



Doctoral Dissertation

Doctoral Program in Information Engineering & Computer Science (30th cycle)

CONVERSATIONAL AGENT FOR HEALTH COACHING

*Design, Development and Pilot Evaluation of A
Conversational Agent Assisted Coaching Platform*

Department of Information Engineering & Computer Science
ICT International Doctoral School - University of Trento

By

Ahmed JUMAAH

Supervisor

Prof. Adolfo VILLAFIORITA
Fondazione Bruno Kessler - University of Trento

*Submitted in part fulfilment of the requirements for the degree of Doctor of
Philosophy in Computer Science at the University of Trento,
Tuesday 19th March, 2019*

Declaration

I hereby declare that the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the personal data security. The thesis is my own work and contains nothing which is the outcome of work done in collaboration with others, except where referenced.

Ahmed JUMAAH
Tuesday 19th March, 2019

Dedication

I would like to dedicate this dissertation to my family. A special feeling of gratitude to my loving parents, whose words of encouragement and push for tenacity ring in my ears. I also dedicate this dissertation to my colleagues who have supported me throughout the process. I will always appreciate all the support they have given me throughout the entire doctorate program.

Acknowledgements

I thank *Prof. Adolfo Villafiorita* for guiding me through my research study. Over the past years you have exposed me to new research areas and challenged me to think critically about how to conduct research in the efficient way and how to proceed in the future. More importantly, I learned why we choose an approach or a direction over another, what is the scientific contribution and value of a research work, when to proceed and when to shift direction through the research journey. I learned how to distinguish opinions from facts and create a strong research contribution. For all that you have imparted on me, I am grateful for the opportunity to have been one of your students. Thank you for listening to my ideas and challenging me to think beyond traditionally held beliefs. I would like to acknowledge each of *Dr. Gianluca Schiavo* and *Dario Betti* for their continuous collaboration and support. I am also very grateful for the support of each *Dr. Giandomenico Nollo* and *Ilaria Simonelli*, who helped me set my initial experimentation with the ALMAC association and test the initial prototype. Their support to get in contact with the organization helped to advance my research. I am also grateful for the initial guidance of *Dr. Giampaolo Armellin* who opened my eyes on the real value and applicability of my research. I thank each of *Pietro Molini* and *Dr. Aaron Ciaghi* for their technical support. Finally, I would like to send special thanks to *Dr. Silvia Gabrielli* for her collaboration on various topics on health coaching. I also truly appreciate all the constructive comments and suggestions from both reviewers. I had a lot of fun writing this thesis and I am proud I had the chance to engage in critical research. That said, my work is not finished yet and there is still much to be done to ensure people with poor lifestyle promote their health. I look forward to the road ahead.

Ahmed JUMAAH
Tuesday 19th March, 2019

Abstract

Poor diet and physical inactivity are two of the biggest healthcare challenges we are facing, and are related to individuals lifestyle. In fact, a poor lifestyle is strongly correlated to chronic diseases, the leading causes for morbidity and mortality. Adhering to a healthy diet and following an active lifestyle are thus necessary to promote the overall health. However, maintaining a healthy diet and physically active lifestyle is hard. This is due to poor health literacy, lack of awareness, motivation and effective intervention support. Recent years have seen a blast of mHealth apps for health promotion, targeting in particular dietary behavior change. However, reviews showed difficulties in effective adoption and use of these applications in long-term health promotion. Contemporary approaches have focused on tracking user condition and few have analyzed aspects of user interaction with the system.

To promote individuals health, users can benefit from some form of tailored guidance or coaching. That said, to ensure adequate users support, personalized care with a human agent in the loop can enhance the care delivery. Due to the increasing demand for continuous care by users and the shortage of caregiver resources, current health services are inefficient relative to user support and decreasing caregivers workload.

Digital health devices can act as a key player in providing interactive health activities (via mobile and telemedicine systems), enhancing self-monitoring (through wearable tracker) and tailored coaching (using either automated or manual coaching systems). However, they're ineffective in providing continuous health services and creating a balance between users support and caregivers workload. In addition, even with the technology existence, there is low motivation to maintain a healthy diet or exercise routines.

Individuals use messaging applications as part of their regular daily routines. We harness the power of messaging chatbot systems to provide behavior change interventions for healthy lifestyle promotion. We particularly introduce the role of

chatbot in task automation and adhering users to a health plan. Thus, in this thesis we present the concept of "Conversational User Interface in Health Coaching Interventions" that consists of a just-in-time health services to users and caregivers. We discuss ways to integrate the chatbot to assist caregivers with their tasks and support users with their condition. We get users to cue themselves to action by attaching the chatbot with users' daily messaging routines. The service will eliminate the technology barrier and impairment for the users i.e., elderly. The chatbot accesses reliable user compliance data, sets adherence reminders by condition, and reports daily individuals adherence. The chatbot alerts the coach through a web application in critical cases. The approach facilitates adherence to health interventions by investigating a human-virtual agent mediated coaching approach on user motivation to adhere to the health promotion plan.

The approach was validated with different experimentation phases. Using multiple research methods, this dissertation has made several contributions to the understanding of user motivation and the role of a semi-automated system with a human and virtual agent in tracking individuals with poor lifestyle. We will discuss the main contributions and experimentation results throughout the thesis.

KEYWORDS: *Conversational Agents, Health & Wellness, Behavior Change, Persuasive Technologies, Health Coaching, Machine Learning, Adaptive Feedback.*

Publications List

1. Fadhil et al. "Coach Me: a platform for promoting healthy lifestyle." *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. ACM, 2016.
2. Fadhil et al. "Addressing challenges in promoting healthy lifestyles: the al-chatbot approach." *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM, 2017.
3. Fadhil, A., Schiavo, G., Wang, Y., & Yilma, B. A. (2018, May). *The Effect of Emojis when interacting with Conversational Interface Assisted Health Coaching System*. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 378-383). ACM.
4. Fadhil, A., & Villafiorita, A. (2017, July). *An adaptive learning with gamification & conversational UIs: The rise of CiboPoliBot*. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (pp. 408-412). ACM.
5. Fadhil, A., Schiavo, G., Wang, Y.: *CoachAI: A Conversational Agent Assisted Health Coaching Platform*. *EAI Endorsed Transactions on Ambient Systems*.
6. Fadhil, A., *Assistive Conversational Agent for Health Coaching: A Validation Study*. *Journal of Methods of Information in Medicine*.
7. Fadhil, A., *Patterns in eHealth Coaching Approaches Based on Persuasive Technology: A Systematic Review*. *Journal of Methods of Information in Medicine*.
8. Yunlong Wang, Lingdan Wu, Jan-Philipp Lange, Ahmed Fadhil, and Harald Reiterer. *Persuasive technology in reducing prolonged sedentary behavior at work: A systematic review*. *Smart Health*, 2018.
9. Wang, Y., Fadhil, A., & Reiterer, H. (2018, May). *Supporting Action Planning for Sedentary Behavior Change by Visualizing Personal Mobility Patterns on Smartphone*. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 396-401). ACM.
10. Wang, Y., Fadhil, A., Lange, J. & Reiterer, H. (2018, October) "Integrating Taxonomies into Theory-Based Digital Health Interventions for Behavior Change: A Holistic Framework." *Journal of Medical Internet Research (JMIR) Publications* (2018).

Contents

List of Figures	ix
List of Tables	x
1 Introduction	1
1.1 Health Promotion & Individual's Lifestyle	1
1.2 Challenges in Lifestyle Promotion	3
1.3 Research Objectives	4
1.3.1 Approach	5
1.3.2 Hypotheses	6
1.4 Research Rationale	6
1.4.1 Observations	6
1.5 Key Findings	8
1.6 Thesis Structure	10
2 Literature Review	12
2.1 Introduction	12
2.2 Health Coaching Approaches	12
2.2.1 Health Coaching Types	12
2.2.2 Virtual Health Coaching	13
2.2.3 Mobile Health Coaching	14

2.2.4	Wearable Health Coaching	15
2.3	Conversational Agents in Health Coaching	16
2.4	Behavior Change Theories	18
2.5	Conclusion	20
3	Patterns in eHealth Coaching	21
3.1	Introduction	21
3.2	Aim	22
3.3	Methods	23
3.3.1	Data Source	23
3.3.2	Data Coding	23
3.4	Results	24
3.4.1	Search and Selection	24
3.4.2	Intervention Studies Description	26
3.4.3	Persuasive Technology Analysis	28
3.4.4	Behavioral Theories/Techniques Analysis	29
3.4.5	Data Analysis and Mapping	30
3.4.6	Health Coaching Interventions Effect	32
3.5	Discussion	34
3.6	Persisting Challenges	34
3.7	Implications	34
3.8	Conclusion	35
4	CoachAI Platform	36
4.1	Introduction	36
4.2	Design Rationale	36
4.2.1	Conceptual Model	37

4.2.2	Coaching Layer	38
4.2.3	User Classifier	41
4.2.4	Tasks Scheduler	44
4.2.5	Dialog Engine	44
4.2.6	Wearable Data Integration	45
4.3	CoachAI High-level Architecture	48
4.3.1	Intervention Delivery	50
4.4	Conclusion	50
5	Initial Prototype and Preliminary Evaluation	52
5.1	Introduction	52
5.2	Objectives	53
5.3	Design Rationale	53
5.3.1	Chatbots and Users	53
5.3.2	Chatbots and Caregiver	53
5.3.3	User-Chatbot Interaction Design	54
5.3.4	Coaching Dashboard	55
5.4	Conceptual Model	55
5.4.1	Dialog Engine	56
5.4.2	Monitoring Layer	57
5.5	User Scenario	58
5.6	Hypothesis	59
5.7	Intervention Design	59
5.7.1	Participants	60
5.7.2	Experimental Design	60
5.7.3	Scale & Duration	61
5.8	Activity Data	61

5.9	Outcome Measures	62
5.9.1	Pre-Survey Results	62
5.9.2	Post-Survey Results	63
5.9.3	Post-it Session Analysis	64
5.10	Results	66
5.11	Persisting Challenges	70
5.12	Conclusion	71
6	Second Prototype and Feature Refinement	72
6.1	Introduction	72
6.2	Objectives	73
6.3	Study Design	73
6.3.1	Participants and Experimental Design	75
6.3.2	Chatbot Dialog	76
6.3.3	Procedure	76
6.3.4	Measures	77
6.4	Results	79
6.4.1	Interaction Time Analysis	80
6.5	Discussion	81
6.6	Persisting Challenges	82
6.7	Methodology	82
6.8	Evaluation	83
6.8.1	Initial User Profiling	84
6.8.2	Interaction Feedback	85
6.8.3	Conversation-flow Dead-ends	86
6.8.4	Task Conclusion	86
6.9	UX Design Implications	87

6.10	Persisting Challenges	88
6.11	Conclusions	89
7	Experimentation and Assessment	90
7.1	Introduction	90
7.2	Study Design	91
7.2.1	Participants & Experimental Design	91
7.2.2	Procedure & Measures	92
7.3	Results and Implications	94
7.3.1	Experiment Results	94
7.3.2	Interaction Time Analysis	95
7.3.3	Implications	96
7.4	Discussion	96
7.5	Persisting Challenges	97
7.6	Methods	98
7.6.1	Measurements	98
7.6.2	Study Design	99
7.6.3	Eligibility Criteria	100
7.6.4	Participants	100
7.6.5	Procedures	101
7.7	Results	103
7.7.1	Perceived Usefulness	103
7.7.2	Perceived Ease-of-use	104
7.7.3	Perceived Fun	104
7.7.4	Attitude and Intention	104
7.7.5	Pragmatic Qualities, Hedonic Qualities, Appeal and Social Qualities	105

7.7.6	The Motivation and Volition Phase	106
7.8	User Adherence Analysis	109
7.9	User Familiarity with Chatbot Analysis	109
7.10	Log Data Analysis	110
7.10.1	Interaction Time Analysis	111
7.10.2	Chatbot Functionalities Use Analysis	111
7.11	Topic Specific Interaction Preferences	113
7.12	Participants – Aspects of CoachAI	114
7.12.1	Positive Aspects	114
7.12.2	Poor Aspects	114
7.13	Discussion	115
7.14	Persisting Challenges	116
7.15	Conclusion	116
8	Conclusion and Future Work	117
8.1	Overview	117
8.2	Contribution	118
8.3	Research Findings	118
8.3.1	Conversational Agent	119
8.3.2	Overall Coaching Effect	119
8.3.3	Perceived Credibility & Acceptance	120
8.3.4	Task Automation	120
8.3.5	Individual-Chatbot Relationship	120
8.3.6	Coaching Effect on Activities	121
8.4	Persisting Challenges	121
8.5	Future Directions	122

References	124
Appendix A Experiments Resources	135
A.1 CoachMe Questionnaire	136
A.2 Finite State Machine (FSM) Architecture	142
A.3 CoachAI Entity Relational Diagram (ERD)	143

List of Figures

3.1	The study selection workflow.	26
3.2	Interventions health focus, according to SDC.	27
3.3	Intervention aim frequency.	27
3.4	Coaching technique frequency.	28
3.5	PSD principles usage frequency.	29
3.6	Behavioral theories/techniques frequency.	30
3.7	Overall mapping of PSD – SDC – BT.	33
4.1	CoachAI - Conceptual model.	38
4.2	CoachAI - The coaching dashboard.	39
4.3	CoachAI - Chatbot agent interface.	40
4.4	CoachAI - Excerpts of coach-user interaction dialog.	41
4.5	CoachAI - Initial user clustering.	43
4.6	CoachAI - High-level Finite State Machine (FSM).	45
4.7	CoachAI - Tracker pairing, data tracking and visualization.	46
4.8	CoachAI - Weekly report generated by Fitbit tracker.	47
4.9	CoachAI - Overall system architecture.	49
4.10	CoachAI - Overall users adherence to health activities.	50
5.1	CoachAI - Chatbot application interface.	54
5.2	CoachAI - Users profile.	55

5.3	CoachAI - User adherence profile.	56
5.4	CoachAI - Conceptual model	57
5.5	User performance on CoachAI web application.	62
5.6	Pre-survey results.	63
5.7	The UEQ scales and participants scale ratings.	66
5.8	Participants percussiveness and post results.	69
5.9	The usability and acceptability score.	70
6.1	An interface view of the four bot versions. From left to right (1) Physical wellbeing with emojis (2) Physical wellbeing with plain text (3) Mental wellbeing with emojis (4) Mental wellbeing with plain text.	74
6.2	Scores for the "confidence" scale.	80
6.3	Coaching intervention focus of each chatbot.	84
6.4	Initial user profiling by each of the chatbots.	85
6.5	Interaction feedback provided by each of the chatbots.	85
6.6	Conversation-flow dead-ends compared to normal flow.	86
6.7	Task conclusion provided by each of the chatbots.	87
7.1	CoachAI - Users log data analysis.	95
7.2	CoachAI - Participants familiarity with technology.	101
7.3	CoachAI - Participants sedentary lifestyle classification by the SVM model vs the chatbot.	102
7.4	CoachAI - Experiment design process.	103
7.5	TAM dimensions (average score).	105
7.6	TAM Post-hoc analyses between weeks.	105
7.7	AttrakDiff questionnaire dimensions (average score).	106
7.8	HAPA - Stage model and users adherence.	107

7.9	Top left (physical activity intention), Top right (Healthy diet intention), Bottom left (Mental wellness intention), Bottom right (Health expectation intention) vs individual's adherence data.	108
7.10	User adherence analysis and TAM dimensions.	109
7.11	Chatbot familiarity dimension vs TAM and AttrakDiff dimensions. .	110
7.12	Interaction time with the bot and human.	111
7.13	Average interacting time with the chatbot.	112
7.14	Avg. participants activity interaction over the month.	112
7.15	Participants – Health tracking preferences by topic.	113
7.16	Participants – Coaching style preferences.	114

List of Tables

3.1	Articles searching terms and combination strategy.	24
3.2	Persuasive System Design principles.	25
3.3	Behavior theories and PSD principles mapping.	30
3.4	Stages of disease continuum and PSD principles mapping.	31
3.5	Intervention effectiveness criteria measure.	33
5.1	CoachAI experiment - Participants demographics.	60
5.2	The guidelines to recommend activities [1].	62
5.3	A sample of pre-survey results.	63
5.4	Post-it session analysis.	65
5.5	Evaluation dimensions and users review.	66
5.6	UEQ results.	68
5.7	The usability and acceptability Mean and SD.	70
6.1	Examples from the two dialogs used in the study.	75
6.2	Characteristics and descriptive statistics for the experimental groups. The Mann-Whitney test was used to test for differences between the two groups.	76
6.3	Scales and items used in the questionnaire. Items followed by an as- terisk were inverted for the analysis. Cronbach's alphas are reported for each scale.	78

6.4	Questionnaire results for the interaction experience. Average values (SDs in parentheses)	79
6.5	Average time taken to complete the dialogs (in seconds)	80
6.6	Chatbots feature comparison.	83
7.1	Participants descriptive statistics.	91
7.2	Chatbot questionnaire topics & examples.	92
7.3	Scales and items used in the USE questionnaire.	93
7.4	Scales and items used in the questionnaire. Items followed by an asterisk were inverted for the analysis. Cronbach's alphas are reported for each scale.	94
7.5	CoachAI - USE and WAI Questionnaire (Scoring). $\oplus$: One sample t-test (middle scale value USE =3, WAI =4)	95
7.6	CoachAI - Study dimensions and their definition	98
7.7	Participants descriptive statistics.	101

Chapter 1

Introduction

1.1 Health Promotion & Individual's Lifestyle

Chronic non-communicable diseases (NCDs) are among the leading causes of morbidity and mortality [1–3]. The disproportional impact of NCDs are evolving into a staggering economic burden [4]. Most NCDs are the results of individual's lifestyle, such as poor diet and sedentary habits. These lifestyle patterns are key risk factors contributing to NCD development, such as type 2 diabetes [5, 6]. In addition, more of the world population suffers from conditions related to eating disorder (e.g., binge eating disorder), rather than from problems related to lack of food [3]. In fact, societies are facing a growing epidemic of overweight and obesity, due to the frequent energetic imbalance between energy-dense food consumption and reduced daily physical exercises. This health transition, favored by demographic changes towards aging populations, is occurring at an increased pace in urban societies widely exposed to lifestyle modernization and sedentary occupation [7].

The lifestyle related risk factors are modifiable and the same in both genders. These risk factors are linked with the lifestyle that relies on modifiable behaviors which can help prevent or delay a range of life-threatening NCDs [8]. However, the difficulty of sustaining a behavior change is well defined in the literature [9, 10]. Individual motivation represents the main obstacle to the adoption of healthy lifestyle, leading to a relapses and inability to change unhealthy lifestyle. Moreover, depending on their age, gender or ecosystem, individuals show different levels of interest towards the same behavior change intervention. For example, the impact of healthy dietary

determinants, such as disease concerns, is a significant motivator for healthy eating attitudes among females, but not among males [11]. Moreover, the effect of social networks on health is a topic of growing interest, particularly in an increasingly connected world [12]. Many people are influenced by others and may change their behavior based on what others think about them or follow advices provided by their peers. For example, PatientsLikeMe¹ is a patient network through which patients connect with others sharing the same disease or condition and track or share their own experiences. Peer pressure or support have an important role to play on promoting healthy diet and physical activity, however, limited research has examined the role of peers in this context [13]. A study [13] examined the relationship between self-efficacy for healthy eating, parent and peer support for healthy and unhealthy eating patterns. Two patterns were identified labelled ‘healthy food intake’ and ‘unhealthy food intake’. Lower self-efficacy for healthy eating and higher peer support for unhealthy eating were associated with ‘unhealthy food intake’. Higher self-efficacy was associated with ‘healthy food intake’. The study concluded that peer support for unhealthy eating may be beneficial in improving dietary choices among adolescents. A long-term follow-up can help individuals achieve a sustainable healthy behavior change. Several approaches emerged to support sustainable behavior change. Researchers have used mobile applications [14], wearables and trackers [15], and other health tracking systems [16] as tools to promote healthy diet and physical activity, to send users reminders about upcoming events and to provide users with feedback on their daily logged data. However, adhering to a health plan is hard to sustain, since it is specific and requires attentive planning and monitoring that goes beyond the traditional approaches relying on automated algorithms or manual tracking. Automated algorithms are unable to respond well to individuals need and can’t accurately tailor the health activity to their preferences. On the other hand, manual approaches put a burden on caregivers shoulders and are inefficient in tracking many individual related health parameters, such as daily diet or steps. Moreover, maintaining user motivation to engage in regular activity routine is hard. On the contrary, structured health activities assigned by a virtual agent and tailored by a human coach has been linked to positive health outcomes [17]. Hence, tackling the issue of poor diet and physical inactivity requires a joint focus to promote novel approaches in this direction [18, 19].

¹<https://www.patientslikeme.com>

1.2 Challenges in Lifestyle Promotion

Although the perceived health awareness of healthy diet and physical activity, many seem unable to reverse existing unhealthy habits. Several barriers influence healthy diet and exercise promotion. For example, individuals lack of motivation, technology barriers and easy access to health activities are barriers for health promotion [20]. Among the factors influencing individuals motivation, self-efficacy and individuals perceived ability, which are a strong predictors of adherence to a healthy lifestyle. Research found that self-efficacy is less exhibited in elderly than other demographics [21]. Moreover, individuals find it hard to integrate mobile apps into their daily routine behavior, and most apps require a steep learning curve [22]. Along with these challenges, a lack of access to timely tailored health and emotional support increases the odds of technology abandonment.

There is a link between the technical approaches, be it automatic or manual, and adverse user reaction towards the technology (e.g., user acceptance of the technology) [23–25]. Approaches used telemedicine systems to track user condition and provide the necessary health intervention. Yet, such systems are cost ineffective to build and still require continuous caregivers follow-up which is not always possible [26]. Oftentimes such systems double caregivers efforts to track individuals condition [23]. Providing care support to individuals with a variety of conditions including advanced age, dementia, and cancer is commonly perceived as a chronic stressor, and caregivers often experience negative psychological, behavioral, and physiological effects on their daily lives and health [27]. Hence, research studies [27, 28] suggest that future interventions should focus on caregivers as well as patients, with particular attention to caregivers' perceptions of patient care, and seek to change both positive and negative effects of caregiving.

The long-term effect of approaches relied on mobile apps to deliver health intervention is still questionable [29]. Most mobile apps are either abandoned or uninstalled after a few uses. This is often due to lack of desired functionalities or inability to use an app [24]. For example, adopting mobile apps by elderly with low technology literacy [30] is a burden associated with health technology abandonment.

Although the portability of wearable technologies, they also suffer from discontinuation of use [25]. Research has also shown that existing health technologies are still unable to completely substitute caregivers in handling complicated health tasks [31]. Research showed that individuals view their caregivers as a role model,

and perceive caregivers practice as more credible and motivating to achieve healthy lifestyle choices [31].

The current rise in computing power revolutionized the way we establish a connection among individuals and software agents to deliver interactive content. Many approaches exist which persuade individuals to promote their diet and physical activity. For example, gamification and serious games [32, 33], and persuasive mobile apps [34] are used to promote user health through engagement and increased motivation. However, current applications are generic and are not customizable to personalize them to fit individuals interest and ability.

Providing a health plan may vary from a caregiver to another and among individuals [35]. Individuals may perceive the same health activity differently, due to their heterogeneity, even though they might share the same health condition [36]. Hence, we need a tailored health activity for the individuals that resonates with their needs and ensures the intervention effectiveness.

Structured health activities have been linked to positive outcome in physical and mental wellbeing [17]. Research also suggests that the more familiar a behavior, the more likely the user is to do it [37]. This is an important aspect in inducing user engagement and enhancing the health task delivery. The effectiveness of this approach is well studied in the literature [37–39].

In summary, current solutions are unable to provide support beyond the defined technical functionalities. These solutions can be grouped into health records to track individuals, and personal self-management. Both approaches proved ineffective in promoting individuals lifestyle. The former increases caregivers workload, whereas the later suffers from a steep learning curve. Moreover, human agents are better than software agents in using the right voice tone and providing the right emotional support to individuals throughout their health promotion journey [40]. In this context, human and virtual agent mediated health coaching can facilitate long-term lifestyle management that requires the adoption of complex behavior change recommendations.

1.3 Research Objectives

Given the aforementioned challenges, we introduce the concept of conversational agents as a tool to track individuals in health promotion applications. This considers

user demographics, input/output approaches that fit with user ability, motivation and action. While for minor health issues, conversational agent can take the burden off the caregivers shoulders, communicating such issues requires compassion and emotions that remain in the exclusive domain of humanity. Hence, a human in the loop can guarantee tailoring the activity and emotional support, while the conversational agent can handle mundane tasks.

This dissertation presents the design and implementation of a semi-automated health promotion tool that relies on a conversational agent to deliver health coaching tasks and to simplify the user-coach interaction. The underlying assumptions on which this research is based are: (1) goal-setting technique is effective in improving the feedback and coaching support; (2) semi-automated approaches can overcome the limitations of both fully automated and fully manual systems; (3) system scalability will help adopt the tool into different domains where a coach-user relation exists; finally, (4) building the tool inline with user daily routine (e.g., messaging habit) will help users cue themselves to action.

1.3.1 Approach

We introduce structured health activity assignment and tracking by a human-virtual agent mediated system. This allows the heterogeneous individual groups to participate in scheduled and supervised training programs by interacting with the chatbot. We are interested in enabling individuals to participate in health activities and motivate them to follow the activities in the long-term. The activities are tailored to each individual based on their personal profile created initially by the chatbot, and presented to the human coach. We study and report the intervention design to understand the effect of human-virtual agent mediated coaching on user adherence to exercise and promote healthy diet.

The main research question we tend to answer is: *how can technology be effectively integrated into the doctor-patient interaction to promote tailored healthy lifestyle?*. To achieve this, we try to answer the following research questions:

- What are the pros/cons of each medium when building the chatbot dialog?
- How, in which circumstances, the human in the loop provides better outcomes, than the virtual agent alone?

- How to decide task division between the human coach and the virtual chatbot, which allows building/providing individual-appropriate micro-activities?

To answer the above questions it is fundamental to understand the relation between users daily diet and physical activity habits, and their social routine of using messaging applications. We aim to understand how technology can pair heterogeneous users with different levels of motivation and ability with their caregiver. The health coaching application provides individuals with structured diet and physical activity tasks, assigned and remotely monitored by a human agent and tracked by the conversational agent. The human coach helps tailor the activities to fit with users condition (via the web platform). The approach was validated with pilot and cohort studies with diverse demographics. The results of the trial, along with secondary intervention studies using our system, reveals the strong relation between individuals daily activities and delivering them through a messaging application. The experiment also validated the system design, such that it fits with individuals daily routine.

1.3.2 Hypotheses

There are different experimental hypothesis we could formulate for this study. From the aforementioned challenges and our research questions, we hypothesize that:

- H1:** Building the technology in-line with users' habit of using messaging applications will increase the efficacy of the system.
- H2:** A human-actor in the loop will help enhance the coaching effect and increase the system credibility.

1.4 Research Rationale

1.4.1 Observations

Existing approaches suffer from limitations in increasing user engagement and preventing health deterioration. We need a user familiar technology medium, regardless of their demographics. Although the possibility to access high quality data with existing wearables and mobile apps, yet sustainable interaction with such technologies

is a hassle [41]. Moreover, such technologies are unable to provide tailored health activities to the user groups [42]. This research aims to facilitate the caregiver-users interaction by means of conversation and request/response interaction. The long-term goal is to engage users into health-related activities by delivering behavior change health interventions through a conversational agent.

Research demonstrates that text-messaging interventions have positive effects on health outcomes and behaviors [43, 44]. With the rise of messaging platforms individuals spend more time chatting, thus building the virtual coaching bot where they spend their time is indeed relevant. Our platform assists the caregiver with initial baseline user classification (profiling) and activity assignment. The chatbot acts as a virtual coach interacting with the users and providing all the data to the human coach. This provides access to timely user data through informed conversational interface. Although the available self-service channels in the form of digital voice assistants and chatbots (e.g., Siri, Alexa, and Google Assistant), they're unable to provide healthcare services, due to their general purpose and lack of specific domain knowledge [45]. Various APIs and server-less platforms are available to engage users in conversation with AI-based systems. Such interactions however represent a risk both for the users and the caregiver, since it is scarcely predictable where the conversation will lead and the health suggestions such systems might provide. Existing health intervention tools and sensors lack the personal monitoring and most of the data are released as anonymous diaries and there is no communication between the caregiver and user.

Several psychological factors affect health promotion, and hence having a way to handle these factors may help in user support. According to [46] users tend to be more engaged to train when their activities are developed and remotely supervised by a human coach. In a similar vein, [47] found individuals who believed to be coached by a human judged the coaching to be more influential. Our research focuses on user motivation, ability and action while involved in health tracking process. The conversational agent is abstract to cover a vast number of domains in lifestyle promotion. The approach focuses on physical activity and diet promotion. That said, the platform can support other contexts in the health domain. The goal is to enhance preventive healthcare through health promoting activities, hence the platform doesn't focus on managing or monitoring of any critical health conditions. Since CoachAI is intended for lifestyle promotion, the research will focus on techniques for health promotion for subjects with poor lifestyle.

1.5 Key Findings

Building a health activity tracker has to consider the caregivers interaction, then design for the user engagement. The literature review showed none of the existing studies have considered caregivers workload to provide the necessary user support. In addition, automating repetitive caregiver tasks by the chatbot is a novel approach that we investigated and applied in this dissertation. This is preferred over the mobile app or wearable trackers, due to its simplicity [51]. We also learned from the research reviews that social persuasive strategies, where a user interacts with a human-virtual agent mediated system are effective in health promotion.

The research concluded that building the technology in user's familiar environment and inline with their daily routine of using messaging application can effectively increase factors of technology sustainability. We conclude the important role of human actor in the loop to provide tailored activities to users and support their emotions. The presence of a human coach (trainer) is important for sensitive training programs. Although automatic sensing is preferred over manual data collection, yet such technology is not accurate for sensitive user interactions (e.g., user preferences). The findings from Chapters 2 and 3 revealed the existing limitations with health coaching approaches and the common persuasive coaching patterns and the behavioral change techniques used by most studies. From the preliminary experimentation with elderly (see Chapter 5), we found that utilizing a chatbot helped them familiarize with the technology and feel more comfortable using it. We conclude that conversational agent helps provide the necessary follow up to users and assist caregivers.

We also concluded several findings from the two studies discussed in Chapter 6. The chatbot dialog cohort study found that when participants had to score an interaction with a chatbot that asks personal questions on their mental wellbeing, they rated the interaction with high scores with respect to enjoyment, attitude and confidence. Differently, participants rated with lower scores a chatbot that uses emojis when asking information on their physical wellbeing compared to a dialogue with plain text. Individuals who communicated with the bot on mental wellness + emojis were perceived more confident than those communicated with a bot on physical activities + emojis. We concluded that individuals emojis had a positive effect on the dialog when it was emotionally bound. Moreover, the dialog comparison study also in Chapter 6 showed some dialog patterns emerged from the chatbot comparison and concluded the study with UX design implications.

The CoachAI initial testing discussed in Chapter 7 showed that reminders use in health intervention delivery were positively perceived by the users. In addition, using a mixed human-virtual coaching approach was preferred by all participants. The prospective intervention study in Chapter 7 showed that users had different experiences with the application when coached with a human agent, a virtual agent, or a combination of the two. Users who performed exercise activities rated the virtual agent coaching style higher, whereas those who performed health diet activities rated the human agent coaching style higher, and finally those who performed mental wellness activities rated the mixed human and virtual coaching style higher.

Overall, the dissertation validates the role of conversational agents as a persuasive element in building a tailored lifestyle promotion plan which helps caregivers and users have a common understanding and better respond to both sides needs. The design, implementation and evaluation of the system along with other research activities have provided information on various aspects of conversational user interface. Nonetheless, designing health agents that are emotionally aware is still poorly understood. Furthermore, the feasibility of automating just-in-time mHealth interventions delivery via a chatbot has not been fully studied.

The lessons learned from the aforementioned experiments validated the usability of social, yet tailored health promoting application for individuals with poor diet or physically inactive. The conceptual model and design guidelines along with our learning from the user studies provide information on various aspects of semi-automated training applications. This research made key contributions to the literature on techniques in user motivation and engagement.

- Identifying habitual patterns to understand user preferences.
- The role of chatbot in delivering health promoting micro activities.
- Building the technology while adhering to individuals daily routine.
- A socio-technical system that fits with the role of conversational agent as an assistive component.
- The role of human agent in the loop to tailor the health coaching task.
- Quantitative and qualitative results obtained from using our system with a diverse population.

The research is interdisciplinary involving behavior change, health promotion and user engagement. The developed system investigated the ability of conversational agent to establish and maintain social-emotional relationships with users, and to determine if these reactions can increase health behavior change. The chatbot should prove to be a complement to the human agent, but not a replacement. This highlights the need for insights into the psychology of e-coaching, as well as the dynamics of human-computer interactions. Situated at the crossroads of research on health psychology, human-computer interaction, artificial intelligence, and ethics, we believe this research area is of importance to understanding scientific issues of behavior change through ICT systems.

1.6 Thesis Structure

This dissertation emphasizes our study, research, design, development and evaluation over the course of the PhD research on socio-technical systems for promoting individuals lifestyle. The rest of the thesis are organized in chapters explaining parts of the research work. We summarize the contribution of each chapter below.

Chapter 2: Literature Review. Discusses the state of research on conversational agents, health coaching and behavioral change theories. We shed the light on different approaches in lifestyle promotion. We will discuss the role of conversational agent as persuasive coaching strategy. Each section will be concluded with a discussion of existing limitations.

Chapter 3: Patterns in eHealth Coaching. Presents a systematic review on studies focused on eHealth coaching, persuasive technology, behavioral change theories, and identifies key Persuasive System Design (PSD) components in different coaching context.

Chapter 4: CoachAI Platform. Discusses our human-virtual agent mediated coaching platform to target diverse users and provide them with tailored activities.

Chapter 5: Initial Prototype and Preliminary Evaluation. Discusses the design, development and preliminary evaluation of our chatbot supported e-coaching

system. It summarizes the findings of our small scale experimentation with elderly people at the ALMAC² association.

Chapter 6: Second Prototype and Feature Refinement. Discusses the findings from two distinct experiments related to chatbot dialog study and the dialog state comparison between CoachAI and two different conversational agent applications. We present the experiment settings and cohort studies together with their findings and design implications.

Chapter 7: Experimentation and Assessment. Discusses two experimentation studies to evaluate CoachAI. First, the initial evaluation of CoachAI with adult participants and the findings from the testing. Then the prospective intervention settings and evaluation of CoachAI with diverse participant demographics. We present the findings and limitations from each experiment.

Chapter 8: Conclusion and Future Work. Summarizes the dissertation, discusses the conclusion of this research work and highlights the contribution, states limitations and the future directions for this line of research.

²<http://www.almaconlus.it/>

Chapter 2

Literature Review

2.1 Introduction

This chapter discusses the literature on health technologies to engage individuals in healthy behavior change. We aim to cover recent research on motivating individuals pursue a healthy lifestyle through ICT approaches. We evaluate existing body of knowledge on various technical and social approaches in health and wellness promotion. We will present health coaching approaches, behavioral change techniques, and conversational agents in the context of lifestyle promotion. We conclude each section by discussing advantages and shortcomings of existing approaches and the limitations we tackle in this research.

2.2 Health Coaching Approaches

2.2.1 Health Coaching Types

Health coaching is defined as the practice of health education and promotion to help individuals achieve a healthier lifestyle through coaching [52]. Existing health coaching approaches are either fully-automated (the application implements a virtual coach) [53]; fully-manual (a human coach tracks and assigns activities) [54], and semi-automated (the application is partially supported by a virtual coach to carry out repetitive human tasks, and notifies the human coach to intervene in critical cases)

[55]. This system is a negotiation between the previous two and substitutes the task between the algorithm and a human agent. A previous work by Fadhil et al., [20] highlighted the important role of a human actor in the loop, when building health coaching systems to track user's condition. In addition, users are more engaged to train when their activities are developed and tracked by a human supervisor [46], and those who believed to be coached by a human judged the coaching to be more influential [47].

Discussion: The literature review shows the different coaching systems used to improve health behavior change. Each approach differs in the way it delivers the intervention. The literature revealed that both manual and automated health coaching systems have their own limitations. For example, manual systems put a lot of burden on caregivers to deliver the intervention, whereas the automated systems totally rely on an algorithm and misses out the emotional support that is essential in supporting individuals disease journey and the role of human agent in tailoring the activity. Fully automated type is more complicated to build and may involve various machine learning algorithms and IoT devices to handle different tasks autonomously. Finally, coaching systems that integrate a human agent in the loop were perceived positively by the users.

2.2.2 Virtual Health Coaching

Addressing challenges in health promotion requires a long-term follow-up and scalable intervention techniques. Many types of sensors, wearables and mobile applications exist to promote individuals overall health. However, to provide continuous follow-up, we should look for autonomous solutions that handle humans repetitive tasks efficiently. Virtual agents can be an alternative to wearables or mobile apps in health activity tracking. For example, Watson et al., [56] recruited 70 participants with a body mass index (BMI) between 25 and 35 kg/m² and had them meet with a virtual coach, an animated computer agent that ran on their computer, set goals and provide personalized feedbacks. The coach was beneficial in maintaining activity level. However, the activities were tracked and assigned by a virtual agent, with no human intervention. Other approaches [57] relied on a virtual coach to support self-regulation of emotions, especially those related to emotional eaters. The study validated persuasive coaching strategies based on behavior change techniques by translating successful behavior change mechanisms, such as dialectical strategies,

into personalized coaching strategies. However, such approaches rely totally on a software agent to provide the emotional support to individuals, which is still a complex task to achieve and unrealistic. Animated health agents are also used to influence self-management. For example, [58] developed iCat agent which shows different facial expressions and provides cooperative feedback following principles from motivational interviewing (MI).

Discussion: The review stresses several aspects and further challenges to address. Although the added value of software agents to track individuals, however individuals are heterogeneous and have diverse interest and emotional needs. Existing software agents are still unable to handle complex human needs and personalize the application experience to their condition. Beside this, the long-term effect of software systems to provide tailored emotional support requires further investigations. Finally, a mixed human-virtual agent coaching have been found to be more effective than a mere human or agent coaching for health activities [40].

2.2.3 Mobile Health Coaching

Although the large number of mHealth applications are designed with specific target groups, such as elderly, many have been designed for a general audience. In addition, many apps are not designed as a part of physicians practice, hence a distinct research gap exists about technology that supports physician-patient relationship [59].

There is a strong relation between user personality type and the effect of persuasive technologies in the context of mHealth. Research highlighted that mHealth technology designers may achieve more success if they consider individual's personality type [34]. To consider user personality, the app should trigger users into the activities offered by the application. Cueing strategies, such as real-time reminders, are among the most effective methods of persuading individuals to perform healthy behaviors, such as exercising or tracking daily diet [60]. However, according to Fogg's behavioral model [37] the application should pick the right moment to trigger the user with reminders/notifications into an activity. In addition, these reminders often represent a task interruption for users engaged in work activities. Therefore, custom notification is important to provide personalized experiences. Bickmore et al., [61] concluded that the degree of perceived interruption politeness is positively correlated with predicted long-term adherence, but negatively correlated with short-term compliance. Among several previously explored interruption coordination strategies empathic

interruptions are superior overall in gaining both short-term and long-term adherence [61].

Discussion: Most of the aforementioned literature studies were carried out in health coaching settings. Some of the studies incorporated persuasive techniques and gaming elements. Unlike apps integrating gamification techniques to engage users, health coaching is more focused on individuals behavior change, rather than fun and gratification, and hence more applicable for intensive health promotion. Studies showed that reminders and notifications can improve the adherence to long-lasting health activities. However, the majority relied on SMS notifications, which is cost-ineffective and can be obtrusive. We address both findings in our research. Finally, users download and use a marginal number of apps, and abandon them after a few uses. Moreover, individuals find it hard to integrate mobile apps into their daily routine behavior, and most apps require a steep learning curve [22]. We will further consider these findings in our research study.

2.2.4 Wearable Health Coaching

Wearable health trackers are widely adopted to support individuals with daily health tracking. They can assist in steps taken, measure sleeping pattern, and improve healthy diet. Despite the wide adoption of such technology, their real life results are still questionable. While some users derive long-term values from their trackers, others face barriers to integrate the tracker into their daily routine. Studies [25, 62, 63] have analyzed technical aspects of these barriers. Moreover, to motivate users keep up with their personal activities, some wearables adapted motivational techniques, such as gamification and social recognition [62, 63]. The design space to leverage tracking data and persuade health-related behavior change involves issues and strategies to capture information, monitor progress, notify feedback, and provide social support [14, 64].

Most studies consider technical aspects of health tracking technologies and their correlation with user adoption [65]. Little is known about user characteristics and their relation to wearable adoption. Shih et al., [66] performed a six-weeks user study with 30 users with physical activity tracker. The study described user pattern implications, such as helping them be mindful of their physical activity tracker, to further articulate gender differences in the wearable use and adoption.

Health trackers suffer from low adoption rate in the long-term [66]. Clawson et

al., [25] performed an analysis on ads of secondary sales of wearables. The study identified motives and rationale for wearable adoption and forsaking, and suggested more work on behavioral theories to investigate psychological effects of health behavior change techniques. A work by Lee et al., [67] presented experimental results on 80 college students with fitness trackers. The study found that 50% of students stopped wearing their tracker in less than three weeks, and 34% have forsaken the device in nine weeks. Recent research has been focused on identifying key components of health interventions that combines self-tracking and coaching, referred to as persuasive e-coaching [68].

Discussion: The literature review identified the main challenges associated with wearable health trackers, such as long-term adoption, user interaction with the wearable UI, and simplicity. Similar to mHealth apps, wearable health trackers also suffer from user abandonment, in the short or long term [25]. Moreover, most wearable trackers share the same tracking pattern and are intended for general audience, which is not personalized and can't fulfill user interest. Finally, trackers alone are not enough to handle complex user behavior and preferences, instead they act as a data tracking devices. We will further discuss these points throughout our research.

2.3 Conversational Agents in Health Coaching

Conversational agents enable individuals to interact with a software via spoken or text messages to access, create and process information. Conversational agents run structurally repetitive tasks inside a messaging application [69]. Conversation is a way to share knowledge, emotion and its part of our communication makeup. However, designing conversational user interface (CUIs) dialog requires tailoring to fit with application and context. There are particular requirements with designing conversational applications. These requirements include user demographics, chatbot domain, data interaction and dialog structure.

Several integrated care systems employ conversational agents in the context of lifestyle promotion. For example, MOPET [70] is designed to support users throughout exercise sessions by guiding them through fitness trails. It tracks user's exercise and motivates them through message feedbacks. While the system can form effective and enjoyable exercise for users, it does not actually motivate them to start exercising.

FitTrack system [71] is another example that provides daily monitoring of users and motivates them to walk more. It determines user steps through the accelerometer and provides them with feedback once the walk session is terminated. The system makes users aware of their exercise level and rewards them with praise, but doesn't form suggestions or other explicit tactics to motivate user activity.

Existing conversational agents match user requests with a suitable answer extracted from a trained model. For example, if a user asks feedback on his/her physical activity, as in "What was my level of physical activity yesterday?", and "I have not checked my activity level yesterday", they will be processed by the model as relevant intents. Training the chatbot model is essential to connect both questions with correct intents and provide the right answer. Conversational flow uses buttons, speech-to-text or speech-to-speech, or a combination when designing chatbot conversations. With these techniques, the conversational bot can negotiate for an enhanced mode of interaction [72]. Recent research considered theories and methods in constructing and evaluating conversational agent dialog systems [73]. Other studies focused on emotional aspect of the conversation and building empathic accuracy as more important than user expressivity [74].

User experience with chatbot is mainly about their interaction with the dialog. It is a relationship connecting user emotion and chatbot empathic ability, user demographics and personalized experiences. Chatbots should analyze what users like/dislike, to avoid building a technology that cannot trigger the right action. For example, using online bots to call for volunteering action [75], or acceptance of embodied agent designed to establish long-term relationships with older adults [76].

Finally, deep learning in chatbots is mainly used to generate responses from scratch with no predefined structure. The Recurrent Neural Network (RNN) technique is among the widely-applied techniques in this context [77]. To predict the next word in a sentence, we should know which words came before it. RNNs are combined with long-short term memory to remember past conversations up to that point. Nonetheless, existing advanced conversational agent still suffer from falling outside the conversation context and not handling the dialog flow properly. RNN suffers from the vanishing gradient problem, as the context length increases, layers in the unrolled RNN also increase.

Discussion: Research suggests that instant text messaging could be a good solution to tackle several health issues, such as medication non-adherence [78, 79]. Unlike apps, messaging platforms, such as chatbots are easy to use and don't require fa-

miliarity with a specific tool. Since elderly are among the population with highest medication consumption and they are as well more affected by low medication adherence [80], a virtual coaching can be a perfect fit to provide an intuitive technology at low cost. That said, designing conversational user interfaces requires tailoring the dialog to fit with user demographics, bot domain and the dialog context. Moreover, building rich vocabulary becomes necessary in health domain due to the sensitive data and interaction involved [73, 81]. The literature showed that providing emotional support is necessary to support individuals in their health journey. Existing chatbot systems can extract some emotions from user conversation, however, they are still unable to provide reactive emotional support that fits with user preferences. Hence, tailoring with conversational agents is still limited to human agents and can't trigger individual preferences accurately.

The literature revealed that advanced chatbot systems using deep learning also suffer from limitations related to dialog context and user intent. In addition, a fixed finite state systems are used where the system has the initiative at each turn. Dialogue management with finite states describes the flow inside a conversation. Finite states transition can switch contexts and subjects more reliably than an advanced model. In summery, complex systems fall outside context, whereas simple systems are unable to handle complex human requests and behaviors, hence a human in the loop is necessary to cover both gaps.

2.4 Behavior Change Theories

Behavioral interventions have been used to determine clinical utility of health coaching to measure direct/indirect effects on health improvement [82]. For example, e-coaching have been tested in cardiac rehabilitation programs supported by behavior change support systems [83]. The system identifies causes of non-adherence by reasoning over a computational model based on established psychological theories. The results show that tailored interventions leads to positive behavior change.

Among the commonly used psychological models and theories of behavior change in health coaching, we mention the Transtheoretical Model of Behavior Change [84], Technology Acceptance Model [85], Health Belief Model [86] and Fogg's Model [87]. Each of these models focus on an aspect of user behavior and approach the concept of behavior change from a different perspective. The Transtheoretical Model

of behavior Change consists of dividing the behavior change process into stages, namely pre-contemplation, contemplation, preparation, action, maintenance and relapse prevention. Taking the example of following a healthy diet, it is argued that in the pre-contemplation stage, the person is unaware of their poor dietary habit as a problem and has no intention of changing. In the contemplation stage, the person starts to think about changing their behavior, but does not commit to try to change. The Technology Acceptance Model (TAM) [85] is an information theory that models how users come to accept and use a specific technology. The TAM model is considered a behavior change model, dedicated to the acceptance of a technology. In the context of mHealth, the TAM is used to analyze the health behavior change. This is because user's adoption of healthy behavior is reflected from the level of engagement and interaction with the health application.

The Health Belief Model (HBM) [88] was spelled out in terms of four constructs representing the perceived threat and benefits of a behavior, perceived susceptibility, perceived severity, benefits, and barriers. The model is based on the understanding that a person will take a health-related action (i.e., use condoms) if that person: (1) feels that a negative health condition (i.e., HIV) can be avoided, (2) has a positive expectation that by taking a recommended action he will avoid a negative health condition (i.e., using condoms will be effective at preventing HIV), and (3) believes that he can successfully take a recommended health action (i.e., he can use condoms comfortably).

The Health action process approach (HAPA) [89] is an open framework of various motivational and volitional constructs that are assumed to explain and predict individual changes in health behaviors such as improving physical activity levels and dietary behaviors. The model suggests that the adoption, initiation, and maintenance of health behaviors should be conceived of as a structured process including a motivation phase and a volition phase. The former describes the intention formation while the latter refers to planning, and action (initiative, maintenance, recovery).

Finally, Fogg's behavior model [37] was proposed as a framework for understanding human behavior. The FBM asserts that for a behavior to be performed, a person must (1) be sufficiently motivated, (2) can perform the behavior, and (3) be triggered to perform the behavior. These three factors must occur at the same moment, or else the behavior will not execute.

Discussion: The literature review in adapting behavior change theories in health coaching applications revealed that although these theories can be used when build-

ing health coaching systems, few studies experimented with such theories to build personalized coaching systems. Most applications are not tailored to individuals skills and there is no human agent in the loop to personalize the health activity. Hence, we invite more use of these frameworks as base when building health coaching systems. We highlight in Chapter 3 the health coaching patterns and address the main behavior change theories emerged in each pattern.

2.5 Conclusion

In this chapter we discussed the literature body on health coaching approaches, and the role of virtual agents, mobile apps and wearable trackers in health coaching. We then defined conversational agent and the recent rise in chatbots for health coaching. Finally, we concluded the chapter with the research on behavioral theories used in behavior change techniques.

The literature study revealed that current health coaching applications are mostly automated and can not tailor the activity to fit with user preferences and emotions. Moreover, most studies rely on the system to handle all user activities, and none have considered the role of a human agent to track users and provide personalized health activities. Systems that relied on a human agent were mainly manual, and fully relied on the human to handle all the tasks. This is inefficient, since it increases caregivers workload.

The use of chatbots in health coaching is novel and growing. Individuals with cognitive impairments can benefit from the chatbot as a daily assistant. However, to successfully interact with the system, we should consider the issue of user acceptance of the system and carefully sketch the dialogue conversation and the data flow. Current chatbots have speech recognition, natural language processing, even sentiment detection capabilities, however, they are still unable to provide tailored support. Moreover, advanced chatbot systems still suffer from falling outside the conversation context.

In Chapter 3 we present our study on patterns in health coaching from the theoretical and technical perspectives. We identify the most prominent coaching techniques used and the gap in the current approaches. The lessons learned from this chapter and the following chapter drives our path on designing our health coaching system with conversational agent.

Chapter 3

Patterns in eHealth Coaching

3.1 Introduction

The advent in technology have created enormous possibilities for health tracking systems that are capable of supporting individuals at various stages in their health journey. Individuals are heterogeneous, due to their lifestyle factors, habits, motivation, and other demographic elements [90]. To build successful eHealth systems that can fulfill individuals need, behavioral change theories and persuasive technologies should be considered to build the system in line with the health domain and individuals' preferences [60]. There exists many intervention techniques and health tracking approaches. For example, studies have developed game-based systems [91], mobile applications [92], sensors and IoT systems [93], and even telephone-based interventions [94] to cope with and promote patients' condition. These approaches targeted diverse demographics and health domain to either promote health, prevent escalation into at-risk or diseases, or delay progression into complications [95].

In this chapter, we present our systematic review on studies focused on eHealth coaching, persuasive technology, behavioral change theories, and identify key Persuasive System Design (PSD) components used in different coaching context and combined with behavioral change theories and techniques to deliver effective health interventions. We classify the analyzed interventions into four distinct stages, namely health promotion, prevent from escalation into at-risk, prevent escalation into disease or delay progression into complications and re-admission. These stages are generally referred to as: Stages of Disease Continuum (SDC) [48] which defines the stages of

a disease and the associated health status of an individual (see Chapter 1 for details). We will discuss the data source comparison and the studies distribution alongside the SDC components.

The chapter identifies persuasive coaching approaches and the techniques used by each approach, be it manual, automatic or semiautomatic approaches. Oinas-Kukkonen and Harjumaa's persuasive system design (PSD) model [96] describes such persuasive system technologies that are increasingly used for positive health behavior change. The PSD divides the persuasive components into 4 main categories, namely: primary task support, dialogue support, system credibility support, and social support. The categories provide components, such as reminder and cooperation to specifically describe each action a component is expected to produce. We build our assumptions on the following hypothesis:

- Different coaching pattern will emerge based on various factors that we will investigate in this study.
- A coaching technique may fit in only one health context and different techniques may be needed to readopt to other contexts.

The research provides two contributions to the eHealth coaching, behavior change techniques and persuasive technologies. (1) we present a pattern analysis on PSD and behavior change techniques use in eHealth coaching; and (2) we analyze specific health domains where either a PSD component, a behavior change technique, or a combination of the two is present.

3.2 Aim

We perform a systematic review on health coaching approaches with persuasive system design and behavioral theories and techniques. This is measured in relation to the stages of disease continuum. To our knowledge, this is the first systematic review of health coaching in relation to PSD, behavioral theories and techniques, and SDC. We will review patterns in eHealth coaching that integrate either PSD, behavioral theories and techniques or a combination and check their health focus with the SDC. We will discuss the data coding, and the review process and results. We then present the emerged patterns from the studies analysis. This research will try to answer the following research questions.

- What are the different patterns of eHealth coaching techniques that could emerge from the PSD-behavioral theories and techniques mapping?
- What is the correlation between PSD, with/out behavioral theories and techniques presence, and the intervention effectiveness criteria?
- How behavioral theories and PSD principles are cooperatively used?

Answering these questions will help understand if there is a correlation between a specific PSD element, a behavior change technique, a behavioral theory and a coaching technique in the context of persuasive eHealth coaching.

3.3 Methods

3.3.1 Data Source

The data collection was started between December 2017 and February 2018. We searched for articles published between January 2015 – February 2018 using IEEE¹, ACM Digital Library², Google Scholar³, and Scopus⁴. The terms and the term combination strategy used to search targeted articles are listed in **Table-3.1**. The resulting corpus consisted of 95 papers, with *IEEE*=4, *ACM*=45, *Google Scholar*=11, and *Scopus*=35. The identified articles were manually searched for further relevant publications.

3.3.2 Data Coding

The data coding was achieved by two researchers performing a parallel systematic review of all the articles. The data coders achieved an agreement accuracy of 97% on PSD and 88% on the behavior change theories/techniques and SDC. The conflicts were later resolved by reaching an agreement on the rest of the data. All the selected interventions were coded according to the PSD principles [96], shown in **Table-3.2**.

¹<https://www.ieee.org>

²<https://dl.acm.org/dl.cfm>

³<https://scholar.google.it>

⁴<https://www.scopus.com/search/form.uri?display=basic>

#	Search Keywords	Coding
1	Virtual coaching	(1 OR 2 OR 3 OR 4 OR 5) AND (6 OR 7 OR 8 OR 9 OR 10)
2	e-coaching	
3	Chatbot	
4	Persuasive	
5	Agent	
6	Wellbeing	
7	Health	
8	mHealth	
9	eHealth	
10	Healthcare	

Table 3.1 Articles searching terms and combination strategy.

3.4 Results

This section discusses the data coding and analysis, the common patterns in health coaching approaches and behavioral change techniques from the persuasive technology point.

3.4.1 Search and Selection

The total search returned 505 articles from all the four digital libraries. After screening the results through the eligibility criteria, 440 articles met the inclusion criteria. The next step was to review the title and abstract of those documents to determine whether they fulfill our objectives based on the formulated research questions. This step excluded 310 papers that were not about health coaching in general or they were about other coaching related domains. The next step checked whether the 130 papers fit our inclusion criteria and relevant to our study. A total of 35 papers were excluded, either because they were not related to our study or because of missing data. Finally, 95 articles were selected and included in the systematic review. **Figure-3.1** shows the flow diagram of this systematic review process. For each article we collected the title, keywords, paper type, PSD and behavior theories/techniques, their SDC focus and other information relevant to the study. All the reviews were performed by two coders and later revised to check for agreement accuracy.

PSD principle	Definition
<i>Primary Task Support</i>	
Reduction (1.1)	System should reduce steps users take when performing target behavior.
Tunneling (1.2)	System should guide users in attitude/ behavior change process by providing means for action.
Tailoring (1.3)	System should provide tailored info for user groups.
Personalization (1.4)	System should offer personalized content and services for individual users.
Self-monitoring (1.5)	System should provide means for users to track their performance or status.
Simulation (1.6)	System should provide means for observing link between cause & effect with regard to users' behavior.
Rehearsal (1.7)	System should provide means for rehearsing target behavior.
<i>Dialogue Support</i>	
Praise (2.1)	System should use praise to provide user feedback based on behaviors.
Rewards (2.2)	System should provide virtual rewards for users to give credit for performing target behavior.
Reminders (2.3)	System should remind users of their target behavior while using the system.
Suggestion (2.4)	System should suggest users carry out behaviors while using the system.
Similarity (2.5)	System should imitate its users in some specific way.
Liking (2.6)	System should have a look & feel that appeals to users.
Social role (2.7)	System should adopt a social role.
<i>System Credibility Support</i>	
Trust-worthiness (3.1)	System should provide info that is truthful, fair & unbiased.
Expertise (3.2)	System should provide info showing knowledge, experience & competence.
Surface credibility (3.3)	System should have competent and truthful look & feel.
Real-world feel (3.4)	System should provide info of the organization/actual people behind it content & services.
Authority (3.5)	System should refer to people in the role of authority.
Third-party endorsements (3.6)	System should provide endorsements from external sources.
Verifiability (3.7)	System should provide means to verify accuracy of site content via outside sources.
<i>Social Support</i>	
Social learning (4.1)	System should provide means to observe others performing their target behaviors.
Social comparison (4.2)	System should provide means for comparing performance with the performance of others.
Normative influence (4.3)	System should provide means for gathering people who have same goal & make them feel norms.
Social facilitation (4.4)	System should provide means for discerning others who are performing the behavior.
Cooperation (4.5)	System should provide means for co- operation.
Competition (4.6)	System should provide means for competing with others.
Recognition (4.7)	System should provide public recognition for users who perform their target behavior.

Table 3.2 Persuasive System Design principles.

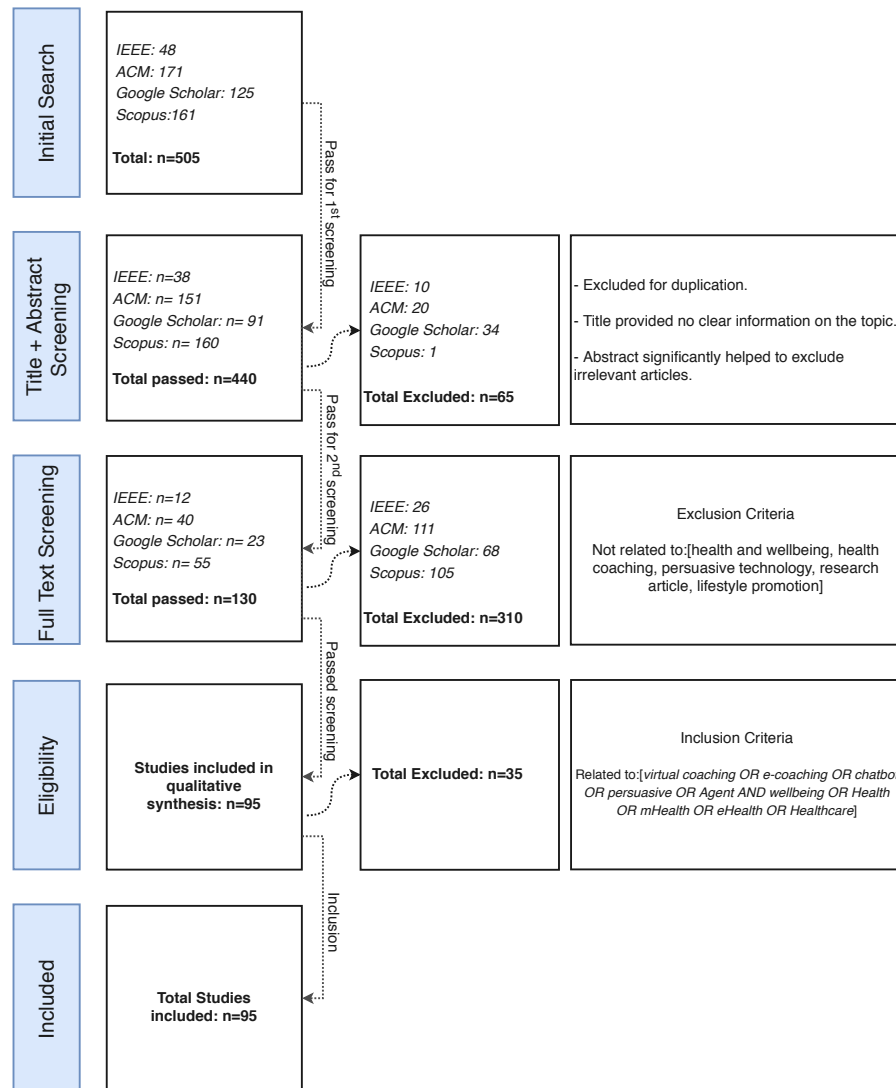


Fig. 3.1 The study selection workflow.

3.4.2 Intervention Studies Description

The selected papers were all related to health coaching and relied on a PSD principles directly or indirectly. Most of the studies were Randomized Controlled Trials (RCTs), single group studies, and the rest were research studies, user centered design, A/B test, focus group sessions, usability interviews and literature reviews. Most studies were targeting adults (63%) and the rest were focused on elderly (19%) and young (18%) as the targeted groups.

The interventions aimed to either promote health, prevent from risk, prevent from

diseases, or prevent from escalation into complication and readmission through health coaching targeting various health conditions (see **Figure-3.2** for the intervention health focus according to the SDC). The SDC distribution showed that most studies either focused on health promotion (first stage) or prevent/delay progression to complication and readmission (last stage). Most of the intervention studies were aimed at physical activity promotion and obesity prevention, or targeting unhealthy behavior change in general (see **Figure-3.3** for the intervention aim distribution).

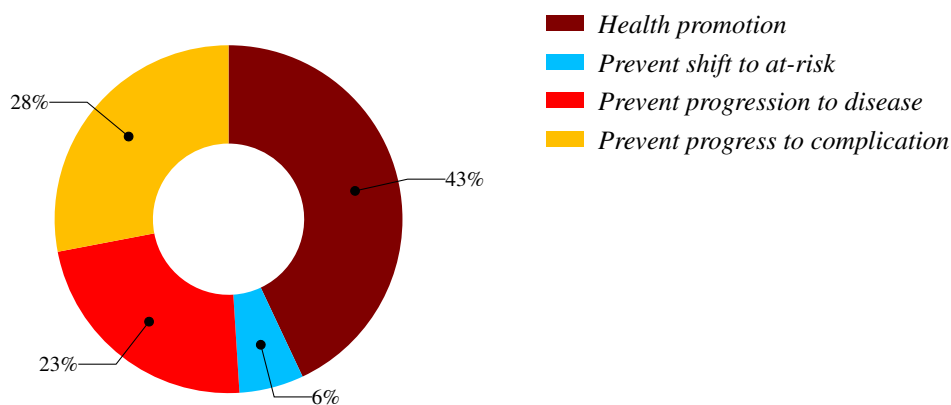


Fig. 3.2 Interventions health focus, according to SDC.

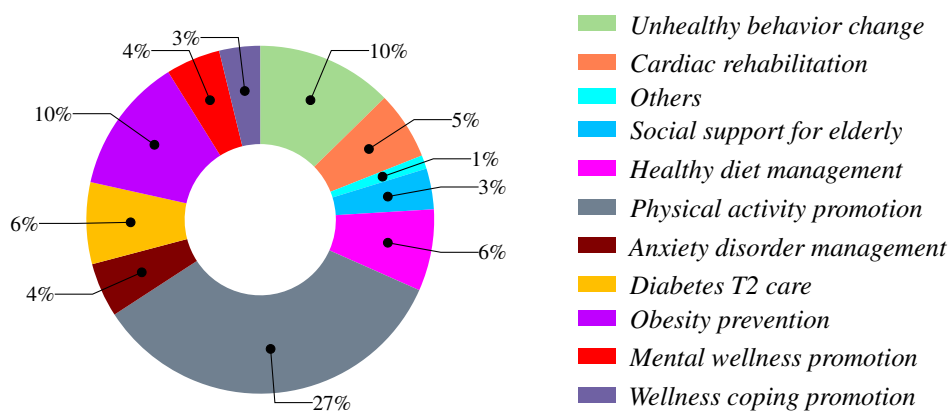


Fig. 3.3 Intervention aim frequency.

Coaching techniques are the tools used by the human agent or the user to deliver or receive health intervention at different stages of the disease. Studies used one or more coaching techniques to deliver the coaching support, either before, during or after the health intervention. The systematic review showed that most studies relied on a mobile app (25%) to deliver the coaching intervention. Moreover, there were

other studies that relied on a human agent to monitor the coaching intervention in-person (16%). Finally, fewer studies used web applications to monitor and track the coaching intervention (11%). **Figure-3.4** shows the coaching technique distribution as was present in the reviewed studies.

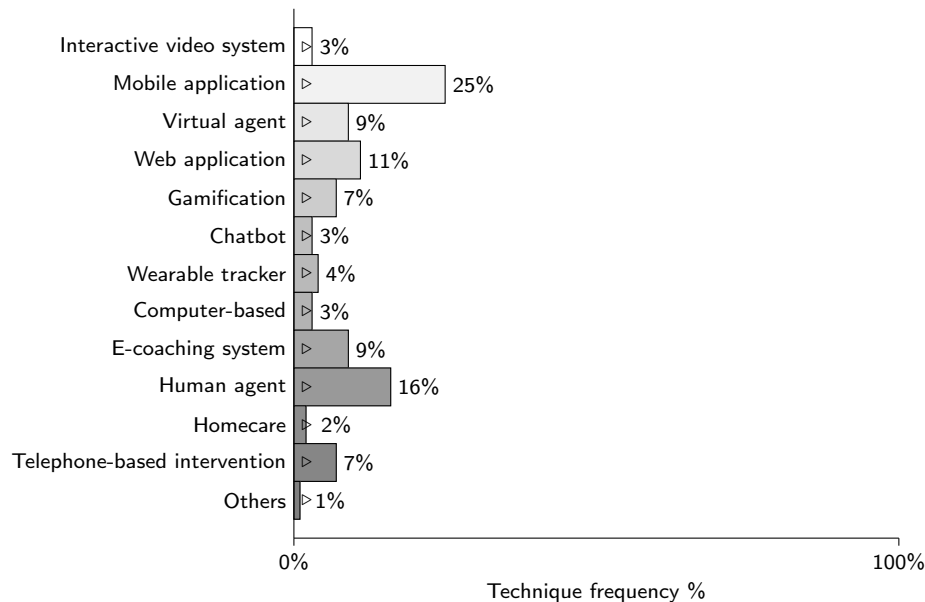


Fig. 3.4 Coaching technique frequency.

3.4.3 Persuasive Technology Analysis

The systematic review analyzed the interventions inclusion of any persuasive system design elements as described in [96]. The analysis showed that Tunneling (e.g., guiding participants to increase water intake by mobile app-based prompt), Personalization (e.g., providing means to personalize users health activities) and Self-monitoring (e.g., providing means to keep track of one's own activity in real-time) were the most frequently used PSD principles among the reviewed studies, while Rewards, Competition, Recognition, Trust-worthiness, Surface credibility, and Liking appeared in only 1% of the studies (see **Figure-3.5**). Reflecting to the four function categories, Primary Task Support (i.e., tunneling, tailoring, personalization, and self-monitoring) was frequently used, while Dialogue Support, system credibility support, and social support appeared not as frequent.

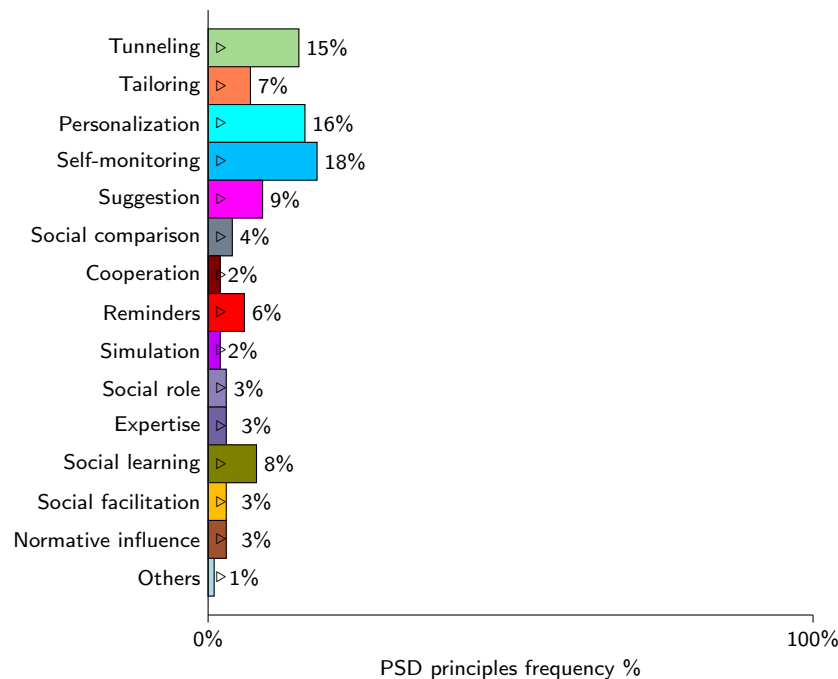


Fig. 3.5 PSD principles usage frequency.

3.4.4 Behavioral Theories/Techniques Analysis

The systematic review considered also the behavioral change theories/techniques applied in the studies. Different behavioral theories appeared among the selected papers. However, many studies did not base their intervention on behavioral change theories/techniques. Since evidence has shown that behavior change interventions based on theories of health behavior are more effective than the non-theory-based ones [60, 97, 98], more theory-integrated sedentary behavior reduction studies are expected. Based on our systematic review, the most frequently applied theories and techniques of behavior change in health coaching included the Fogg's behavior model [37], the Transtheoretical model (TTM) [99] and the Motivational interviewing (MI) [100]. We would encourage more research studies applying these theories into the health coaching domain. Whereas, theories such as, installation theory [101], behavior change support system [102], technology acceptance model (TAM) [26], the nudge theory [103], and elaboration likelihood model [104] were among the least used theories by most intervention studies (see **Figure-3.6**). We believe this variation in using some behavioral change theories or models over others is related to the nature of the interventions and the focus on persuasive technologies. Hence,

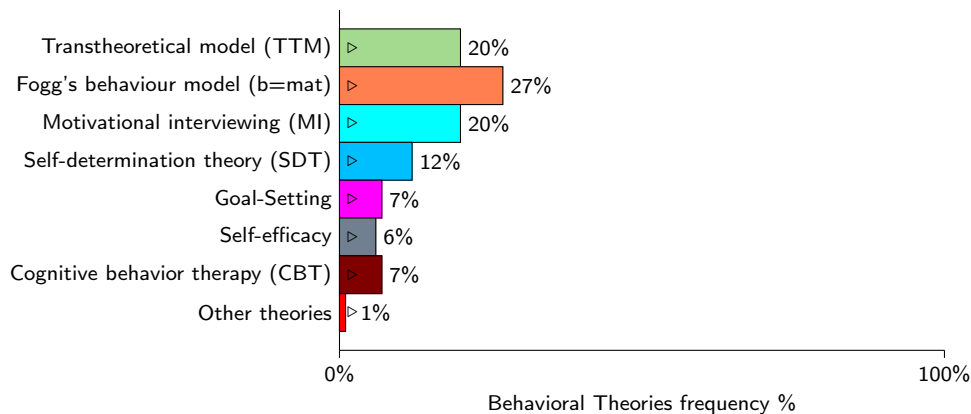


Fig. 3.6 Behavioral theories/techniques frequency.

theories such as, Fogg's behavior model are specifically focused on behavior change by persuasive technologies.

3.4.5 Data Analysis and Mapping

Based on the analyzed data, we performed a mapping between the PSD principles and the most frequent behavior theories, and the stages of disease continuum for the interventions.

PSD-BT Mapping

This consisted of analyzing the studies and checking whether they integrated any PSD principles or based their study on a behavior change theory, or used both. Although the various PSD principles and behavioral theories used by the studies, we focus on the most frequently used PSD principles and behavior change theories, as given above. In **Table-3.3** we show the behavior theories and PSD principles mapping. The mapping revealed that studies based on the Transtheoretical model (TTM) most

Behavioral Theories/techniques	PSD Principles
Transtheoretical model (TTM)	Tunneling, Tailoring, Personalization, Self-monitoring, Suggestion
Fogg's behavior model (b=mat)	Tunneling, Personalization, Reminders, Social facilitation, Social learning, Normative influence, Self-monitoring
Motivational interviewing (MI)	None

Table 3.3 Behavior theories and PSD principles mapping.

frequently used the Tunneling, Tailoring, Personalization, Self-monitoring, and Suggestion principles, while studies based on the Fogg's behavior model (b=mat) most frequently used the Tunneling, Personalization, Reminders, Social facilitation, Social learning, Normative influence, and Self-monitoring. Interestingly, none of the studies based on the Motivational interviewing (MI) used PSD principles. This could be due to the study focus which was on health coaching, however, it did not consider any persuasive techniques in its design or implementation. We invite researchers to further investigate this point in the future.

PSD-SDC Mapping

This consisted of checking the stage focus for each study and then map the stage with the PSD principles they used. Most studies focused on either the health promotion or prevention from progression to complication, as mentioned earlier. The PSD-SDC mapping confirmed that and showed the different PSD approaches in studies focused on health promotion. Studies in the health promotion stage found to mostly use the tunneling, tailoring, personalization and self-monitoring principles, whereas those in the prevent shift to at-risk stage frequently used self-monitoring and suggestion. Studies in the prevent progression to diseases stage mostly used Reminders, tailoring, tunneling, and personalization. Finally, studies in the prevent progression to complications & re-admission stage most frequently used the self-monitoring, Personalization, Tunneling and suggestion (see **Table-3.4**). Overall, studies in the health

Stages of Disease Continuum	PSD Principles
Health promotion	Tunneling=8, Tailoring=5, Personalization=11, Self-monitoring=11, Social learning, Suggestion, Reminders =4, Social facilitation=3, Normative influence, Expertise, Social comparison =2, Cooperation, Surface credibility, Social role, Reduction, Rewards, Liking, Reduction =1
Prevent shift to at-risk	Self-monitoring, Suggestion=2, Personalization, Tunneling =1
Prevent progression to diseases	Self-monitoring, Simulation, Social facilitation, Social role, Trustworthiness, Normative influence, Social comparison, Cooperation, Competition, Recognition, Social learning =1, Reminders, Tailoring, Tunneling, Personalization=2
Prevent progression to complications & re-admission	Self-monitoring=5, Personalization, Tunneling =3, Reminders, Suggestion=2, Simulation, Expertise, Social role, Social learning, Social comparison, Social facilitation =1

Table 3.4 Stages of disease continuum and PSD principles mapping.

promotion and prevent progression to complications stages adopted PSD principles more frequently than the other stages. In addition, Tunneling, Personalization and

Self-monitoring were the most frequently used PSD principles among the reviewed studies in the health promotion and prevent progression to complication stages. This confirms our PSD principle analysis mentioned above.

Overall PSD–SDC–BT Mapping

After analyzing the PSD in relation to the SDC and BT and checking its frequency in each of the two, we now discuss the overall mapping of the PSD, BT and SDC. For this step we consider only the PSD principles most frequently used, and the highly frequent behavior theories as emerged during the systematic review. We then map these two with the stages of disease continuum. This step helps to clarify the PSD principles usage pattern and their stage focus and the behavioral theories they're based on. The most frequently used PSD principles were the Tunneling ($f=8$), Self-monitoring ($f=11$) and Personalization ($f=11$) and these were extensively focused on the Health Promotion stage. Among the frequently used behavior theories were the Fogg's behavior model ($b=mat$), the Transtheoretical model (TTM) and Motivational Interviewing (MI). The behavior theories were extensively focused on the Health Promotion and Prevent progression to complications (see **Figure 3.7** for the mapping).

3.4.6 Health Coaching Interventions Effect

The intervention effectiveness was categorized based on Morrison [105] and Lentferink [68]. This divides the intervention outcome into three, namely highly effective, low effective, and ineffective interventions (see **Table-3.5** for the effectiveness criteria). We listed the interventions with a significant effect on individuals health aspect. Twenty-five interventions (check the source for papers list⁵) reported a positive effect on health outcome. Fifteen interventions led to statistically significant improvement on at least one outcome (e.g., improved quality-of-life and promote fruit and vegetable consumption and decrease unhealthy snack intake), compared with control groups. These interventions were either more effective or as effective as comparison groups, or they were more effective than control groups without an intervention or waiting lists. Other studies concluded some approaches as more effective or proven effective when used in the context of health coaching for various health aspects (e.g.,

⁵Source: https://www.researchgate.net/profile/Ahmed_Fadhil4

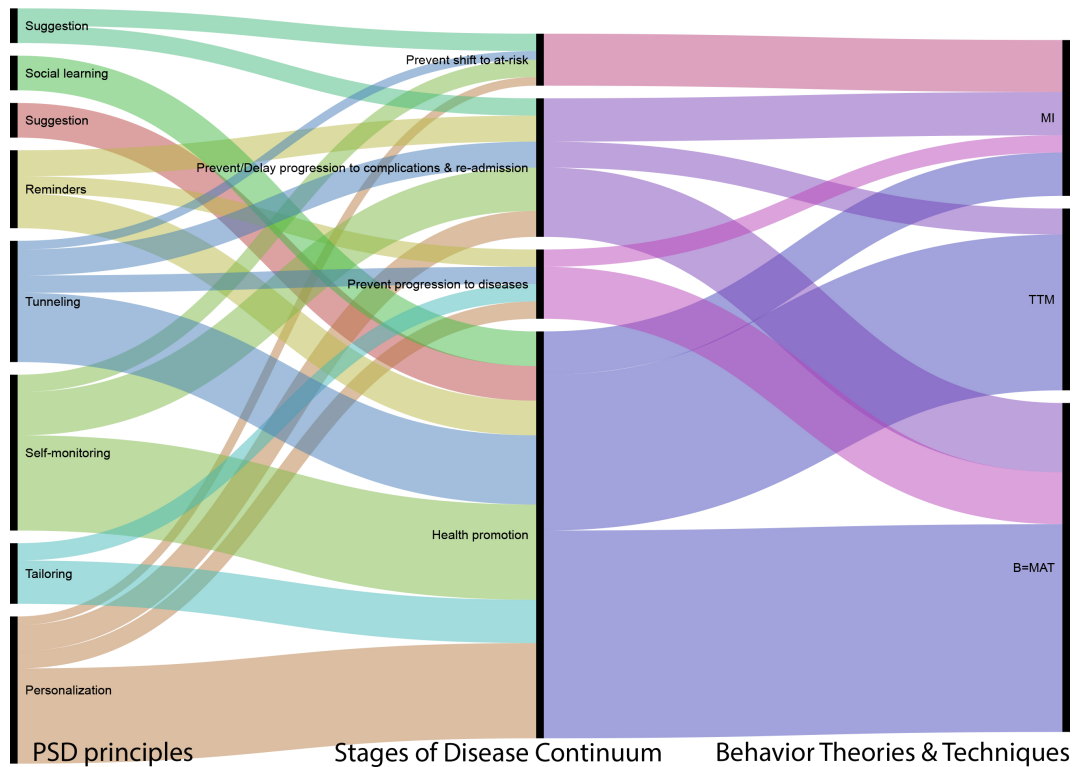


Fig. 3.7 Overall mapping of PSD – SDC – BT.

users are more engaged to train when their trainings are developed and remotely supervised by a Human coach).

Effectiveness	Criteria
<i>High effective</i>	The intervention led to statistically significant improvement on the majority of outcome measures.
	The intervention was more effective or as effective as comparison groups.
	The intervention was more effective than control groups without an intervention or waiting lists.
<i>Low effective</i>	The intervention led to statistically significant improvement on the minority of outcome measures.
	The intervention was as effective or less effective than comparison groups.
	The intervention was more effective than control groups without an intervention or waiting lists.
<i>Ineffective</i>	The intervention led to no statistically significant improvements on any of the outcome measures.
	The intervention was no more effective than control groups without an intervention or waiting lists.

Table 3.5 Intervention effectiveness criteria measure.

3.5 Discussion

Health coaching is a multidisciplinary concept and may integrate one or more persuasive techniques or behavior theories to engage users into diverse health aspects. We performed research analysis on the different PSD, BT and SDC patterns emerged during our systematic analysis of interventions related to health coaching in different health domains and for diverse demographics. More specifically, we study and report on the design, technical and behavioral patterns that might emerge from checking for PSD and BT in the systematic review. Our study confirms that different health coaching patterns exist when analyzing the PSD principles, the SDC and the BT. The findings showed that studies used some PSD principles more frequently than others and focused on a specific SDC. In addition, the intervention studies mostly used few behavior theories more frequently than others. The analysis revealed encouraging results and motivates further research into the various health coaching patterns.

3.6 Persisting Challenges

Several limitations emerged during the systematic review. First, although the study analyzed the most and least frequent PSD principles and BTs, we only focused on the most frequently used once when performing the PSD-SDC-BT mapping. Second, we focused on only studies that led to statistically significant improvement on at least one outcome, compared with control groups. Other less significant studies might provide interesting insights into health coaching patterns from the PSD and BT perspectives. We acknowledge these as limitations for our study and invite future researchers to further investigate these points.

3.7 Implications

Based on our review results, we provide the following implications:

- The majority of health coaching interventions with explicit involvement of persuasive technology are still lacking. Therefore, we encourage interdisciplinary cooperation in this field.

- When applying PSD principles in intervention studies, the underlying behavior change theories/techniques should be reported explicitly. As system credibility or social support in reviewed studies are rarely applied, it should be further explored in future studies.
- Researchers of persuasive technology or PSD should further investigate the health coaching patterns emerged from this study and investigate why certain PSD principles are mostly used in a specific SDC stage.

3.8 Conclusion

In this chapter, we presented our systematic review on patterns in eHealth coaching that involved PSD, SDC and BT applications and elements. This systematic review is the first to examine the effective use of persuasive technology (namely, PSD principles) in health coaching patterns. We showed the different PSD elements used in a health intervention targeting a specific domain and demographics. In addition, we checked the most frequently used behavior change theories/techniques in these interventions and revealed the lack of theory-based intervention considerations in the selected studies. The systematic review revealed that Tunneling (n=15%), Personalization (n=16%), and Self-monitoring (n=18%) as the most frequently used PSD elements, while the principles of system credibility support and social support were seldom deployed. Most interventions targeted health promotion and prevent progression to complications & re-admission stages. Finally, the TTM, Fogg's behavior model and MI were the most frequently used behavioral change theories/techniques by most interventions.

More studies are expected to explicitly report the details following the PSD principles to conclude powerful systematic review and meta-analysis. More theory-based behavior change interventions using persuasive technology are required to enable comprehensive meta-analysis. Longitudinal studies are also required to evaluate the long-term effects of health coaching interventions using PSD principles. The key findings identified in this study led us to the design of our health coaching platform. In the following Chapter 4 we will discuss the design and implementation of our coaching platform driven by the findings from this chapter and the literature review in Chapter 2.

Chapter 4

CoachAI Platform

4.1 Introduction

We applied the lessons learned from the previous chapters on our health coaching system. This chapter presents the design and implementation of CoachAI, a conversational agent to assist the caregiver and provide follow up to users. We discuss the approach to build the components of the platform. This consists of the coaching layer, users classification, dialog engine, wearable data integration, and health intervention delivery channel. This chapter makes the following contributions: (1) automate repetitive tasks by predicting outcomes on data, (2) enhance human decision-making by handling their tasks and providing insights about their users, and (3) present a human-virtual agent mediated coaching approach. CoachAI is domain independent and general in terms of architecture to integrate it into any domain falling into the context of health promotion. We demonstrate how using human-virtual chatbot mediated system to deliver health recommendations is a sound approach and can provide tailored and social support to individuals.

4.2 Design Rationale

To answer the research questions posed in Chapter 1 and consider the findings from the systematic review in Chapter 3, we defined the system design and architecture. The application consisted of designing the Graphical User Interface (GUI) for the

web application and the Conversational User Interface (CUI) for the chatbot. The chatbot architecture design consists of the dialog engine which defines the interaction between the user and the chatbot. This interaction consists of collecting user data, tracking their activities and adherence. The chatbot provides notifications to users whenever a plan is detected or a feedback is required. In addition, the bot provides notifications about positive reinforcement wherever users reach a steps goal or a distance. The data about user steps or distance are all aggregated from a wearable tracker, if paired with CoachAI.

Several approaches are built betting users will do what they make them do, such approaches fail to change user behaviors because they often ask users to learn new routines and unfamiliar actions instead of making old routine easier. Our approach leverages on familiar user routines and focuses on building the service inline with their routine (e.g., daily messaging routine).

The dialog flow was initially structured as a decision tree including the necessary topics the chatbot supports. This was achieved after the literature analysis on building dialog models and existing approaches on designing for conversational user interfaces. We then integrated the dialog structure into the Dialogflow platform¹. This was useful to test our conversation flow between the user and chatbot. However, the Dialogflow suffered from falling outside the conversation context. Hence, we switched to a finite state machine architecture and with the help of a healthcare professional, we validated the dialog flow and the required user interactions to provide them with tailored health activities.

4.2.1 Conceptual Model

The conceptual model was sketch after several iterations and rapid prototyping and consists of more conversational features and automation in addition to user monitoring and tracking. In **Figure-4.1** we present the conceptual model of CoachAI with all its components. The design process was agreed with a dietician who deals with individuals with poor dietary conditions at a health ambulatory clinic. We investigated various approaches applicable in the context of health promotion. Since the platform performs specific tasks and expects certain results from the users, we excluded defining a free style conversation and focused more on building specific

¹<https://dialogflow.com>

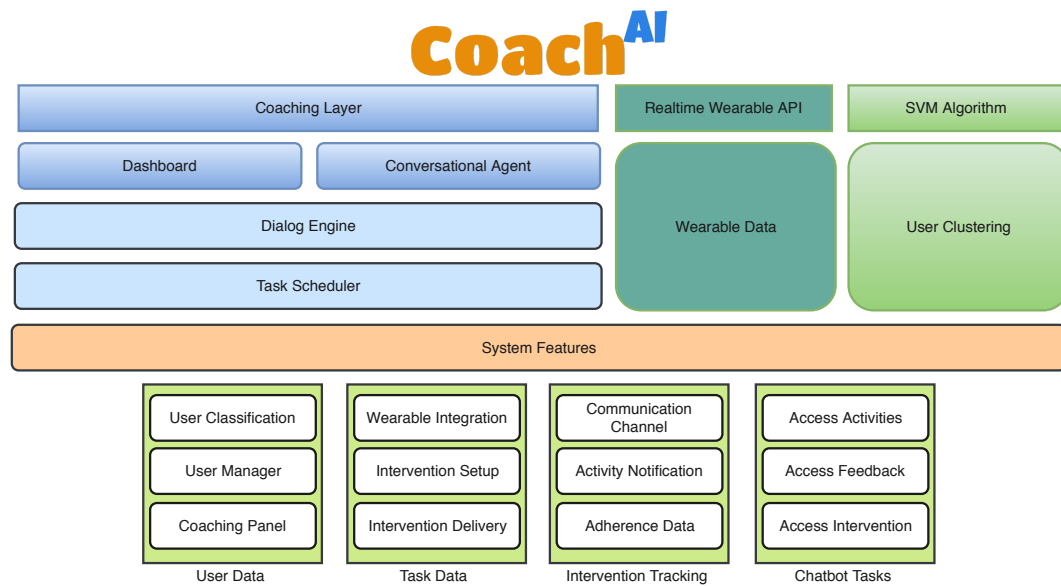


Fig. 4.1 CoachAI - Conceptual model.

context dialogs. Therefore, we focused on building the dialog-states and enriched it with task-automation.

4.2.2 Coaching Layer

This refers to the two main layers forming the coaching layer in CoachAI, namely coaching dashboard and the conversational agent.

Coaching Dashboard

This is the coaching panel to register and track users. It contains all users data about their steps, activities performed, sleeping pattern, and their active and passive days of the week. The dashboard allows the coach to assign initial tasks to the users and check their adherence overtime. The tasks are formed by plans, with each plan containing a set of activities. Moreover, the machine learning model helps the coach to cluster new users into distinct groups. The assigned plans have an associated time, performance validation, and adherence measure. The plan is assigned to the task scheduler to perform the periodic plan assignment and feedback collection. User data about steps, distances and sleeping pattern can optionally be aggregated from a tracker wore by the user. This helps access more accurate data by the coach and

ease the data logging by the user (see **Figure-4.2** for the coaching portal). All the data are aggregated by the chatbot operating inside the task scheduler. It is important

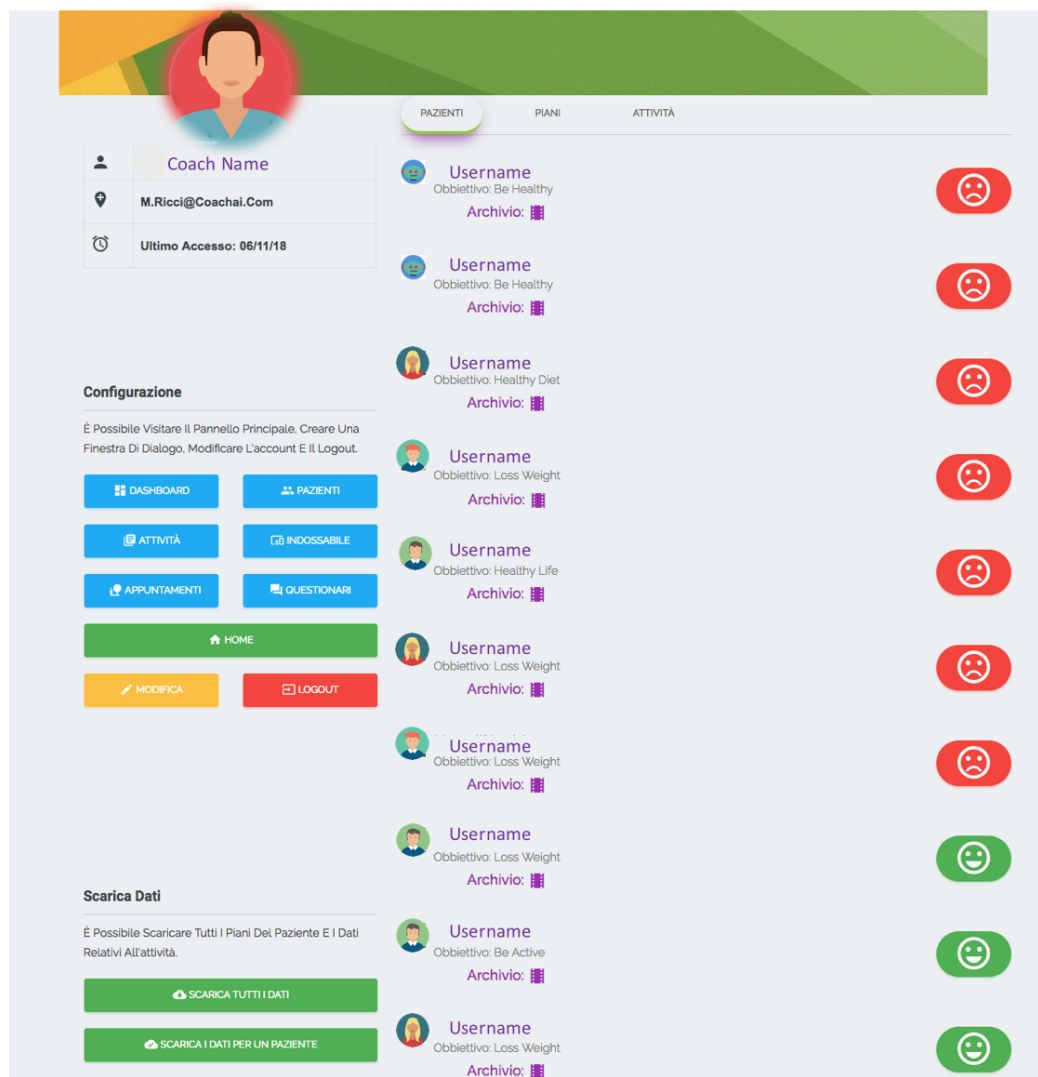


Fig. 4.2 CoachAI - The coaching dashboard.

to consider the limitations of self-organized exercise training and diet management. When individuals choose a health promotion plan, they often set high, unrealistic goals due to inaccurate self-estimations and high motivation initially. Although they tend to execute the plan, but later the probability of deviation from the plan in terms of missed training units increases because it is difficult to keep up with the schedule. At this point, a human coach might adapt the schedule according to observations and

the feedback he/she receives from the subject. These adaptations might be in terms of changing the difficulty of the training units, recommending different units or both.

Conversational Agent

The concept of conversational agents have already been applied to improve aspects of user lifestyle, through face-to-face conversation, with an actual agent using speech and hand gesture [106]. The conversational agent in CoachAI is based on text conversation and simple graphical elements embedded into the conversation. Once the user accesses the chatbot, they interact with the bot on specific topics defined within the hierarchical finite state machine architecture. We gather data at this step about user's BMI, physical activity, diet and sleep. The chatbot performs two separate phases, namely information gathering and feedback collection. All user feedback are submitted to the coach, to notify them about user condition. The chatbot tracks users adherence in general and categorizes them into three categories, namely sedentary, mild or vigorous users. The coach is notified on the platform in case of any user condition deteriorations. The user interacts with the chatbot through a guided dialog structurally defined by the finite state machine. In **Figure-4.3** we provide an illustration of the user-chatbot interaction dialog. The chatbot

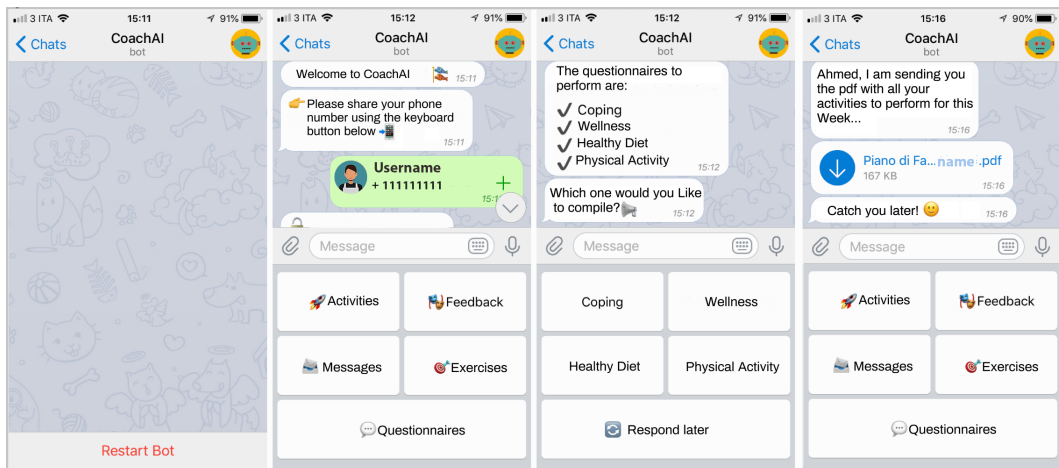


Fig. 4.3 CoachAI - Chatbot agent interface.

handles tasks about activities assigned, user feedback to each activity, exercises, coach messages, and health intervention questionnaires (see **Figure-4.4** for excerpts of coach-user interaction dialog). The dialog agent in CoachAI is based on a semi-automated approach that divides the care delivery task between a human agent and

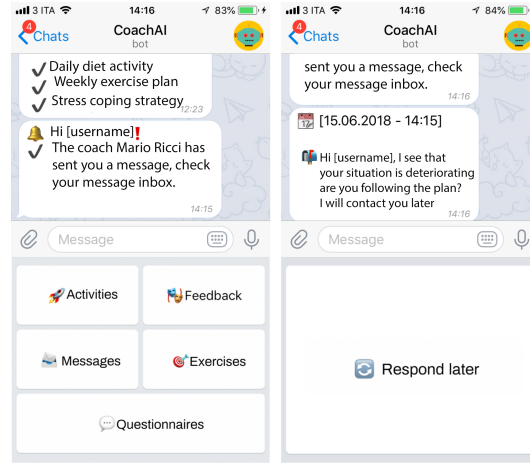


Fig. 4.4 CoachAI - Excerpts of coach-user interaction dialog.

a chatbot agent. More specifically, the platform assigns the activity definition and assignment to the human agent, and handling activity notification and feedback collection to the chatbot agent. This is important to ensure the activities are tailored to individuals lifestyle and hence overcome the limitation of activity repetitiveness that goes against user engagement. Finally, the research relied on conversational user interface approach to solve user interaction problems from UX perspective. Instead of relying on a natural language interaction, we focused with the current architecture on a more dynamic user experience, driven by the dialog states structure.

4.2.3 User Classifier

We used a multi-class SVM classifier to separate new users into three distinct groups based on their initial activity level. The dataset used to train the model was anonymized user data provided by an ambulatory healthcare clinic. The dataset consisted of real patients' data (age range between 22-74, from both genders) and in the form of health record. The data contained information about their demographics, their dietary pattern, their health condition (if any), their health goal, to mention a few. Based on the given data, we constructed a health questionnaire and collected data from 375 participants. The overall dataset consisted of 25 features with dependent (e.g., active or inactive individual) and independent variables (e.g., individuals' age, work, daily sedentary time, how much physical activity they perform). The independent variables cause a change in the dependent variables. The features used

were mainly about user's physical activity and sedentary lifestyle. We plot each data item as a point in n-dimensional space (where n is the number of features) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the three classes. Based on users' replies to the conversational agent questions they are categorized into three distinct groups (namely, vigorous, mild, and sedentary users). The model considered some significant features in the data to check for when classifying the users. We used a category=3 and a response variable about the physical activity of a person. The data randomly selected samples when split into train and test. The algorithm to perform user clustering is highlighted in the **Algorithm-1** code snippet below.

Algorithm 1: Initial user clustering.

```

1  userClustering ( $P, C$ )
   | inputs :  $P = \{p_1, p_2, \dots, p_n\}$  (set of users to be clustered)
2   |  $K$  (number of clusters);
3   |  $MaxIters$  (Limit of iteration);
   | output :  $C = \{c_1, c_2, \dots, c_k\}$  (set of cluster centroids)
   |  $\{L = \{l(p) \mid p = 1, 2, \dots, n\}$  (set of cluster labels of  $P$ );
4   | };
5   | foreach  $c_i \in C$  do
6   | |  $c_i \leftarrow p_i \in P$ ;
7   | end
8   | foreach  $p_i \in P$  do
9   | |  $l(p_i) \leftarrow \text{argminDistance}(p_i, c_j) j \in \{1, \dots, k\}$ ;
10  | end
11  | Changed  $\leftarrow$  false;
12  | Iter  $\leftarrow$  0;
13  | repeat;
14  | foreach  $c_i \in C$  do
15  | | UpdateCluster ( $c_i$ );
16  | end
17  | foreach  $p_i \in P$  do
18  | | minDist  $\leftarrow \text{argminDistance}(p_i, c_j) j \in 1, \dots, k$ ;
19  | | if minDist  $\neq l(p_i)$  then
20  | | |  $l(p_i) \leftarrow$  minDist;
21  | | | Changed  $\leftarrow$  true;
22  | | end
23  | end
24  | iter ++;
25  | Until Changed = true && iter MaxIters;
26

```

1. The algorithm defines initial user group centroids, which is achieved in different strategies. A common approach is assigning random values for centroids of all user groups. However, since lifestyle information provides a strong predictor about users condition, before checking other features for the first two clustering approaches, we rely on users adherence to assign user groups centroid (heigh adherence and different adherence).

2. This step assigns each user to a cluster with the closest centroid. To find the cluster with similar centroid, the algorithm should calculate the distance between all users and each centroid.
3. This step recalculates the values of the centroids. The value fields are updated, by averaging the values of user's attributes that are part of the cluster.
4. Repeat steps 2 and 3 iteratively until users groups are fixed and stable.

For a new user, we find the difference with the mean of the clusters and then the closest cluster is chosen. To find the differences we use the lifestyle data and find nearest Euclidean distance from each mean of previous clusters (see below **Figure-4.5** for an illustration of user classification). Other classification, such as classifying

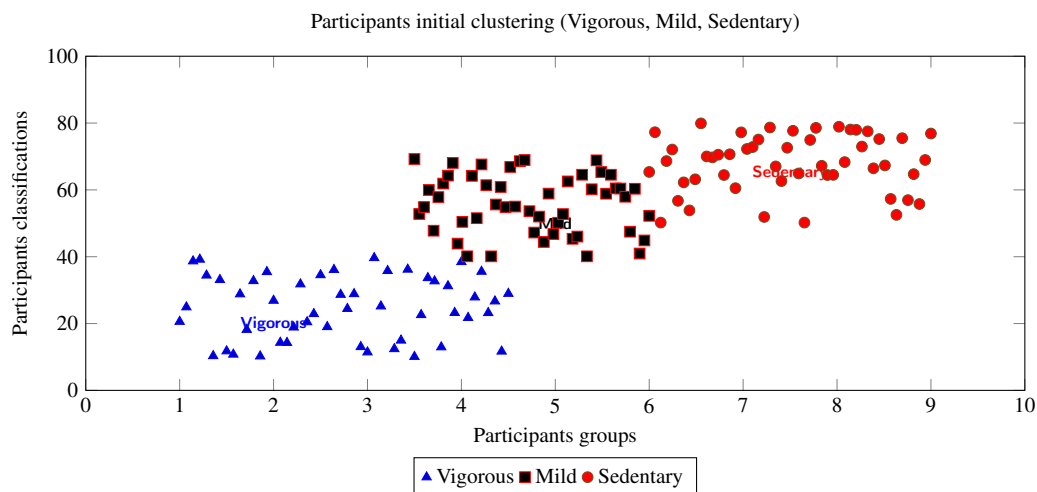


Fig. 4.5 CoachAI - Initial user clustering.

users based on their dietary type and emotional state would be helpful for the coach while building users plan. However, accessing users data was a limitation, and hence we left this part for the future work. Finally, the motivation behind constructing groups of participants based on their daily physical activity is twofold: First, we want to explore identifying specific patterns in participants that could describe exercise profiles of different individuals. Second, we aim to generalize to new participants and tackle the issue of insufficient data to proceed. Profiling via clustering will facilitate this goal, since assigning a new user to a group will allow providing feedback based on their exercise profile. That said, the classifier can be a way to provide some information about system users, but that is not enough to predict precisely their

preferences. Collecting more user data will help the classifier provide pattern about their preferences.

4.2.4 Tasks Scheduler

The scheduler is the mechanic that steers the dialog engine with all the tasks it performs. The scheduler handles the state transition when the coach or the user are interacting with the chatbot. This may include defining a plan with the duration and duties and tracking the plan expiration date and user adherence to the plan. The task scheduler also handles all sorts of notifications between the coaching platform and the conversational agent. The scheduler notifies users when a plan has been assigned and sends them the list with the duration. The scheduler provides the appropriate plan and feedback notifications based on the time chosen for the notification. The periodic tasks are handled with a time-based background job scheduler which handles all the periodic events in the system.

4.2.5 Dialog Engine

The dialog is based on a structured finite state machine that offers the capabilities to define the structure for a domain specific dialog model. Based on the defined pattern, the state machine allows constructing conversation patterns about various topics involved in the conversation (**Figure-4.6** presents the skeleton of the finite state machine). The finite state machine has unique states that guide the user through a series of steps to achieve their daily goal. The states use the result of previous states and its results are used by the next state. The choice of FSM or automata over the similar approaches was because the number of user tasks are limited. We followed a standard design pattern to structure the most relevant states for health dialog models [107]. For instance, to exercise more or achieve healthy diet, there has to be a finite set of steps a user has to fulfill to reach the end of the dialog. Hence, the conversation is a composition of finite state machine based dialog that delivers the services through a text based chatbot. The high-level dialog engine states are defined as methods following the structure:

- *request*: the user text message request to the chatbot.
- *reply*: the chatbot message feedback to user.

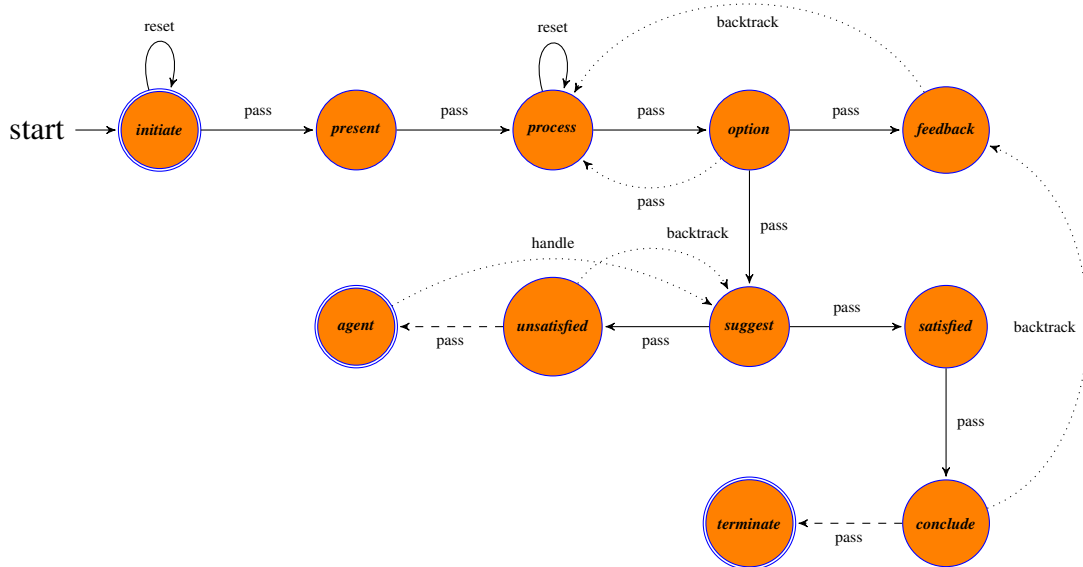


Fig. 4.6 CoachAI - High-level Finite State Machine (FSM).

- *go*: trigger next state into the same machine, after fulfilling some conditions.
- *return*: close the dialog, returning to the caller dialog (main menu).
- *call*: call an external dialog or a sequence of dialogs. This includes switching to another branch in the dialog.

The state transition happens when certain conditions are met. When the chatbot is in the present state, the user is presented with the main functionalities the bot can perform. The user has to select a button or type the functionality they wish to perform to go to options. The option state provides users with information available or asks them to provide their input. Based on user action, the state will take a certain path, i.e., either feedback or suggest state. Unless the state condition is satisfied, the chatbot will not allow the user to change the state. Once the conditions are satisfied, the user can conclude the dialog and terminate the interaction.

4.2.6 Wearable Data Integration

Wearable integration provides automated data collection about steps, distances, sleeping pattern, the active and passive days of the week and the days where the user meets a steps or distance goal set by the coach. For example, if the coach sets 10,000

steps a day for 3 days within a week, then the plan will be communicated to the user through the bot. Whereas, the data about users steps per day will be automatically aggregated from the wearable and presented on to the CoachAI platform. If the user meets the goal set, the bot should trigger a notification and send a relevant message to the user, and the opposite should also apply. Hence, the main purpose of integrating the wearable is two folds: first aggregate steps and distance data without asking users to manually log them. Second, collect other data about their active and passive days and amount of sleep that deems relevant to the coach. In **Figure-4.7** we present pairing the wearable with CoachAI platform, where in #1 allows the coach to pair the user with Fitbit tracker, in #2 the coach builds a health activity, in #3 the coach decides the activity duration, type and amount, and in #4 the coach visualizes user compliance to the activity as aggregated from their wearable tracker. Integrating



Fig. 4.7 CoachAI - Tracker pairing, data tracking and visualization.

the wearable tracker is optional and should be treated as such. That said, if the user has no tracker, then the chatbot should be able to handle assigning steps, distances and other activities and later validate their application through feedback. In this case, the user will get and communicate the activities through the chatbot channel by conversing with the agent.

Whereas, in case of wearable integration, the CoachAI platform should allow the coach to define steps and distance related activities directly and communicate them to the user through the chatbot channel. The activity data are automatically aggregated through the wearable device and later communicated to the coach. If the user meets the daily or weekly goal, the data communicated to the platform will trigger a

notification to the user that is communicated through the chatbot channel. This notification consists of positive reinforcement and a summary of their steps data. In the contrary, the platform will trigger a notification to notify the user through the chatbot channel about their failure to adhere to the plan, then provide them with hints on how to improve. The main topics the coach defines are related to diet and exercise, which may consists of the steps and distance taken within a specified timeframe. Therefore, we envisage the platform to behave as such in case of steps and distance tracking with or without the wearable integration. In **Figure-4.8** below we present the graph data aggregated from the wearable tracker.

Once the integration and data flow is achieved, the next step is to work on the logical

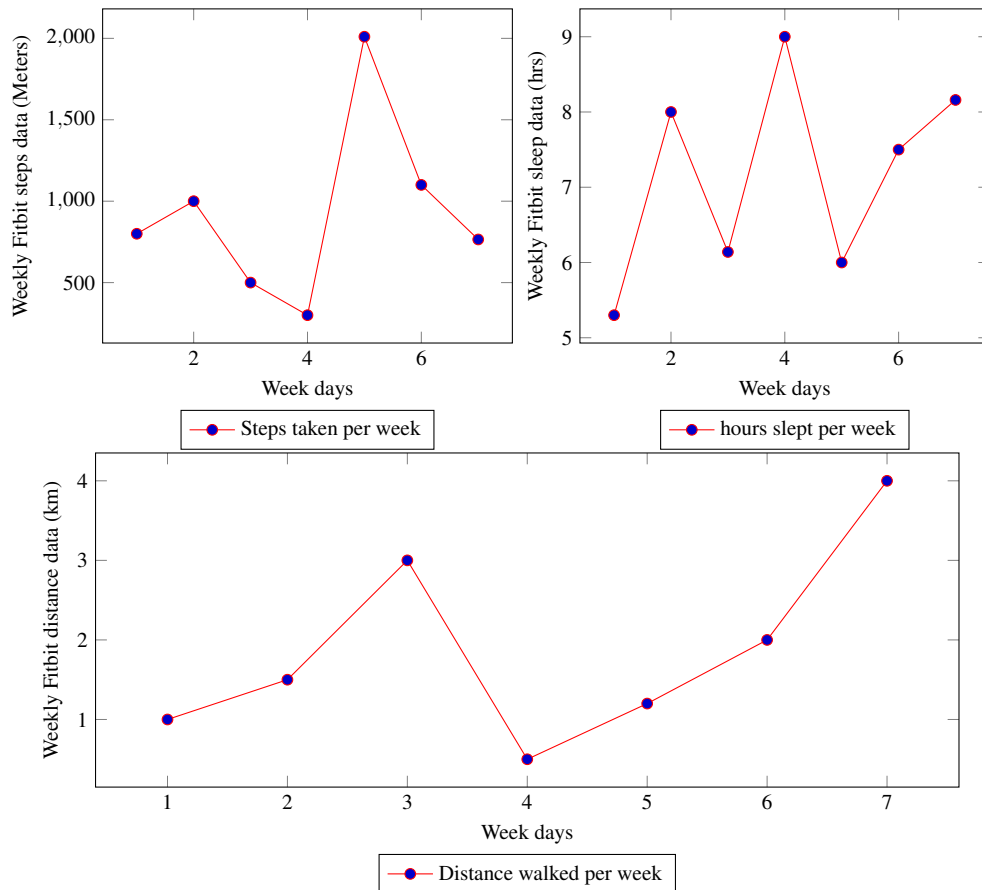


Fig. 4.8 CoachAI - Weekly report generated by Fitbit tracker.

layer of the application. The core tasks include presenting the Fitbit data on CoachAI platform, defining the behavior of the system when a coach builds a steps or distance related activity, then pushes it to the user and the mechanism to aggregate the data from the wearable and check if the condition to achieve the goal has been fulfilled.

The coaching platform should successfully trigger a notification and communicate it to the user through the chatbot channel in either cases (when the user meets the goal or when they fail to achieve the goal set). Finally, the coach can define the amount of steps to be taken and the duration (e.g., daily or weekly). The Fitbit is the service or data provider, whereas the CoachAI platform is the consumer of this data.

4.3 CoachAI High-level Architecture

Putting all components together we form the overall coaching platform that handles physicians workload and provides tailored care to the user. In **Figure-4.9** we show the overall system architecture with all the components and actors highlighted. The architecture consists of the coaching layer, the dialog engine, the user clustering component, the conversational agent and wearable integration.

The current version of CoachAI supports only text and the bot would respond back with either an appropriate message or provide some files to view. Hence this version is not concerned with the aspects like ASR, speech recognition, speech to text, text to speech etc. The chatbot maintains the conversation state and respond to the user request by checking the FSM tasks and validating if the user has fulfilled them. Part of this dialog state performs a supervised intent classification model that is trained on varieties of user physical activity data. This classifier clusters users based on their exercise level.

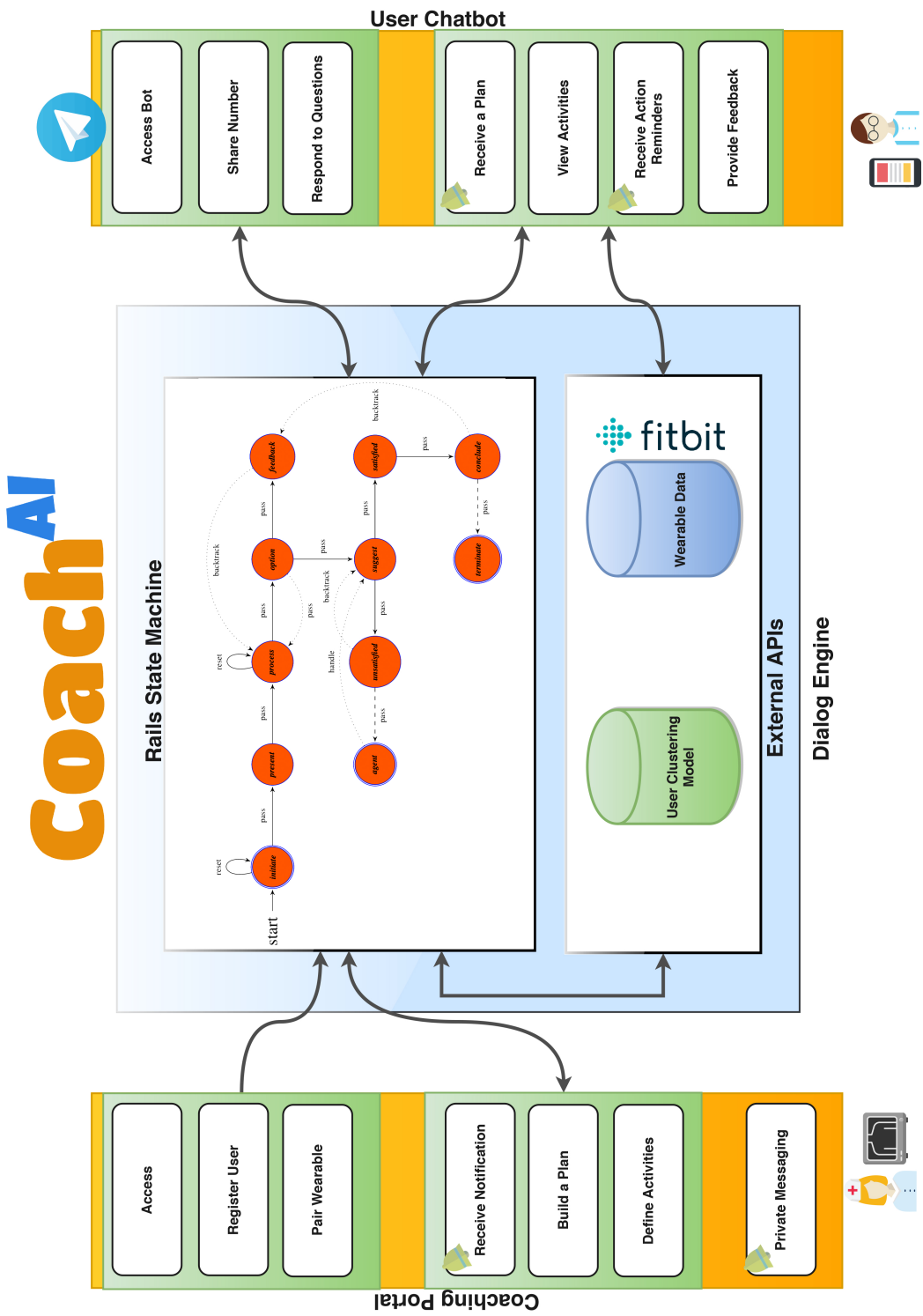


Fig. 4.9 CoachAI - Overall system architecture.

4.3.1 Intervention Delivery

The platform classifies users into 3 distinct groups after the initial assessment. The assessment also collects data about user dietary habits, physical activity, emotional state, and more. These data help the coach build a plan and assign it to the user. However, the coach can also deliver standard or customized interventions through the platform. The intervention may consist of some follow-up dialogs with the user or a structured questionnaire for the user to perform to evaluate their condition. The questionnaire consists of a set of questions with their answers and a score for each answer (see **Figure-4.10** for a list of users with their overall health adherence). This approach is based on the Kaiser Pyramid model of care (KP) [108]. The user answers are scored and a total score is given for each user responses. The platform can deliver any type of health and wellness interventions, such as cognitive behavioral therapy (CBT) [109].

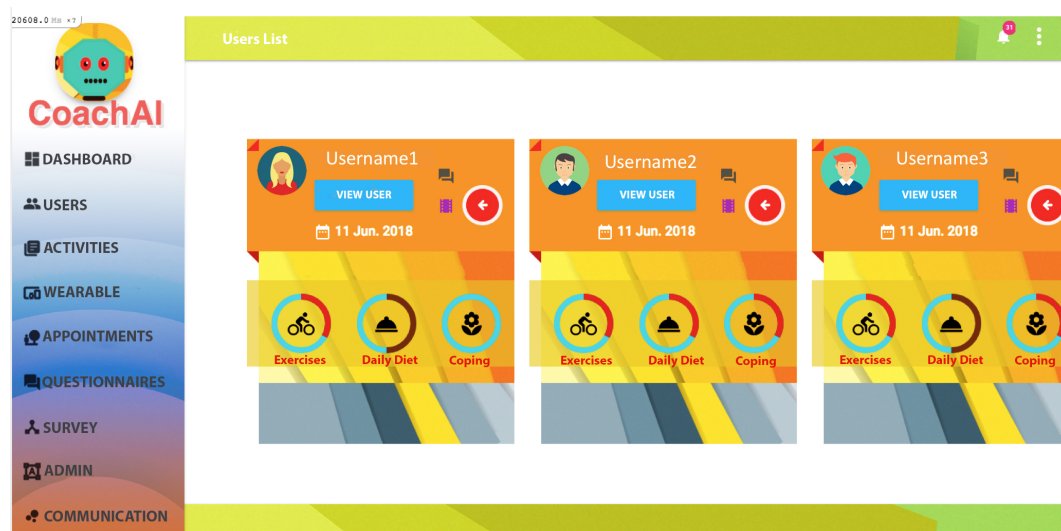


Fig. 4.10 CoachAI - Overall users adherence to health activities.

4.4 Conclusion

With rising costs of healthcare services, the current focus have shifted from health treatment to health promotion. Healthcare chatbots currently seem to be a mix of both patient-only (apps that help a patient track and make sense of health data) and

patient-caregivers applications (apps that connect the two groups, for diagnosis or treatment). This chapter discussed the refined approach of our coaching system that mainly focuses on assessing the quality of the coaching delivered by a messaging chatbot. We focused on the efficiency of intervention delivery by facilitating coaches tasks and providing user follow up.

We presented the various components forming the CoachAI platform and their integration to form a meaningful medium for the coach, the conversational agent, and the user. In the same vein, we highlighted the value of a human-in-the-loop, that helps avoid the shortcomings of pure agent-based control, affording an experience in which the system responds appropriately to both technical and emotional needs (e.g., user preferences). We described one concrete instance of a text-based healthcare chatbot (THCB) system that was designed to support the coach and provide users with tailored activities. The results of the prototype preliminary evaluation (see Chapter 5), the cohort and comparison studies (see Chapter 6), and the platform testing and prospective intervention (see Chapter 7) will discuss the application validation and the findings emerged from each experiment.

Chapter 5

Initial Prototype and Preliminary Evaluation

5.1 Introduction

Several studies have tackled the issue of poor lifestyle and its health determinants [110]. In this chapter we present the first prototype of our coaching application¹ [20] for diet and physical activity promotion. We then discuss our findings from the three weeks experimentation with elderly and highlight the lessons learned and persisting challenges.

Initial Prototype

We test whether human and conversational agents mediated health coaching systems are effective in maintaining user adherence to a healthy lifestyle. The main contribution is introducing a conversational agent into the health coaching loop to handle tasks and provide supports to both the caregiver and user. The developed bot follows a rule-based reasoning and allows more than a user to interact with and provides users activity to the coach on a web application.

¹<http://ict4g.org/coachme/>

5.2 Objectives

The application focuses on simple activities and follows "is able to do?" instead of "have to do!" approach. We introduce a new interaction paradigm where caregiver-user interact through the console. Based on common paradigm the coach provides standard personalized model specific for each user. To achieve this, we introduce the role of task-automation offered by the chatbot to fulfill specific users need, and handle some tasks repeatedly carried by the coach.

5.3 Design Rationale

5.3.1 Chatbots and Users

Chatbots can be available around the clock to offer users access to all kinds of health support including medication advice, test results, and tasks, such as scheduling an appointment. Users could also experience a service similar to what an in-home caregiver offers by using a post-discharge chatbot. By understanding users input the bot can analyze and provide valuable insights to the coach.

Based on these rationales and the literature findings, we decided to build a simple dialog interaction for the chatbot to make individuals interaction as smooth as possible. Hence, the chatbot consisted of only two main functionalities, namely "Daily Activity" & "Individual's Feedback".

5.3.2 Chatbots and Caregiver

Caregiver refers to any person who is providing care and training to other individuals. Although coaching in our context is targeting individuals health promotion, other domains are also applicable, as long as a caregiver-user relation exists. The chatbot can help deliver better care and the coach can interact with the bot to get the latest information about users adherence. This includes details recollected before user's appointment. Ultimately, the chatbot aids the health coach in enhancing productivity and enables them to respond to users quickly. For example, a coach can interact with a diet chatbot to recommend the appropriate meal to their users. Users could also engage with this chatbot to get instant information about their dietary choices.

Based on the above assumption and the literature findings, we decided to develop a dashboard for the coach to build plans and activities and order the bot to deliver these plans and activities to the users.

5.3.3 User-Chatbot Interaction Design

The user-chatbot interaction was designed and structured based on the tasks assigned by the coach and the dialog structure for the bot. The communication from/to the chatbot is structured and follows a rule-based dialog. The users can interact with the chatbot and all their performance is tracked and reported to the web application in real-time. The dialog engine handles specific commands assigned by the coach or collected from the users.

There are many hurdles to overcome when building dialog experience and flow optimization. The challenge in building a bot is both technical and an issue of user experience [111]. The chatbot interaction in the prototype was programmed to perform repetitive tasks and hardly interact with the user in a free-style conversation. In **Figure-5.1** we provide an overview of the user interaction with the chatbot agent. The figures: #1 provides the main dialog page, #2 introduces the user to the bot functionalities, #3 presents the coach assigned health activities, and #4 reminds user about their compliance feedback. The chatbot enables a very targeted

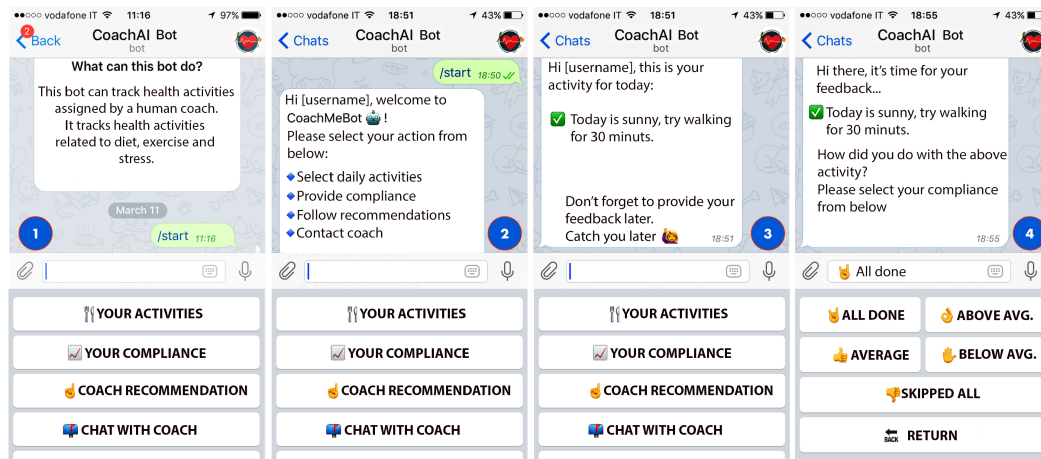


Fig. 5.1 CoachAI - Chatbot application interface.

individual interaction and a very personal approach. The bot knows the user during the interaction and on the other end the human agent can use this information and

react dynamically. This offers personalized services to the user and proactively approach them.

5.3.4 Coaching Dashboard

Having a human coach can enhance the personalization aspect in the health promotion process. Unlike automated systems, real coaches can communicate the language of the user. Existing automated coaching systems are still limited with respect to user emotional states. Having a coach can provide more flexibility to better understand users preferences and hence tailor the recommendation to their interest. The methodology focuses on positive reinforcement for future activities. The system feedback is inserted via health expert who monitors users activities through a dashboard. In **Figure-5.2** we present the users profile on the coaching dashboard. This dashboard contains users personal information and adherence to the activities. The compliance

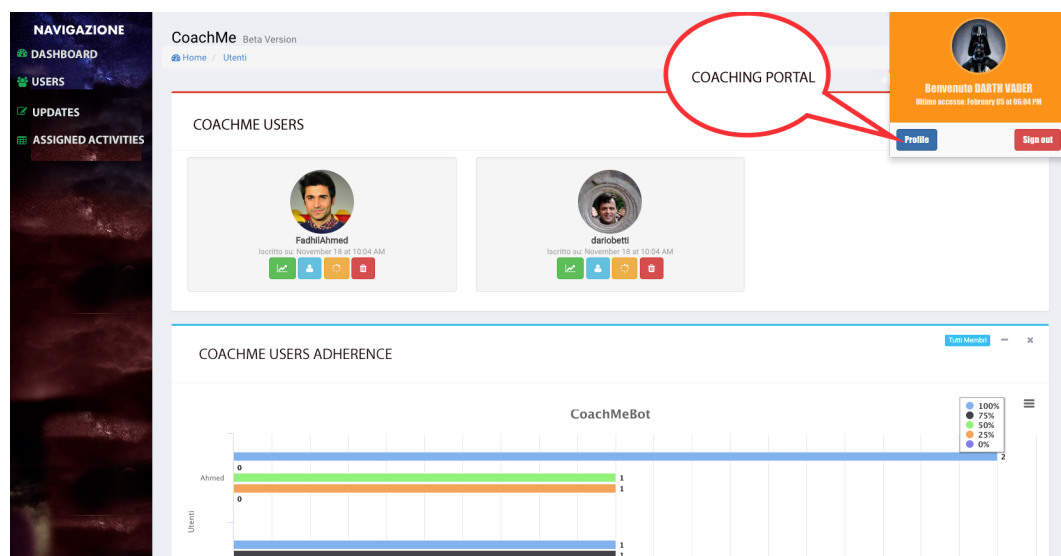


Fig. 5.2 CoachAI - Users profile.

can be visualized independently per user, as provided in **Figure-5.3**. The coach inserts daily activities for the user and tracks their adherence through the chatbot. Hence, with a human in the loop augmented with the chatbot, we aim to bring the best of both approaches for the users.

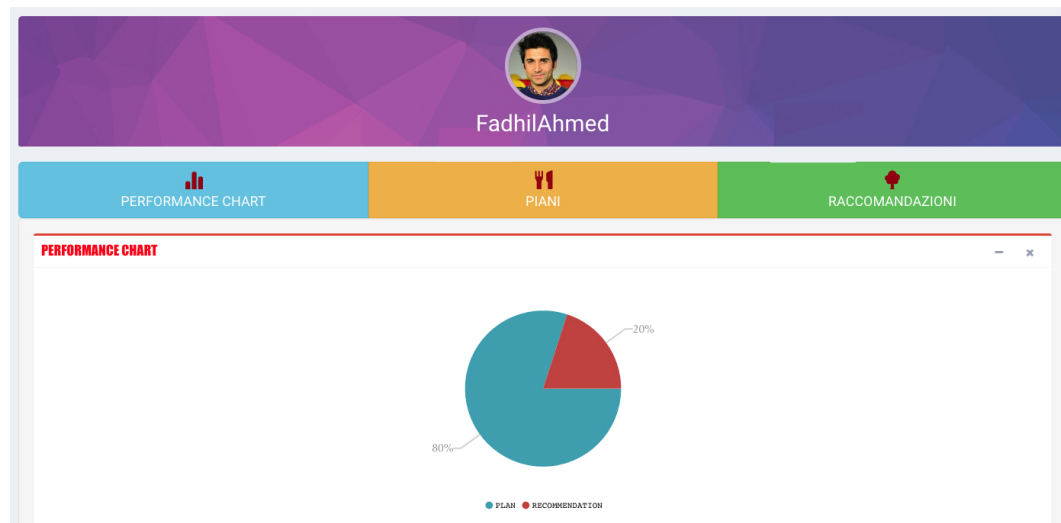


Fig. 5.3 CoachAI - User adherence profile.

5.4 Conceptual Model

The design of CoachAI focused on building a sustainable user interaction with the coaching application. The initial prototype was developed based on the findings from Chapter 2 and Chapter 3 and our conceptual model both for the web and chatbot framework. The conceptual model consisted of a coaching dashboard with users profile, users activities, and their adherence data to a given activity. The coach-user interaction is achieved through their interaction with the chat engine which communicates with the Telegram Bot Platform², where users personal data and bot application are hosted. The user then interacts with the chatbot to access the various activities assigned for them by the expert coach. In **Figure-5.4** we show the conceptual model of CoachAI platform with the web and chatbot application.

The conceptual model consists of the coaching portal, and the application architecture that forms the CoachAI web application. This in turn communicates with the Telegram bot platform to access and provide data to the users interacting with the chatbot, hosted on their Telegram application.

²<https://core.telegram.org/bots>

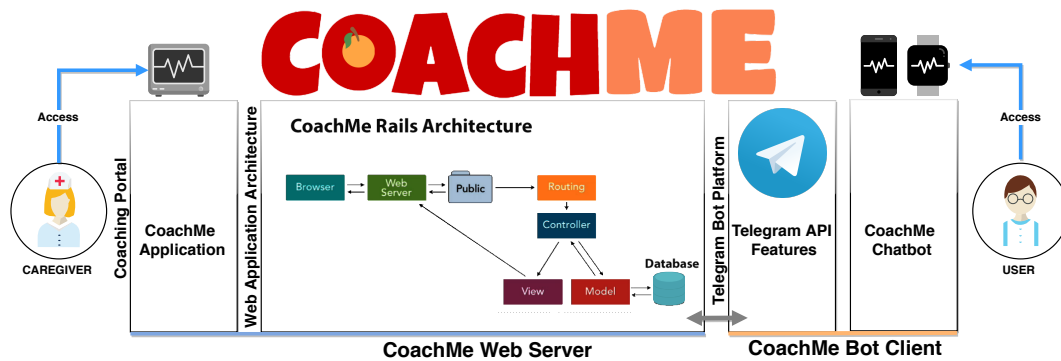


Fig. 5.4 CoachAI - Conceptual model

5.4.1 Dialog Engine

The dialog consisted of building the services the platform is supposed to deliver within the conversation. We had only three components in this prototype to embed into the conversation, namely user activities, user feedbacks and the coach recommendation. We mainly used rule-based reasoning to build the dialog for these components and Telegram bot platform was used to handle the chatbot interaction. The activities planned and assigned were later tuned by the coach to enhance their personalization aspect. The chatbot is formed by a structured dialogue and design elements. We used a chat scripting language to define the dialog engine structure³. The dialog flow was simple and focused on defined rules to trigger the next conversation.

All the possible responses of the bot are predefined and intended for a specific purpose. The bot uses the message and context of conversation to select the respond from a predefined list of messages. All the dialog data used to build the dialog engine are based on research validated studies [73, 112, 113]. The output of this research validated whether using bot for task management in user-coach interaction as a substitute to an automated system will create a difference in adhering users to follow a healthy lifestyle.

³<https://github.com/bwilcox-1234/ChatScript>

5.4.2 Monitoring Layer

The majority of caregivers are already using messaging applications to communicate with users. Integrating such system within a health application will provide quick and timely follow-up with users. The coaching layer is intended to build a plan formed by activities and forward it to users, then later notify users about the feedback and collect their performance data. The process is monitored by a health coach to effectively track all users condition. The functionalities in the coaching layer were limited to providing activities, collecting user adherence data, and tracking their performance. Being a prototype, the functionalities presented in this version focused on achieving the minimum valuable product (MVP). This was decided to develop, test and iterate quicker than focusing on building a more advanced system that might take more time and will still require several iterations.

5.5 User Scenario

To better define the system architecture and components interaction, we illustrate a use case scenario where the caregiver interacts with a patient: *Sara is a caregiver and works with many patients either by visiting them at home or at her clinical ambulatory. Many of Sara's patients are elderly and require regular care and monitoring. Sara finds it hard to keep track of all the patients and perform similar tasks repeatedly. In addition, most of her patients are with limited technical abilities, although they use smartphone for simple operations, such as chatting or calling. Since Sara deals with diverse user groups, understanding each individual needs is another hurdle. For example, to provide daily physical exercise to elderly, Sara has to check with each patient and define the activities to perform. However, she has neither the time nor the resources to keep track of their adherence and progress. She checks their performance by monthly appointments with each patient.*

Sara starts using CoachAI tool to manage her care activities and enhance patient tracking. She registers all the patients in the application and begins profiling them. Sara is now more informed about her patients and can assign a set of activities to more than a patient. The activities will be handled by the chatbot and Sara can dedicate more time to patients in need. Meanwhile, the chatbot can handle the activity assignment and patient adherence tracking, hence help Sara with her tasks.

Preliminary Evaluation

We tested the prototype with real users at the ALMAC Association⁴ for a duration of 3 weeks. The experiment consisted of 4 participants who used the chatbot application and were provided with daily lifestyle promotion activities. The users were given a notification by the bot at the end of the day to provide their compliance with the activity. A researcher took the role of the coach, and provided clinically validated diet and physical activity promotion plans to each of the participants [114, 1]. The study tested whether introducing a human-virtual bot helps to adhere users to follow a given activity. The outcome of the study was used to consider answering the following research questions:

- Was bot inclusion effective in user-coach communication?
- Did the inclusion of bot promoted changes in user diet or increased their exercise?

The questions were confirmed via a questionnaire with the users after the experimentation was ended.

5.6 Hypothesis

This research aims to answer the most routine issues faced by users with low adherence, including exercise, nutrition and sleeping pattern. With this experiment we initially hypothesized that:

- H1:** Delivering health activities with the chatbot positively correlates with user compliance.
- H2:** Assigning more tasks to the bot optimizes coaches task while providing efficient services to the user

⁴<http://www.almaconlus.it/>

5.7 Intervention Design

We followed the framework described by Laugwitz et al., [115] to evaluate user experience through a questionnaire. The experiment evaluation was carried out by two researchers and took place at the ALMAC Association for 3 weeks. All the participants have had a previous cardiac conditions. The participants were given instructions to familiarize them with the chatbot application and the purpose of the experiment. They were then given access to the chatbot and each of the functionalities were described by one of the experimenters. The experiment provided general recommendations to promote lifestyle among the elderly familiar with smartphone. The targeted group were above 55 years old with a sedentary lifestyle and low adherence to healthy diet. Moreover, the activities were intended to measure the user interaction with the bot and see if such approach outpaces mobile applications in terms of user acceptance and ease of use.

5.7.1 Participants

The targeted group was formed by one female and three male with the age range (≥ 55 & ≤ 76) (see **Table-5.1**). No reimbursement was provided to the participants. At the end of the instruction, participants were given a list to fill their information and familiarity with the use of smartphone. The choice of elderly with previous health condition was because we wanted to validate if chatbot system is effective to adhere user with limited or no technology background. Moreover, participants had access to smartphone that they used mostly to chat with or call their families. There were however some associated limitations with the study. Since this was the initial experiment and due to the difficulty to obtain a larger user base, we stucked to this low user base to evaluate our prototype. Our primary target audience of the system according to the Transtheoretical Model [116] would consist of people who are in the action stage of change.

5.7.2 Experimental Design

The experiment measured user acceptance and the role of human coach powered with the chatbot on adhering users to follow a healthy lifestyle. This is useful to validate the features and improve them iteratively. Because the potential number

Participants	Gender	Age Range
P1	Male	≥ 65
P2	Male	65
P3	Male	≤ 65
P4	Female	≤ 65

Table 5.1 CoachAI experiment - Participants demographics.

of participants is low, we used a within-subjects design for the experiment. This increases the chances of finding statistically significant results. The experiment design focused on two phase survey questionnaires. The first phase tested users technology awareness and familiarity with smartphone apps, specifically messaging applications. Whereas, the second experiment measured and validated aspects of participants interaction and the functionalities within the chatbot application.

5.7.3 Scale & Duration

The experiment scale was limited to the number of participants and the actual coach, and the experiment duration. Participants had to have their smartphone to interact with the chatbot and the coach had to have access to the web application to assign plans and track performance. Overall, this allowed for a maximum of 4 participants within a period of three weeks to describe the experiment, collect data and analyze the UX findings. The experiment ran for 7 days of the week, so participants received plans even during the weekend to follow and submit their adherence.

5.8 Activity Data

The coaching application records user's dietary and physical activity data that is received from the chatbot application. This data can be analyzed to possibly find patterns or to corroborate or contradict answers given by the participants in the interviews. A noticeable limitation with the system was the fully availability of the physical coach for the initial experiment. Therefore, to convey with clinically validated data, we used general health recommendation data provided by the world health organization [114, 1]. We specifically referenced the Australian Dietary Guidelines to achieve a healthy weight, be physically active and choose amounts

of nutritious food and drinks (see **Table-5.2** for a sample of the guidelines). The

A sample of the Australian Dietary Guidelines	
Older people should eat nutritious foods and keep exercising to help maintain muscle strength and a healthy weight.	Enjoy a wide variety of nutritious foods from the five groups of food every day.
Plenty of vegetables, including different types and colors, and legumes/beans.	Grain (cereal) foods, mostly wholegrain and/or high cereal fibre varieties, such as breads, noodles, polenta, couscous, oats, quinoa and barley.
Lean meats and poultry, fish, eggs, tofu, nuts and seeds, and legumes/beans.	Milk, yoghurt, cheese and/or their alternatives, mostly reduced fat (reduced fat milks are not suitable for children under the age of 2 years).
Drink plenty of water.	Limit intake of foods high in saturated fat, such as many biscuits, processed meats, crisps and other savory snacks.
Limit intake of foods and drinks containing added salt.	Limit intake of foods and drinks containing added sugars such as confectionary, sugar-sweetened soft drinks and sports drinks.

Table 5.2 The guidelines to recommend activities [1].

activities are assigned by the coach and have a type (e.g., healthy diet or physical activity), a duration (e.g., daily or weekly) and a description (e.g., activity task description). The coach assigns the activities to the chatbot and an activity can be reused to form different plans for diverse users. In **Figure-5.5** we provide a sample of the activities submitted to the participants

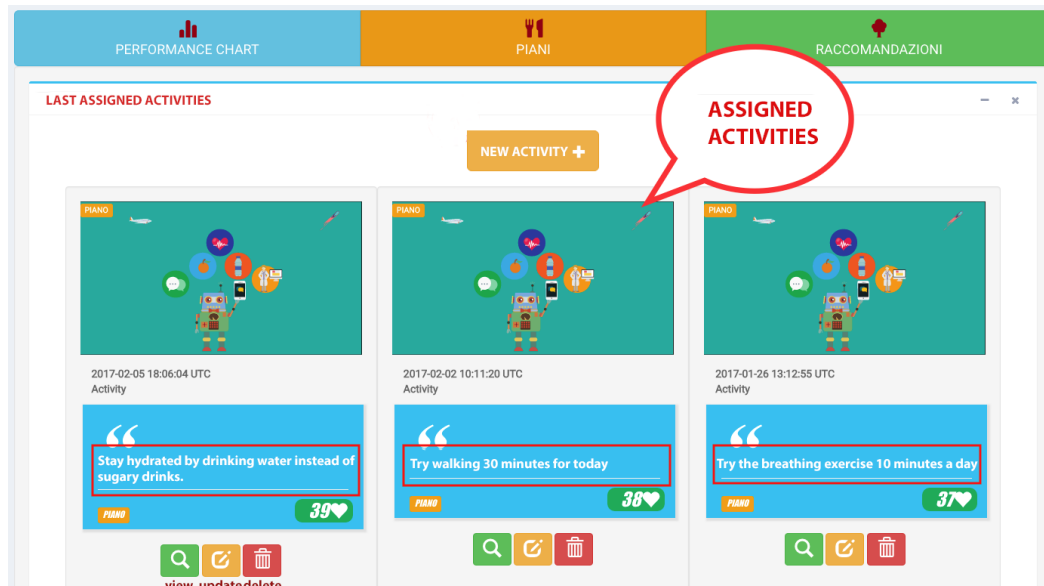


Fig. 5.5 User performance on CoachAI web application.

5.9 Outcome Measures

5.9.1 Pre-Survey Results

The intake survey at the start of the experiment meant to collect some background information about each participant. Participants were asked to fill in their age, weight, height and gender. Additionally, they were asked to choose from a list of options the situation regarding physical activity that most closely resembles their own. This is to determine participant's stage of change. Finally, participants are asked to indicate how much experience they have in several categories relevant to the experiment. These categories are: smartphones/PDAs, ECAs, human coaching, non-interactive self-help products, digital coaching systems and activity measurement systems. Participants were asked whether they are using smartphone and the purpose of having a smartphone. The majority gave higher score for using the smartphone to text messages (see **Table-5.3** and **Figure-5.6** for the pre-survey results).

Pre-Survey Results				
(5 point scale, 1= not at all, 5= very much)				
Questionnaires	P1	P2	P3	P4
<i>I use the phone to make calls</i>	4	4	4	5
<i>I use the phone to make video calls</i>	1	2	3	1
<i>I use the phone to organize the calendar</i>	5	3	3	1
<i>I use the phone to text messages (SMS)</i>	4	3	3	1
<i>I use the phone to text messages (whatsApp / Telegram)</i>	4	5	4	5

Table 5.3 A sample of pre-survey results.

5.9.2 Post-Survey Results

This survey consisted of participants filling out about their experience with the application. We used the User Experience Questionnaire (UEQ) model as described by Laugwitz et al., [115] to evaluate user experience with the chatbot application and measure aspects of their interaction with various functionalities and design elements. Two researchers have distributed the questionnaires and defined the way to respond by providing some example cases. One participant was not present the last day, and we had to send him a digital copy of the questionnaire. The complete version of each

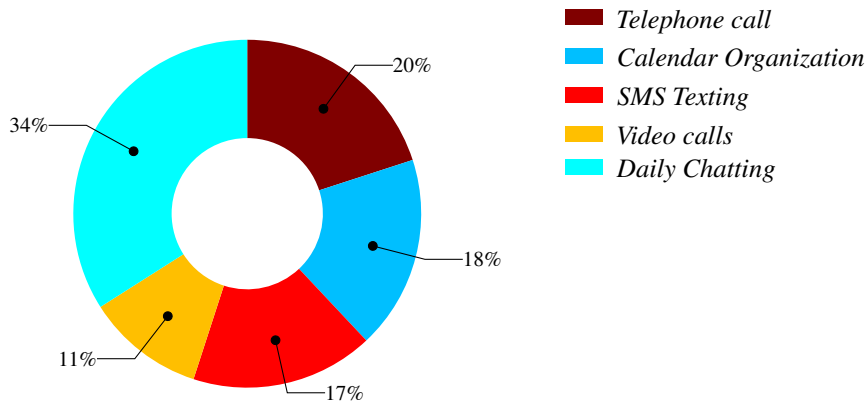


Fig. 5.6 Pre-survey results.

of the surveys as presented to participants can be found in Appendix-A (in Italian). Both questionnaires contain several standardized validated questions, to maximize reliability. Each questionnaire is directly related to one of the research questions posed in Chapter 1. The UEQ contains 6 scales with 26 items. According to the framework, among the scales, attractiveness is a pure valence dimension, whereas perspicuity, efficiency and dependability are pragmatic quality aspects (goal-directed), while stimulation and novelty are hedonic quality aspects (not goal-directed) [117]. This usability questionnaire provides rough indicators for certain application features [117]. The questionnaire uses a seven stage semantic differential to reduce the well-known central tendency bias for such types of items. The final questionnaire contains the scales attractiveness (six items), perspicuity, dependability, efficiency, novelty and stimulation (four items each). We calculated the scale means from the collected questionnaires.

5.9.3 Post-it Session Analysis

After the third week conclusion of the experiment, the two researchers performed a post-it session and usability testing with the think aloud protocol. The post-it session was to analyze aspects of the technology the participants liked/disliked and justify their decision. Whereas, the think-aloud testing was to test the chatbot interface with the users. The researchers asked the participants to think out loud, describe their expectations and feelings about the experience. The two sessions analyzed users experience and collected data about their experience through the post-it to check

for evidence of usability issues and user preferences (see **Table-5.4** for the users response outcome from the post-it session).

Focus Group Analysis	
What I Like	What I don't Like
<i>Get in contact with the coach</i>	<i>Sometimes the app did not respond</i>
<i>Good performance</i>	<i>The task arrived late</i>
<i>The mobile is always with you</i>	<i>Messages not always clear</i>
<i>Speed (of interaction)</i>	<i>Lack of notifications</i>
<i>Simple interaction</i>	<i>Lack of communication between peers</i>
<i>Graphic</i>	<i>More interaction between coach and the users</i>

Table 5.4 Post-it session analysis.

User Experience

Participants were asked to indicate (on a scale 1 - 7) which of the two adjectives they feel more appropriately describes the system, and how strongly they feel about it. Where #1 stands for very strong association with the first adjective and #7 represents feelings that are less strong, and whereas #4 meaning both adjectives are equally applicable (middle score).

Because this test has a strong psychological characteristics, participants are asked not to spend a lot of time thinking about their answers, but rather to choose to give their responses based on their initial feelings and instincts. Participants are also asked to give an answer on each of the adjective pairs, even if they intellectually feel that neither of the adjectives apply to the system. In these cases their (possibly subconscious) evaluations may still be valuable. A number of the adjective pairs were reversed, so that there isn't a "positive" and a "negative" column. This is to get participants to interpret the concepts properly, and avoid acquiescence bias.

We distributed 15 questions to analyze usability, acceptability, and design scales of the applications. The scales considered 15 dimensions within the data. Usability considered the perceived efficacy, perceived efficiency, pleasantness of use, ease of use, learnability, memorability dimensions of the app, whereas acceptability considered the comprehension, fun, daily usage, social, similarity, future usage, familiarity, and utility. Finally, the design scale considered the emoji dimension in the application.

The user experience questionnaire scoring was used to estimate to which properties

of the application are users particularly responsive. That said, the questionnaire helps estimate which feelings or attitudes of users are caused by the application and what are the typical reactions of users during or after usability studies. In **Figure-5.7** we show the UEQ Scales and participants scale ratings. The interview outline focused

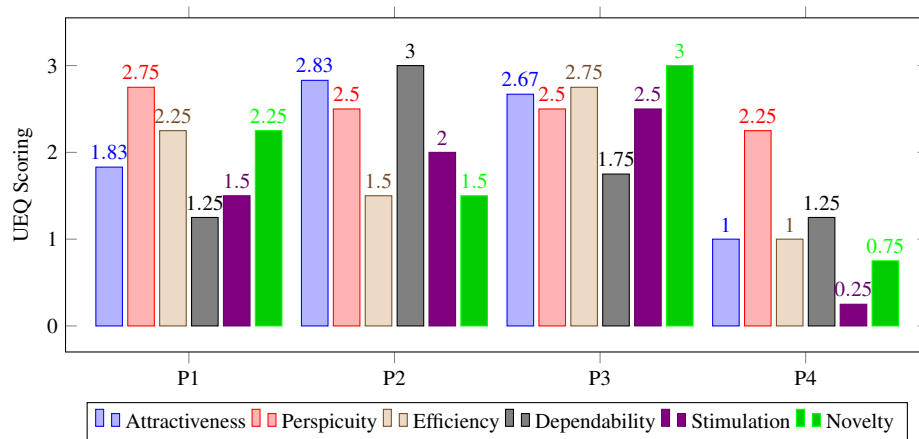


Fig. 5.7 The UEQ scales and participants scale ratings.

on analyzing habits and technology use in everyday engagement adherence. After the first week of using the CoachAI application, we analyzed the use of notification and reminders. The evaluation dimensions focused on the usability, acceptability, ethics and unobtrusiveness aspects of the interaction. In **Table-5.5** we highlight the evaluation dimensions and users review relevant to each one. e

Evaluation dimensions	
Usability	The coaching service allows me to improve my lifestyle
	The commitment I am asked for by the coaching service seems proportionate to me
	I learned to use the app easily
	I can easily remember how to use the app
Acceptability	I am happy to use the coaching service
	I understand what is needed and how the coaching service works
Ethics	I believe privacy and the protection of my private life are guaranteed by the app
	I trust in the coaching service
Unobtrusiveness	I think the number of notifications is proportionate to the activity
	Responding to notifications does not take me too long
	Responding to notifications does not obstruct my daily activity

Table 5.5 Evaluation dimensions and users review.

5.10 Results

The UEQ does not produce an overall score for the user experience. The values for the single items are listed to allow detecting outliers in the evaluations. If an item shows big deviations to the evaluations of the other items of the same scale this can be a hint that the item is misinterpreted by a higher number of participants. Values between -0.8 and 0.8 represent a neutral evaluation of the corresponding scale, values > 0.8 represent a positive evaluation and values < -0.8 represent a negative evaluation. The range of the scales is between -3 (horribly bad) and +3 (extremely good). It is due to the calculation of means over a range of different persons with different opinions and answer tendencies, for example the avoidance of extreme answer categories, extremely unlikely to observe values above +2 or below -2 (see **Table-5.6** for the UEQ results).

The focus group analysis performed after the experimentation with the post-it session helped us to analyze users review and extract meaningful insights about what they liked/disliked about the interaction, design and functionalities within the application. We summarized the participants percussiveness, the features and participants posts result in **Figure-5.8** obtained from the post-it session. This session provided useful insights on how participants perceived a feature and design element of interaction pattern in the chatbot application.

The users showed general impression with the system in contexts related coaching, application portability, design and interaction. Most participants perceived the availability of human coaching as positive in tracking their condition, as in: *"It is easy to contact the coach"*. The application portability was also perceived as positive by users. The portability referred mainly to having the chatbot in the messaging application or smartphone, as in: *"Having the chatbot in your messaging application is comfortable"*. The interaction feature referenced by users was mainly about the performance and simplicity, as in: *"It is simple to use and interact with"*. Finally, the design aspect perceived positively was about the chatbot graphics, which includes the buttons and emojis used throughout the interaction within the chatbot dialogue. However, the practical problems experienced varied from technical to interaction issues. For example, users highlighted connection problems, app not being accessible throughout the interaction, as in: *"Sometimes the bot did not respond!"*. Whereas, the interaction issues were about the lack of a desired feature, such as users community, as in: *"Lack of communication between peers"*.

Item	Mean	Variance	Std. Dev.	No.	Left	Right	Scale
1	2,7	0,3	0,6	3	<i>regrettable</i>	<i>welcoming</i>	Attractiveness
2	2,7	0,3	0,6	3	<i>incomprehensible</i>	<i>comprehensible</i>	learnability
3	3	0	0	3	<i>creative</i>	<i>devoid of imagination</i>	Originality
4	2,7	0,3	0,6	3	<i>easy to understand</i>	<i>hard to understand</i>	learnability
5	1	1	1	3	<i>of great value</i>	<i>of little value</i>	Stimulation
6	1,7	0,3	0,6	3	<i>boring</i>	<i>exciting</i>	Stimulation
7	2,3	1,3	1,2	3	<i>not interesting</i>	<i>Interesting</i>	Stimulation
8	1,7	5,3	2,3	3	<i>unpredictable</i>	<i>predictable</i>	Controllability
9	2,3	1,3	1,2	3	<i>fast</i>	<i>slow</i>	Efficiency
10	0,7	10,3	3,2	3	<i>original</i>	<i>conventional</i>	Originality
11	2,3	1,3	1,2	3	<i>obstructive</i>	<i>support</i>	Controllability
12	2,7	0,3	0,6	3	<i>good</i>	<i>bad</i>	Attractiveness
13	2	0	0	3	<i>complicated</i>	<i>easy</i>	learnability
14	1,7	2,3	1,5	3	<i>repellent</i>	<i>attractive</i>	Attractiveness
15	2,7	0,3	0,6	3	<i>common</i>	<i>modern</i>	Originality
16	2,3	0,3	0,6	3	<i>unpleasant</i>	<i>enjoyable</i>	Attractiveness
17	1,3	8,3	2,9	3	<i>sure</i>	<i>not sure</i>	Controllability
18	3	0	0	3	<i>activating</i>	<i>soporific</i>	Stimulation
19	2,7	0,3	0,6	3	<i>expectation</i>	<i>not expectation</i>	Controllability
20	2,7	0,3	0,6	3	<i>inefficient</i>	<i>efficient</i>	Efficiency
21	3	0	0	3	<i>clear</i>	<i>confused</i>	learnability
22	0,7	10,3	3,2	3	<i>not pragmatic</i>	<i>pragmatic</i>	Efficiency
23	3	0	0	3	<i>ordered</i>	<i>overload</i>	Efficiency
24	2,7	0,3	0,6	3	<i>attractive</i>	<i>not attractive</i>	Attractiveness
25	2,7	0,3	0,6	3	<i>nice</i>	<i>unpleasant</i>	Attractiveness
26	2,7	0,3	0,6	3	<i>conservative</i>	<i>innovative</i>	Originality

Table 5.6 UEQ results.

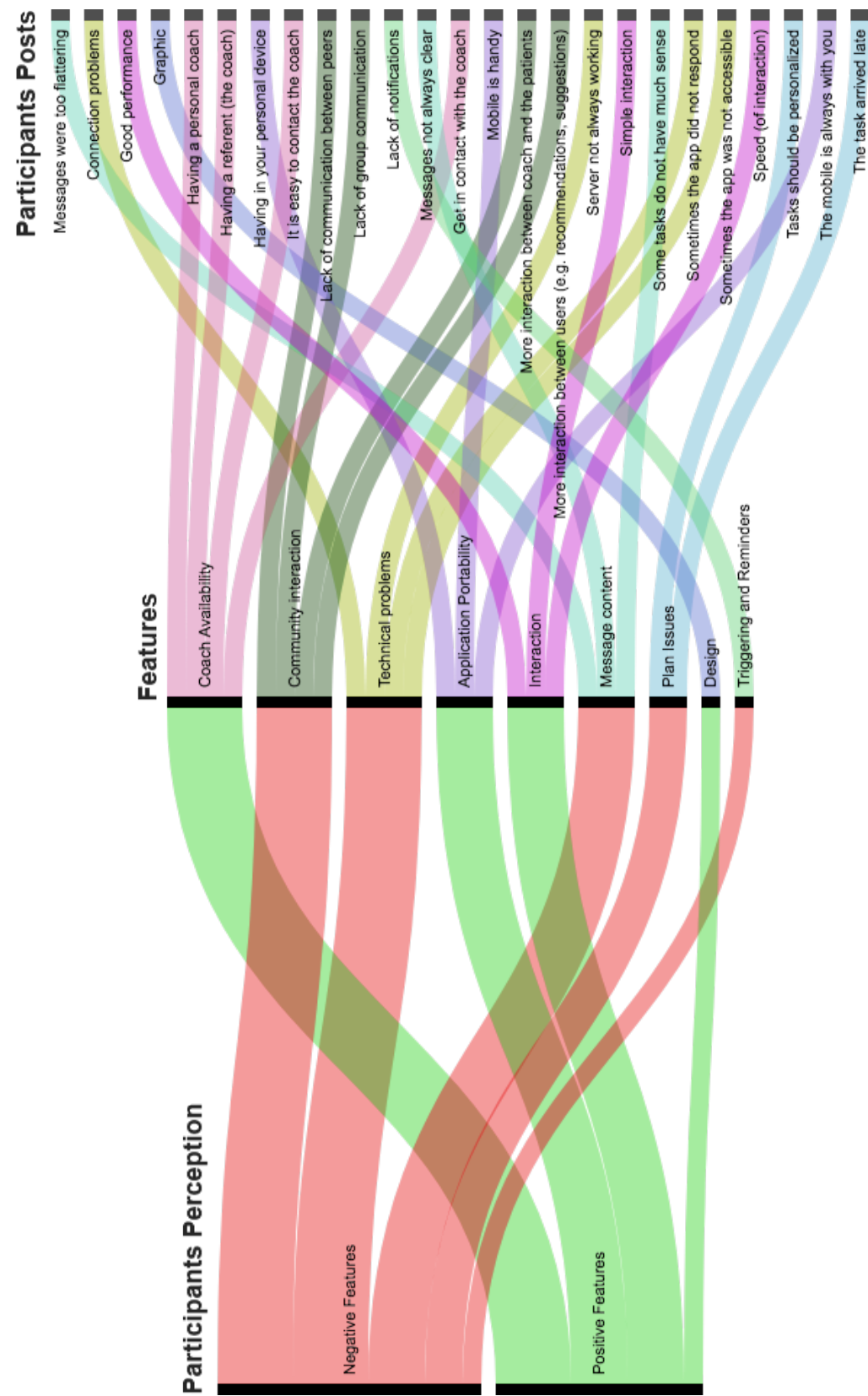


Fig. 5.8 Participants percussiveness and post results.

The findings have shown that the usability and acceptability were the highly rated scales among all the scales. The mean score for Usability=6.5 and Acceptability=5.7. The scale and dimension(SD) for both were Usability=0.4, Acceptability=0.7. **Table-5.7** and **Figure-5.9** show the usability and acceptability average for all participants. Across the experiment, there was strong support for hypothesis H1; delivering

	Mean Score	SD
<i>Usability</i>	6.5	0.4
<i>Acceptability</i>	5.7	0.7

Table 5.7 The usability and acceptability Mean and SD.

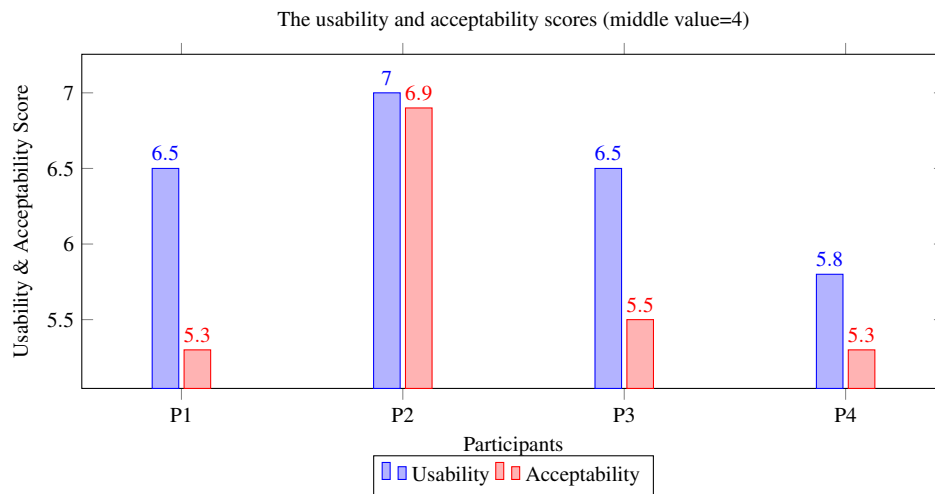


Fig. 5.9 The usability and acceptability score.

health activities with the chatbot positively correlates with user compliance. The research concluded that delivering the health activities with the chatbot was perceived positively by the participants. However, there was little or no support for hypothesis H2; assigning more tasks to the bot did not show any coach task optimization while providing services to the user.

5.11 Persisting Challenges

The challenges obtained from the design and preliminary evaluation revealed limitations in terms of robust dialogue system that can provide more flexibilities when interacting with the user. This is still limited even with advanced systems where

they use machine learning or neural network techniques, such as Recurrent Neural Network (RNN) which suffers from falling outside the conversation context [118]. Existing chatbot platforms are still limited with respect to the rich CUI to build interactive chatbots. The challenge associated with CoachAI prototype is creating a sustainable interaction between the users and the coach. Another limitation is related to data logging, which is often a tedious process for the user. In addition, triggering users to reuse the tool is essential for the success of the bot. Therefore, we have decided to implement the bot into a mobile device and enrich it with notifications, and automate aspects of data logging while keeping the interaction minimum. Based on the above results, some technical and design improvements needed to be taken into consideration.

5.12 Conclusion

In this chapter, we discussed the general dietary and physical activity recommendations used in CoachAI prototype experimentation with elderly participants who had previous cardiac conditions. The chapter presented the prototype design and implementation and the conclusions derived from the user experiment. We started by discussing original research questions, then the execution of experimental process from discussing the application, executing the experimentation, and analyzing the findings. We performed a pre-experiment questionnaire with participants about their general knowledge on technology and using messaging applications. We then performed a post-it session and post-experiment questionnaire to analyze their experience with the chatbot and gather some data for the post-analysis. The suggestions and recommendations regarding the experiment were used in the iterative refinement of the application. Participants have revealed the features they liked and disliked within the app, which was useful to readopt the current version for user's needs. We concluded that using a chatbot to provide general dietary and physical activity recommendations to elderly is a sound approach. Future data analysis will validate if such approach is effective in providing long-term coaching support to promote individuals' overall lifestyle and health condition.

This work was partially supported by the [Associazione Per La Lotta Alle Malattie Cardiovascolari](#). We thank all the participants and the caregivers who supported this study.

Chapter 6

Second Prototype and Feature Refinement

6.1 Introduction

The renaissance in conversational user interfaces is evident in many domains to help handle tasks, support users and accomplish goals. In this chapter we discuss two distinct studies and experimentations related to CoachAI dialog comparison and flow. We begin by discussing the domain specific chatbot dialog comparison with/out emojis. We then discuss the CoachAI dialog state transition and task completion. We describe the process along with relevant information about the experiments. We conclude by presenting the results of and implications from each study.

Chatbot Dialog - Cohort Study

Dialog systems and conversational agents can bring a lot into healthcare to reduce cost, increase efficiency and provide continuing care, albeit its infancy and complexity with regard to building natural dialogs. However, the design guidelines to design dialogs for conversational agents are usually based on common knowledge, and less frequently on empirical evidence. For example, the use of emojis in conversational agent dialogs is still a debated issue, and the added value of including such graphical elements is mainly anecdotal. In this work, we present an empirical

study comparing users feedback when interacting with chatbot applications that use different dialog styles, i.e., plain text or text with emojis, when asking different health related questions. We believe this work can contribute to the research on integrating conversational agents in the health and wellbeing context and can serve as a guidance in the design and development of interfaces for text-based dialog systems.

6.2 Objectives

Our objective is to discuss the effect of communication style of a conversational system architecture in the context of health and wellness. We particularly focus on the role of emoji in textual conversational agents for physical activity and mental wellness data gathering. For that, we will describe our system and the result of a user study with 58 participants. The study consisted of two chatbot versions, namely emoji based dialog and plain text dialog, to converse with users about their physical activity and mental wellbeing. We hypothesized that using emojis will benefit the most to conversational dialog where user emotion and empathy are relevant, such as mental wellness. Whereas, it is less relevant to conversation dialog intended for physical activity data collection. Regardless, emojis could make the generated replies more anthropomorphic and interesting, which might enhance user experience with the dialog system.

6.3 Study Design

This study compares interactions with chatbot conversational agents used to collect information on two different types of personal health data: mental and physical wellbeing. While mental wellbeing is usually evaluated with questions on the person's emotional, affective and cognitive status (e.g., *"How are you feeling now?"*), physical activity is assessed through objective and factual information (e.g., *"How many steps did you take today?"*). In order to recreate these two conditions in the experimental study, we developed two different dialogs based on two widely used questionnaires created to assess wellbeing in general population: the General Health Questionnaire - GHQ [119] and the Global Physical Activity Questionnaire - GPAQ [120]. The GHQ is a screening questionnaire for identifying minor psychiatric disorders in

the general population, including non-clinical population [119]. The GHQ comprises 12 items with a four-point Likert response scale. The GPAQ is a questionnaire developed by the World Health Organization for monitoring physical activity measuring sedentary and active behavior in different contexts (daily routine, work and leisure time) [120]. GPAQ originally includes 16 items, but only 12 of them were used in this study in order to match the length of the two dialogs.

Both questionnaires were used as bases for creating two comparable chatbot dialogs, and for each dialog a plain text and an emoji version were created (see **Figure-6.1** for a visual view of the different chatbot interfaces and **Table 6.1** for excerpts from the different dialogs). The plain text version displayed the questions in textual format. For the emoji version, emojis were inserted at the end of each sentence as decorations [121]. In assigning the emojis, we considered the main concept expressed in the question and inserted a related emoji at the end of the sentence. Emojis were also displayed on the response options, providing some variations (see **Table 6.1** for the two examples). The emojis did not replace any text and were only added at the end of each question and answer option. The questions were indeed the same in the two dialogs, the difference was in the presence/absence of the emojis. The **Figure-6.1**

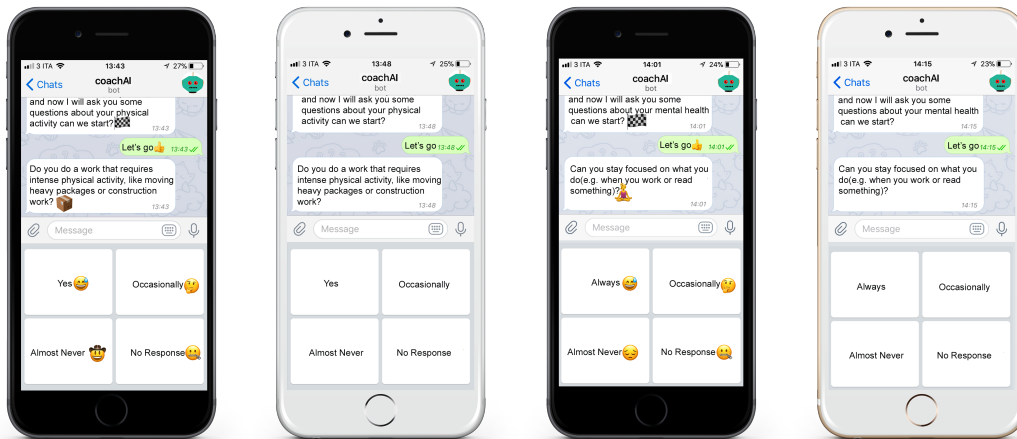


Fig. 6.1 An interface view of the four bot versions. From left to right (1) Physical wellbeing with emojis (2) Physical wellbeing with plain text (3) Mental wellbeing with emojis (4) Mental wellbeing with plain text.

provides a visual comparison for each of the four different chatbots. (1) *Physical wellbeing with emojis* is asking the user whether their work requires intense effort or mostly sitting, (2) *Physical wellbeing with plain text* performs the same dialog but only with plain text, (3) *Mental wellbeing with emojis* is asking if the user is able to

focus on their task while working, and (4) Mental wellbeing with plain text performs the same dialog, but with plain text.

Source	Emoji example	Plain text example
<i>Mental Wellbeing - GHQ</i>	Have you recently lost much sleep over worry? 🇸🇬 • More than usual 😫 • As usual 😊 • Less than usual 😊 • I prefer not to answer 🙅	Have you recently felt constantly under strain? • More than usual • As usual • Less than usual • I prefer not to answer
<i>Physical Wellbeing - GPAQ</i>	Do you walk or use a bicycle to get to and from places? 🚲 • Very often 🏃 • Sometimes 🚲 • Rarely 🚫 • I prefer not to answer 🚫	Does your work involve vigorous-intensity activity that causes large increases in breathing or heart rate? • Very often • Sometimes • Rarely • I prefer not to answer

Table 6.1 Examples from the two dialogs used in the study.

6.3.1 Participants and Experimental Design

The study involved 58 participants randomly assigned to one of the two conditions. The two conditions are consistent with the 2x2 design, with topic of the dialog domain (physical vs. mental wellbeing) and style of communication (with emoji vs. text only) as independent variables. All participants interacted with both dialog systems (physical and mental information), one was characterized by the use of the emoji in the dialog and one with plain text. The order of the presentation was counterbalanced across participants.

The overall sample consisted of 28 male and 30 female participants. Their age ranged from 18 to 60 years old ($M = 29.9$, $SD = 11.9$). Half of the participants responded to the mental wellbeing dialog enriched with emojis and the physical wellbeing dialog in plain text (Group A), while for the other half the combination of type of dialog and style of communication was inverted (Group B). The two groups did not differ in terms of age, level of self-confidence, trust propensity and emoji familiarity (all tested with the Mann-Whitney nonparametric test) (see **Table-6.2**).

	Group A	Group B	Comparison
<i>Emoji enriched dialog</i>	<i>Mental wellbeing (GHQ)</i>	<i>Physical wellbeing (GPAQ)</i>	
<i>Plain text dialog</i>	<i>Physical wellbeing (GPAQ)</i>	<i>Mental wellbeing (GHQ)</i>	
<i>Gender</i>	F=15, M=14	F=15, M=14	
<i>Age</i>	29.9 (12.8)	29.8 (11.1)	Z= -0.6; p=.55
<i>Self-confidence</i>	3.9 (0.8)	3.9 (0.9)	Z= -0.01; p=.93
<i>Trust propensity</i>	3.7 (0.9)	3.4 (0.7)	Z= -0.9; p=.33
<i>Familiarity with emojis</i>	4.3 (1.3)	4.2 (1.5)	Z= -1.2; p=.90

Table 6.2 Characteristics and descriptive statistics for the experimental groups. The Mann-Whitney test was used to test for differences between the two groups.

6.3.2 Chatbot Dialog

The dialog architecture was structured using ChatScript dialog engine¹. ChatScript is a rule-based engine, where rules are created in program scripts through a process called dialog flow scripting. It uses a scripting metalanguage as their source code. We defined the bot's dialog flows with a script stored as a normal text file. Topics are contained in specific files. For example, the ~exercise topic code is contained in exercises.top file. Writing the scripts as a text files gave us full control over the dialog flows. The whole procedure took about 20 min, with an average time of 4 minutes of interaction per dialog. Here what a ChatScript scripting file looks like:

```
s: ( I like juggling )
Are you a fan of outdoor juggling?
a: ( ~yes )
Wow me too Ahmed. Did you try doing it this week?
*   b: ( ~no ) Ok no problem. What about we organize a session?.
*   b: ( ~yes ) You probably are already active...
a: ( ~no ) What exercises do you like?
*   b: Yoga and meditation.
*   b: None, I don't exercise...
```

6.3.3 Procedure

After the welcoming and the instructions on how to perform the testing, participants were asked to respond to the questionnaire for measuring individual self-confidence and trust. After completing the questionnaire, participants started the first dialog

¹<https://github.com/bwilcox-1234/ChatScript>

with the chatbot using the Telegram desktop application. We chose the desktop version of Telegram to make using the chatbot as easy as possible and display it on a computer screen rather than a small smartphone screen. Conversational agents in health domain can directly initiate the conversation (system-initiated) [122], let the user to initiate the conversation (patient-initiated) [122, 123], or provide a mixture of the two (mixed-initiative) [122]. All the chatbots dialog in our study followed a user oriented dialogs (user-initiated): users triggered all the contents and the dialog content changed based on their selected options. At the end of each interaction participants completed a questionnaire for estimating engagement, attitude and confidence in the conversational agent. After the last interaction, participants were asked to answer the questions comparing the experience with the two dialogs.

6.3.4 Measures

The measures consisted in a questionnaire for assessing individual self-confidence and trust, and for estimating engagement, attitude and confidence in the chatbot. Finally, participants were asked to answer few questions comparing the experience with the two dialogs.

Individual self-disclosure and trust propensity. Self-disclosure and trust propensity measures were collected to control for differences between groups of participants in revealing information about themselves to the conversational agent. Self-disclosure has been defined as the propensity of an individual to communicate information about him/herself to another one [124]. Self-disclosure was measured using three dimensions of the Wheelless's scale [125]. Specifically, the three dimensions of self-disclosure were measured using a 6-point scale for each, namely: (i) amount of disclosure, (ii) positive/negative nature of disclosure, and (iii) honesty of disclosure. Moreover, four items from the Mayer and Davis's Trust Propensity Scale [126] were used to measure dispositional trust. For both self-confidence and trust propensity, scale values were calculated as the average of the items.

Familiarity with emojis. Three items, rated on a 6-point scale, were used to assess participants' familiarity with the use of emojis. The items were: i) I like to read text with emojis; ii) I use emoji in chatting apps such as Telegram and Whatsapp; iii) I think emojis enrich the communication. Also these items were used to control for differences between the groups.

Interaction experience: enjoyment, attitude and confidence. The interaction was eval-

uated using three scales, i.e. engagement, attitude and confidence, each composed of four items with a 6-point Likert rating scale (from "strongly agree" to "strongly disagree"). The items used in the scales were adapted from [127]. The *enjoyment scale* comprises items related to the enjoyment and entertainment in the interaction. The *attitude scale* measures user's attitude towards the dialog and the interaction with the agent. Lastly, the *confidence scale* evaluates user perception of the privacy and intrusiveness of the questions in the dialog. Items and scale reliability scores are reported in **Table 6.3**.

Dialog comparison. Three items, rated on a 6-point scale (from "strongly agree" to

Scale	Item
<i>Enjoyment</i> $\alpha = .91$	I enjoyed chatting with the conversational agent during the interaction
<i>Attitude</i> $\alpha = .87$	I felt comfortable answering the questions
	The more I interacted with the agent, the more I liked the experience
	I would like to chat again with this conversational agent in the future
	I found a kind of "emotional connection" between myself and the conversational agent
<i>Confidence</i> $\alpha = .77$	I found the dialog with the conversational agent to be realistic
	I noticed a negative emotional change in myself during the interaction *
	I found the dialog to be coherent. In other words, the sequence of responses of the conversational partner made sense
	The agent asked very personal questions *
	I trust the conversational agent
	I found the questions to be very intrusive *
	I answered the questions with honesty

Table 6.3 Scales and items used in the questionnaire. Items followed by an asterisk were inverted for the analysis. Cronbach's alphas are reported for each scale.

"strongly disagree"), were asked immediately following the last questionnaire for comparing participants' experience with the two dialogs. The sentences were: i) The two dialogs were identical; ii) I liked more the dialog with the emojis; iii) The dialog without emojis was easier to read.

6.4 Results

Data was analyzed through a mixed-design ANOVA model, with "communication style" (emoji vs. plain-text) as within-subject factor and "experimental group" (group A vs. group B) as between-subject factor. The difference between the two groups was that in the former participants saw the mental wellbeing dialog with the emoji and the physical wellbeing dialog in plain text, while in the latter the conditions were inverted. Regarding the interaction experience, descriptive measures from the three scales are reported in **Table-6.4**. The ANOVA found a significant effect of the

Dimension	<i>Mental wellbeing</i>		<i>Physical wellbeing</i>	
	Emoji	Text	Emoji	Text
<i>Enjoyment</i>	4.7 (1.1)	4.1 (1.4)	4.2 (1.3)	4.8 (1.0)
<i>Attitude</i>	4.0 (0.9)	3.4 (1.3)	3.5 (1.2)	4.0 (0.7)
<i>Confidence</i>	4.6 (0.8)	4.2 (1.0)	4.7 (0.8)	5.0 (0.8)

Table 6.4 Questionnaire results for the interaction experience. Average values (SDs in parentheses)

between subject factor for all the three scales. When participants had to score the mental wellbeing dialog with emoji and the physical wellbeing dialog with plain text, they assigned higher scores for both enjoyment ($F(1,56)= 4.4$; $p<.05$), attitude ($F(1,56)= 4.2$; $p<.05$) and confidence ($F(1,56)= 8.9$; $p<.01$). Moreover, a significant interaction between factors has been observed specifically for the confidence scales ($F(1,56)= 21.44$; $p<.01$). This effect indicates that the profile of confidence ratings when the dialog had emojis or plain text were different depending on the type of dialog. Post-hoc analysis revealed that the confidence for the dialog with emojis were similar between physical and mental wellbeing, however, compared with the plain text version, the scores increased for the physical wellbeing dialog and decreased for the mental wellbeing (see **Figure-6.2**).

Regarding the items comparing the two dialogs, the ANOVA found a difference only on the third item of the scale. Both groups noticed a difference in the two dialogs and tended to prefer the dialog with emojis. However, group A found the dialog without emojis easier to read compared to group B ($F(1,56)= 4.36$, $p<.05$).

Confidence in the chatbot

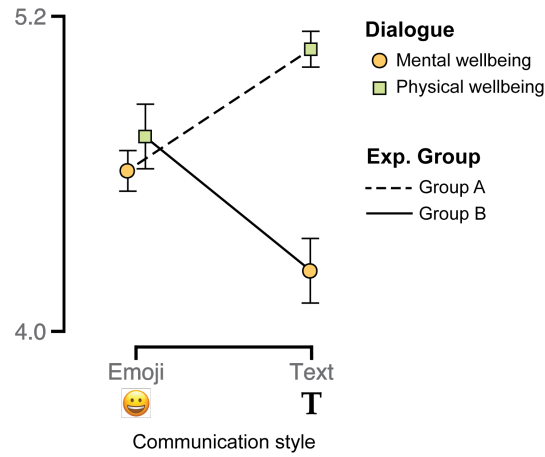


Fig. 6.2 Scores for the "confidence" scale.

6.4.1 Interaction Time Analysis

In order to assess possible differences in the time taken to interact with the dialog systems, we ran a mixed-design ANOVA on interaction time (time taken to complete the two dialogs measured in seconds, see **Table 6.5**). The analysis suggested no differences in interaction time between the two communication styles ($F(1,56)=1.68$; $p>.05$). There was a significant effect of the between subject factor ($F(1,56)=13.57$; $p<.01$), showing that group A required more time ($M=365$ sec, $SD=105$) to complete the two dialogs with respect to group B ($M=206$ sec, $SD=102$). A possible explanation is that when users are engaged in the conversation, and have a positive attitude and high confidence toward the chatbot, as in the group A, they spent more time interacting with the agent.

	<i>Mental wellbeing</i>		<i>Physical wellbeing</i>	
	Emoji	Text	Emoji	Text
<i>Time</i>	383 (217)	190 (71)	223 (134)	346 (134)

Table 6.5 Average time taken to complete the dialogs (in seconds)

6.5 Discussion

The result of the study showed how the use of emoji can influence user experience with chatbot applications for health information tracking. The findings point out that emojis can benefit enjoyment, attitude and confidence with the conversational agent when the dialog is around emotional and personal topics, such as mental wellbeing, while it might hinder the conversation experience when the agent asks questions about objectives and factual information, such as physical activity. This effect is noticeable for the confidence scale: participants were confident in sharing information with the agent about their mental wellbeing when the dialog included emojis, while they were less confident when the dialog was in plain text only. The effect was inverted for the dialog about physical wellbeing, with a decrease in confidence when the emoji were used. Our findings suggest that the use of emojis in some health context might be a promising strategy to influence user's engagement, as it was also observed in other studies on health text message intervention [128]. These findings might guide practitioners in designing chatbot dialogs and interfaces, providing indications on whether to include emojis in the communication for health intervention.

Chatbot services might benefit from having a nuanced understanding of emotional content in text represented by emojis and it would be interesting to replicate this research to test a chatbot integrating an AI-based emoji generator (e.g., using a deep learning approach). Another interesting aspect for future research is to improve the personalization within chatbots by creating a user model, before the user actually interacts with the chatbot (e.g., the user could fill in a personality questionnaire, such that the chatbot adapts to the user). This includes analyzing how users will perceive the use of self-prepared sets of emojis that resembles their emotion. Moreover, there are many health issues that could have their own emoji representation to graphically express ideas and feelings about the context. We found several studies [129, 130] analyzing the use of emojis to express feelings in e-health applications. However, there were no empirical studies that provided a mapping of emojis to a specific health domain, or particularly health dialogs. Future work will explore methods to incorporate emojis into specific health oriented dialog models (e.g., for diet management, and physical or mental wellness).

6.6 Persisting Challenges

The study has also some limitations worth noting. The dialog with the conversational agents was mainly based on questions and answers. We can hypothesize that different effects of the use of emoji might emerge with different types of interactions (for example, information retrieval or command execution). This work focused on a specific application context, the health domain, that covers personal and private aspects of a person's life. We can expect that the effect of emojis might vary with respect to different contexts and settings. The current work explored the use of emojis with a decorative use, while different patterns of use (e.g., stand-in or reaction use of emojis) might yield different results. Lastly, while there were no significant statistical differences regarding the age of the two experimental groups, most of the participants (76%) were less than 30 years and were young adults. These considerations may constrain the generalization of the results to different user population.

CoachAI - Dialog State Comparison

The purpose of this section is to compare the dialog structure and task processing of CoachAI with two distinct chatbots, namely Woebot² and Lark³. Where the former is a chatbot for mental wellness and cognitive therapy and the later is a chatbot to track individuals overall health and lifestyle. We tested only the free version of the applications. Each of CoachAI and the other two chatbots follow a distinct approach to track aspects of individuals health and provide them with the feedback to help improve their condition.

6.7 Methodology

We conducted a comprehensive review of quality attributes for each of the chatbots and identified appropriate quality approaches. The comparison consisted of in-person user testing to perform a conversation with each of the chatbots on the topic they support. However, since CoachAI is general in terms of topic focus, the user

²<https://woebot.io/>

³<https://www.web.lark.com/>

performed both mental wellness and physical activity conversation in the chatbot. The features extracted from each of the chatbots were listed and compared among all the three chatbots (see **Table-6.6** for the various features found in each of the chatbots). The selected two chatbots were based on two criteria. First, their focus on an aspect of health coaching and second their focus on lifestyle promotion, rather than disease condition management.

Features	CoachAI	Woebot	Lark
NLP	Task-oriented	Slot-based	Task-oriented
Initial Profiling	✓	✓	-
Dialog Structure (Matching technique)	Finite State Machine (FSM)	Deep Learning	Slot filling
Conversation Dead-ends	-	✓	-
Emotional Intelligence	-	✓	✓
Conversation Flow	Linear	Non-linear	Linear
Chatbot Personality	✓	✓	✓
User Content Personalisation	✓	-	-
User Condition Tracking	✓	✓	✓
Dialog Focus	Lifestyle Coaching	Virtual Therapist	Lifestyle Coaching
Coaching Approach	Semi-automated (Human + Bot)	Automated (Bot)	Automated (Bot)
Corpus(knowledge-base)	Coach defined	CBT related corpus	Coach defined
Application Type	Lifestyle Coach	Health Assistance	Lifestyle Coach
Conversation Type (Coach-User)	1-many	1-1	1-1
Dialog Model	Retrieval-based model,	Deep learning model	Pattern-based heuristics

Table 6.6 Chatbots feature comparison.

6.8 Evaluation

We compared the dialog output of CoachAI with the Woebot for cognitive behavioral therapy (CBT) and Lark for physical activity and diet promotion. We discovered each of the chatbots focus on providing health coaching at a different stage of either the CBT (CoachAI and Woebot), and physical activity and diet promotion (CoachAI and Lark) (see **Figure-6.3**). To measure various aspects of the chatbots dialog, the participant had to converse with the bots about irrelevant context. However, this was not possible with Lark since it is strictly button oriented dialog. The study investigated the dialog initiation (by user or the bot), whether the dialog is linear or

not, and whether it is text-based, button-based or both.

The evaluation consisted of the elements that form a chatbot dialog based on the approach followed to develop it. Our study indicates that although all the chatbots are focused on achieving the same goal, health and wellness promotion, the comparison show that the chatbots vary in many points based on their design, the dialog engine, the platform and the overall system architecture. From **Table-6.6** and **Figure-6.3**

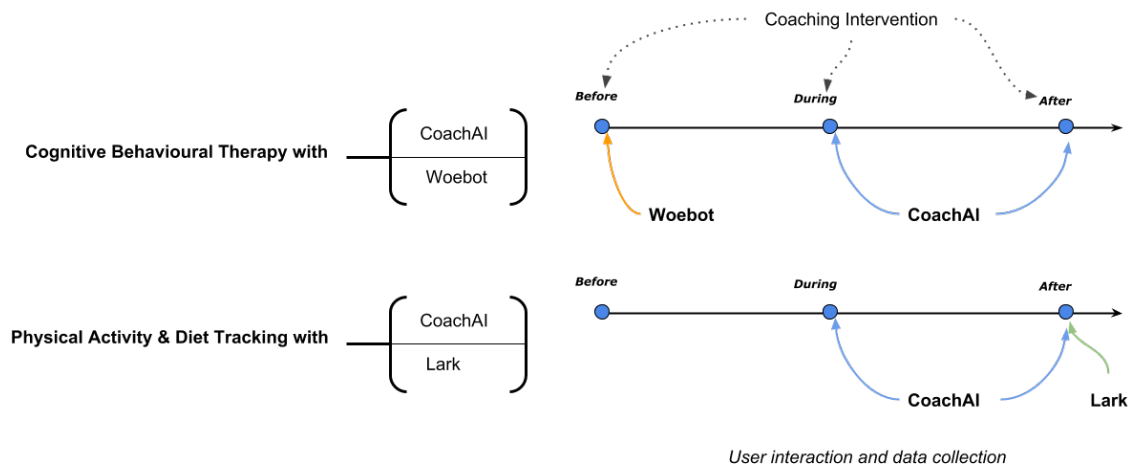


Fig. 6.3 Coaching intervention focus of each chatbot.

we see the differences between the three chatbot applications. Some differences are related to the approach, the design or the implementation and platform choice. Although the pros and cons of each approach, however we can not conclude which approach is better in terms of tasks handling and sustaining user engagement. This is due to the short evaluation of each application and to provide conclusions on which system is better in an aspect rather than an other, a long-term randomized clinical trial (RCT) is needed. That said, we highlight some interesting distinctions found in each application after we tracked the dialog flow for each one.

6.8.1 Initial User Profiling

Based on the dialog flow analysis, we found CoachAI and Woebot to perform initial user profiling to either collecting information or perform initial analysis of their condition. For example, CoachAI performs initial data collection to analyze each user and perform clustering on them based on their physical activity level. In **Figure-6.4** we show the three chatbots with their initial information gathering.

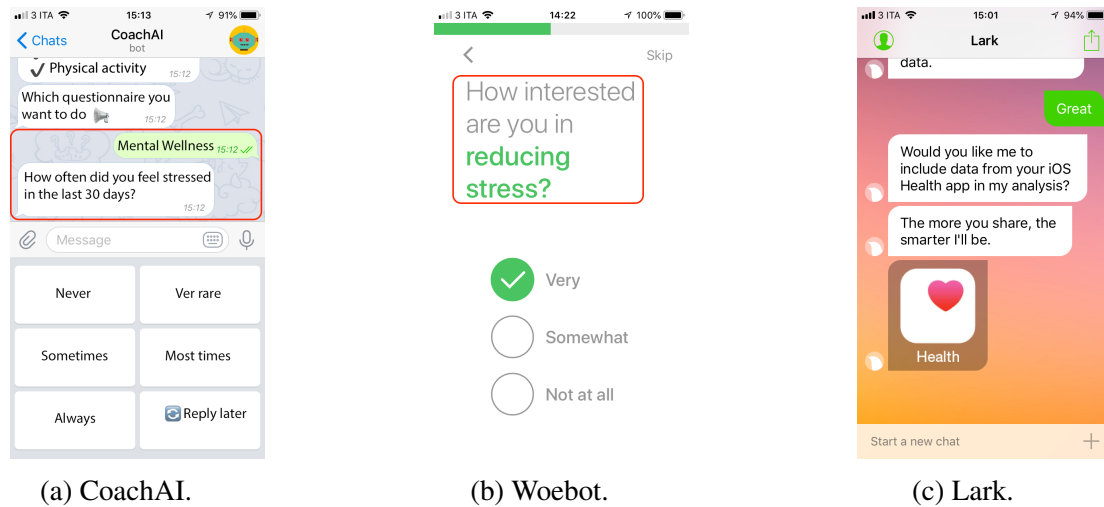


Fig. 6.4 Initial user profiling by each of the chatbots.

6.8.2 Interaction Feedback

All three chatbot systems provide feedback on users interaction. This is either based on their interaction with the application and the feedback they provide on aspects of their health (e.g., Woebot), or it is about their adherence to the activity assigned and their level of compliance to these activities (e.g., CoachAI), or it is to provide various visualization on their level of physical activity, sleeping pattern and dietary compliance, either gathered from the smartphone or provided by the user on daily bases (e.g., Lark). **Figure-6.5** illustrates the interaction feedback by the chatbots.



Fig. 6.5 Interaction feedback provided by each of the chatbots.

6.8.3 Conversation-flow Dead-ends

This happens when the conversational agent falls into a state where it cannot proceed or it provides irrelevant responses to user utterance. Based on the evaluation and analysis of each dialog flow, we found that Woebot sometimes provided irrelevant responses to the user. We believe this is because unlike the other two applications, Woebot uses generative model conversation (deep learning) in their dialog structure, whereas CoachAI uses Finite State Machine and Lark relies on task-orientation to guide the conversation. In **Figure-6.6** we show the three chatbots with their dialog structure.



Fig. 6.6 Conversation-flow dead-ends compared to normal flow.

6.8.4 Task Conclusion

All the three chatbot systems provide task conclusion when interacting with the users. This step refers to task accomplishment and end of conversation. Each of the conversational agents found to use a different mechanism to conclude the dialog. For example, CoachAI focuses on plan assignment and activity feedback tracking or coach-user message communication. Once these tasks are fulfilled the bot takes the user to the conclusion state and waits for new tasks to start over. However, Woebot focuses on the user-bot interaction and slot filling, once the required user interaction data are collected and the chatbot fulfills its tasks, the conversation is terminated. In **Figure-6.7** we provide an overview of the dialog conclusion for each of the chatbots.



Fig. 6.7 Task conclusion provided by each of the chatbots.

6.9 UX Design Implications

Our purpose was not to merely compare different chatbot applications in the health context, but rather to extract meaningful insights from their design that will give an indication on challenges to consider when defining or developing chatbots in the health context, as well as similar context of use. The analyzed chatbots were different in their capabilities and specific domain focus. While the study focused on the three chatbot applications, the design implications provided are generic and applicable for similar platforms and applications.

A combined text-based interface with buttons and media elements. The chatbots used a mixture of text, and other graphical elements in their interaction. The analysis found this natural and engaging by the user. The user expressed dissatisfaction when the bot relied on only text or passed on the control to an external interface, such as opening a browser window. Providing a way to allow such a mixed text-graphic interaction and providing inline link view will help enhance user experience with the chatbot [131].

Provide flexible user input. The chatbot should reduce the interaction effort by users. It should allow editing of user last input or cancelling their request. This is important when the edit changes a single parameter in a search query (such as changing price in a shopping/travel chatbot). This feature was present in CoachAI, which allows users to backtrack in the finite states.

Provide persistent view on chatbot capabilities and context. To avoid mismatching chatbot efforts and user expectations, the chatbot should present a persistent view of their capabilities to the user [131]. For example, Telegram restricts bot developers to provide a description of the chatbots capabilities at the start of the interaction. This will help the chatbot and the user to have the same state-of-mind.

Give the chatbot a personality. Bot personality is relevant to build strong UX bound with the user. Since bots feature a text-only interface, vocabulary, syntax, formatting, and line spacing all take on a new relevance in crafting its personality. This personality crafting is important in twofold. First, personality sets the tone for the user experience. Second, chatbot personality is the distinguishing feature amongst different chatbots.

Choose the right dialog approach. Each chatbot serves a specific purpose, hence their dialog should be crafted accordingly. Chatbots can be either system initiated or user initiated. System initiated dialogs are completely controlled by the chatbot, called single initiate. The drawback of this approach is that real human-human dialog involves give and take. Hence, single initiate dialog will be limited in the natural dialog flow. On the other hand, neural net based models provide both direction initiation, however they require a lot of data, and are often less accurate in formatting their response. Hence, finding a common ground where goal-based and chatbot-based dialogs can integrate to provide the best UX in their interaction.

6.10 Persisting Challenges

This is the first study towards understanding the interaction style and usage experience among chatbots. There are several limitations worth mentioning. First, the participant took part in the analysis was highly educated, and familiar with chatbots and their use. This population is at one end of the spectrum, albeit those more likely to be early-adopters, and enthusiastic to play with chatbots for extended durations. Hence, there might be idiosyncrasies to this first-time chatbot users group that might not extend to other groups. Second, the sample size was limited to one person. A

larger number of participants is required to identify broader trends. Finally, the study was limited to our curated list of only three chatbots in the health context.

6.11 Conclusions

Using conversational agents in various domains is increasing in popularity. One complexity with conversation oriented systems is building the dialog. Empirical evidence is needed to inform the design of such dialogs and to measure the effectiveness of conversation technology. In this chapter, we highlighted the results of two distinct experiments related to CoachAI dialog domain and flow.

The first study investigated the effect of chatbot applications that use different dialog styles when collecting information on user wellness. The result of the study showed that a simple task with question/answer might require using emojis to engage users when the conversation covers personal aspects, such as mental wellness. When the dialog focuses on physical activity and factual information, the use of plain text would be preferable. These results provide an initial guidance in the design and development of conversational interfaces that use emojis in their dialog model.

The second study compared the dialog state transition and task processing of CoachAI with two similar conversational agents focused on mental wellness and lifestyle promotion. All the three applications are focused on health coaching. However, there were interesting distinctions to highlight in their design, implementation and approach. We believe these findings can support the research on integrating conversational agents in the context of health and wellbeing and it can be harnessed to contribute to the development of effective mobile health coaching systems for health intervention delivery. We have considered the findings from these studies into our CoachAI system refinement and applied them into the two experiments discussed in Chapter 7.

Chapter 7

Experimentation and Assessment

7.1 Introduction

Assisting healthcare providers with their intervention may help in decreasing their burnout and increasing the quality of care. It is of utmost interest to identify ways to reallocate the limited resources of health professionals to only users that need personal, face-to-face care and where the health application is likely to fail.

In this chapter, we present two experimental studies of CoachAI with users over a different period of time. We first discuss the process of three weeks experiment with 15 adult participants. The proposed approach leverages the chatbot agent to assist the caregivers (e.g., data collection, reminders, user follow-up) by engaging the users in pre-built scripts and conversations. Although CoachAI doesn't offer natural conversation, as in "Siri" or "OK Google" agents that rely on frame based model, the dialog engine provides a structured conversational style that is domain specific and can be tailored to specific demographics. We then discuss a one month intervention study with 19 participants. We measure different functionalities and features of CoachAI and conclude both studies by presenting the results and the limitations.

CoachAI - Platform Testing Study

7.2 Study Design

We tested the tailored activity recommendation and tracking by a conversational agent to automate some caregiver's tasks. While setting up the experiment, a researcher referenced the general guidelines by the FITT (Frequency, Intensity, Time, Type) guidelines from American College of Sports Medicine ACSM's Guidelines for Exercise Testing [132] and recommendations by WHO to build a set of demographic specific activities. The activity listing and assignment was validated with a dietitian. We used emojis as quasi-nonverbal cues throughout the conversation (while performing the questionnaire and activity recommendation and feedback) [133]. We enriched the conversation with empathic feedback whenever user feedback was collected. All the dialogue models were handled by a finite state machine architecture.

7.2.1 Participants & Experimental Design

The study involved 15 participants randomly selected (Male = 13, Female = 2, Std. Deviation = 7.5, range = 27.00) (see **Table-7.1**) and asked to provide their phone number and age for the registration. All subjects were healthy, with no health condition and were motivated to participate in the experiment. There was no reimbursement provided to participants to compensate their participation. All participants interacted with the bot and performed the initial user profiling. This step consists of answering to specific bot questions that are related to participant's diet, physical activity, mental wellness, and coping skills (see **Table-7.2** for excerpts from the different questionnaire topics and an example per topic). Based on par-

Descriptive Statistics		
	Participants	Age
<i>Valid</i>	15	15
<i>Missing</i>	0	0
<i>Mean</i>		29.47
<i>Std. Deviation</i>		7.5
<i>Range</i>		27.00

Table 7.1 Participants descriptive statistics.

participant's demographic information and following the FITT and WHO guidelines and recommendations for physical activities, we defined the activities to assign to each participant. A researcher took the role of the coach and created then assigned the activities to the participants. Participants then interacted with the conversational agent for the duration of the assigned activities. Participants received an activity reminder at the beginning of the day and a notification to provide their feedback in the evening. At the end of each day, participants feedback per activity was collected by the chatbot.

Topics	Example
<i>Coping</i>	How would you rate the amount of stress in your life?
<i>Mental health</i>	How often did you feel nervous in the last 30 days?
<i>Physical activity</i>	Does your job require sitting or moving?
<i>Diet</i>	How would you define your eating habits?

Table 7.2 Chatbot questionnaire topics & examples.

7.2.2 Procedure & Measures

After participant's interaction with the chatbot and submission of the feedback, they were asked to fill in the USE questionnaire [134] to evaluate the usefulness, satisfaction, and ease of use and learning of their interaction with the chatbot. We have also used the Working Alliance Inventory (WAI) questionnaire [135] to evaluate participant's enjoyment and attachment bond with the system. This is necessary to evaluate their overall experience with the application. The measures also consisted of a questionnaire for assessing individual self-confidence and trust, and for estimating enjoyment, attitude and confidence in the chatbot.

Usefulness, satisfaction, and ease of use and learning: The USE questionnaire was adopted to evaluate the usefulness, satisfaction, and ease of use and learning of the chatbot as perceived by the participants. The interaction was evaluated with a 5-point Likert rating scale (from "strongly agree" to "strongly disagree"). The items used in the scales were adapted from [134]. The Usefulness scale comprises items related to the usefulness of the chatbot in the interaction. The Satisfaction scale measures user's satisfaction with the dialogue and the interaction with the agent. Lastly, the ease of use and learning scale evaluates

user experience and learning while interacting with the chatbot. The items are reported in **Table-7.3**.

USE Questionnaire	
Scale	Item
<i>Usefulness ($p=0.2$)</i>	It helps me be more effective.
	It helps me be more productive.
	It saves me time when I use it.
<i>Satisfaction ($p=0.4$)</i>	I am satisfied with it.
	I would recommend it to a friend.
	I feel I need to have it
<i>Ease of use ($p<0.5$)</i>	It is user friendly.
	Using it is effortless.
	It is simple to use.
<i>Ease of learning ($p<0.5$)</i>	I learned to use it quickly.
	I easily remember how to use it.
	It is easy to learn to use it.

Table 7.3 Scales and items used in the USE questionnaire.

Individual self-confidence and trust propensity: Self-disclosure has been defined as any message about the self that an individual communicates to another [124]. Self-disclosure was measured using three dimensions of the Wheelless's scale [125]. Specifically, the three dimensions of self-disclosure were measured using a 5-point scale for each, namely: (i) amount of disclosure, (ii) positive/negative nature of disclosure, and (iii) honesty of disclosure. Moreover, four items from the Mayer and Davis's Trust Propensity Scale [126] were used to measure dispositional trust. For both self-confidence and trust propensity, scale values were calculated as the average of the items.

Interaction experience (enjoyment, attitude and confidence): The interaction was evaluated using three scales, i.e. enjoyment, attitude and confidence, each composed of four items with a 6-point Likert rating scale (from "strongly agree" to "strongly disagree"). The items used in the scales were adapted from [136]. The enjoyment scale comprises items related to the enjoyment and entertainment in the interaction. The attitude scale measures user's attitude towards the dialogue and the interaction with the agent. Lastly, the confidence scale evaluates user perception of the privacy and intrusiveness of the questions in the dialogue. The items are reported in **Table-7.4**.

Scale	Item
<i>Enjoyment</i> $\alpha = .61$	I enjoyed chatting with the conversational agent during the interaction
	I felt comfortable answering the questions
	The more I interacted with the agent the more I liked the experience
	I would like to chat again with this conversational agent in the future
<i>Attitude</i> $\alpha = .59$	I found a kind of "emotional connection" between myself and the conversational agent
	I found the dialog with the conversational agent to be realistic
	I noticed a negative emotional change in myself during the interaction *
	I found the dialog to be coherent. In other words, the sequence of responses of the conversational partner made sense
<i>Confidence</i> $\alpha = .70$	The agent asked very personal questions *
	I trust the conversational agent
	I found the questions to be very intrusive *
	I answered the questions with honesty

Table 7.4 Scales and items used in the questionnaire. Items followed by an asterisk were inverted for the analysis. Cronbach's alphas are reported for each scale.

Working Alliance Inventory (WAI): Alliance refers to the achievement of a collaborative relation between caregiver and user; meaning that there is a consensus and willingness on the part of both parties to engage in and do the work that leads to improvement [135]. There is a broad agreement that this collaborative engagement involves consensus over the goals of treatment, a sense of confidence and commitment to the kinds of activities that the helper (coach) and helpee (user) engage in as part of the helping journey, and the relationship or engagement is in a context of mutual trust and confidence. The WAI questionnaire was evaluated with a 7-point Likert rating scale (from "strongly agree" to "strongly disagree"). From the participant's responses, the WAI scores were statistically different from the middle value 4 throughout all the questionnaire. The items are reported in **Table-7.5**.

7.3 Results and Implications

7.3.1 Experiment Results

The overall average values and standard deviations were measured for both the USE dimensions and the WAI. We have tested for differences to the scale middle values

(3 for USE and 4 for WAI). The findings for the USE scale showed that scores for the Ease of use and Ease of learning are different from the middle scale (they're higher). Whereas, the WAI score is statistically different from the middle value 4 (see **Table-7.5** for the mean, SD, t-test and p values for the USE and WAI questionnaires).

USE - WAI	Mean	SD	$\oplus$
<i>USEFULNESS</i>	3.43	1.13	t(11)=1.33, p=0.2
<i>EASE OF USE</i>	3.70	0.88	t(11)=2.93, p<0.5
<i>EASE OF LEARNING</i>	4.48	0.68	t(11)=7.97, p<0.1
<i>SATISFACTION</i>	3.29	1.28	t(11)=0.84, p=0.4
<i>WAI</i>	4.78	0.99	t(11)=3.01, p<0.5

Table 7.5 CoachAI - USE and WAI Questionnaire (Scoring). $\oplus$: One sample t-test (middle scale value USE =3, WAI =4)

7.3.2 Interaction Time Analysis

We analyzed the log data for our user-chatbot interaction. Logging messages makes it possible to know how many are sent everyday, as well as the average number of messages per conversation per day, and the number of messages per conversation. For our coaching bot, we analyzed three log data, namely the average users interaction with the chatbot about daily activities and their feedbacks, and the overall users interaction time with the chatbot. In general, we analyzed the average time took the users to chat with the bot about a specific functionality. In **Figure-7.1** we present the log data analysis of all the users for over the three weeks experiment. Where

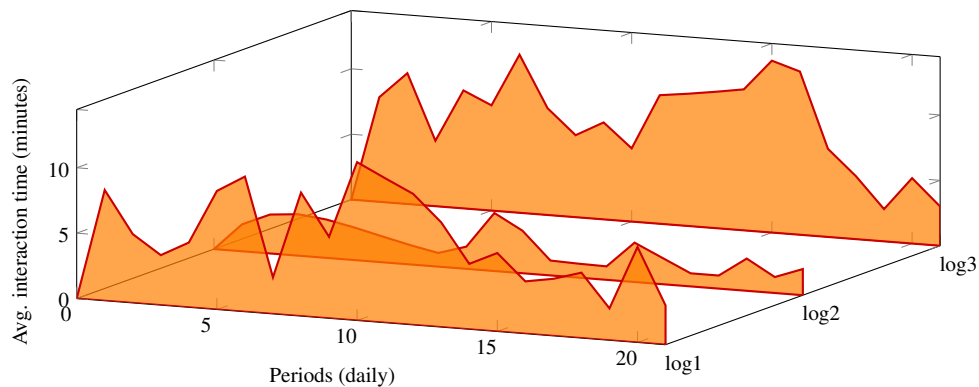


Fig. 7.1 CoachAI - Users log data analysis.

log1=average users interaction with the chatbot about daily activities, log2=average

users interaction with the chatbot about their feedback, and log3=overall users interaction time with the chatbot. The findings showed that users spent less time to interact with the bot when providing their feedback on the activities, rather than accessing health activities.

7.3.3 Implications

Several implications and insights emerged from leveraging this experiment:

- Of all the functions the chatbot provided, reminders were the functions appreciated the most by the users. This is a useful guide for corresponding public intervention designers and providers.
- Using mixed human-virtual coaching approach was perceived positively by all participants. Across all the experiment, there was strong support for hypothesis H1 and H2 as listed in **Chapter 1**.

7.4 Discussion

Using a chatbot tool to suggest and track various physical activity tasks, this study examined (1) the use of a human-virtual agent mediated coaching approach for physical activity promotion, (2) the classification of users into distinct groups based on their adherence, and (3) the handling of some caregiver tasks by the chatbot, while tailoring user activities. The result of this study showed how using conversational agent can handle providing and tracking health information.

CoachAI provides a full control of personal health data and a generic framework to manipulate the design and communication style of the conversational agent and adopt it to different domain settings. The platform allows caregivers to tune the health activity per user. To our knowledge, few studies have considered assisting caregivers with the burden caused by user's continuous need of care. This includes handling repetitive tasks that consume a vast amount of caregiver's time.

7.5 Persisting Challenges

The findings of this study must be interpreted in view of its limitations. First, although the activities were clinically validated and clearly guided, they were assigned by a researcher and not a healthcare provider. Moreover, the participants were healthy subjects, this was because of the study focus, which was on validating participants experience with the conversational agent. Second, a long-term follow up is required with a health professional to clinically validate the platform efficacy in assisting the caregiver and supporting their users through tailored health activities.

CoachAI - Prospective Intervention Study

Studies [56, 137] compared the effect of a physical activity application with virtual coaching vs applications without coaching, and concluded that trainees with a virtual coach adhered longer to the training program. However, coaching in this form does not provide the social support that a human coach can provide. In addition, virtual coaching approaches are still unable to handle user preferences efficiently. Our approach combines both virtual and human agent and leverages the chatbot agent to levies the caregiver of menial works by engaging the individuals in pre-built script conversations. Rather than relying on SMS services, we used a natural conversation to interact with users, delivered by the conversational agent and handled the dialogue with a state machine architecture to communicate with users about their daily health tasks. In this section we highlight the outcome of our one-month intervention study with users. We present the intervention settings and findings from the study. The main research question we tend to answer is: *how can technology be effectively integrated into the doctor-patient interaction to promote tailored healthy lifestyle?*. To achieve this, we try to answer the following points:

- P1** What is the overall user experience of using CoachAI system?
- P2** Will user experience change over time when using CoachAI over 4 weeks?
- P3** Will CoachAI change users' intention or behavior?
- P4** How much do users prefer a direct human agent support or to what extent does the chatbot agent satisfy their needs?

7.6 Methods

In this section, we define the study aim, the experiment settings, participants recruitment and the intervention design.

7.6.1 Measurements

We aim to enhance user interaction with the system during all the stages of health intervention. The study investigated the design and interaction aspects, listed below, with the participants interaction. We used three models throughout the intervention to validate the dimensions listed below. The Health Action Process Approach (HAPA) [89] was used to validate the health behavior engagement. HAPA is a two-layer model: The continuum version is used for analyses and predictions, the stage version is used for interventions, hence we focused on the stage model during the experiment. The technology acceptance model (TAM) [85] was used to model how participants come to accept and use the system. Finally, the AttrakDiff model [138] was used to evaluate how participants rate the usability and design of CoachAI (see **Table-7.6**). We will further elaborate these points throughout the chapter and highlight the emerging patterns from individual's participation in the intervention.

Dimensions	Definition
<i>Perceived usefulness (TAM)</i>	This evaluates how useful the users perceived the application.
<i>Perceived ease-of-use (TAM)</i>	This evaluates how easy it was for the users to use the system.
<i>Perceived fun (TAM)</i>	This measures users' engagement and entertaining aspects while interacting with the system.
<i>Attitude (TAM)</i>	This measures users' attitude towards the system.
<i>Intentions (TAM)</i>	This measures the future aspects of users to reuse the system.
<i>Pragmatic Qualities (AttrakDiff)</i>	Pragmatic factors are, for example, usefulness and usability.
<i>Hedonic Qualities (AttrakDiff)</i>	Hedonic factors include emotional needs, such as curiosity and identification. The resulting attractiveness is based on the combination of pragmatic and hedonic factors.
<i>Appeal (AttrakDiff)</i>	This is to measure how appealing the system is to users.
<i>Social Qualities (AttrakDiff)</i>	This was added by us to check for the social qualities the system provides to the users.
<i>The Motivation Phase (HAPA)</i>	This phase is to measure user's intention to either adopt a precaution measure or change risk behaviors in favor of other behaviors.
<i>The Volition Phase (HAPA)</i>	This phase describes how hard users try and how long they persist to a behavior change.

Table 7.6 CoachAI - Study dimensions and their definition

7.6.2 Study Design

The study initially gathered basic demographic information from participants through a text conversation with the chatbot. The information contained participants personal data and personality test about their daily exercise, dietary pattern and stress. The intervention study lasted for one-month and participants interacted with the chatbot daily.

Participants behavior was measured over the four weeks validation study. Participants were divided into two conditions, baseline (week 1) and intervention (weeks 2, 3, and 4). In the first week, participants were asked to chat with the bot and provide their baseline physical activity and healthy diet intention, together with their stress evaluation. The initial questionnaire and interaction with the chatbot provided an estimated data about each individuals exercise, diet and stress level. These data were considered when providing the individual plan per participant and later measured with the intervention condition (the rest of the weeks).

The experimental design used between subjects (different subjects in each group) and within subjects (same subjects exposed to different interventions/control) measure to analyze the collected data with a multivariate testing (more than one dependent variable and one or more independent variables). The study focused on qualitative analysis to measure participants behavior change between the baseline and intervention condition. The overall user experience with CoachAI was measured with the validation questionnaires (TAM or AttrakDiff) which was targeting H1, whereas their experience change with CoachAI was measured over their feedback provided to the chatbot (3 weeks) which also targeted H1, and their adherence to the weekly plans was used to measure the H2 (HAPA).

All the data collected from participants were anonymized for the whole study duration. Participants adherence was tracked to decide what plan or activities to suggest and to communicate with the participants when necessary. The emphasis of this study was on participants interaction with CoachAI and trying to follow the instructions asked by the chatbot. Moreover, participants received at least a questionnaire to perform about their overall experience. There were also the private messages communicated to the participants at least once a week to report their adherence and provide them with motivational messages. The messages were fixed and tailored to each participant based on their adherence to the physical activity, healthy diet, and mental wellness activities. Assigned plans on healthy diet, physical activity

and stress management were all validated and structured. All the activities assigned were based on the ChooseMyPlate [139] general recommendations and guidelines. A researcher structured the list of activities to be assigned to the participants. The activities were validated with healthcare provider based on individuals baseline information.

7.6.3 Eligibility Criteria

The eligibility criteria are a set of predefined characteristics to identify subjects who will be included in the study. To respond to the study objective and optimize the external and internal validity of the study, we tried to ensure homogeneity of the sample population and increase the likelihood of finding a true association between intervention and outcomes. Therefore, the inclusion criteria selected individuals with sorting in common with each other. For example, the participants were all individuals with sedentary lifestyle and had a poor dietary lifestyle. However, the participants were motivated to pursue a healthy lifestyle. Most participants were university students or have just obtained their degree.

7.6.4 Participants

The study involved 22 participants recruited from online social networks and selected based on the above criteria. At the beginning 3 participants decided to drop out for personal reasons. The total number of participants was 19 (Male = 11, Female = 8, aged 19-53, Std. Deviation = 9.371, range = 34.00) (see **Table-7.7**), and were asked to provide their phone number and age for the registration. All subjects were healthy, with no health condition and were motivated to participate in the experiment and perform health promotion activities. Participants were either native Italian speakers or spoke Italian fluently. This was important due to the language-based nature of the study. A reimbursement of 5€ was given to all participants in the form of Amazon coupon. All participants interacted with the bot and performed the initial user profiling. The participants were asked to indicate their familiarity with messaging applications, social media, internet and other uses of their smartphone. The majority of participants showed their familiarity with messaging apps and mentioned their daily use of different types of messaging applications (see **Figure-7.2**). In the next phase, participants were asked to wait for the activities to be assigned by the coach

Descriptive Statistics			
	Participants	Age	Gender
<i>Valid</i>	19	19	male=11, female=8
<i>Missing</i>	0	0	0
<i>Mean</i>		28.53	
<i>Std. Error of Mean</i>		2.150	
<i>Median</i>		25.00	
<i>Std. Deviation</i>		9.371	
<i>Variance</i>		87.82	
<i>Range</i>		34.00	
<i>Minimum</i>		19.00	
<i>Maximum</i>		53.00	

Table 7.7 Participants descriptive statistics.

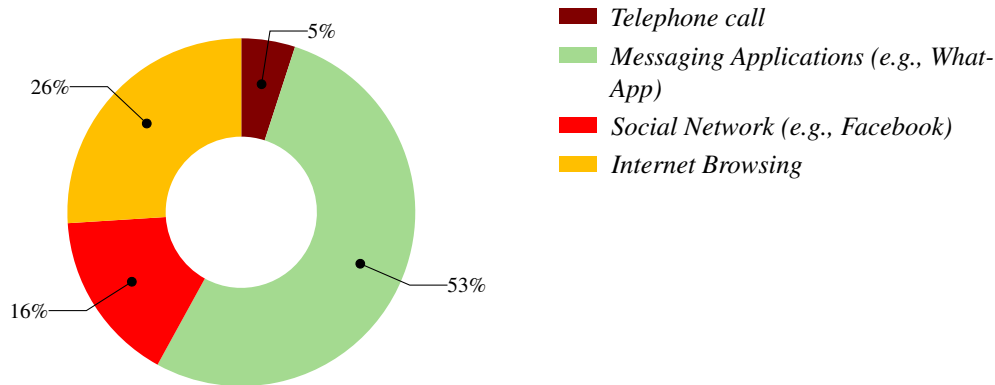


Fig. 7.2 CoachAI - Participants familiarity with technology.

based on their initial interaction with the chatbot. Participants physical activity was aggregated from their interaction with the chatbot and some were used as features to evaluate their daily exercise level. The machine learning model was used to classify users into three distinct groups (namely, vigorous, mild, and sedentary users), this was compared with the users reply to the chatbot when asked about their sedentary lifestyle (see **Figure-7.3** for the result of the SVM classifier and users respond to the chatbot).

7.6.5 Procedures

The procedure starts with the human agent registering the users, then the chatbot asks them to perform the initial questionnaire to evaluate aspects of their diet,

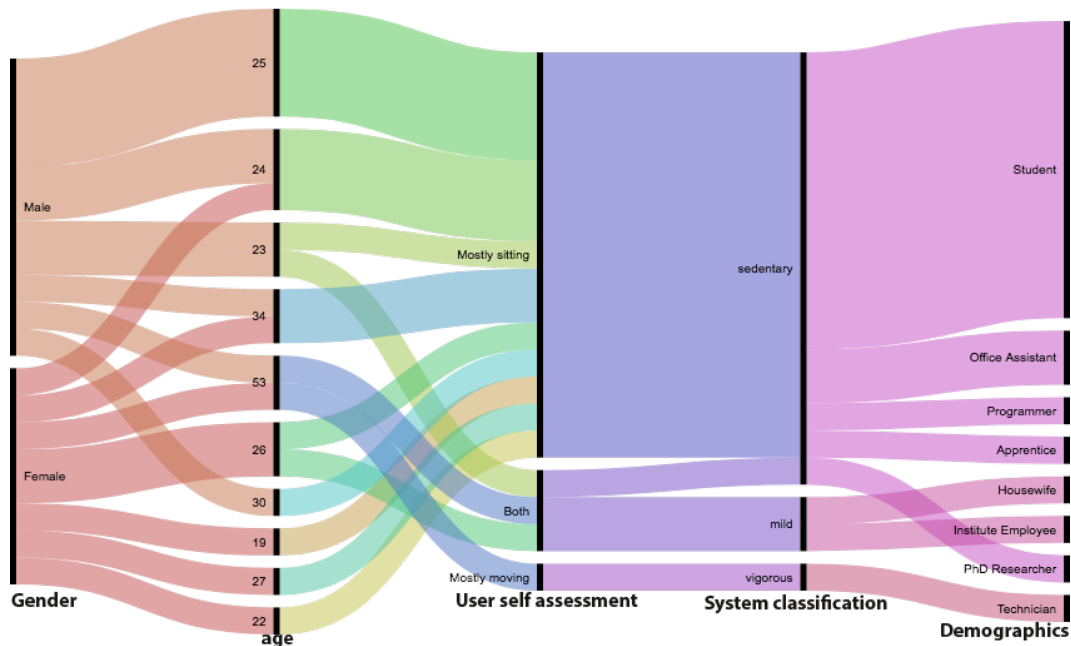


Fig. 7.3 CoachAI - Participants sedentary lifestyle classification by the SVM model vs the chatbot.

physical activity and other health parameters (baseline). On concluding the initial questionnaire by users, the human agent receives a notification indicating user profile completion. The human coach can then either define a new health plan adherent with user condition or assign a pre-built plan from the activity pool. All the plans have a category, a description, periodic event triggering time and expiration time. Upon plan submission, the chatbot notifies the users about the plan availability, and collects their feedback. Overtime, the chatbot also tracks user's adherence to the plan by measuring the overall average adherence to the activities defined. For example, in a weekly plan the chatbot measures user's average adherence to all the activities. The adherence is either collected manually by the chatbot or aggregated from user wearable, in case paired with CoachAI.

By the end of each week the coach performs a follow-up with the users through the private messaging channel. The follow-up consists of a message sent to the users which contains motivational messages and a feedback on their overall compliance to the plans. In addition, 3 weekly questionnaires (namely, the TAM, the HAPA and the AttrakDiff) are sent to the users. These questionnaires measure aspects of user's experience, expectations, and other interaction related parameters with CoachAI. This procedure lasted for 4 weeks and was terminated at the end of the fourth week.

Figure-7.4 below shows the experiment setup we followed throughout CoachAI experiment design process.

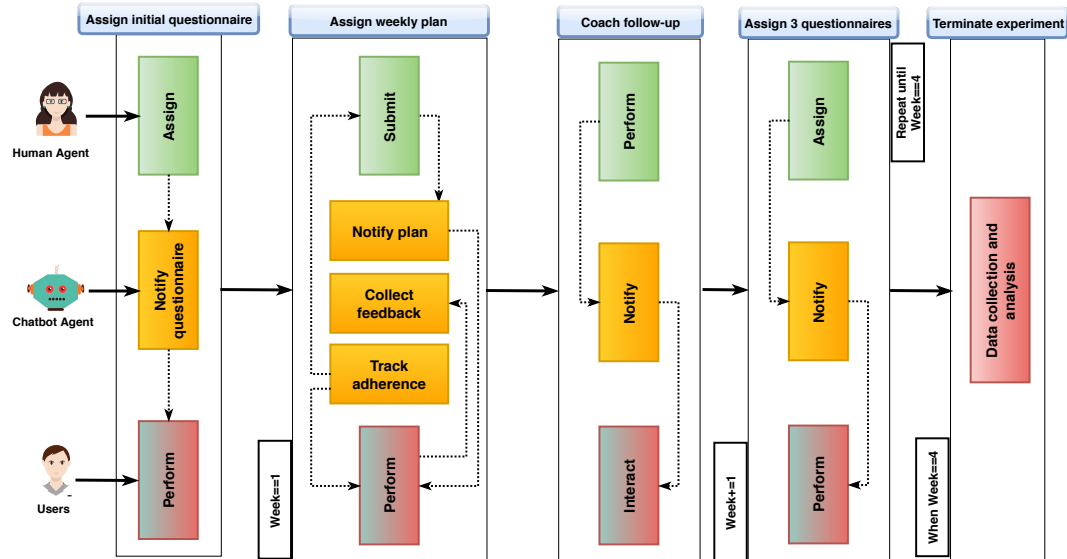


Fig. 7.4 CoachAI - Experiment design process.

7.7 Results

7.7.1 Perceived Usefulness

To understand the added value of using the system, we asked the users to report on their experience regarding the perceived usefulness of some CoachAI features. The features that were experienced as useful by most participants included having a human coach to assign and monitor the activities, the reminders to notify the activities and ask for the feedback and the weekly private follow up messages by the human coach. The private Coach-User messaging channel is considered a social interaction feature allowing the participants to communicate with the coach. These features were highly regarded by all participants.

7.7.2 Perceived Ease-of-use

This dimension was above the middle score ($t(17)=4.9$; $p<.01$). Although the ease of use and interaction with the chatbot, some features were deemed less friendly and somehow difficult to use. Among the least user-friendly features was downloading the activity pdf in the chatbot, which was perceived tedious by most users, instead they suggested providing the activities directly via the dialog channel.

7.7.3 Perceived Fun

Although participants rated higher the aspects of personalization and tailoring the activities, which avoids introducing repetitiveness in the system, most users rated lower the fun aspect of using the system. This might be because of the nature of the application, which is health promotion. This puts the application among the most task-oriented and least entertainment-oriented applications. Future work will investigate integrating aspects of persuasiveness, such as social interaction and gamification into the application.

7.7.4 Attitude and Intention

Participants highly rated both attitude and intention towards using the system. Most participants gave a rating above average to both dimensions. Considering the questionnaire on the dimensions of the Technology Acceptance Model (TAM) (see **Figure-7.5**), one-sample t-test was used to compare the means for each scale to the scale middle value (score =4). The scales "Ease of Use", "Attitude" and "Intention" are significantly higher than the middle score (respectively: $t(17)=4.9$; $p<.01$, $t(17)=2.5$; $p<.05$, $t(17)=3.1$; $p<.01$). A repeated measure ANOVA (with time as a within factor) showed a difference only for the scale "Intention" between the three different weeks. Post-hoc analyses reveal that the scores for this scale at week 1 were higher than week 2 and week 3 (see **Figure-7.6**).

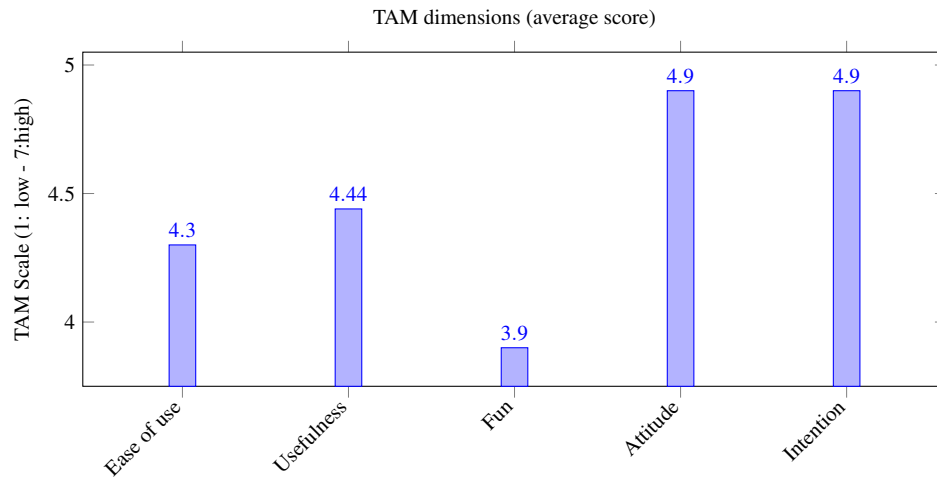


Fig. 7.5 TAM dimensions (average score).

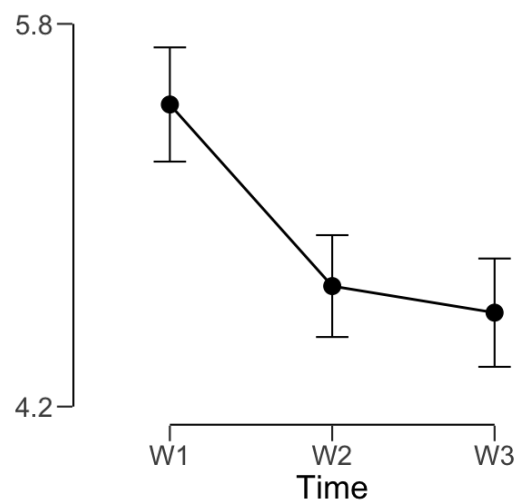


Fig. 7.6 TAM Post-hoc analyses between weeks.

7.7.5 Pragmatic Qualities, Hedonic Qualities, Appeal and Social Qualities

This was performed over four weeks. We analyzed these qualities based on users' responses. The results of the AttrakDiff questionnaire are summarize in **Figure-7.7** below: A one-sample t-test was used to compare the mean scores for each dimension to the scale middle value corresponding to "Neutral" (score = 4). The test shows that the average scores are statistically higher than 4 for each dimension: Pragmatic

($t(17)=5.41$, $p<.01$), Hedonic ($t(17)= 3.4$, $p <.01$), Appealing ($t(17)= 4.2$, $p <.01$), Social ($t(17)= 2.6$, $p <.05$).

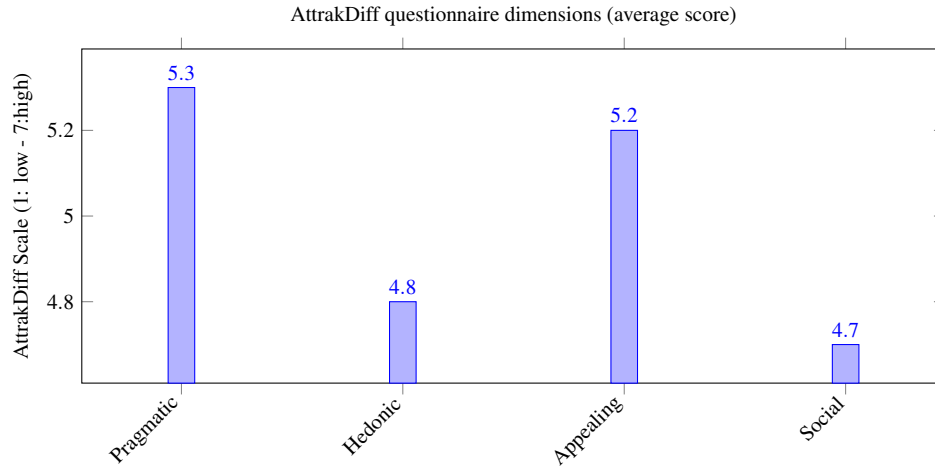


Fig. 7.7 AttrakDiff questionnaire dimensions (average score).

7.7.6 The Motivation and Volition Phase

Participants intention to change risky behavior or adopt a new measure was evaluated through the motivation and volition phases of the HAPA model [140, 89]. The HAPA Stage model was adopted to measure participants intention to change improve self-efficacy, outcome expectancy and risk perception. This model divides users into three distinct groups, namely non-intenders, intenders, and actors (see **Figure-7.8**). Users in this model can either progress their behavioral intention to change or they may also fall into relapses and recycle through the stages. We compared these stages with individual's overall adherence to the activities. A repeated measure ANOVA (with time as a within factor) showed no different between the three weeks, the scores remained unchanged. Participants were asked to provide their answers to the HAPA questionnaire over the month. We analyzed individual's data about physical activity, healthy diet and overall health expectation and mapped these data with the perceived self-efficacy, outcome expectancy and risk perception.

Physical activity intention Users data about their intention to pursue an active lifestyle was collected at the end of each week. The longitudinal datasets were collected through questionnaires at three points. We mapped individual's adherence and their data from the HAPA model to measure their risk perception,

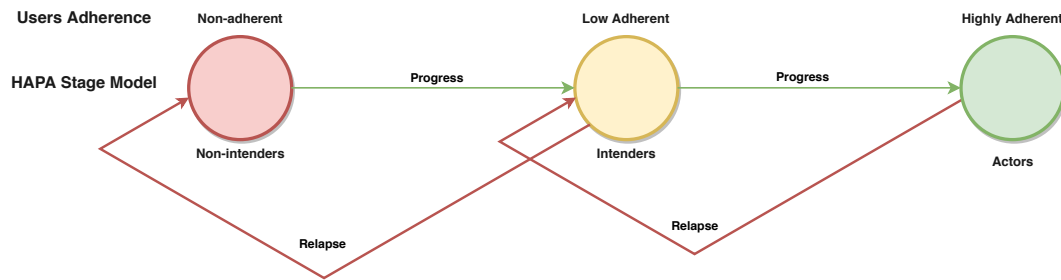


Fig. 7.8 HAPA - Stage model and users adherence.

positive outcome expectancies, action self-efficacy, and behavioral intentions. The analysis revealed no statistically significant results when analyzed within subject and between subject effects. **Figure-7.9a** below presents the comparison between HAPA data and individuals physical activity adherence.

Healthy diet intention The study examined the role of self-efficacy and outcome expectancy in the context of dietary behaviors. The model includes three predictors of the intention to eat a healthy diet (action self-efficacy, outcome expectancies and health risk). A longitudinal field study was designed to examine the interrelationships of these factors with individual's adherence to healthy diet activities. The analysis revealed no statistically significant results when analyzed within subject and between subject effects. **Figure-7.9b** below presents the comparison between HAPA data and individuals adherence to healthy dietary activities.

Mental wellness intention The study examined the role of self-efficacy and outcome expectancy in the context of mental wellness. The model includes three predictors of the intention to perform stress management skills (action self-efficacy, outcome expectancies and health risk). A longitudinal field study was designed to examine the interrelationships of these factors with individual's adherence to a coping skill. The analysis revealed no statistically significant results when analyzed within subject and between subject effects. **Figure-7.9c** below presents the comparison between HAPA data and individuals adherence to healthy dietary activities.

Overall health expectation Health expectancy data were collected with questionnaires at three points. We mapped individual's health expectancy data with their adherence to measure their risk perception, positive outcome expectan-

cies, action self-efficacy, and behavioral intentions. The analysis revealed no statistically significant results when analyzed within subject and between subject effects. **Figure-7.9d** below presents the comparison between health expectancy and individuals overall adherence.

The high and low variables refer to individual's overall adherence to the plans over the three weeks. We measured them by calculating the average of user's weekly activities and dividing them by the total number of activities assigned. We compared these two variables against users' responses to each of the physical activity, healthy diet, mental wellness and health expectancy measures.

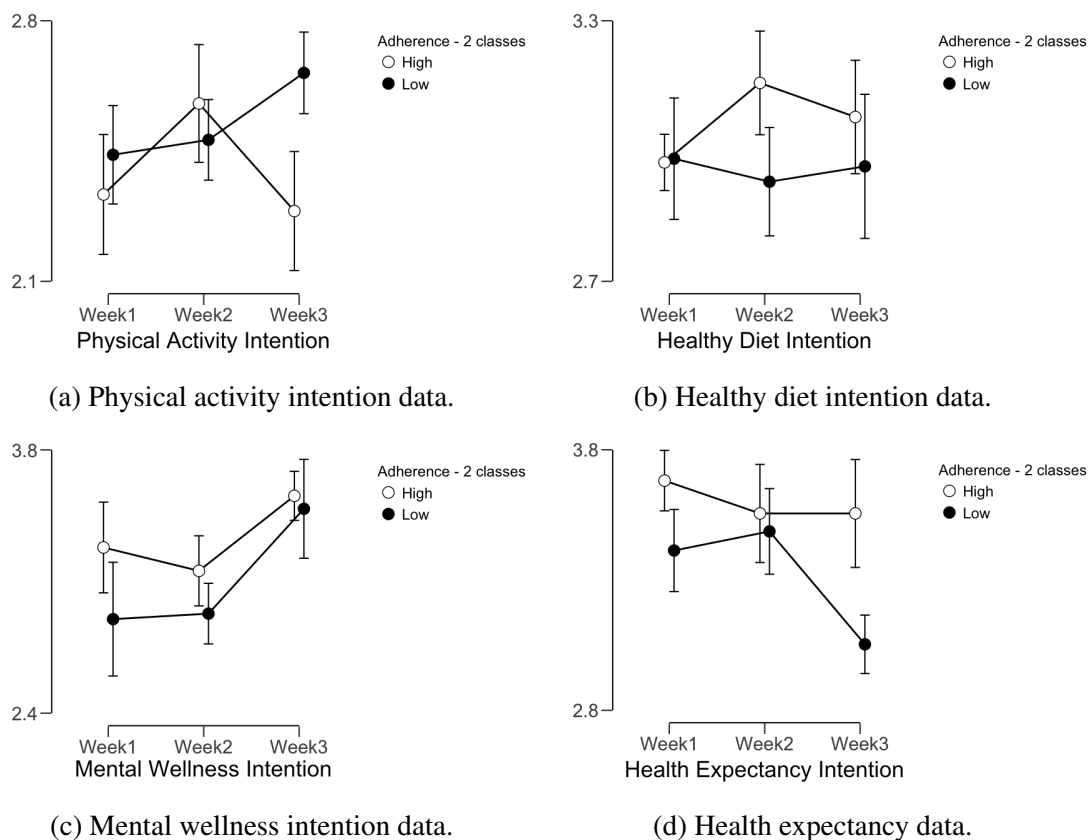


Fig. 7.9 Top left (physical activity intention), Top right (Healthy diet intention), Bottom left (Mental wellness intention), Bottom right (Health expectation intention) vs individual's adherence data.

7.8 User Adherence Analysis

The system calculated participants overall adherence to the plan and reported their total adherence at the end of each plan expiration. The adherence was categorized into high and low adherence groups. The adherence measure is calculated by taking the average of all the activities performed within a plan. Users performing above the threshold were categorized as highly adherent, and vice versa.

Participants were categorized according to their adherence to the plan in two categories in high and low adherence, with respectively 10 and 9 participant each. The TAM dimensions were analyzed with a Multivariate ANOVA (MANOVA) with "Adherence level" as between subject factor. The MANOVA shows a significant effect of the between subject factor for Usefulness ($F(1,16)= 6.5, p<.01$), Fun ($F(1,16)= 4.5, p<.01$) and Attitude ($F(1,16)= 6.9, p<.01$). No differences were found for the Ease of Use and Intention. **Figure-7.10** presents the result of the user adherence vs the TAM dimensions.

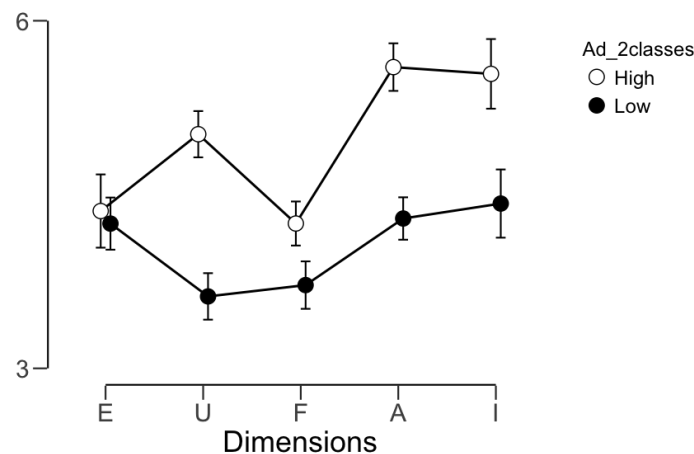


Fig. 7.10 User adherence analysis and TAM dimensions.

7.9 User Familiarity with Chatbot Analysis

We ran a multivariate ANOVA test to measure this dimension against the AttrakDiff and TAM dimensions. Based on self-rated scale, participants were categorized by their familiarity with chatbots and the frequency of use of such technology. Partic-

ipants were grouped in two categories: High familiarity (11 participants) and low familiarity (7 participants). A multivariate ANOVA was used to test for differences between these two categories with AttrakDiff and TAM dimensions. Regarding AttrakDiff, the analysis shows a significant effect of the between factor for the Pragmatic ($F(1,16)=4.9, p<.01$), Hedonic ($F(1,16)=7.8, p<.01$) and Appealing ($F(1,16)=4.8, p<.05$) dimensions. Participants more familiar with chatbot applications scored higher in all the dimensions. Considering the TAM dimensions, the analysis showed an effect for the Usefulness ($F(1,16)=6.2, p<.05$) and Fun ($F(1,16)=8.4, p<.05$) dimensions. Participants with high familiarity with chatbot technology reported higher scores in these categories (see **Figure-7.11a** and **Figure-7.11b** for chatbot familiarity dimension vs the TAM and AttrakDiff dimensions, respectively).

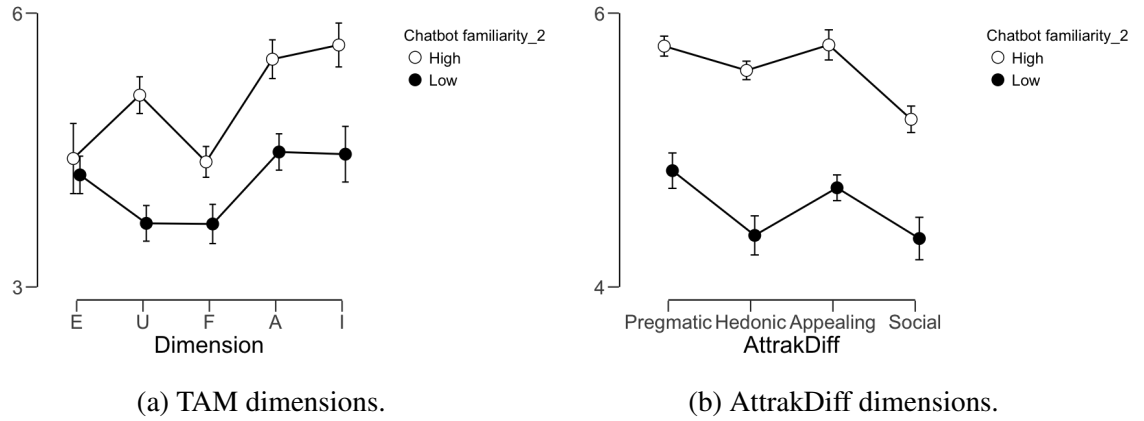


Fig. 7.11 Chatbot familiarity dimension vs TAM and AttrakDiff dimensions.

7.10 Log Data Analysis

We gathered the log data for all the participants interaction with the system over the four weeks. The log data was analyzed to measure participants overall interaction time with the chatbot and the human agent and the frequency with which participants accessed activities, provided feedback, and accessed private messages provided by the coach. This step is necessary to identify patterns of use among participants.

7.10.1 Interaction Time Analysis

Participants interaction with the human coach and the chatbot was measured by checking the time they accessed and spent on the system. The analysis showed that participants spent more time interacting with the chatbot and relatively less time interacting with the human coach. We believe this is due to the nature of the experiment, which emphasized the chatbot role and let the human coach intervene only on specific times. For example, during the experiment, the coach provided private messages only during the weekend days, which included a feedback and motivation messages to each participant based on their adherence (see **Figure-7.12** for participants interaction time with the bot and the human coach).

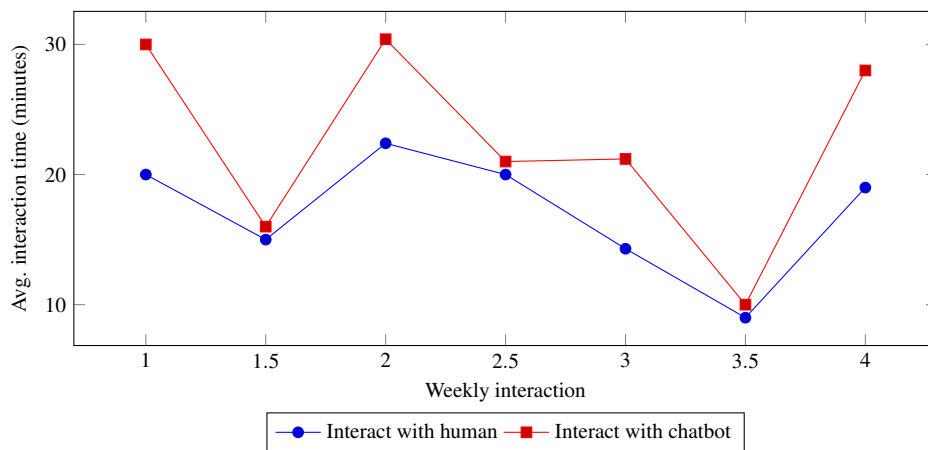


Fig. 7.12 Interaction time with the bot and human.

7.10.2 Chatbot Functionalities Use Analysis

We analyzed the log data with respect to participants average interaction time with the chatbot to access activities, provide feedback and access private messages. The analysis showed that participants mostly used the chatbot to access the activities and provide their feedbacks. The data revealed very low access rate to private messages given by the coach. As mentioned above, this might be due to the way the messages were provided. While the health activities and feedbacks were performed daily, even over the weekend, the private messages were provided once at the end of the week by the human coach (see **Figure-7.13**). We also analyzed participants average interaction frequency with respect to physical activity, diet and mental wellness.

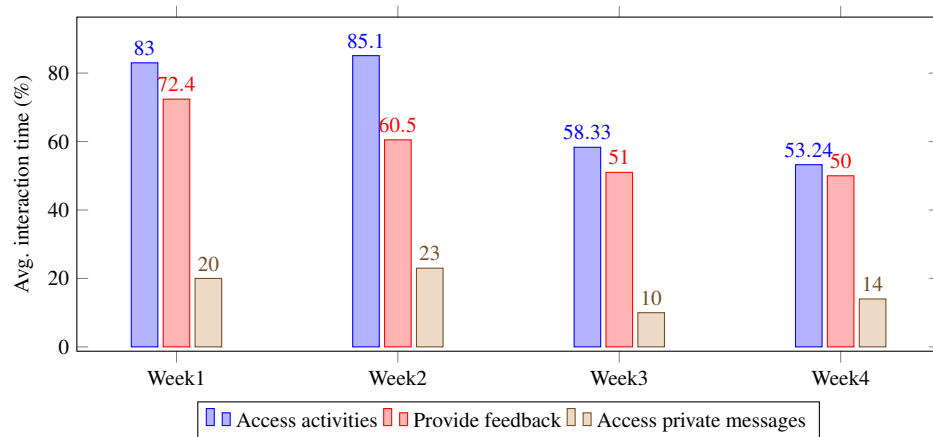


Fig. 7.13 Average interacting time with the chatbot.

Although the same amount of activities assigned to all participants, there were some differences in the average interaction and feedback submission per activity. On average, participants mostly accessed and interacted with the bot when the activities were about exercise and diet, and there were few interactions with activities about mental wellness (see **Figure-7.14**).

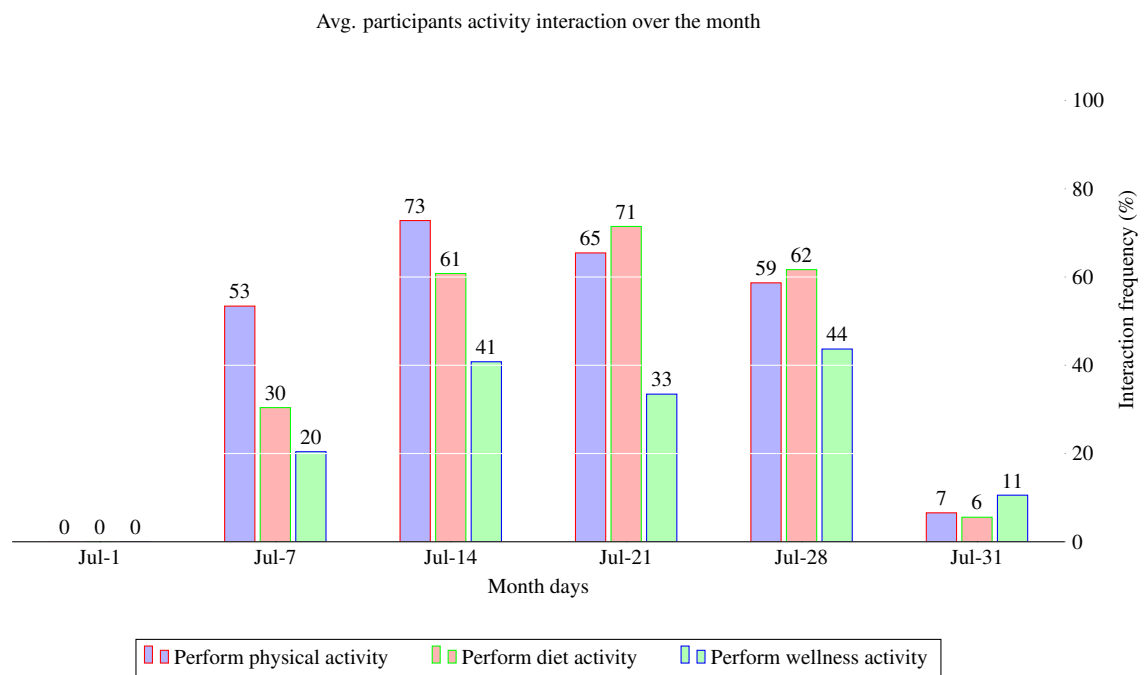


Fig. 7.14 Avg. participants activity interaction over the month.

7.11 Topic Specific Interaction Preferences

We performed a self-evaluation at the end of users feedback to validate their interaction with the chatbot and measured their experience with the system when performed a topic specific activity. For example, if the user performed a physical activity plan, they're asked to indicate their preferences to be coached by a human agent, a virtual agent, or a combination of the two. A fixed set of validation questionnaires were asked to participants through the chatbot channel.

Interestingly, participants following a physical activity plan indicated their preferences to a virtual coach to track their physical activities. Whereas, those in the healthy diet group preferred having a human agent to provide them with daily dietary activities and monitor their progress. Finally, participants in the mental wellness group chose a combination of a human agent and a virtual agent to monitor and assist their anxiety and stress coping (see **Figure-7.15** below for the differences). We believe this result is due to the nature of each topic which may defer in terms of needs and user expectations. In addition, some topics are more emotional bound than others and vary in the level of health complexity. For example, users in the mental wellness group may need support beyond the technological capabilities, hence having a human in the loop will provide them with the emotional support and can change the communication tone which is crucial in such domains [20].

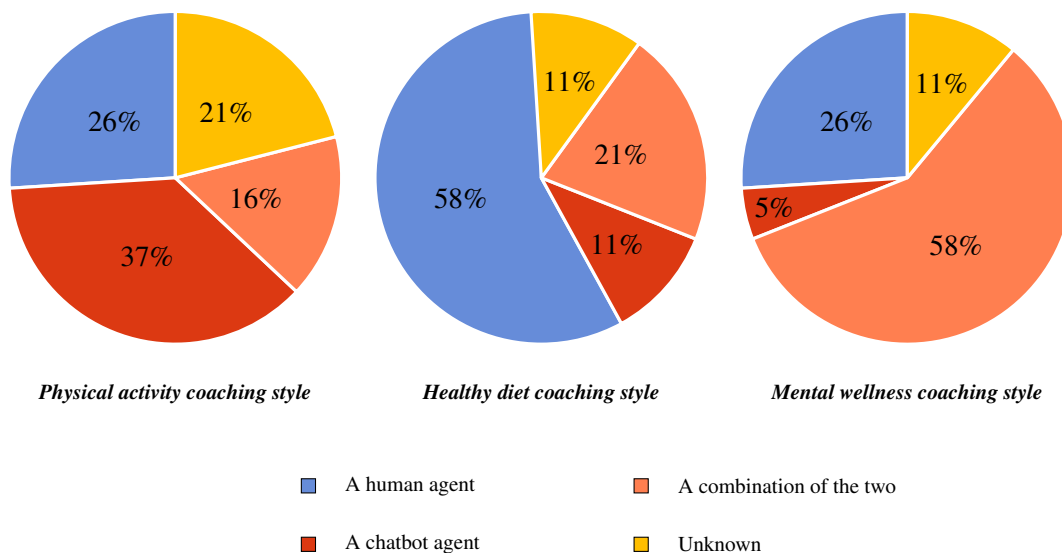


Fig. 7.15 Participants – Health tracking preferences by topic. Left (physical activity group). Middle (healthy diet group). Right (mental wellness group).

7.12 Participants – Aspects of CoachAI

7.12.1 Positive Aspects

At the end of the experiment, participants were asked to provide at least three features they found motivating, useful and fun. Participants highly rated the "ease of use" stating that the chatbot was very easy to use. There were other features perceived as positive by most participants, such as "activity reminder" and "on-time feedback". In addition, participants positively rated the coach motivational and feedback messages sent through the private communication channel. These messages were sent to encourage participants to follow the plan. **Figure-7.16** below provides a representation of the features perceived as positive.

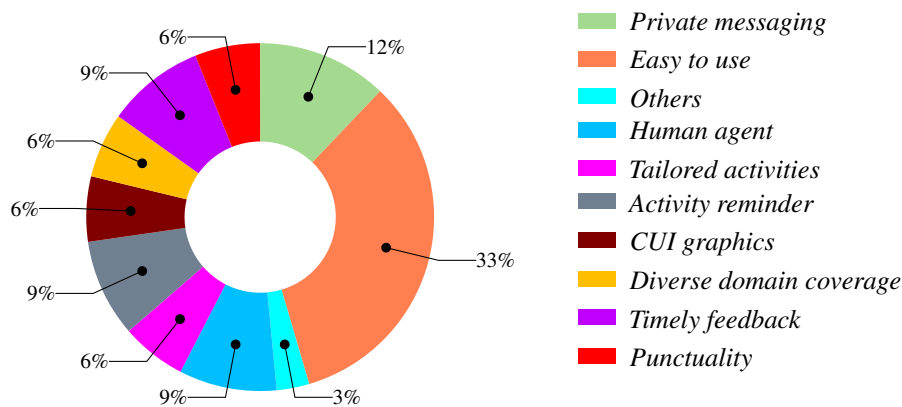


Fig. 7.16 Participants – Coaching style preferences.

7.12.2 Poor Aspects

We asked participants to report their negative experiences with the application. Some design, technical and user experience related issues emerged by the participants feedback. Participants complained that the chatbot dialog was not very clear when they communicated with the bot. Some participants complained that the activities were not arriving with the message, but rather were listed in the pdf file which they were expected to download to access the activities. This was considered laborious and they suggested providing the activities directly in the chatbot dialog. Participants complained that notifications are annoying sometimes, especially when provided in an inappropriate timing. This led us to investigate this issue and future work will

consider studying the fingerprint for each individual through their smartphone to tailor the notification delivery time based on their most optimum time. Finally, a participant mentioned it was difficult to track his progress and suggested adding more features into the system to enable progress tracking.

7.13 Discussion

Text-messaging based mobile application is a powerful tool for behavior change interventions, due to its widely availability, inexpensive, and ease of use [141, 142]. A substantial body of evidence spanning several years of research demonstrates that text-messaging interventions have positive effects on health outcomes and behaviors [141]. CoachAI is a messaging-based conversational agent built to support the development, classification and delivery of both individual and group-based health interventions. The health activities were sent recurrently (e.g. daily, weekly, on weekends or weekdays) or at a scheduled date and time decided by the coach.

The results validated our hypothesis and provided interesting insights when using conversational agents in health coaching systems. Across the experiment there was strong support for some of the research questions. Overall, the experiment results clearly led to a positive answer to the P1. There was some noticeable user experience emerged from using CoachAI. Users experience changed and varied while performing their activities, which confirmed our P2. However, there was little or no support for P3, user's intention towards physical activity, healthy diet, or mental wellness activities remained unchanged over the three weeks experiment. Finally, there was a strong support for P4, individuals performing different health activities revealed their preferences to a direct human agent support, or the chatbot agent, or a combination of the two. The research concluded that building the technology in a user familiar environment is an effective way to increase factors of technology sustainability. The users positively perceived the activity sessions when they are planned by a human coach. Hence, the support of a human agent is clearly more motivating and engaging with respect to the virtual trainer that reveals itself alone insufficient to fulfill user's needs.

7.14 Persisting Challenges

Several limitations worth mentioning emerged from the intervention study. Although the activities were clinically validated, they were assigned by a researcher and not a caregiver. A long-term follow up is required with a health professional to clinically validate the efficacy of the platform in assisting the caregiver and supporting their users.

The other limitation was with the finite variability within the system which becomes less engaging because they eventually become predictable. The set of coaching activities, coaching messages and chatbot follow ups with users are renewed in every time and iteration, which minimizes the repetition involved, hence decreases the finite variability in the system. Nonetheless, since the state machine steering the dialog and tasks is determined and finite, this might impact user's engagement in the long-term. Future work will investigate whether there is any difference between a coaching system with and without a human coach.

7.15 Conclusion

In this chapter we highlighted the results of two experimental studies to validate our health coaching platform. The first experiment was conducted with real users over three weeks. The experiment focused on assessing the quality of the coaching delivered by a chatbot. Using conversational agents as a tool to facilitate caregiver's tasks and perform user follow-up can significantly improve tailoring the activity to fit user preferences. The user evaluation revealed participant's positive enjoyment, attitude, and confidence towards the use of chatbot. Participants also positively perceived the usefulness, ease of use, and satisfaction from the chatbot interaction. The second experiment was conducted with real users over one month. The users interacted with the chatbot where they received activities and provided their feedback. The chatbot forwarded to and later on collected the activities from each of the users, the chatbot also performed intervention questionnaires periodically. The findings revealed users who were highly engaged were also more adherent to chatbot activities. Future improvements will consider building the activity recommender based on user's interaction data and integrating user's dietary pattern and emotional wellbeing into the initial user clustering by leveraging ICT techniques (e.g., machine learning).

Chapter 8

Conclusion and Future Work

8.1 Overview

Chronic non-communicable diseases are the result of poor diet and physical inactivity, which are related to modifiable behaviors. Although existing approaches induce healthy behavior change, user engagement in such activities is still poor and it is evident that existing approaches lack the elements to achieve a sustainable user engagement. This dissertation have addressed the problem domain of healthy lifestyle promotion, through diet and physical activity adherence from the technology perspective. We leveraged upon an interdisciplinary approach to research, design and development to combine research in HCI, behavior change techniques and health informatics in a single endeavor towards extending the research on socio-technical systems to manage health and wellness promotion. This chapter summarizes the contributions of our dissertation by presenting the findings, acknowledging the limitations and highlighting the future directions we envision for this line of research.

We investigated conversational agents role in health promotion systems to track user daily diet and physical activity and provide follow-up to promote these aspects in the long-term. Conversational agents can persuade individuals to change their attitude or behavior through conversation. In the context of health, persuasive conversational agents converse with users to track their condition, assist with their care, provide them with instant supports and reminders, and also assist caregivers throughout the care process. Chatbots have shown a significant positive effect on user experience and interaction with health support systems [71, 70].

In this thesis, we explicitly implemented tactics into the chatbot that supports the CoachAI platform with the process of providing coaching tailored activities and tracking user adherence overtime. Although the study didn't target specific demographics, we consider elderly people as potential demographic fit for CoachAI. This is based on the results of our study described in Chapter 5. In addition, it enables users to overcome the confusing interfaces found in the GUI for mobile and web applications.

8.2 Contribution

Taking into account what we discussed in previous chapters, the contribution of this dissertation can be summarized as follows:

- Identifying habitual patterns to understand user preferences.
- The role of chatbot in delivering health promoting micro activities.
- Building the technology while adhering to individuals daily routine.
- A socio-technical system that fits with the role of conversational agent as an assistive component to caregiver, rather than a substitute.
- The role of human agent in the loop to tailor the health coaching task.

8.3 Research Findings

Across all the experiments discussed in previous chapters, there was strong support for hypothesis H1; building the technology inline with individual's routine of using messaging application did result in significant increase in the quality of the relationship between subjects and the conversational agent. The research concluded that building the technology in a user familiar environment and goes with their routine is an effective way to increase factors of sustainability of the technology. There was indeed a strong correlation between the inclusion of human agent in the coaching loop and individuals perceiving the system as credible, as was listed in hypothesis H2. This was concluded in the findings from both the CoachAI testing

and prospective intervention studies discussed in Chapter 7. Involving the human agent with individuals to set tailored exercise and follow-up on their progress has a significant impact on individuals behavior towards the system.

8.3.1 Conversational Agent

The design, implementation and evaluation of our system along with relevant research activities have provided us information on various aspects of conversational user interface as a persuasive element applied in health coaching. However, designing health dialog agent that is emotionally aware is still poorly studied. We found that different chatbot agent design and implementation pattern may fit well in some context rather than others. For example, a chatbot for health information tracking (as in CoachAI) should be task oriented and provide empathic support whenever relevant.

8.3.2 Overall Coaching Effect

Since we're testing a coaching system, another relevant measure we considered is the coaching quality. Although homogeneous participants received identical coaching activities and feedback, quality of coaching pertains to the entirety of the coaching experience received by the participants. There were some differences between the feedback version pretended to be sent by the bot and those sent by the human coach. This could be because participants expected a more personal connection with the human coach and hence judge certain aspects of the coaching differently. To measure the coaching effect we prepared questionnaires and asked participants to indicate, using a seven-point scale, how often the coach exhibits some behaviors. We omitted several points in the survey, because they pertained to qualities we can not reasonably assume our software to convey.

Although the tool was not used directly by a health coach, however, the design, implementation and evaluation parts were all supervised by a dietitian. The human coach reviewed the initial design and gave hints on how to improve the design to fit with their daily routine. The routine tasks were all given priority to be assigned to the chatbot agent.

8.3.3 Perceived Credibility & Acceptance

To measure perceived credibility, a combination of 3 questionnaires were distributed to the participants. We distributed the AttrakDiff (a seven-point scale) to participants to choose between them. One of the research questions concerns the potential beneficial effect of a chatbot on the long-term use of the system. We used Technology Acceptance Model (TAM) for this purpose. The TAM model presents a questionnaire for determining several aspects relevant to the expected acceptance and use of technologies. It is divided into a number of different constructs dealing with these aspects. Each construct consists of a small number of statements. The participant is asked to indicate his or her level of agreement with these statements on a seven-point scale.

8.3.4 Task Automation

Delivering activities related to individuals health and tracking their performance by the chatbot is among the goals of this research. Task automation is responsible for handling the main communications between the coach-chatbot-users. We found the activity definition and delivery to be effective in engaging users in the health activity. However, there was no clear indication on the effect of task automation on perceiving the activity as effective.

8.3.5 Individual-Chatbot Relationship

Hypothesis H1 was among the important goals of this dissertation. Demonstrating that building the approach inline with individuals routine of messaging application use is a prerequisite to demonstrate an agent can maintain a relationship with the user. We hypothesized this to impact the task outcome, and our research has accomplished that. Analyzing all subjects interaction with the system and their replies to the questionnaires were shown to have significant impact on the bond dimension-liking of CoachAI.

The fact that we didn't find significant changes in individuals interaction measure over time indicates that the conversational agent was able to maintain the sense of relationship with subjects over the entire month of the intervention. CoachAI focused on the goal represented by the activities to be achieved for a predefined period of

time. Having the agent track and remember things about subjects also allowed the dialog to be more tailored to the users. For example, suggesting that individuals reward themselves for a period of time and by doing things they love, when they perform well in the activities.

8.3.6 Coaching Effect on Activities

Even though few significant between-group differences were found, all groups increased their level of interaction with the system (on several measures) during the intervention period. However, there was a decrease in their activity level following removal of the intervention. Some plausible reasons for the increased activity may be because of being in a study or having to report in everyday on behavior are motivating in and of themselves. We used several intervention techniques throughout the experiment, including self-monitoring, goal-setting, and education, thus the increase in activity (as well as self-efficacy) for all groups during the intervention period most likely reflects the effectiveness of these proven techniques.

8.4 Persisting Challenges

By applying multiple methods, naturally there are several limitations to consider for this research, some of them have already been introduced. One is the multiple coding stages performed during the study. It originated mainly due to the nature of this work, which was that most of the coding process was conducted by a single researcher, only with direct assistance of a second coder for a test of the codebook. Another limitation of this dissertation is that the initial experiment lacked the participation of young people, instead they participated at the later experimentation. Perhaps setting an experiment with mixed participants would have resulted in substantially more interesting results.

A second limitation has to do with an unbalance between the effort devoted to the design, preparation and execution of the field studies; and the effort devoted to the analysis of its outcomes. Moreover, the delay to do initial or long-term experimentation was also a limitation. The former was substantially more limiting than the later and the former was also the cause for the later limitation. Therefore, the study duration was between 3 weeks - 1 month, which was mainly due to the

difficulties to recruit participants and obtain the legal confirmation to setup the experiment. Although the results obtained are interesting to provide hints on the effectiveness of the approach, however a long-term pilot study is needed to validate the effectiveness of the approach in the long-term [143, 144].

A third limitation has to do with the non-implemented features of CoachAI, which we believe would have had an important impact in the sessions development. For example, adding a layer of emotional intelligence into the CoachAI bot could make the bot very powerful to detect various users' emotional states from the conversation. The support for clinical NLP and NLU in the chatbot for Italian language could have transformed the tool in a much more effective instrument for supporting the interventions. These features will be considered for the future work and once we have the data to accomplish them.

8.5 Future Directions

Although the future suggestions we provided previously, this section describes future directions for this research. To overcome some of the aforementioned limitations, one future improvement is involving multiple researchers in data coding process, thus increase the consistency with the research questions, and consequently increase results reliability. Moving on to the discussion on persuasiveness, focused research should be conducted to widely identify specifications to incorporate persuasive principles into the system. Further research is needed to help characterize the requirements to develop a robust system. For example, the activity scheduler may remind the users about upcoming sessions. Some future CoachAI improvements may include: integrating a knowledge-base in the platform on specific coaching topics, customizable with questions/answers provided by the users in addition to automatic learning ability from previous conversations. Therefore, on model we consider for the future improvement is building the chatbot dialog using the RASA Core¹ dialog engine to build contexts and add AI assistant into the chatbot dialog. Employing CoachAI in the health context should reduce caregivers calls, provide timely help and quick responses, and decrease users need for in-person visits. Integrating a medical knowledge-base into CoachAI will have a significant impact on reducing the burden on caregiver. Users can also seek any general advices and maintain a healthy

¹<https://rasa.com/docs/core/>

lifestyle. This was not considered in the current study, since it falls outside the scope of this research and requires a research study of its own. Other promising future directions for CoachAI are detecting user intents from their inputs and providing them with timely information.

There are still open and promising spaces for research and innovation in integrating virtual agents in different health contexts. We have thoroughly explored one of such experiences, but there is still room for other contributions. For example, building a dialog model to tackle mental diseases. We envision a future where more health coaching systems will integrate virtual agents to support tracking users and collaborating with the coach to tailor the activity. A new algorithm could leverage real time activity suggestion to the coach based on users activities and adherence level. Finally, research studies could investigate other motivational techniques, such as using gamification to induce user engagement with the application and integrating this into the dialog model to enhance long-term user engagement.

References

- [1] Amalia Waxman. Why a global strategy on diet, physical activity and health? In *Nutrition and Fitness: Mental Health, Aging, and the Implementation of a Healthy Diet and Physical Activity Lifestyle*, volume 95, pages 162–166. Karger Publishers, 2005.
- [2] Maria Inês Schmidt, Bruce Bartholow Duncan, Gulnar Azevedo e Silva, Ana Maria Menezes, Carlos Augusto Monteiro, Sandhi Maria Barreto, Dora Chor, and Paulo Rossi Menezes. Chronic non-communicable diseases in brazil: burden and current challenges. *The Lancet*, 377(9781):1949–1961, 2011.
- [3] Abdesslam Boutayeb and Saber Boutayeb. The burden of non communicable diseases in developing countries. *International journal for equity in health*, 4(1):2, 2005.
- [4] David Bloom, Elizabeth Cafiero, Eva Jané-Llopis, Shafika Abrahams-Gessel, Lakshmi Bloom, Sana Fathima, Andrea Feigl, Tom Gaziano, Ali Hamandi, Mona Mowafi, et al. The global economic burden of noncommunicable diseases. Technical report, Program on the Global Demography of Aging, 2012.
- [5] Michele Cecchini, Franco Sassi, Jeremy A Lauer, Yong Y Lee, Veronica Guajardo-Barron, and Daniel Chisholm. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. *The Lancet*, 376(9754):1775–1784, 2010.
- [6] Stephen S Lim, Theo Vos, Abraham D Flaxman, Goodarz Danaei, Kenji Shibuya, Heather Adair-Rohani, Mohammad A AlMazroa, Markus Amann, H Ross Anderson, Kathryn G Andrews, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. *The lancet*, 380(9859):2224–2260, 2012.
- [7] Ashkan Afshin, Mohammad H Forouzanfar, Marissa B Reitsma, Patrick Sur, Kara Estep, Alex Lee, Laurie Marczak, Ali H Mokdad, Maziar Moradi-Lakeh, Mohsen Naghavi, et al. Health effects of overweight and obesity in 195 countries over 25 years. *The New England journal of medicine*, 377(1):13–27, 2017.
- [8] Helaine E Resnick and Barbara V Howard. Diabetes and cardiovascular disease. *Annual review of medicine*, 53(1):245–267, 2002.
- [9] Falk Müller-Riemenschneider, Thomas Reinhold, Marc Nocon, and Stefan N Willich. Long-term effectiveness of interventions promoting physical activity: a systematic review. *Preventive medicine*, 47(4):354–368, 2008.
- [10] Michael L Dansinger, Athina Tatsioni, John B Wong, Mei Chung, and Ethan M Balk. Meta-analysis: The effect of dietary counseling for weight losseffect of dietary counseling on weight loss. *Annals of internal medicine*, 147(1):41–50, 2007.

- [11] Rita O Orji, Julita Vassileva, and Regan L Mandryk. Modeling gender differences in healthy eating determinants for persuasive intervention design. In *International Conference on Persuasive Technology*, pages 161–173. Springer, 2013.
- [12] Emil Chiauzzi, Pronabesh DasMahapatra, Elisenda Cochín, Mikele Bunce, Raya Khoury, and Purav Dave. Factors in patient empowerment: a survey of an online patient research network. *The Patient-Patient-Centered Outcomes Research*, 9(6):511–523, 2016.
- [13] Amanda Fitzgerald, Caroline Heary, Colette Kelly, Elizabeth Nixon, and Mark Shevlin. Self-efficacy for healthy eating and peer support for unhealthy eating are associated with adolescents' food intake patterns. *Appetite*, 63:48–58, 2013.
- [14] Predrag Klasnja and Wanda Pratt. Healthcare in the pocket: mapping the space of mobile-phone health interventions. *Journal of biomedical informatics*, 45(1):184–198, 2012.
- [15] Shyamal Patel, Hyung Park, Paolo Bonato, Leighton Chan, and Mary Rodgers. A review of wearable sensors and systems with application in rehabilitation. *Journal of neuroengineering and rehabilitation*, 9(1):21, 2012.
- [16] Upkar Varshney. Pervasive healthcare and wireless health monitoring. *Mobile Networks and Applications*, 12(2-3):113–127, 2007.
- [17] Waneen W Spirduso and D Leilani Cronin. Exercise dose–response effects on quality of life and independent living in older adults. *Medicine & science in sports & exercise*, 2001.
- [18] Feng J He, Katharine H Jenner, and Graham A MacGregor. Wash?world action on salt and health. *Kidney international*, 78(8):745–753, 2010.
- [19] Perviz Asaria, Dan Chisholm, Colin Mathers, Majid Ezzati, and Robert Beaglehole. Chronic disease prevention: health effects and financial costs of strategies to reduce salt intake and control tobacco use. *The Lancet*, 370(9604):2044–2053, 2007.
- [20] Ahmed Fadhil, Cristina Matteotti, Giampaolo Armellin, Adolfo Villafiorita, and Dario Betti. Coach me: a platform for promoting healthy lifestyle. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 1077–1080. ACM, 2016.
- [21] Edward M Phillips, Jeffrey C Schneider, and Greg R Mercer. Motivating elders to initiate and maintain exercise1. *Archives of physical medicine and rehabilitation*, 85:52–57, 2004.
- [22] Gabe Zichermann and Christopher Cunningham. *Gamification by design: Implementing game mechanics in web and mobile apps*. " O'Reilly Media, Inc.", 2011.
- [23] Enrique J Gómez, M Elena Hernando, A Garcia, Francisco Del Pozo, J Cermenó, R Corcoy, Eulalia Brugués, and Alberto De Leiva. Telemedicine as a tool for intensive management of diabetes: the diabetel experience. *Computer methods and programs in biomedicine*, 69(2):163–177, 2002.
- [24] Elizabeth L Murnane, David Huffaker, and Gueorgi Kossinets. Mobile health apps: adoption, adherence, and abandonment. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 261–264. ACM, 2015.

- [25] James Clawson, Jessica A Pater, Andrew D Miller, Elizabeth D Mynatt, and Lena Mamykina. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 647–658. ACM, 2015.
- [26] Paul J Hu, Patrick YK Chau, Olivia R Liu Sheng, and Kar Yan Tam. Examining the technology acceptance model using physician acceptance of telemedicine technology. *Journal of management information systems*, 16(2):91–112, 1999.
- [27] Margaret Bevans and Esther M Sternberg. Caregiving burden, stress, and health effects among family caregivers of adult cancer patients. *Jama*, 307(4):398–403, 2012.
- [28] Ross E O’Hara, Jay G Hull, Kathleen D Lyons, Marie Bakitas, Mark T Hegel, Zhongze Li, and Tim A Ahles. Impact on caregiver burden of a patient-focused palliative care intervention for patients with advanced cancer. *Palliative & supportive care*, 8(4):395–404, 2010.
- [29] Qing Wang, Bjørge Egelandstal, Gro V Amdam, Valerie L Almlie, and Marije Oostindjer. Diet and physical activity apps: perceived effectiveness by app users. *JMIR mHealth and uHealth*, 4(2), 2016.
- [30] Andreas Holzinger, Gig Searle, and Alexander Nischelwitzer. On some aspects of improving mobile applications for the elderly. In *International Conference on Universal Access in Human-Computer Interaction*, pages 923–932. Springer, 2007.
- [31] Felipe Lobelo and Isabel Garcia de Quevedo. The evidence in support of physicians and health care providers as physical activity role models. *American journal of lifestyle medicine*, 10(1):36–52, 2016.
- [32] Luciano Gamberini, Giacinto Barresi, A Maier, and Fabiola Scarpetta. A game a day keeps the doctor away: A short review of computer games in mental healthcare. *Journal of CyberTherapy and Rehabilitation*, 1(2):127–145, 2008.
- [33] Ellen Brox, Luis Fernandez Luque, Gunn J Evertsen, and Juan Emilio González Hernández. Exergames for elderly: Social exergames to persuade seniors to increase physical activity. In *Pervasive Computing Technologies for Healthcare (Pervasive-Health), 2011 5th International Conference on*, pages 546–549. IEEE, 2011.
- [34] Sajanee Halko and Julie Kientz. Personality and persuasive technology: an exploratory study on health-promoting mobile applications. *Persuasive technology*, pages 150–161, 2010.
- [35] Joel C Perry, Julien Andureu, Francesca Irene Cavallaro, Jan Veneman, Stefan Carmien, and Thierry Keller. Effective game use in neurorehabilitation: user-centered perspectives. In *Handbook of research on improving learning and motivation through educational games: multidisciplinary approaches*, pages 683–725. IGI Global, 2011.
- [36] Wijnand Ijsselsteijn, Henk Herman Nap, Yvonne de Kort, and Karolien Poels. Digital game design for elderly users. In *Proceedings of the 2007 conference on Future Play*, pages 17–22. ACM, 2007.
- [37] Brian J Fogg. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*, page 40. ACM, 2009.
- [38] Brian J Fogg and Jason Hreha. Behavior wizard: a method for matching target behaviors with solutions. In *International Conference on Persuasive Technology*, pages 117–131. Springer, 2010.

- [39] Harri Oinas-Kukkonen and Marja Harjumaa. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems*, 24(1):28, 2009.
- [40] Eva van Veen, Jeske FM Bovendeert, Frank JG Backx, and Bionka MA Huisstede. E-coaching: New future for cardiac rehabilitation? a systematic review. *Patient Education and Counseling*, 2017.
- [41] Laura Dennison, Leanne Morrison, Gemma Conway, and Lucy Yardley. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. *Journal of medical Internet research*, 15(4), 2013.
- [42] Lindsey Dayer, Seth Heldenbrand, Paul Anderson, Paul O Gubbins, and Bradley C Martin. Smartphone medication adherence apps: potential benefits to patients and providers. *Journal of the American Pharmacists Association: JAPhA*, 53(2):172, 2013.
- [43] Jen-Her Wu and Shu-Ching Wang. What drives mobile commerce?: An empirical evaluation of the revised technology acceptance model. *Information & management*, 42(5):719–729, 2005.
- [44] Inseong Lee, Boreum Choi, Jinwoo Kim, and Se-Joon Hong. Culture-technology fit: Effects of cultural characteristics on the post-adoption beliefs of mobile internet users. *International Journal of Electronic Commerce*, 11(4):11–51, 2007.
- [45] Tobias Kowatsch, Marcia Nißen, Chen-Hsuan I Shih, Dominik Rüegger, Dirk Volland, Andreas Filler, Florian Künzler, Filipe Barata, Severin Hung, Dirk Büchter, et al. Text-based healthcare chatbots supporting patient and health professional teams: Preliminary results of a randomized controlled trial on childhood obesity. In *Persuasive Embodied Agents for Behavior Change (PEACH2017)*. ETH Zurich, 2017.
- [46] Ludovico Boratto, Salvatore Carta, Fabrizio Mulas, and Paolo Pilloni. An e-coaching ecosystem: design and effectiveness analysis of the engagement of remote coaching on athletes. *Personal and Ubiquitous Computing*, 21(4):689–704, 2017.
- [47] Bart A Kamphorst, Michel CA Klein, and Arlette van Wissen. Human involvement in e-coaching: Effects on effectiveness, perceived influence and trust. In *International Workshop on Human Behavior Understanding*, pages 16–29. Springer, 2014.
- [48] Ville-Petteri Mäkinen, Pasi Soininen, Carol Forsblom, Maija Parkkonen, Petri Ingman, Kimmo Kaski, Per-Henrik Groop, and Mika Ala-Korpela. 1h nmr metabonomics approach to the disease continuum of diabetic complications and premature death. *Molecular systems biology*, 4(1):167, 2008.
- [49] Janet F Grant, Catherine R Chittleborough, Anne W Taylor, Eleonora Dal Grande, David H Wilson, Patrick J Phillips, Robert J Adams, Julianne Cheek, Kay Price, Tiffany Gill, et al. The north west adelaide health study: detailed methods and baseline segmentation of a cohort for selected chronic diseases. *Epidemiologic Perspectives & Innovations*, 3(1):4, 2006.
- [50] Reisa A Sperling, Paul S Aisen, Laurel A Beckett, David A Bennett, Suzanne Craft, Anne M Fagan, Takeshi Iwatsubo, Clifford R Jack, Jeffrey Kaye, Thomas J Montine, et al. Toward defining the preclinical stages of alzheimer’s disease: Recommendations from the national institute on aging-alzheimer’s association workgroups on diagnostic guidelines for alzheimer’s disease. *Alzheimer’s & dementia*, 7(3):280–292, 2011.

- [51] Maged N Kamel Boulos, Ann C Brewer, Chante Karimkhani, David B Buller, and Robert P Dellavalle. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online journal of public health informatics*, 5(3):229, 2014.
- [52] Melinda Huffman. Health coaching: A new and exciting technique to enhance patient self-management and improve outcomes. *Home Healthcare Now*, 25(4):271–274, 2007.
- [53] Michael F Roizen, Mehmet C Oz, and Jeffrey D Roizen. Automated coaching, July 15 2010. US Patent App. 12/649,661.
- [54] Margarite J Vale, Michael V Jelinek, James D Best, Anthony M Dart, Leeanne E Grigg, David L Hare, Betty P Ho, Robert W Newman, and John J McNeil. Coaching patients on achieving cardiovascular health (coach): a multicenter randomized trial in patients with coronary heart disease. *Archives of Internal Medicine*, 163(22):2775–2783, 2003.
- [55] Harm op den Akker, Valerie M Jones, and Hermie J Hermens. Tailoring real-time physical activity coaching systems: a literature survey and model. *User modeling and user-adapted interaction*, 24(5):351–392, 2014.
- [56] Alice Watson, Timothy Bickmore, Abby Cange, Ambar Kulshreshtha, and Joseph Kvedar. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. *Journal of medical Internet research*, 14(1), 2012.
- [57] Aranka Dol, Olga Kulyk, Hugo Velthuijsen, Lisette van Gemert-Pijnen, and Tatjana van Strien. The use of dialectical dialogues in a personalized virtual coach for obese emotional eaters.
- [58] Olivier A Blanson Henkemans, Paul JM van der Boog, Jasper Lindenberg, Charles APG van der Mast, Mark A Neerincx, and Bertie JHM Zwetsloot-Schonk. An online lifestyle diary with a persuasive computer assistant providing feedback on self-management. *Technology and Health Care*, 17(3):253–267, 2009.
- [59] Julie Barnett, Michelle Harricharan, Dave Fletcher, Becky Gilchrist, and Jane Coughlan. mypace: An integrative health platform for supporting weight loss and maintenance behaviors. *IEEE journal of biomedical and health informatics*, 19(1):109–116, 2015.
- [60] Yunlong Wang, Lingdan Wu, Jan-Philipp Lange, Ahmed Fadhil, and Harald Reiterer. Persuasive technology in reducing prolonged sedentary behavior at work: A systematic review. *Smart Health*, 2018.
- [61] Timothy Bickmore, Daniel Mauer, Francisco Crespo, and Thomas Brown. Persuasion, task interruption and health regimen adherence. In *International Conference on Persuasive Technology*, pages 1–11. Springer, 2007.
- [62] Sunny Consolvo, Katherine Everitt, Ian Smith, and James A Landay. Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 457–466. ACM, 2006.
- [63] Tammy Toscos, Anne Faber, Shunying An, and Mona Praful Gandhi. Chick clique: persuasive technology to motivate teenage girls to exercise. In *CHI'06 extended abstracts on Human factors in computing systems*, pages 1873–1878. ACM, 2006.
- [64] Rebecca E Grinter, Katie A Siek, and Andrea Grimes. Wellness informatics: towards a definition and grand challenges. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 4505–4508. ACM, 2010.

- [65] Ashok J Bharucha, Vivek Anand, Jodi Forlizzi, Mary Amanda Dew, Charles F Reynolds, Scott Stevens, and Howard Wactlar. Intelligent assistive technology applications to dementia care: current capabilities, limitations, and future challenges. *The American journal of geriatric psychiatry*, 17(2):88–104, 2009.
- [66] Patrick C Shih, Kyungsik Han, Erika Shehan Poole, Mary Beth Rosson, and John M Carroll. Use and adoption challenges of wearable activity trackers. *IConference 2015 Proceedings*, 2015.
- [67] Hyunho Lee and Youngseok Lee. A look at wearable abandonment. In *Mobile Data Management (MDM), 2017 18th IEEE International Conference on*, pages 392–393. IEEE, 2017.
- [68] Aniek J Lentferink, Hilbrand KE Oldenhuis, Martijn de Groot, Louis Polstra, Hugo Velthuijsen, and Julia EWC van Gemert-Pijnen. Key components in ehealth interventions combining self-tracking and persuasive ecoaching to promote a healthier lifestyle: A scoping review. *Journal of medical Internet research*, 19(8), 2017.
- [69] Chayan Chakrabarti and George F Luger. A semantic architecture for artificial conversations. In *Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th International Conference on*, pages 21–26. IEEE, 2012.
- [70] Fabio Buttussi and Luca Chittaro. Mopet: A context-aware and user-adaptive wearable system for fitness training. *Artificial Intelligence in Medicine*, 42(2):153–163, 2008.
- [71] Timothy W Bickmore, Daniel Mauer, and Thomas Brown. Context awareness in a handheld exercise agent. *Pervasive and mobile computing*, 5(3):226–235, 2009.
- [72] Alice Kerly, Phil Hall, and Susan Bull. Bringing chatbots into education: Towards natural language negotiation of open learner models. *Knowledge-Based Systems*, 20(2):177–185, 2007.
- [73] Timothy Bickmore and Toni Giorgino. Health dialog systems for patients and consumers. *Journal of biomedical informatics*, 39(5):556–571, 2006.
- [74] Timothy W Bickmore, Laura M Pfeifer, and Michael K Paasche-Orlow. Health document explanation by virtual agents. In *International Workshop on Intelligent Virtual Agents*, pages 183–196. Springer, 2007.
- [75] Saiph Savage, Andres Monroy-Hernandez, and Tobias Höllerer. Botivist: Calling volunteers to action using online bots. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 813–822. ACM, 2016.
- [76] Timothy W Bickmore, Lisa Caruso, Kerri Clough-Gorr, and Tim Heeren. ?it?s just like you talk to a friend?relational agents for older adults. *Interacting with Computers*, 17(6):711–735, 2005.
- [77] Matthew Henderson, Blaise Thomson, and Steve Young. Word-based dialog state tracking with recurrent neural networks. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 292–299, 2014.
- [78] Cristian Pop-Eleches, Harsha Thirumurthy, James P Habyarimana, Joshua G Zivin, Markus P Goldstein, Damien De Walque, Leslie Mackeen, Jessica Haberer, Sylvester Kimaiyo, John Sidle, et al. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. *AIDS (London, England)*, 25(6):825, 2011.

- [79] Xiaoqiu Liu, James J Lewis, Hui Zhang, Wei Lu, Shun Zhang, Guilan Zheng, Liqiong Bai, Jun Li, Xue Li, Hongguang Chen, et al. Effectiveness of electronic reminders to improve medication adherence in tuberculosis patients: a cluster-randomised trial. *PLoS Med*, 12(9):e1001876, 2015.
- [80] Eduardo Sabaté. *Adherence to long-term therapies: evidence for action*. World Health Organization, 2003.
- [81] Timothy Bickmore and Daniel Schulman. The comforting presence of relational agents. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, pages 550–555. ACM, 2006.
- [82] Allison A Vorderstrasse, Geoffrey S Ginsburg, William E Kraus, Maj Carlos J Maldonado, and Ruth Q Wolever. Health coaching and genomics?-potential avenues to elicit behavior change in those at risk for chronic disease: Protocol for personalized medicine effectiveness study in air force primary care. *Global Advances in Health and Medicine*, 2(3):26–38, 2013.
- [83] Bart A Kamphorst, Michel CA Klein, and Arlette Van Wissen. Autonomous e-coaching in the wild: empirical validation of a model-based reasoning system. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 725–732. International Foundation for Autonomous Agents and Multiagent Systems, 2014.
- [84] James O Prochaska. Transtheoretical model of behavior change. In *Encyclopedia of behavioral medicine*, pages 1997–2000. Springer, 2013.
- [85] Viswanath Venkatesh and Fred D Davis. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2):186–204, 2000.
- [86] Godfrey Hochbaum, Irwin Rosenstock, and Stephen Kegels. Health belief model. *United States Public Health Service*, 1952.
- [87] BJ Fogg. Bj fogg’s behavior model. *A behavior model for persuasive design URL: http://bjfogg.com/fbm_files/page4_1.pdf [accessed 2014-01-09][WebCite Cache]*, 2011.
- [88] Nancy K Janz and Marshall H Becker. The health belief model: A decade later. *Health education quarterly*, 11(1):1–47, 1984.
- [89] Ralf Schwarzer, Sonia Lippke, and Aleksandra Luszczynska. Mechanisms of health behavior change in persons with chronic illness or disability: the health action process approach (hapa). *Rehabilitation psychology*, 56(3):161, 2011.
- [90] Iman Khaghani-Far, Svetlana Nikitina, Marcos Baez, Ekaterina A Taran, and Fabio Casati. Fitness applications for home-based training. *IEEE Pervasive Computing*, (4):56–65, 2016.
- [91] Maggie Sourial, Ahmed Elnaggar, and Dirk Reichardt. Development of a virtual coach scenario for hand therapy using leap motion. In *Future Technologies Conference (FTC)*, pages 1071–1078. IEEE, 2016.
- [92] Chris Van Hoof. The virtual personal health coach: Technology and data analytics join forces to disrupt preventive health. In *Advances in Sensors and Interfaces (IWASI), 2017 7th IEEE International Workshop on*, pages 233–233. IEEE, 2017.

- [93] Kimiko Ryokai, Florian Michahelles, Mareike Kritzler, and Suhaib Syed. Communicating and interpreting wearable sensor data with health coaches. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, 2015 9th International Conference on, pages 221–224. IEEE, 2015.
- [94] Stephen M Schueller and David C Mohr. Initial field trial of a coach-supported web-based depression treatment. In *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*, pages 25–28. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2015.
- [95] Heiko Braak, Estifanos Ghebremedhin, Udo Rüb, Hansjürgen Bratzke, and Kelly Del Tredici. Stages in the development of parkinson’s disease-related pathology. *Cell and tissue research*, 318(1):121–134, 2004.
- [96] Harri Oinas-Kukkonen and Marja Harjumaa. Persuasive systems design: key issues, process model and system features. In *Routledge Handbook of Policy Design*, pages 105–123. Routledge, 2018.
- [97] Rachel Davis, Rona Campbell, Zoe Hildon, Lorna Hobbs, and Susan Michie. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health psychology review*, 9(3):323–344, 2015.
- [98] Karen Glanz and Donald B Bishop. The role of behavioral science theory in development and implementation of public health interventions. *Annual review of public health*, 31:399–418, 2010.
- [99] Carlo C DiClemente and James O Prochaska. Toward a comprehensive, transtheoretical model of change: Stages of change and addictive behaviors. 1998.
- [100] Stephen Rollnick, William R Miller, Christopher C Butler, and Mark S Aloia. Motivational interviewing in health care: helping patients change behavior, 2008.
- [101] Saadi Lahlou. Social representations and social construction: The evolutionary perspective of installation theory. *The Cambridge handbook of social representations*, 193:210–223, 2015.
- [102] Harri Oinas-Kukkonen. Behavior change support systems: A research model and agenda. In *International Conference on Persuasive Technology*, pages 4–14. Springer, 2010.
- [103] Pelle Guldberg Hansen and Andreas Maaløe Jespersen. Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation*, 4(1):3–28, 2013.
- [104] Richard E Petty and John T Cacioppo. The elaboration likelihood model of persuasion. In *Communication and persuasion*, pages 1–24. Springer, 1986.
- [105] Leanne G Morrison, Lucy Yardley, John Powell, and Susan Michie. What design features are used in effective e-health interventions? a review using techniques from critical interpretive synthesis. *Telemedicine and e-Health*, 18(2):137–144, 2012.
- [106] Timothy W Bickmore, Laura M Pfeifer, Donna Byron, Shaula Forsythe, Lori E Henault, Brian W Jack, Rebecca Silliman, and Michael K Paasche-Orlow. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. *Journal of health communication*, 15(S2):197–210, 2010.
- [107] Ahmed Fadhil. Domain specific design patterns: Designing for conversational user interfaces. *arXiv preprint arXiv:1802.09055*, 2018.

- [108] *Osakidetza. Chronicity in Basque country.* http://www.osakidetza.euskadi.eus/r85-gkgnr100es/contenidosinformaciondocumentos_cronicoses_cronicadjuntosChronicityBasqueCountry.pdf, 2010.
- [109] Barbara Olasov Rothbaum, Elizabeth A Meadows, Patricia Resick, and David W Foy. Cognitive-behavioral therapy. 2000.
- [110] Michael Marmot. Social determinants of health inequalities. *The Lancet*, 365(9464):1099–1104, 2005.
- [111] Steve Jones. How i learned to stop worrying and love the bots. *Social Media+ Society*, 1(1):2056305115580344, 2015.
- [112] Jeffrey P Migneault, Ramesh Farzanfar, Julie A Wright, and Robert H Friedman. How to write health dialog for a talking computer. *Journal of biomedical informatics*, 39(5):468–481, 2006.
- [113] Timothy W Bickmore, Daniel Schulman, and Candace L Sidner. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. *Journal of biomedical informatics*, 44(2):183–197, 2011.
- [114] World Health Organization et al. Fiscal policies for diet and prevention of noncommunicable diseases: technical meeting report, 5-6 may 2015, geneva, switzerland. 2016.
- [115] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group*, pages 63–76. Springer, 2008.
- [116] James O Prochaska and Wayne F Velicer. The transtheoretical model of health behavior change. *American journal of health promotion*, 12(1):38–48, 1997.
- [117] Martin Schrepp. User experience questionnaire handbook. *All you need to know to apply the UEQ successfully in your project*, 2015.
- [118] Guozhen An, Mehrnoosh Shafiee, and Davood Shamsi. Improving retrieval modeling using cross convolution networks and multi frequency word embedding. *arXiv preprint arXiv:1802.05373*, 2018.
- [119] Matthias Romppel, Elmar Braehler, Marcus Roth, and Heide Glaesmer. What is the general health questionnaire-12 assessing? *Comprehensive Psychiatry*, 54(4):406–413, 2013.
- [120] Timothy Armstrong and Fiona Bull. Development of the world health organization global physical activity questionnaire (GPAQ). *Journal of Public Health*, 14(2):66–70, 2006.
- [121] Henning Pohl, Christian Domin, and Michael Rohs. Beyond just text: Semantic emoji similarity modeling to support expressive communication 👫📲😃. *ACM Trans. Comput.-Hum. Interact.*, 24(1):6:1–6:42, March 2017.
- [122] Timothy Bickmore, Daniel Schulman, and Langxuan Yin. Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 24(6):648–666, 2010.
- [123] Jingjing Ren, Timothy Bickmore, Megan Hempstead, and Brian Jack. Birth control, drug abuse, or domestic violence? what health risk topics are women willing to discuss with a virtual agent? In *International Conference on Intelligent Virtual Agents*, pages 350–359. Springer, 2014.

- [124] Paul C Cozby. Self-disclosure: a literature review. *Psychological bulletin*, 79(2):73, 1973.
- [125] Lawrence R Wheelless and Jams Grotz. Conceptualization and measurement of reported self-disclosure. *Human Communication Research*, 2(4):338–346, 1976.
- [126] Roger C Mayer, James H Davis, and F David Schoorman. An integrative model of organizational trust. *Academy of management review*, 20(3):709–734, 1995.
- [127] Marcin Skowron, Mathias Theunis, Stefan Rank, and Arvid Kappas. Affect and social processes in online communication—experiments with an affective dialog system. *IEEE Transactions on Affective Computing*, 4(3):267–279, 2013.
- [128] Jessica Fitts Willoughby and Shuang Liu. Do pictures help tell the story? an experimental test of narrative and emojis in a health text message intervention. *Computers in Human Behavior*, 79:75–82, 2018.
- [129] Alexander Hogenboom, Daniella Bal, Flavius Frasinca, Malissa Bal, Franciska de Jong, and Uzay Kaymak. Exploiting emoticons in sentiment analysis. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pages 703–710. ACM, 2013.
- [130] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, 4(2):e19, June 2017.
- [131] Mohit Jain, Pratyush Kumar, Ramachandra Kota, and Shwetak N Patel. Evaluating and informing the design of chatbots. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018*, pages 895–906. ACM, 2018.
- [132] American College of Sports Medicine et al. *ACSM’s health-related physical fitness assessment manual*. Lippincott Williams & Wilkins, 2013.
- [133] Joseph B Walther and Kyle P D’Addario. The impacts of emoticons on message interpretation in computer-mediated communication. *Social science computer review*, 19(3):324–347, 2001.
- [134] Arnold M Lund. Measuring usability with the use questionnaire12. *Usability interface*, 8(2):3–6, 2001.
- [135] Adam O Horvath and Leslie S Greenberg. Development and validation of the working alliance inventory. *Journal of counseling psychology*, 36(2):223, 1989.
- [136] Marcin Skowron, Mathias Theunis, Stefan Rank, and Arvid Kappas. Affect and social processes in online communication—experiments with an affective dialog system. *IEEE Transactions on Affective Computing*, 4(3):267–279, 2013.
- [137] Ron Kwok Chi-Wai, Tania Mak So-Ning, Ken Lee Wing-Kuen, Stanley Sai-Chuen Hui, Peter Wu Ka-Shun, and Clara Choi-Ki Wong. Can mobile virtual fitness apps replace human fitness trainer? In *Information Science and Service Science (NISS), 2011 5th International Conference on New Trends in*, volume 1, pages 56–63. IEEE, 2011.
- [138] Marc Hassenzahl, Michael Burmester, and Franz Koller. Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. In *Mensch & Computer 2003*, pages 187–196. Springer, 2003.

- [139] United States Department of Agriculture. Choose myplate. *How many vegetables are needed daily or weekly?*, 2012.
- [140] Ralf Schwarzer. Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied psychology*, 57(1):1–29, 2008.
- [141] Carlos H Schenck, Thomas D Hurwitz, Scott R Bundlie, and Mark W Mahowald. Sleep-related eating disorders: polysomnographic correlates of a heterogeneous syndrome distinct from daytime eating disorders. *Sleep*, 14(5):419–431, 1991.
- [142] Ahmed Fadhil and Silvia 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*, pages 261–265. ACM, 2017.
- [143] Verena K Günther, P Schäfer, BJ Holzner, and GW Kemmler. Long-term improvements in cognitive performance through computer-assisted cognitive training: a pilot study in a residential home for older people. *Aging & mental health*, 7(3):200–206, 2003.
- [144] Ms DENISE MAIERLE and Polly Ryan. Evaluation of satisfaction and use of electronic intervention for behavior change. *Computers, informatics, nursing: CIN*, 29(11):622, 2011.

Appendix A

Experiments Resources

A.1 CoachMe Questionnaire

Questionario Progetto *Coach me*

Istruzioni per la compilazione

Nel questionario ti verranno chieste delle domande sull'applicazione CoachMe.

Per rispondere basta fare una crocetta (x) sulla risposta che più si avvicina alla tua opinione. Se non trovi, tra le risposte possibili, proprio quella che vorresti dare, dovrai indicare quella che corrisponde di più al **tuo punto di vista**. Tieni presente che **non esistono risposte giuste** o sbagliate.

Nella maggior parte delle domande ti sarà chiesto di indicare quanto sei d'accordo con una data affermazione. In questi casi le opzioni di scelta vanno dal livello più **basso**, che significa che sei **"del tutto in disaccordo"**, fino al livello più **alto**, che significa che sei **"del tutto d'accordo"** con l'affermazione riportata.

L'immagine qui sotto mostra tutte le opzioni di scelta possibili

☐	☐	☐	☐	☐	☐	☐
Del tutto in disaccordo	In disaccordo	Parzialmente in disaccordo	Indeciso	Parzialmente d'accordo	D'accordo	Del tutto d'accordo

Buona compilazione!

Esprimi il tuo livello di accordo o disaccordo con le affermazioni riportate.

1. Con l'applicazione CoachMe posso migliorare il mio stile di vita

☐	☐	☐	☐	☐	☐	☐
Del tutto in disaccordo			Indeciso	Del tutto d'accordo		

2. L'impegno che mi è stato chiesto nell'usare CoachMe è proporzionato rispetto ai vantaggi offerti

☐	☐	☐	☐	☐	☐	☐
Del tutto in disaccordo			Indeciso	Del tutto d'accordo		

3. Ho usato volentieri CoachMe

☐	☐	☐	☐	☐	☐	☐
Del tutto in disaccordo			Indeciso	Del tutto d'accordo		

4. CoachMe è un'applicazione molto facile da usare

☐	☐	☐	☐	☐	☐	☐
Del tutto in disaccordo			Indeciso	Del tutto d'accordo		

5. Ho imparato facilmente ad usare CoachMe

☐	☐	☐	☐	☐	☐	☐
Del tutto in disaccordo		Indeciso				Del tutto d'accordo

6. Mi ricordo perfettamente come usare CoachMe e i suoi comandi

☐	☐	☐	☐	☐	☐	☐
Del tutto in disaccordo		Indeciso			Del tutto d'accordo	

7. Ho compreso a cosa serve CoachMe

☐	☐	☐	☐	☐	☐	☐
Del tutto in disaccordo		Indeciso				Del tutto d'accordo

8. Mi sono divertito nell'usare CoachMe

☐	☐	☐	☐	☐	☐	☐
Del tutto in disaccordo		Indeciso				Del tutto d'accordo

9. Mi sono ricordato ogni giorno di accedere CoachMe

□ □ □ □ □ □ □

Del tutto in
disaccordo

Indeciso

Del tutto
d'accordo

10. Ho mostrato CoachMe ad amici e parenti

☐☐☐☐☐☐☐

Del tutto in
disaccordo

Indeciso

Del tutto
d'accordo

**11. L'uso dei simboli grafici (come questo 🙌) rendevano
l'interazione con CoachMe più piacevole**

☐☐☐☐☐☐☐

Del tutto in
disaccordo

Indeciso

Del tutto
d'accordo

12. Usare CoachMe è come usare Whatsapp

☐☐☐☐☐☐☐

Del tutto in
disaccordo

Indeciso

Del tutto
d'accordo

13. Continuerei ad usare CoachMe anche in futuro

☐☐☐☐☐☐☐

Indica il tuo giudizio

Per valutare l'applicazione ti chiediamo di compilare il seguente questionario. Si tratta di coppie di caratteristiche, in antitesi tra loro, che il prodotto può avere.

In questa valutazione di esempio si dichiara di stimare il prodotto più attraente che non attraente.

Esempio:

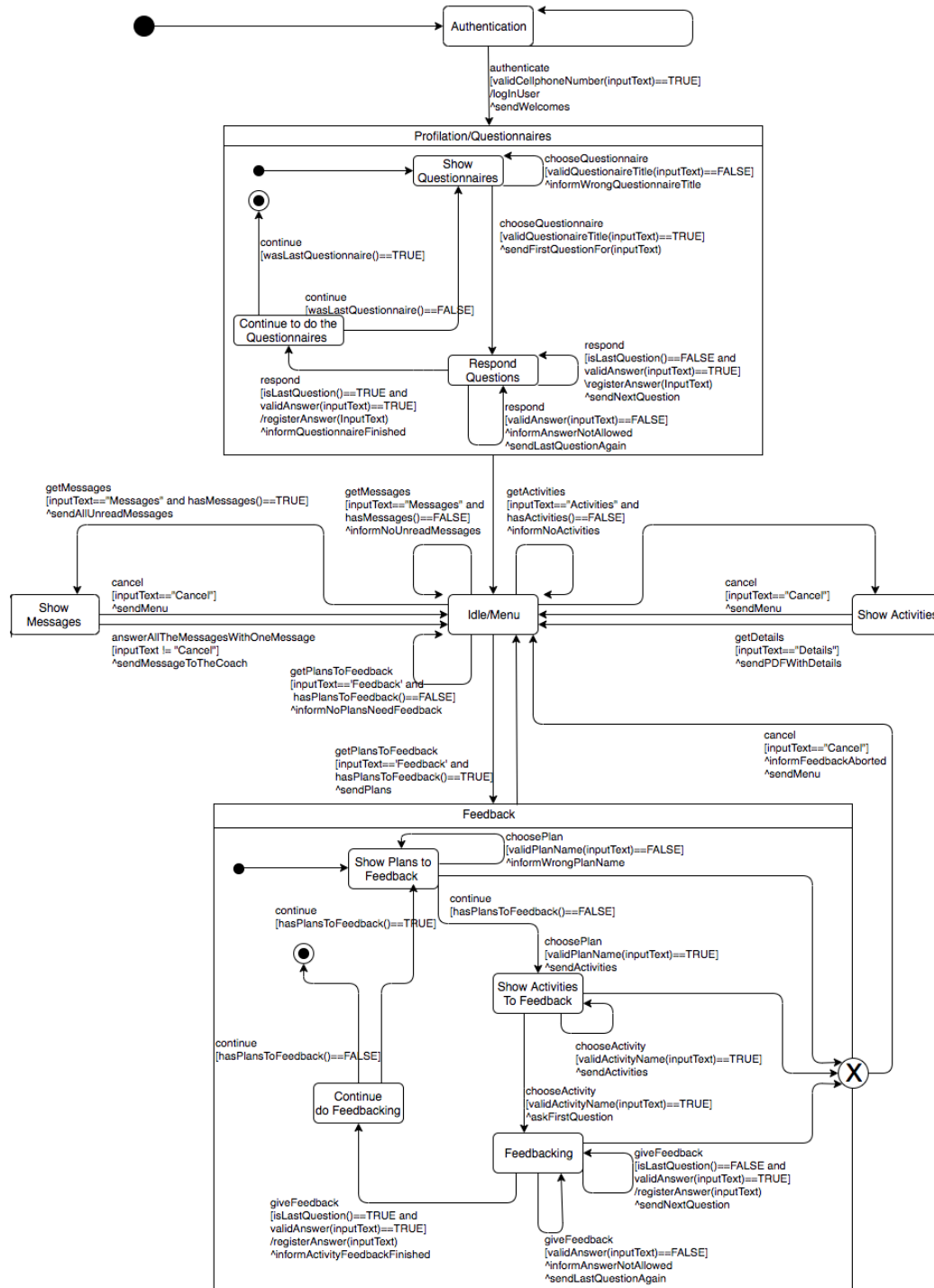
attraente	☐	☒	☐	☐	☐	☐	☐	non attraente
-----------	-----------------------	----------------------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	---------------

Scegli in modo spontaneo senza troppi ragionamenti e ricorda che non esiste una risposta "giusta" o "sbagliata", ma è importante la tua opinione personale

Ora inserisci la tua valutazione del prodotto. Per ogni coppia di aggettivi è possibile dare una sola risposta.

	1	2	3	4	5	6	7		
sgradito	☐	☐	☐	☐	☐	☐	☐	gradito	1
incomprensibile	☐	☐	☐	☐	☐	☐	☐	comprensibile	2
creativo	☐	☐	☐	☐	☐	☐	☐	privo di fantasia	3
facile da capire	☐	☐	☐	☐	☐	☐	☐	difficile da capire	4
di grande valore	☐	☐	☐	☐	☐	☐	☐	di poco valore	5
noioso	☐	☐	☐	☐	☐	☐	☐	appassionante	6
non interessante	☐	☐	☐	☐	☐	☐	☐	interessante	7
imprevedibile	☐	☐	☐	☐	☐	☐	☐	prevedibile	8
veloce	☐	☐	☐	☐	☐	☐	☐	lento	9
originale	☐	☐	☐	☐	☐	☐	☐	convenzionale	10
ostruttivo	☐	☐	☐	☐	☐	☐	☐	di supporto	11
bene	☐	☐	☐	☐	☐	☐	☐	male	12
complicato	☐	☐	☐	☐	☐	☐	☐	facile	13
repellente	☐	☐	☐	☐	☐	☐	☐	attraente	14
usuale	☐	☐	☐	☐	☐	☐	☐	moderno	15
sgradevole	☐	☐	☐	☐	☐	☐	☐	piacevole	16
sicuro	☐	☐	☐	☐	☐	☐	☐	insicuro	17
attivante	☐	☐	☐	☐	☐	☐	☐	soporifero	18
rispetta le aspettative	☐	☐	☐	☐	☐	☐	☐	non rispetta le aspettative	19
inefficiente	☐	☐	☐	☐	☐	☐	☐	efficiente	20
chiaro	☐	☐	☐	☐	☐	☐	☐	confuso	21
non pragmatico	☐	☐	☐	☐	☐	☐	☐	pragmatico	22
ordinato	☐	☐	☐	☐	☐	☐	☐	sovraccarico	23
attraente	☐	☐	☐	☐	☐	☐	☐	non attraente	24
simpatico	☐	☐	☐	☐	☐	☐	☐	antipatico	25
conservativo	☐	☐	☐	☐	☐	☐	☐	innovativo	26

A.2 Finite State Machine (FSM) Architecture



A.3 CoachAI Entity Relational Diagram (ERD)

