, 2017.

[81] R. Voigt, D. Jurgens, V. Prabhakaran, D. Jurafsky, and Y. Tsvetkov, "Rtgender: A corpus for studying differential responses to gender," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[82] H. Jhamtani, V. Gangal, E. Hovy, and E. Nyberg, "Shakespearizing modern language using copy-enriched sequence-to-sequence models," *arXiv preprint arXiv:1707.01161*, 2017.

[83] S. M. Noar, N. G. Harrington, S. K. V. Stee, and R. S. Aldrich, "Tailored health communication to change lifestyle behaviors," *American Journal of Lifestyle Medicine*, vol. 5, no. 2, pp. 112–122, 2011.

[84] J. A. Uribe, J. F. Duitama, and N. G. Gómez, "Personalized message emission in a mobile application for supporting therapeutic adherence," in *2011 IEEE 13th International Conference on e-Health Networking, Applications and Services*, pp. 15–20, IEEE, 2011.

[85] T. R. Fried, C. A. Redding, M. L. Robbins, A. L. Paiva, J. R. O'Leary, and L. Iannone, "Development of personalized health messages to promote engagement in advance care planning," *Journal of the American Geriatrics Society*, vol. 64, no. 2, pp. 359–364, 2016.

[86] D. J. O'keefe, *Persuasion: Theory and research.* Sage Publications, 2015.

[87] F. Muench, K. van Stolk-Cooke, J. Morgenstern, A. N. Kuerbis, and K. Markle, "Understanding messaging preferences to inform

development of mobile goal-directed behavioral interventions," *Journal of Medical Internet Research*, vol. 16, no. 2, p. e14, 2014.

[88] M.-C. Ko and Z.-H. Lin, "Cardbot: A chatbot for business card management," in *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, IUI '18 Companion, (New York, NY, USA), Association for Computing Machinery, 2018.

[89] R. Kocielnik, D. Avrahami, J. Marlow, D. Lu, and G. Hsieh, "Designing for workplace reflection: A chat and voice-based conversational agent," in *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, (New York, NY, USA), p. 881–894, Association for Computing Machinery, 2018.

[90] D. Maulsby, S. Greenberg, and R. Mander, "Prototyping an intelligent agent through wizard of oz," in *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, (New York, NY, USA), p. 277–284, Association for Computing Machinery, 1993.

[91] Y. Choi, T.-J. K. P. Monserrat, J. Park, H. Shin, N. Lee, and J. Kim, "Protochat: Supporting the conversation design process with crowd feedback," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, Jan. 2021.

[92] T. Hu, A. Xu, Z. Liu, Q. You, Y. Guo, V. Sinha, J. Luo, and R. Akkiraju, *Touch Your Heart: A Tone-Aware Chatbot for Customer Care on Social Media*, p. 1–12. New York, NY, USA: Association for Computing Machinery, 2018.

[93] L. Zhou, J. Gao, D. Li, and H.-Y. Shum, "The design and implementation of xiaoice, an empathetic social chatbot," *Computational Linguistics*, vol. 46, no. 1, pp. 53–93, 2020.

[94] R. G. Athreya, A.-C. Ngonga Ngomo, and R. Usbeck, "Enhancing community interactions with data-driven chatbots–the dbpedia chatbot," in *Companion Proceedings of the The Web Conference 2018*, pp. 143–146, 2018.

[95] T. J.-J. Li and O. Riva, "Kite: Building conversational bots from mobile apps," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 96–109, 2018.

[96] E. Cambria, S. Poria, A. Gelbukh, and M. Thelwall, "Sentiment analysis is a big suitcase," *IEEE Intelligent Systems*, vol. 32, no. 6, pp. 74–80, 2017.

[97] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao, "Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey," *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 15169–15211, 2019.

[98] N. Rosruen and T. Samanchuen, "Chatbot utilization for medical consultant system," in *2018 3rd Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*, pp. 1–5, 2018.

[99] J. Simonsen and T. Robertson, *Routledge International Handbook of Participatory Design.* New York, NY, USA: Routledge, 2012.

[100] F. Sleeswijk Visser and V. Visser, "Re-using users: Co-create and co-evaluate," *Personal Ubiquitous Comput.*, vol. 10, p. 148–152, Jan. 2006.

[101] C. S. Pinhanez, "Design methods for personified interfaces.," in *CHIRA*, pp. 27–38, 2017.

[102] V. D'Andrea and M. Teli, "Teaching participatory design: A participatory approach," in *Proceedings of the 11th Biennial Participatory Design Conference*, PDC '10, (New York, NY, USA), p. 223–226, Association for Computing Machinery, 2010.

[103] E. B.-N. Sanders, E. Brandt, and T. Binder, "A framework for organizing the tools and techniques of participatory design," in *Proceedings of the 11th biennial participatory design conference*, pp. 195–198, 2010.

[104] F. Tellez-Bohorquez, A. Tellez, and J. Gonzalez-Tobon, "Empathic design as a framework for creating meaningful experiences," vol. 2, 11 2019.

[105] M. Steen, M. Manschot, and N. De Koning, "Benefits of co-design in service design projects," *International Journal of Design*, vol. 5, no. 2, 2011.

[106] P. Jonell, M. Bystedt, F. I. Dogan, P. Fallgren, J. Ivarsson, M. Slukova, U. Wennberg, J. Lopes, J. Boye, and G. Skantze, "Fantom: A crowdsourced social chatbot using an evolving dialog graph," *Proc. Alexa Prize*, 2018.

[107] T.-H. K. Huang, A. Azaria, and J. P. Bigham, "Instructable-crowd: Creating if-then rules via conversations with the crowd," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, (New York, NY, USA), p. 1555–1562, Association for Computing Machinery, 2016.

[108] W. S. Lasecki, R. Wesley, J. Nichols, A. Kulkarni, J. F. Allen, and J. P. Bigham, "Chorus: A crowd-powered conversational assistant," in *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, (New York, NY, USA), p. 151–162, Association for Computing Machinery, 2013.

[109] X. Liang, R. Ding, M. Lin, L. Li, X. Li, and S. Lu, "Ci-bot: a hybrid chatbot enhanced by crowdsourcing," in *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data*, pp. 195–203, Springer, 2017.

[110] J. Sedoc, D. Ippolito, A. Kirubarajan, J. Thirani, L. Ungar, and C. Callison-Burch, "Chateval: A tool for chatbot evaluation," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pp. 60–65, 2019.

[111] Y. Choi, H. Shin, T.-J. K. Monserrat, N. Lee, J. Park, and J. Kim, "Supporting an iterative conversation design process," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, (New York, NY, USA), p. 1–8, Association for Computing Machinery, 2020.

[112] J. L. McAdam, S. Arai, and K. A. Puntillo, "Unrecognized contributions of families in the intensive care unit," *Intensive Care Medicine*, vol. 34, pp. 1097–1101, Jun 2008.

[113] N. K. Arora, L. J. Finney Rutten, D. H. Gustafson, R. Moser, and R. P. Hawkins, "Perceived helpfulness and impact of social support provided by family, friends, and health care providers to

women newly diagnosed with breast cancer," *Psycho-Oncology*, vol. 16, no. 5, pp. 474–486, 2007.

[114] J. xWagner, M. Burg, and B. Sirois, "Social support and the transtheoretical model: Relationship of social support to smoking cessation stage, decisional balance, process use, and temptation," *Addictive behaviors*, vol. 29, p. 1039—1043, July 2004.

[115] C. Snyder, "Assessing the subjective impact of caregiving on informal caregivers of cancer patients," *Outcomes assessment in cancer: Measures, methods, and applications*, pp. 329–345, 2005.

[116] C. B. Wortman, "Social support and the cancer patient," *Cancer*, vol. 53, no. S10, pp. 2339–2360, 1984.

[117] D. Popenoe, "American family decline, 1960-1990: A review and appraisal," *Journal of Marriage and the Family*, pp. 527–542, 1993.

[118] M. Bates, "Health care chatbots are here to help," *IEEE Pulse*, vol. 10, pp. 12–14, May 2019.

[119] S. Sundar, *Self as source: Agency and customization in interactive media*, vol. 9780203926864, pp. 58–74. Routledge Taylor & Francis Group, 5 2008.

[120] L. Dindo, J. R. Van Liew, and J. J. Arch, "Acceptance and commitment therapy: a transdiagnostic behavioral intervention

for mental health and medical conditions," *Neurotherapeutics*, vol. 14, no. 3, pp. 546–553, 2017.

[121] T. W. Bickmore, S. E. Mitchell, B. W. Jack, M. K. Paasche-Orlow, L. M. Pfeifer, and J. O'Donnell, "Response to a relational agent by hospital patients with depressive symptoms," *Interacting with Computers*, vol. 22, pp. 289–298, 09 2010.

[122] M. Jordan, H. Lanham, B. Crabtree, P. Nutting, W. Miller, K. Stange, and R. McDaniel, "The role of conversation in health care interventions: Enabling sensemaking and learning," *Implementation Science*, vol. 4, 4 2009.

[123] E. Broadbent, V. Kumar, X. Li, J. Sollers, 3rd, R. Q. Stafford, B. A. MacDonald, and D. M. Wegner, "Robots with display screens: A robot with a more humanlike face display is perceived to have more mind and a better personality," *PLOS ONE*, vol. 8, pp. 1–9, 08 2013.

[124] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. te Boekhorst, and K. L. Koay, "Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion," *Autonomous Robots*, vol. 24, pp. 159–178, Feb 2008.

[125] T. L. Smestad and F. Volden, "Chatbot personalities matters: Improving the user experience of chatbot interfaces," 2019.

[126] J. Złotowski, E. Strasser, and C. Bartneck, "Dimensions of anthropomorphism: From humanness to humanlikeness," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '14, (New York, NY, USA), p. 66–73, Association for Computing Machinery, 2014.

[127] T. Conde, W. Tambellini, and D. Thalmann, "Behavioral animation of autonomous virtual agents helped by reinforcement learning," in *International Workshop on Intelligent Virtual Agents*, pp. 175–180, Springer, 2003.

[128] J. Alvarez Perez, H. Garcia Goo, A. Sánchez Ramos, V. Contreras, and M. Strait, "The uncanny valley manifests even with exposure to robots," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '20, (New York, NY, USA), p. 101–103, Association for Computing Machinery, 2020.

[129] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *J. Artif. Int. Res.*, vol. 30, p. 457–500, Nov. 2007.

[130] G. Roffo, C. Giorgetta, R. Ferrario, and M. Cristani, "Just the way you chat: Linking personality, style and recognizability in

chats," in *International Workshop on Human Behavior Understanding*, pp. 30–41, Springer, 2014.

[131] G. Roffo, C. Giorgetta, R. Ferrario, W. Riviera, and M. Cristani, "Statistical analysis of personality and identity in chats using a keylogging platform," in *Proceedings of the 16th International Conference on Multimodal Interaction*, pp. 224–231, 2014.

[132] M.-A. Cornier, D. Dabelea, T. L. Hernandez, R. C. Lindstrom, A. J. Steig, N. R. Stob, R. E. Van Pelt, H. Wang, and R. H. Eckel, "The Metabolic Syndrome," *Endocrine Reviews*, vol. 29, pp. 777–822, 12 2008.

[133] F. D. Davis, *A technology acceptance model for empirically testing new end-user information systems: Theory and results.* PhD thesis, Massachusetts Institute of Technology, 1985.

[134] P. Legris, J. Ingham, and P. Collerette, "Why do people use information technology? a critical review of the technology acceptance model," *Information & Management*, vol. 40, no. 3, pp. 191 – 204, 2003.

[135] B. G. Glaser, A. L. Strauss, and E. Strutzel, "The discovery of grounded theory; strategies for qualitative research," *Nursing research*, vol. 17, no. 4, p. 364, 1968.

[136] H. T. Reis and N. Collins, “Measuring relationship properties and interactions relevant to social support,” *Social support measurement and intervention: A guide for health and social scientists*, pp. 136–192, 2000.

[137] C. Torrey, S. Fussell, and S. Kiesler, “How a robot should give advice,” in *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, HRI ’13, p. 275–282, IEEE Press, 2013.

[138] L. Bloom and M. Lahey, “Language development and language disorders.,” 1978.

[139] S.-S. Lee, J. Chae, H. Kim, Y.-k. Lim, and K.-p. Lee, “Towards more natural digital content manipulation via user freehand gestural interaction in a living room,” in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp ’13, (New York, NY, USA), p. 617–626, Association for Computing Machinery, 2013.

[140] “Tlk.io,” 2011.

[141] C. Lewis, *Using the” thinking-aloud” method in cognitive interface design*. IBM TJ Watson Research Center Yorktown Heights, NY, 1982.

[142] L. Clark, N. Pantidi, O. Cooney, P. Doyle, D. Garaialde, J. Edwards, B. Spillane, E. Gilmartin, C. Murad, C. Munteanu, V. Wade, and B. R. Cowan, "What makes a good conversation? challenges in designing truly conversational agents," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, (New York, NY, USA), p. 1–12, Association for Computing Machinery, 2019.

[143] T. J. Burleigh, J. R. Schoenherr, and G. L. Lacroix, "Does the uncanny valley exist? an empirical test of the relationship between eeriness and the human likeness of digitally created faces," *Computers in Human Behavior*, vol. 29, no. 3, pp. 759 – 771, 2013.

[144] E.-M. MERZ, N. S. CONSEDINE, H.-J. SCHULZE, and C. SCHUENGEL, "Wellbeing of adult children and ageing parents: associations with intergenerational support and relationship quality," *Ageing and Society*, vol. 29, no. 5, p. 783–802, 2009.

[145] K. Lukoff, T. Li, Y. Zhuang, and B. Y. Lim, "Tablechat: Mobile food journaling to facilitate family support for healthy eating," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, Nov. 2018.

[146] Y. Fukukawa, S. Tsuboi, N. Niino, F. Ando, S. Kosugi, and H. Shimokata, "Effects of social support and self-esteem on depressive symptoms in japanese middle-aged and elderly people," *Journal of epidemiology*, vol. 10, p. S63—9, Apr. 2000.

[147] A.-M. Rosland, E. Kieffer, B. Israel, M. Cofield, G. Palmisano, B. Sinco, M. Spencer, and M. Heisler, "When is social support important? the association of family support and professional support with specific diabetes self-management behaviors," *Journal of general internal medicine*, vol. 23, p. 1992—1999, December 2008.

[148] E. Litwak, *Helping the elderly: The complementary roles of informal networks and formal systems.* Guilford Press, 1985.

[149] M. B. Sussman, "Family, bureaucracy, and the elderly individual: An organizational/linkage perspective," *Family, bureaucracy, and the elderly*, pp. 2–20, 1977.

[150] P. A. Thomas, H. Liu, and D. Umberson, "Family Relationships and Well-Being," *Innovation in Aging*, vol. 1, 11 2017. igx025.

[151] T. Nadarzynski, O. Miles, A. Cowie, and D. Ridge, "Acceptability of artificial intelligence (ai)-led chatbot services in healthcare: A mixed-methods study," *DIGITAL HEALTH*, vol. 5, p. 2055207619871808, 2019. PMID: 31467682.

[152] P. Srivastava and N. Singh, "Automatized medical chatbot (medibot)," in *2020 International Conference on Power Electronics IoT Applications in Renewable Energy and its Control (PARC)*, pp. 351–354, 2020.

[153] S. A. Hussain, F. Ogundimu, and S. Bhattarai, "Mobile phone-based chatbot for family planning and contraceptive information," in *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Healthcare Applications* (V. G. Duffy, ed.), (Cham), pp. 342–352, Springer International Publishing, 2019.

[154] A. Fadhil and S. Gabrielli, "Addressing challenges in promoting healthy lifestyles: The al-chatbot approach," in *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '17, (New York, NY, USA), p. 261–265, Association for Computing Machinery, 2017.

[155] J. Casas, E. Mugellini, and O. A. Khaled, "Food diary coaching chatbot," in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, UbiComp '18, (New York, NY, USA), p. 1676–1680, Association for Computing Machinery, 2018.

[156] C. Huang, M. Yang, C. Huang, Y. Chen, M. Wu, and K. Chen, "A chatbot-supported smart wireless interactive healthcare system for weight control and health promotion," in *2018 IEEE Inter-*

national Conference on Industrial Engineering and Engineering Management (IEEM), pp. 1791–1795, 2018.

[157] R. Crutzen, G.-J. Y. Peters, S. D. Portugal, E. M. Fisser, and J. J. Grolleman, "An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: An exploratory study," *Journal of Adolescent Health*, vol. 48, no. 5, pp. 514–519, 2011.

[158] J. Pereira and O. Díaz, "Using health chatbots for behavior change: A mapping study," *J. Med. Syst.*, vol. 43, p. 1–13, May 2019.

[159] P. Mathur and L. Mascarenhas, "Lifestyle diseases: Keeping fit for a better tomorrow," *The Indian journal of medical research*, vol. 149, no. Suppl 1, p. S129, 2019.

[160] F. H. Kanfer, "Self-monitoring: Methodological limitations and clinical applications.," 1970.

[161] W. H. Organization, *Diet, nutrition, and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation*, vol. 916. World Health Organization, 2003.

[162] F. Cordeiro, D. A. Epstein, E. Thomaz, E. Bales, A. K. Jagannathan, G. D. Abowd, and J. Fogarty, "Barriers and negative nudges: Exploring challenges in food journaling," in *Proceedings*

of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, (New York, NY, USA), p. 1159–1162, Association for Computing Machinery, 2015.

[163] S. Gabrielli, K. Marie, and C. D. Corte, "Slowbot (chatbot) lifestyle assistant," in *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '18, (New York, NY, USA), p. 367–370, Association for Computing Machinery, 2018.

[164] A. Panicker, K. Basu, and C.-F. Chung, "Changing roles and contexts: Symbolic interactionism in the sharing of food and eating practices between remote, intergenerational family members," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, May 2020.

[165] L. R. Pina, S.-W. Sien, T. Ward, J. C. Yip, S. A. Munson, J. Fogarty, and J. A. Kientz, "From personal informatics to family informatics: Understanding family practices around health monitoring," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, (New York, NY, USA), p. 2300–2315, Association for Computing Machinery, 2017.

[166] Y. Kim, S. Ji, H. Lee, J.-W. Kim, S. Yoo, and J. Lee, """my doctor is keeping an eye on me!": Exploring the clinical applicability of a mobile food logger," in *Proceedings of the 2016 CHI Conference*

on Human Factors in Computing Systems, CHI '16, (New York, NY, USA), p. 5620–5631, Association for Computing Machinery, 2016.

[167] Y. Luo, P. Liu, and E. K. Choe, "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*, CHI '19, (New York, NY, USA), p. 1–13, Association for Computing Machinery, 2019.

[168] E. Thomaz, I. Essa, and G. D. Abowd, "A practical approach for recognizing eating moments with wrist-mounted inertial sensing," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, (New York, NY, USA), p. 1029–1040, Association for Computing Machinery, 2015.

[169] A. B. M. S. U. Doulah, T. Ghosh, D. Hossain, M. H. Imtiaz, and E. Sazonov, ""automatic ingestion monitor version 2" — a novel wearable device for automatic food intake detection and passive capture of food images," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2020.

[170] S. Sen, V. Subbaraju, A. Misra, R. Balan, and Y. Lee, "Annapurna: Building a real-world smartwatch-based automated food journal," in *2018 IEEE 19th International Symposium on "A*

World of Wireless, Mobile and Multimedia Networks" (WoWMoM), pp. 1–6, 2018.

[171] F. Cordeiro, E. Bales, E. Cherry, and J. Fogarty, "Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, (New York, NY, USA), p. 3207–3216, Association for Computing Machinery, 2015.

[172] P. P.-T. Hsu, J. Zhao, K. Liao, T. Liu, and C. Wang, "Allergybot: A chatbot technology intervention for young adults with food allergies dining out," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, (New York, NY, USA), p. 74–79, Association for Computing Machinery, 2017.

[173] Y. Zhang and A. G. Parker, "Eat4thought: A design of food journaling," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, (New York, NY, USA), p. 1–8, Association for Computing Machinery, 2020.

[174] J. Casas, E. Mugellini, and O. A. Khaled, "Food diary coaching chatbot," in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and*

Ubiquitous Computing and Wearable Computers, pp. 1676–1680, 2018.

[175] "Lark," 2020.

[176] "Lysa," 2017.

[177] "Forksy," 2016.

[178] "Tasteful," 2020.

[179] E. B. Kahn, L. T. Ramsey, R. C. Brownson, G. W. Heath, E. H. Howze, K. E. Powell, E. J. Stone, M. W. Rajab, and P. Corso, "The effectiveness of interventions to increase physical activity: a systematic review," *American journal of preventive medicine*, vol. 22, no. 4, pp. 73–107, 2002.

[180] N. Countdown *et al.*, "Ncd countdown 2030: worldwide trends in non-communicable disease mortality and progress towards sustainable development goal target 3.4," *Lancet (London, England)*, vol. 392, no. 10152, pp. 1072–1088, 2018.

[181] T. C. Clarke, B. W. Ward, G. Freeman, and J. S. Schiller, "Early release of selected estimates based on data from the january–march 2015 national health interview survey," *National Center for Health Statistics*, pp. 1–120, 2015.

[182] S. Consolvo, D. W. McDonald, and J. A. Landay, "Theory-driven design strategies for technologies that support behavior change

in everyday life," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, (New York, NY, USA), p. 405–414, Association for Computing Machinery, 2009.

[183] B. S. McEwen and E. Stellar, "Stress and the individual: Mechanisms leading to disease," *Archives of internal medicine*, vol. 153, no. 18, pp. 2093–2101, 1993.

[184] M. F. Scheier, J. K. Weintraub, and C. S. Carver, "Coping with stress: divergent strategies of optimists and pessimists.," *Journal of personality and social psychology*, vol. 51, no. 6, p. 1257, 1986.

[185] K. M. Richardson and H. R. Rothstein, "Effects of occupational stress management intervention programs: a meta-analysis.," *Journal of occupational health psychology*, vol. 13, no. 1, p. 69, 2008.

[186] Y.-C. Lee, N. Yamashita, and Y. Huang, "Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, May 2020.

[187] G. I. Hess, G. Fricker, and K. Denecke, "Improving and evaluating emma's communication skills: a chatbot for managing medication," *Stud Heal Technol Inf*, vol. 259, pp. 101–4, 2019.

[188] A. Fadhil, G. Schiavo, and Y. Wang, "Coachai: A conversational agent assisted health coaching platform," *CoRR*, vol. abs/1904.11961, 2019.

[189] R. MUGGE, J. P. SCHOORMANS, and H. N. SCHIFFERSTEIN, "17 - product attachment: Design strategies to stimulate the emotional bonding to products," in *Product Experience* (H. N. Schifferstein and P. Hekkert, eds.), pp. 425 – 440, San Diego: Elsevier, 2008.

[190] J. Vincent, "Emotional attachment to mobile phones: An extraordinary relationship," *Mobile World: Past, Present and Future*, p. 95, 2006.

[191] J. Vincent, "Emotional attachment and mobile phones," *Knowledge, Technology & Policy*, vol. 19, no. 1, pp. 39–44, 2006. Specialising in the social shaping of mobile communications, she is also studying for a PhD. Prior to joining DWRC in 2001 she worked for twenty years in the European mobile communications industry.

[192] E. S. Anderson, R. A. Winett, and J. R. Wojcik, "Self-regulation, self-efficacy, outcome expectations, and social support: Social cognitive theory and nutrition behavior," *Annals of Behavioral Medicine*, vol. 34, no. 3, pp. 304–312, 2007. The study conducted at the Center for Research in Health Behavior, Department of

Psychology, Virginia Tech, Blacksburg, was supported by a grant from the National Cancer Institute (RO1CA79469). The trial is registered in the National Cancer Institute's PDQ trial registry (VPISU-05-209).

[193] B. Graf, M. Krüger, F. Müller, A. Ruhland, and A. Zech, "Nombot: simplify food tracking," in *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, pp. 360–363, 2015.

[194] G. H. Beaton, J. Milner, P. Corey, V. McGuire, M. Cousins, E. Stewart, M. de Ramos, D. Hewitt, P. V. Grambsch, N. Kassim, and J. A. Little, "Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation," *The American Journal of Clinical Nutrition*, vol. 32, pp. 2546–2559, 12 1979.

[195] O. Perski, A. Blandford, R. West, and S. Michie, "Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis," *Translational behavioral medicine*, vol. 7, p. 254—267, June 2017.

[196] C. E. Short, A. DeSmet, C. Woods, S. L. Williams, C. Maher, A. Middelweerd, A. M. Müller, P. A. Wark, C. Vandelanotte, L. Poppe, M. D. Hingle, and R. Crutzen, "Measuring engagement

in ehealth and mhealth behavior change interventions: Viewpoint of methodologies," *J Med Internet Res*, vol. 20, p. e292, Nov 2018.

[197] J. Gratch, A. Okhmatovskaia, F. Lamothe, S. Marsella, M. Morales, R. J. van der Werf, and L.-P. Morency, "Virtual rapport," in *Proceedings of the 6th International Conference on Intelligent Virtual Agents*, IVA'06, (Berlin, Heidelberg), p. 14–27, Springer-Verlag, 2006.

[198] Y. Hwang, S. Song, D. Shin, and J. Lee, "Linguistic features to consider when applying persona of the real persona to the text-based agent," in *Proceedings of the 22th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '20 Adjunct, (New York, NY, USA), ACM, 2020.

[199] K. B. Adamo, J. A. Rutherford, and G. S. Goldfield, "Effects of interactive video game cycling on overweight and obese adolescent health," *Applied Physiology, Nutrition, and Metabolism*, vol. 35, no. 6, pp. 805–815, 2010.

[200] T. Baranowski, J. Baranowski, D. Thompson, R. Buday, R. Jago, M. J. Griffith, N. Islam, N. Nguyen, and K. B. Watson, "Video game play, child diet, and physical activity behavior change: A randomized clinical trial," *American journal of preventive medicine*, vol. 40, no. 1, pp. 33–38, 2011.

[201] A. Direito, Y. Jiang, R. Whittaker, and R. Maddison, "Apps for improving fitness and increasing physical activity among young people: the aimfit pragmatic randomized controlled trial," *Journal of medical Internet research*, vol. 17, no. 8, p. e210, 2015.

[202] J.-N. Kramer, F. Künzler, V. Mishra, B. Presset, D. Kotz, S. Smith, U. Scholz, and T. Kowatsch, "Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: Protocol of a microrandomized trial," *JMIR research protocols*, vol. 8, no. 1, p. e11540, 2019.

[203] M. Piao, H. Ryu, H. Lee, and J. Kim, "Use of the healthy lifestyle coaching chatbot app to promote stair-climbing habits among office workers: Exploratory randomized controlled trial," *JMIR mHealth and uHealth*, vol. 8, no. 5, p. e15085, 2020.

[204] T. N. Stephens, A. Joerin, M. Rauws, and L. N. Werk, "Feasibility of pediatric obesity and prediabetes treatment support through tess, the ai behavioral coaching chatbot," *Translational behavioral medicine*, vol. 9, no. 3, pp. 440–447, 2019.

[205] E. Ruane, A. Birhane, and A. Ventresque, "Conversational ai: Social and ethical considerations.," in *AICS*, pp. 104–115, 2019.

[206] A. Satariano, "Gdpr, a new privacy law, makes europe worlds leading tech watchdog," 2018.

[207] R. E. Evans and P. Kortum, "The impact of voice characteristics on user response in an interactive voice response system," *Interacting with Computers*, vol. 22, no. 6, pp. 606–614, 2010.

[208] K. Saffarizadeh, M. Boodraj, T. M. Alashoor, *et al.*, "Conversational assistants: Investigating privacy concerns, trust, and self-disclosure.," in *ICIS*, 2017.

[209] J. Weizenbaum, "Eliza—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.

Appendix

In this appendix, the study materials we used for the evaluation are presented. These materials include survey questions used during the experiments in the study 1, 2 and 3.

Survey Used in Study 1

1. 나는 해당 챗봇이 유용하다고 느낀다.

1	2	3	4	5	6	7
전혀 아니다						매우 그렇다

2. 나는 해당 챗봇이 사용하기 쉽다고 느낀다.

1	2	3	4	5	6	7
전혀 아니다						매우 그렇다

3. 나는 해당 챗봇에 신뢰를 느낀다.

1	2	3	4	5	6	7
전혀 아니다						매우 그렇다

4. 나는 해당 챗봇에 친밀감을 느낀다.

1	2	3	4	5	6	7
전혀 아니다						매우 그렇다

Survey Used in Study 2

1. 나이, 성별, 직업, 함께 참여한 사람과의 관계를 기입해주세요.

2. 평소 스마트폰으로 얼마나 자주 지인들과 대화를 하시나요?

3. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "문장을 끊어서 말하는 정도"가 그 사람의 대화의 특수성을 얼마나 잘 나타내나요? ex. 동생 뭐해? vs 동생/뭐해/?

1	2	3	4	5	6	7
거의 나타내지 않는다						매우 많이 나타낸다

4. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "단어의 변형"이 그 사람의 대화의 특수성을 얼마나 잘 나타내나요? ex. 학교 갔어요 vs 학교감?

1	2	3	4	5	6	7
거의 나타내지 않는다						매우 많이 나타낸다

5. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "문장의 구조"가 그 사람의 대화의 특수성을 얼마나 잘 나타내나요? ex. 완전문장 (나 어제 도서관 가서 공부했어) vs 불완전문장 (함 ㅎㅎ)

1	2	3	4	5	6	7
거의 나타내지 않는다						매우 많이 나타낸다

6. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "나를 부르는 호칭"이 그 사람의 대화의 특수성을 얼마나 잘 나타내나요?

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

7. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "이모티콘의 사용"이 그 사람의 대화의 특수성을 얼마나 잘 나타내나요?

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

8. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "욕설의 사용빈도"가 그 사람의 대화의 특수성을 얼마나 잘 나타내나요?

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

9. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "강조표현"이 그 사람의 대화의 특수성을 얼마나 잘 나타내나요? ex. 진짜, 정말, 대박 등

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

10. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "감정

언어"가 그 사람의 대화의 특수성을 얼마나 잘 나타내나요? ex. 우울하다, 행복하다 등

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

11. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "미사여구"가 그 사람의 대화의 특수성을 얼마나 잘 나타내나요? ex. 음,.... 아.... 등

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

12. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "특수기호의 사용"이 그 사람의 대화의 특수성을 얼마나 잘 나타내나요? ex. !!!, , ??? 등

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

13, 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "줄임말 사용"이 그 사람의 대화의 특수성을 얼마나 잘 나타내나요?

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

14. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "걍

탄사의 사용"이 그 사람의 대화의 특수성을 얼마나 잘 나타내나요? ex. 와우!, 오

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

15. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "내용의 전달력"이 그 사람의 대화의 특수성을 얼마나 잘 나타내나요? ex. 그 사람만이 사용하는 특수한 단어들

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

16. 그 사람 만이 사용하는 특수한 단어들을 생각나는데로 적어주세요

17. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "답장이 오는 시간"이 그 사람의 대화의 특수성을 얼마나 잘 나타내나요?

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

18. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 "돌려말하기 기법"이 그 사람의 대화의 특수성을 얼마나 잘 나타내나요?

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

19. 평소 카카오톡이나 문자로 자주 대화하는 지인과의 대화에서 ”문장의 완성도”가 그 사람의 대화의 특수성을 얼마나 잘 나타내나요? ex¿ 완전문장(나 어제 도서관 가서 공부했어) vs 불완전문장 (함 ㅎㅎ)

1	2	3	4	5	6	7

거의 나타내지 않는다 매우 많이 나타낸다

20. 그 밖에 지인이 자주 사용하는 화법이 있으면 생각나는 대로 적어주세요.

Survey Used in Study 3

User Experience

1. 나는 해당 챗봇을 사용하기 전에 관련 헬스케어 태스크를 지속적으로 수행해 왔다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

2. 나는 해당 챗봇을 사용하는 동안에 관련 헬스케어 태스크를 지속적으로 수행하였다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

3. 전반적으로 나는 해당 챗봇이 태스크를 수행하기에 유용하다고 생각한다.

1 2 3 4 5 6 7

전혀 아니다 매우 그렇다

4. 전반적으로 나는 해당 챗봇이 태스크를 수행하기에 사용하기 쉽다고 느꼈다.

1 2 3 4 5 6 7

전혀 아니다 매우 그렇다

5. 챗봇을 사용하는 동안 나는 해당 챗봇을 신뢰하였다.

1 2 3 4 5 6 7

전혀 아니다 매우 그렇다

6. 챗봇을 사용하는 동안 나는 해당 챗봇에 친밀감을 느꼈다.

1 2 3 4 5 6 7

전혀 아니다 매우 그렇다

7. 전반적으로 나는 해당 챗봇을 사용하는 것에 몰입하였다.

1 2 3 4 5 6 7

전혀 아니다 매우 그렇다

8. 챗봇을 사용하는 동안 나는 해당 챗봇에게 감정적 애착을 느꼈다.

1 2 3 4 5 6 7

전혀 아니다 매우 그렇다

9. 챗봇을 사용하는 동안 나는 사람과 상호작용한다고 느꼈다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

10. 챗봇을 사용하는 동안 나는 챗봇에게 불쾌한 감정을 느꼈다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

User Perception

1. "나를 부르는 호칭"이 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

2. "답장이 오는 시간"이 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

3. "문장의 완성도"는 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

4. ”욕설의 사용빈도”는 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

5. ”특수기호의 사용”은 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

6. ”감탄사의 사용”은 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

7. ”단어의 변형”은 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

8. 챗봇의 ”전달력”은 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

9. ”강조 표현”은 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1	2	3	4	5	6	7

전혀 아니다 매우 그렇다

10. ”미사여구”는 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1 2 3 4 5 6 7

전혀 아니다 매우 그렇다

11. ”줄임말”은 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1 2 3 4 5 6 7

전혀 아니다 매우 그렇다

12. ”감정의 전달”은 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1 2 3 4 5 6 7

전혀 아니다 매우 그렇다

13. ”돌려말하기”는 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1 2 3 4 5 6 7

전혀 아니다 매우 그렇다

14. ”문장의 구조”는 내가 챗봇을 더 나의 실제 가족구성원처럼 인지하게 하였다.

1 2 3 4 5 6 7

전혀 아니다 매우 그렇다

국문초록

디지털 헬스케어(Digital Healthcare) 기술의 발전은 일상 헬스케어 영역에서의 혁신을 주도 하고 있다. 이는 의학 전문가들의 정확한 진단, 질병의 치료를 도울 뿐만 아니라 사용자가 스스로 일상에서 자기관리를 할 수 있도록 돕는다. 디지털 헬스케어 기술의 대표적인 목표 중 하나는 효과적으로 헬스케어 서비스를 개인화 시키는 것인데, 이러한 측면에서 대화형 인공지능(Conversational AI)은 사용하기 쉽고 효율적인 비용으로 개인화된 서비스를 제공할 수 있기에 주목받고 있다.

기존의 개인화된 케어 서비스들의 경우는 대부분 의료기관의 질병치료 서비스들에 포함되었는데, 대화형 인공지능은 이러한 개인화된 케어 서비스의 영역을 일상에서의 질병 예방을 위한 관리로 확장하는데 기여한다. 일대일 대화를 통해 맞춤형 교육, 테라피, 그외의 관련 정보 등을 제공할 수 있다는 측면에서 일상 헬스케어에 적합한 디지털 헬스케어 기술로의 활용도가 높다. 이러한 이점으로 인해 다양한 역할을 가진 대화형 인공지능들의 개발이 이루어지고 있다.

그러나, 이러한 대화형 인공지능들에게 사용자의 선호도에 적합한 페르소나를 부여하는 연구는 드물게 이루어 지고 있다. 대화형 인공지능의 주요 기능인 자연어 기반 상호작용은 CASA 패러다임(CASA Paradigm)에서 제기하는 사용자가 시스템을 의인화하는 경향을 높인다. 때문에 페르소나에 대한 사용자의 선호도가 지속적인 대화형 인공지능의 사용과 몰입에 영향을 미친다. 또한 대화형 인공지능의 장기적인

사용을 위해서 적절한 사용자와의 사회적, 감정적 상호작용을 디자인 해주어야 하는데, 인지된 페르소나에 대한 사용자의 선호도가 이 과정에도 유의미한 영향을 미친다. 때문에 지속적인 참여가 결과에 큰 영향을 미치는 일상 헬스케어 영역에서 대화형 인공지능을 활용하는데 개인화된 페르소나 디자인이 긍정적인 사용자 경험 및 사용자 건강 증진의 가능성을 높일 것으로 본 연구는 가정한다. 개인화된 페르소나 디자인을 위해 사용자와 현실에서 친밀한 관계에 있는 실존인물(호스트)의 페르소나를 대화형 인공지능에 적용하고 평가하는 것이 본 연구의 핵심적인 아이디어이다.

이를 검증하기 위해서 해당 학위 논문은 총 세 가지의 세부 연구를 포함한다. 첫째는 실존인물의 페르소나 중에서도 일상 건강관리에 적합한 호스트의 페르소나를 탐색하는 연구이다. 둘째는 호스트의 페르소나를 대화형 인공지능에 적용하기 위해 고려해야 할 언어적 요소들을 정의하는 연구이다. 마지막으로는 위의 과정을 통해 개발된 실존하는 인물의 페르소나를 가진 대화형 인공지능이 일상 헬스케어 영역에서 실제 효과를 보이는지를 평가하는 연구이다. 또한 해당 학위논문은 일련의 연구들에서 발견한 결과들을 바탕으로 사용자와 친밀한 관계에 있는 페르소나를 일상 헬스케어를 위한 대화형 인공지능에 적용할 때 고려해야할 디자인 함의점들을 도출하고 가이드라인을 제시한다.

주요어: 대화형 인공지능, 디지털 헬스케어, 페르소나, 사용자 경험
학번: 2017-39859