DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics: Games Engineering

Supporting Actionable Knowledge: A Conversational AI Chat Assistant for Dietary Monitoring

Anton Steuer

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics: Games Engineering

Supporting Actionable Knowledge: A Conversational AI Chat Assistant for Dietary Monitoring

Unterstützung von umsetzbarem Wissen: ein konversationsfähiger KI-Chat-Assistent für die Beobachtung von Ernährungsgewohnheiten

Author:	Anton Steuer
Supervisor:	Prof. Dr. Georg Groh
Advisors:	Monika Wintergerst, Nadja Leipold
Submission Date:	09.02.2022

I confirm that this bachelor's thesis in informatics: games engineering is my own work and I have documented all sources and material used.

Munich, 09.02.2022 Anton Steuer

Acknowledgments

I would like to thank Prof. Dr. Georg Groh for giving me the opportunity to write my Bachelor's Thesis in a subject area combining my interest in health and nutrition with my studies in Informatics: Games Engineering.

A great thank to my advisors Monika Wintergerst and Nadja Leipold for their great support and valuable feedback throughout the development of the prototype and writing of this thesis.

To my family and friends who supported me throughout my studies: Thank you really much for your support! Whenever I was doubting my capabilities, you were there for me and motivated me to keep going.

Special thanks to all study participants for your time and feedback! You have contributed a very important part to this thesis.

Abstract

Overweight and obesity provide a high risk for cardiovascular diseases and type 2 diabetes. However, the percentage of people, who are overweight or obese, continues to rise at a high rate. The main reasons for gaining and maintaining an unhealthy weight are the lack of physical activity and an unhealthy and unbalanced diet. To change a person's dietary behavior, the flaws in the current habits must be identified and replaced by healthy eating habits. Behavior change is a complex process over a long period of time. In this thesis we present fundamental behavior change theories and techniques supporting behavior change. We describe conventional diet and nutrition applications as well as their flaws and barriers and present a different approach to support behavior change by using conversational AI chat-bots. Based on analysis of existing systems, related work, and the theoretical background we present RAINA: a conversational AI chat assistant for dietary monitoring. RAINA implements a coarser approach for tracking food intake by using servings of food groups as measurement instead of precisely counting calories. She provides feedback in form of a Nutrition Pyramid, visualizing the consumed foods in an easily understandable way. In the seven-day user study we conducted, RAINA was perceived well by the participants. Findings indicate that the approach positively impacts a person's awareness about their diet.

Contents

1. Introduction

Cardiovascular diseases, type 2 diabetes, and the increased possibility to suffer from a severe course of infectious diseases, such as COVID-19, are only three examples of health consequences from being overweight or even obese (World Health Organization 2013; World Obesity Federation 2021). The percentage of overweight people is increasing at an alarming rate. According to the WHO, "the worldwide prevalence of obesity nearly tripled between 1975 and 2016" (World Health Organization 2021b). About 1.9 billion (39%) people were overweight in 2016, and the majority of the world's population is living in countries in which overweight and obesity are responsible for more deaths than underweight and hunger (World Health Organization 2021b; World Health Organization 2021a).

Overweight and obesity are not caused by a few unhealthy meals, but by consuming more energy than one needs over an extended period of time. A generally unbalanced diet with incorrect or nonexistent focus on energy and nutrients may lead to an unhealthy weight and nutrient deficiencies. The unhealthy dietary behavior and habits, a person got used to and subsequently integrated into their lifestyle subconsciously, need to be identified and changed, so that patients can regain a healthy weight and thus minimize the risk for the diseases mentioned.

Changing dietary behavior is a long and complex process that is often supported and accompanied by professional nutritionists. However, the ongoing digitalization and development of smartphones as well as wearable devices like smartwatches, open new opportunities to provide and enhance healthcare treatments. Since nutritional advice is expensive and time binding, many people already turn to smartphone applications to monitor and improve their diet. The large number of apps and the associated download figures suggest that there is a great amount of interest in nutrition and diet smartphone apps (Franco, Fallaize, et al. 2016). Although those applications can have a positive effect on weight management and physical activity, they come with certain barriers including a great focus on weight loss and require a lot of effort to entry foods (Vasiloglou, Christodoulidis, et al. 2021; Franco, Fallaize, et al. 2016; Achananuparp, Lim, et al. 2018). This focus on weight loss most of the time is not supported by professional recommendations, but rather with generic diet plans and a goal weight and date the users are setting themselves (Franco, Fallaize, et al. 2016). These barriers discourage many potential users – some even state that they are afraid of developing an eating disorder if they use such apps (Cordeiro, Epstein, et al. 2015).

With the emerging technologies of Artificial Intelligence (AI) and Natural Language Processing (NLP), a new approach is taking shape in recent years: conversational AI chat assistants. Using NLP, users can communicate with those systems via text or speech. Conversational AI chat assistants can be used to promote behavior change: either in combination with healthcare professionals or on their own, by interacting directly with the user without human supervision (Zhang, Oh, et al. 2020). This interactivity enables developers to incorporate empathy in the chat assistants and integrate social support in order to promote behavior change (Milne-Ives, De Cock, et al. 2020). Since chat bots can be integrated in familiar messenger applications, it makes them available at a large scale. They are available at all times and respond with almost no delay, making them a cost-effective and very efficient solution compared to individual nutritionists who are bound to office hours. Especially tasks which can be automated like answering frequently asked questions, shaping knowledge, or providing feedback on a patient's food diary can be handled by a chat assistant.

In this thesis we start by presenting an underlying behavior change theory in health, called the *Transtheoretical Model of health behavior change (TTM)* by Prochaska and Velicer as well as describing frequently used techniques to support behavior change listed in the *Behavior Change Technique Taxonomy (v1)* by Michie, Richardson, et al. Subsequently, we evaluate barriers in conventional nutrition and diet applications and present related work in the domain of chat assistants in nutrition and weight management. Based on the presented basis of background knowledge, we introduce RAINA – a conversational AI chat assistant to support dietary monitoring. After describing the architecture and the implementation of the system, we present the results of our seven-day user study and discuss our findings and limitations.

2. Theoretical Background

Designing a system to support people, who are trying to change their eating behavior, requires a deeper understanding of the underlying theories and methodologies. In this chapter describes the main topics on which this Bachelor's Thesis is based: the theory of behavior change in health, techniques for changing behavior, as well as the concept of the nutrition pyramid by the *Bundeszentrum für Ernährung (BZfE)*[1].

2.1. Health Behavior Change

When talking about changing behavior, in this work's case: balancing one's diet, change is often assessed as an event at a distinct point of time. However, Prochaska and Velicer identified behavior change to be a process over a period in time, which can be divided into stages of change. Based on the seven identified stages, they designed the *Transtheoretical Model of health behavior change (TTM)* (Prochaska and Velicer 1997). Even before talking about the actual change in behavior, Prochaska and Velicer identified three stages over which a person gradually changes their attitude towards changing their behavior.

In the *Precontemplation* phase, there is no intention to change habits and behavior in the next six months. This may be due to insufficient awareness of the health consequences the current lifestyle leads to. Prochaska stated that decision making depends strongly on the personal arguments for and against the change of habits. In the *Precontemplation* phase, the cons for changing the behavior outweigh the pros (Prochaska 2008).

Once there is an equilibrium in the pros and cons, the person transitions to the *Contemplation* phase in which they plan to take action in the next six months, but do not have a precise plan for action just yet. As soon as the person is aiming to change their behavior in the next four weeks, they move onto the *Preparation* phase. From this phase onward the arguments for a behavior change will outweigh the arguments against it (Prochaska 2008).

The phase in which a person belonging to an at-risk population group takes action to change their behavior is called the *Action* phase. People in this stage modified their

[1] `https://www.bzfe.de`

lifestyle significantly in the last half year. Prochaska and Velicer emphasized, that "people must attain a criterion that scientists and professionals agree is sufficient to reduce risks for disease." (Prochaska and Velicer 1997) (page 39). At the end of the *Action* phase less changes to the behavior are made, and people start to subconsciously integrate their newly learned habits into their lives.

As soon as a person stops making any major changes to their lifestyle, they have reached the *Maintenance* phase in which they solely try to prevent relapsing to old behaviors. The behavior change is complete when the person has no desire to fall back into their old behavioral patterns - they have reached the *Termination* stage (Prochaska and Velicer 1997). Prochaska and Velicer state that people who reach *Termination* have "zero temptation and 100% self-efficacy" (Prochaska and Velicer 1997).

The term *self-efficacy* regarding health behavior change describes a "situation-specific confidence people have that they can cope with high-risk situations without relapsing to their unhealthy or high-risk habit." (Prochaska and Velicer 1997) (page 40). Bandura introduced this concept in his work: *Self-efficacy: Toward a unifying theory of behavioral change* (Bandura 1977). Self-efficacy in combination with motivation is often being used in work related to behavior change to assess progress towards learning and implementing healthy habits (Kocielnik, Xiao, et al. 2018; Strong, Parks, et al. 2008).

2.2. Behavior Change Techniques

The effectiveness of changing behavior, especially in the *Action* phase, is influenced by many factors and is a complex process. There are numerous techniques to support behavior change, which Michie, Richardson, et al. clustered into groups and summarized in their *Behavior Change Technique Taxonomy (v1)* (Michie, Richardson, et al. 2013). They identified a total of 93 distinct *Behavior Change Techniques (BCT)* and grouped those into 16 fields. According to Michie, Richardson, et al., it is the "first consensus-based, cross-domain taxonomy of distinct BCTs to be published, with reliability data for the most frequent BCTs." (Michie, Richardson, et al. 2013) (page 19). It enables an accurate replication of BCTs and provides a reliable list of techniques for identifying and assessing intervention content. Behavior Change Techniques can either be delivered by an interventionist e.g., a dietitian or medical professional, or can be self-delivered when for example, using a self-help book (Michie, Richardson, et al. 2013).

Frequently used groups of Behavior Change Techniques are *Goal Setting and Planning*, *Feedback and Monitoring*, *Social Support* and *Shaping Knowledge*. When setting a goal and planning one's behavior, both agreeing on a pre-defined goal and setting an individual goal are valid approaches (Michie, Richardson, et al. 2013). However, in the case of behavior change in health, according to Prochaska and Velicer the change must fulfill the criterion to significantly reduce the risk for disease (Prochaska and Velicer 1997).

In terms of *Feedback and Monitoring* both the behavior itself as well as the outcomes of the behavior can be monitored, with optional feedback. *Social support* can be integrated in form of practical support e.g., having a partner to remind a person to take their medication or emotional support, like taking a friend to the doctor's appointment (Michie, Richardson, et al. 2013). The BCTs in the group of *Shaping Knowledge* are especially important in the *Contemplation* phase to transition to the *Action* phase (Prochaska and Velicer 1997): providing information and shaping a person's knowledge supports the development of pro arguments for a behavior change, which is important to progress through the stages (Prochaska and Velicer 1997; Prochaska 2008).

There are indications, that the more groups of BCTs are used as once, the greater the effect of the intervention will be (Hankonen, Sutton, et al. 2015). However, Hankonen, Sutton, et al.'s findings, of a study observing the impact of BCTs on diabetes-patients, suggest that the BCTs *Goal Setting*, *Goal Review* and *Social Support*, have the greatest impact on BMI-reduction (Hankonen, Sutton, et al. 2015). This is also in line with Strong, Parks, et al. who interviewed college students on their reasons for eating unhealthy food or not exercising. When investigating reasons for weight gain by young adults, the students reported that they were more likely to eat more fruits and vegetables when they were eating with friends e.g., receive some kind of social support or social control (Strong, Parks, et al. 2008). However, they perceived the availability of unhealthy foods to be higher than the availability of fruits and vegetables on their colleges' campus. Furthermore, Strong, Parks, et al. emphasize the importance of self-regulatory skills such as *Goal Setting* as well as *Self-Monitoring*. In addition, social and environmental support can "facilitate adherence and the long-term adoption of healthy behaviors" (Strong, Parks, et al. 2008).

When using *Goal Setting*, *Feedback* and *Self-Monitoring* as Behavior Change Techniques they are often associated with reflection. Baumer, Khovanskaya, et al. identified an implicit assumption that "feedback constitutes reflection" (Baumer, Khovanskaya, et al. 2014), when evaluating works which claim to be designed about reflection. However, most of the time it was assumed that simply providing access to collected data will initiate reflection, without further advice on action itself (Baumer, Khovanskaya, et al. 2014). Although Baumer, Khovanskaya, et al. identified a lack of specific advice on future actions, this might not even be necessary to trigger reflection and change a persons behavior.

The so called *mere-measurement effect* describes a phenomenon in which only asking about an intention or behavior influences a person's behavior (Williams, Block, et al. 2004; Sherman 1980). Williams, Block, et al. found out that asking a person questions about their behaviors, increases both healthy and unhealthy behaviors (Williams, Block, et al. 2006). This suggests that asking questions about a person's behavior might be a sufficient approach to persuade users subconsciously to improve their eating habits.

2.3. Dietary Monitoring with the Nutrition Pyramid

Monitoring one's diet is a crucial part for changing dietary behaviors. On the one hand, when preparing to take action, the current behavior has to be observed to be able to identify possible flaws and act accordingly. On the other hand, being aware of what foods have been consumed in a day can help persons maintain their newly learned habits. Therefore, dietary monitoring is a powerful tool in the *Preparation, Action* and *Maintenance* phase of the transtheoretical model of behavior change (Prochaska and Velicer 1997).

Keeping a food diary has become an easy task, supported by many diet-tracking apps. According to Statista, 31% of adults interviewed in 2017 stated to have been using or are using apps for tracking their diet. 41% mentioned they are not using such apps just yet but could think of using them in the future (Statista 2017). Applications for tracking foods have a high accuracy of nutrient values, but require a very precise input of the foods and amounts consumed, which results in a barrier for many potential users. This as well as other barriers will be described in section 3.1.

The *Bundeszentrum für Ernährung (BZfE)* has a different and broader approach for implementing a healthy diet in everyday life, without weighing ingredients and dishes: the *BZfE Nutrition Pyramid* (Seitz 2020; Brüggemann 2018). A pyramid is commonly used to visualize the distribution of food groups: the lower the recommended amount of food, the higher in the pyramid it will be displayed (Willett 2005; Nutrition Australia 2014; Seitz 2020; Brüggemann 2018). However, the pyramids may vary slightly, since there will be new information discovered regarding nutrition constantly (Willett 2005). The amounts of the foods are measured in *servings* most of the time, to make the pyramid easy to remember and integrate in the daily diet (Flothkötter 2021; Brüggemann 2018; Nutrition Australia 2014). In contrast to the *Healthy Eating Pyramid* by Nutrition Australia (Figure 2.1), the *BZfE Nutrition Pyramid* (Figure 2.2) visualizes the servings directly in the graphic of the pyramid in form of boxes, with an icon representing each of the food groups.

Comparing Figure 2.1 and Figure 2.2, the rankings of the different foods in the pyramids are almost identical. The beverage of choice should always be water, which is integrated in the *BZfE Nutrition Pyramid* in Figure 2.2 as an explicit field, while Nutrition Australia used it as side information with no precise number of servings (Nutrition Australia 2014). The foundation for a healthy diet in both cases are fruits and vegetables, followed by sources for carbohydrates such as bread, rice or pasta. Animal

Figure 2.1.: Healthy Eating Pyramid by Nutrition Australia (retrieved from `https://nutritionaustralia.org/fact-sheets/healthy-eating-pyramid`, accessed: 08.01.2022)

Figure 2.2.: Nutrition Pyramid by the Bundeszentrum für Ernährung (BZfE) (retrieved from: `https://www.bzfe.de/ernaehrung/die-ernaehrungspyramide/die-ernaehrungspyramide-eine-fuer-alle`, accessed: 08.01.2022)

products in form of dairy or meat as well as alternatives for vegetarians or vegans are listed more to the top of the pyramids. Furthermore, the BZfE specifies to consume more servings of dairy products into the daily diet than meat, while Nutrition Australia does not recommend a specific number of servings, but rather a coarse overview of the general composition of a balanced diet. On the top of the pyramid are healthy fats, which the BZfE splits into oils like olive- or nut-oils and fats such as butter or margarine. In the Nutrition Pyramid by the BZfE there is also a field for sweets on the top, which Nutrition Australia did not list specifically.

As mentioned before, the food pyramids usually do not display the number of servings directly in the graphic and do not describe the size of the servings in an easy way to integrate them into the daily diet (Willett 2005; Nutrition Australia 2014). Instead of using grams or milliliters, the BZfE makes use of a person's hands as a tool for measuring food (Brüggemann 2018; BZfE 2020). This approach allows the pyramid to be applicable to persons of all age-groups and sizes, since the hands represent each individual's serving sizes. For measuring a portion of fruits, vegetables, carbohydrates such as rice or pasta, both hands are being used: both hands are put together to form a "bowl". The amount that fit into this "bowl" represents one serving. Liquids, like water, juices or milk, are measured in a glass, since one can hold it with a hand: one glass equals one field in the pyramid for the corresponding food group. A portion of bread or cheese equals the area of one hand with its fingers, whereas a serving of meat can be placed on the palm of the hand. The recommendation by the BZfE for oils and fats, is a combined amount of two to three tablespoons per day. For the sweets or extras on the top of the pyramid, the amount which fits in one hand represents one serving. A glass of an alcoholic beverage or salty snacks such as chips or fries also fall into the category of extras (Brüggemann 2018; BZfE 2021; BZfE 2020).

We decided to use the *BZfE Nutrition Pyramid* as the basis for our proof-of-concept chat assistant for the following reasons: Firstly, the visual representation of one's diet in the form of a pyramid is easy to understand without any prior knowledge about nutrition and the image is memorable. Secondly, the approach of tracking servings of the specified food groups is easier to integrate into everyday life than counting calories, which will be described in more detail in section 3.1. Last but not least, the measurement using your own hands scales applies and adapts to each person, independently from age, gender or size. In the next chapter, we present conventional diet and nutrition applications and their drawbacks and introduce related work in the domain of chat assistants supporting behavior change.

3. Related Work

Supporting diet behavior change is a complex and protracted endeavor. Professional nutritionists for individual or group nutrition counseling sessions are expensive and not accessible to every population group. The trend of *mobile Health (mHealth)* systems offers a promising opportunity to provide cost efficient and effective health services at a large scale. Especially in developing countries the healthcare systems are overloaded, with too few medical professionals for a large population (Latif, Rana, et al. 2017). Mobile health systems have the potential to relieve the health care systems and medical professionals in developing countries (Latif, Rana, et al. 2017).

In this section, we describe mHealth systems in form of convectional smartphone applications as well as the AI chat assistant approach in the domain of nutrition behavior change. At this point, it is important to notice that there is a variety of other health sectors in which mHealth systems are being used to support behavior change, such as fighting an addiction, managing stress, or integrating physical activity into daily life (Calvaresi, Calbimonte, et al. 2019; Park, Choi, et al. 2019; Luo, Aguilera, et al. 2021). Last of which is often used in combination with a behavior change in diet to support overall weight loss (Maher, Davis, et al. 2020; Hankonen, Sutton, et al. 2015).

3.1. Conventional Nutrition Apps and their Drawbacks

The large number of nutrition related mobile applications and the high amount of downloads indicate a great interest in the usage of apps to monitor and change ones diet (Franco, Fallaize, et al. 2016). Health education via the smartphone as a medium is cheap and efficient, since in today's society almost every person owns a device (Latif, Rana, et al. 2017). On the one hand for example, the elderly population, which is generally not as used to smartphones as younger generations, is less likely to use those apps, than persons with a higher technical understanding (Vasiloglou, Christodoulidis, et al. 2021). On the other hand, adolescents tend to assess health and nutrition apps as being "the last resort to staying healthy" and "were meant only for adults" (Chan, Kow, et al. 2017) (page 51). However, according to Chan, Kow, et al., teenagers who use applications to manage a medical condition, assess them as very helpful (Chan, Kow, et al. 2017).

Recent study findings by West, Belvedere, et al. suggest, that the usage of applications, related to diet and nutrition, support the change of behavior by increasing the motivation to eat a healthy diet as well as the nutrition self-efficacy (West, Belvedere, et al. 2017). Participants also reported an increase in their knowledge on how a healthy diet can be achieved as well as formation of awareness regarding the benefits of a healthy diet (West, Belvedere, et al. 2017).

Nutrition apps like *MyFitnessPal*[1], *YAZIO*[2] or *Fddb*[3] usually have similar functionality: they use huge databases with food data i.e., amount of calories and distribution of macro-/micro-nutrients, out of which the user has to find the consumed food, and insert it into the food diary. This is done via text search or bar-code scanning to find the food and text input to specify the portion size in a metric unit like grams or milliliters (Franco, Fallaize, et al. 2016). The apps are also capable of logging physical activity, either via manual input of the exercise or by tracking it directly with sensors of the phone or connected wearable devices like smart watches. Taking the energy expenditure (i.e., physical activity) and the energy intake (i.e., logged foods) into account, the systems are able to precisely calculate the energy balance, given the user input is correct. Most of the time there is a focus on weight loss, which means the caloric intake should be lower than the expenditure (Vasiloglou, Christodoulidis, et al. 2021; Franco, Fallaize, et al. 2016; Achananuparp, Lim, et al. 2018).

This focus on weight loss can be problematic, as Franco, Fallaize, et al. reported: out of the nine tested apps, only one had a limit when it comes to the rate of losing weight or reaching a person's target weight (Franco, Fallaize, et al. 2016). For users without prior knowledge of nutrition, this might encourage excessive training and unhealthy diets up to an eating disorder (Chan, Kow, et al. 2017; Vasiloglou, Christodoulidis, et al. 2021). Vasiloglou, Christodoulidis, et al.'s findings suggest, that more people of the normal- and underweight group use nutrition applications than people of the overweight or obese group (Vasiloglou, Christodoulidis, et al. 2021). The task of counting calories is assessed as tiring and taking much time and effort, which presents a barrier for average smartphone users to download and use nutrition applications (Cordeiro, Epstein, et al. 2015). Especially regional foods are often not covered in the databases of the apps and the calories or portion sizes are not estimated correctly, which are criticisms frequently reported by users (Vasiloglou, Christodoulidis, et al. 2021). Although, the databases are often writable, allowing every user to insert new foods, dishes, and products; the entries are not checked for errors. Packaged foods on the other hand, are easy to enter since scanning the bar-code is a matter of seconds and the nutrients are always the same, unlike a self-prepared meal (Cordeiro, Epstein, et al. 2015).

[1] `https://www.myfitnesspal.com`
[2] `https://www.yazio.com`
[3] `https://fddb.info`

Achananuparp, Lim, et al. analyzed data from MyFitnessPal users to investigate their eating behaviors. Their findings show, that despite the tracking of food intake, the overall diet is not significantly healthier than the diet of the general public (Achananuparp, Lim, et al. 2018). The analysis of average caloric intake suggested that the users were dieting, but the choices of foods did not meet the dietary recommendations by professionals (Achananuparp, Lim, et al. 2018). A reason for this would be that the tracking of calories does not take the quality of foods into account and packages and pre-made foods are easier to enter into the food journal (Achananuparp, Lim, et al. 2018; Franco, Fallaize, et al. 2016; Cordeiro, Epstein, et al. 2015).

Research agrees on the importance of learning how a healthy diet can be put together, without the major points of focus being calorie counting, energy balance and weight loss (Franco, Fallaize, et al. 2016; West, Belvedere, et al. 2017; Chan, Kow, et al. 2017). Developers should integrate "design configurations that emphasize the provision of knowledge to shape attitudes and beliefs" (West, Belvedere, et al. 2017) (page 8) into their applications to support behavior change in nutrition (West, Belvedere, et al. 2017). Chan, Kow, et al. state that interactivity and personalization can increase the appeal of nutrition applications to adolescents. A specific example are virtual coaches with personalized exercises and dietary recommendations (Chan, Kow, et al. 2017).

3.2. Supporting Behavior Change with AI Assistants

According to West, Belvedere, et al., smartphone nutrition and diet applications focusing on increasing motivation, self-efficacy, goal-setting abilities and shaping knowledge worked the best when changing behavior (West, Belvedere, et al. 2017).

In contrast to the conventional nutrition apps, there is a different approach to support behavior change in health-related topics: virtual coaches and AI assistants. The most popular areas of application are physical and mental health as well as nutritional and metabolic disorders (Pereira and Díaz 2019). While there are chat bots for supporting smoking cessation (Perski, Crane, et al. 2019; Calvaresi, Calbimonte, et al. 2019) and stress management (Park, Choi, et al. 2019), this thesis focuses on behavior change in nutrition and weight management. Behavior change in nutrition is often combined with support for increasing physical activity in existing systems since they are equally important for maintaining a healthy weight (Maher, Davis, et al. 2020). In this section related work and prototypes for virtual assistants in form of AI chat bots are introduced and described, to provide an overview of the current research in the field of behavior change in weight management.

This approach of integrating conversational agents in the domain of behavior change is making use of emerging technologies including *Natural Language Processing (NLP)*, the *Internet of Things (IoT)* and *Artificial Intelligence (AI)*. Firstly, IoT and AI provide the possibility to collect large amounts of data using for example wearables such smart watches, analyze the data and provide personalized interventions and recommendations (Maher, Davis, et al. 2020; Kocielnik, Xiao, et al. 2018; El Kamali, Angelini, et al. 2018). Secondly, NLP enables chat bot systems to interact with users directly via speech or text, like in familiar messenger apps instead of selecting pre-defined messages to send to the bot (Pereira and Díaz 2019). On the on hand, NLP requires a large amount of training data for the system to understand free text input. Poor understanding of input, caused by limited knowledge of vocabulary or voice recognition is a commonly reported issue when using conversational agents (Nadarzynski, Miles, et al. 2019; Milne-Ives, De Cock, et al. 2020). On the other hand, pre-defined messages limit the freedom of user input and the interactivity of the systems. In addition to interactivity, empathy and personality of the chat assistant are often reported to be liked by users (Milne-Ives, De Cock, et al. 2020). A familiar form of communication via text or speech can be used by a larger group of users since no learning of how to use a program or app is required. Research suggests that this approach is also likely to be adopted by an older population group, which are not digital natives (Maher, Davis, et al. 2020). In combination with a personality, which can be designed by the developers, the chat assistant approach opens the possibility to integrate additional behavior change techniques compared to static nutrition applications and food diaries. Those BCTs include emotional support paired with empathy, which are perceived as positive by users (Milne-Ives, De Cock, et al. 2020; Pereira and Díaz 2019).

According to Ho, Hancock, et al. opening up about personal information, has positive effects on psychological outcomes, regardless of the perceived identity of the conversation partner i.e., a conversational agent or a human (Ho, Hancock, et al. 2018). Those findings are in line with the *Computer are Social Actors (CASA)* paradigm by Nass, Steuer, et al., which states that users "[...] apply social rules to their interaction with computers, even though they report that such attributions are inappropriate." (Nass, Steuer, et al. 1994) (page 77). In contrast to the *CASA*-paradigm the *uncanny valley effect (UVE)* described by Mori, Macdorman, et al. argues that if a robot appears too human like, this might cause discomfort (Mori, Macdorman, et al. 2012). While Mori, Macdorman, et al. use the appearance of a robot as an example, the behavior of a chat assistant can be assessed in the same way (Ta, Griffith, et al. 2020).

Zhang, Oh, et al. findings suggest that users respond best to AI chat assistants if their identity is clearly presented. It is important to find a balance in human-like behavior in order to develop a conversational agent which mimics empathy and provides social support, but without triggering the *UVE* (Zhang, Oh, et al. 2020). This is particularly important in the later phases of the transtheoretical model of behavior change for maintaining the change in behavior a person learned in the action phase (Prochaska and Velicer 1997). Gentner, Neitzel, et al. state that the emotions and behavior of the chat assistant are equally important to the actual advice it provides to the users. A suggestion to implement emotions in chat assistants is including emojis, photos or videos in responses (Gentner, Neitzel, et al. 2020).

In addition to social support, frequently used behavior change techniques in conversation AI chat assistants include *Goal Setting*, *Feedback*, *Self-Monitoring* and *Shaping of Knowledge* (Luo, Aguilera, et al. 2021). Chat assistants can send personalized motivational messages or exercise and diet tips as interventions (Prasetyo, Achananuparp, et al. 2020; Zhang, Oh, et al. 2020). El Kamali, Angelini, et al. presented a system design with the goal to engage the user in "emotionally rich conversations". The *NESTORE e-Coach* should be a system, tailored to users of an older age to be a source for information on diet and physical activity as well as social and cognitive abilities. The architecture proposed included a so-called *Tangible Coach* which should be placed at home to which the user is likely to establish an emotional connection. As an example El Kamali, Angelini, et al. used a "pet plant" as an embodiment of the e-Coach. The *Tangible Coach* should communicate via voice control if the user is in the near environment and use a chat bot if the user is not at home. The user would then subconsciously connect all messages and interventions received by the chat bot to its physical embodiment (El Kamali, Angelini, et al. 2018). This approach is making use of the IoT by building a system with multiple devices collecting data and communicating with the user.

Using external tools, such as smart watches is a common approach to collect user data without explicit input. The *Reflection Companion*, developed by Kocielnik, Xiao, et al., uses a Fitbit[4] activity tracker to collect physical activity data from the user. This data is then aggregated and presented to the user combined with questions triggering reflection. Reflection leads to benefits such as increased motivation and mindfulness. Participants reported that typing their answers to the reflective questions felt like someone was reading the answers, even though it was just an artificial intelligence, which led to being more conscious about their behavior (Kocielnik, Xiao, et al. 2018).

However, the Reflection Companion did not have goal setting for physical activity explicitly implemented in the system, rather than reflecting over raw collected data. Since all participants were Fitbit users, the assumption can be made that they have already set personal step and activity goals using the Fitbit App. When looking at nutrition assistants, they are often more explicit when it comes to goal setting. A frequently used approach is to pre-set goals for users, using evidence-based recommendations on which foods and how much of certain foods should be consumed.

[4] https://www.fitbit.com

According to Michie, Richardson, et al. "BCTs can be delivered by an 'interventionist' or self-delivered" (Michie, Richardson, et al. 2013) (appendix: "BCT Taxonomy (v1): 93 hierarchically-clustered techniques", page 1). In the domain of behavior change in nutrition an *interventionist* might be a medical professional, nutritionist or in this case: a chat assistant. Most chat bots deliver a broad range of behavior change techniques by themselves acting as the interventionist.

Prasetyo, Achananuparp, et al. presented *Foodboot* who sends the user so-called *JIT Interventions (just-in-time Interventions)*, reminding the user of their dietary goals before their meals to plan accordingly if the current intake deviates from pre-set goal by a certain threshold (Prasetyo, Achananuparp, et al. 2020). Unlike conventional nutrition apps, which are designed for tracking nutrients as precise as possible, several systems make take a different approach by using the number of servings as a measurement for food intake instead. Examples for this approach are *FoodBot* by Prasetyo, Achananuparp, et al., *Rupert le nutritionniste* by Casas, Mugellini, et al., and *Paola* by Maher, Davis, et al. While *Foodbot* and *Paola* use evidence-based recommendations for the amount of servings as default goals, Casas, Mugellini, et al.'s system lets the user select their goal themselves, based on their current self-assessed consumption of meat or fruits and vegetables. The users have the possibility of either reducing their intake of meat or increasing their intake of fruits and vegetables. Casas, Mugellini, et al. reported, that only 11% of the study participants completed their goals, stating the goals were either too ambitious or the users lacked motivation. Nonetheless, 65% of users improved their consumption of the selected foods and 70% assessed the system as more efficient than precise nutrient tracking (Casas, Mugellini, et al. 2018). These findings suggest that guidance when setting goals might be useful, if not even fully pre-set goals according to state-of-the-art dietary recommendations.

The chat systems *CoachAI* developed by Fadhil, Schiavo, et al. and *Tess* designed by X2AI[5] pair the use of chat assistants with healthcare and medical professionals. *Tess* addresses mental health such as depression or anxiety via self-help chats. Fulmer, Joerin, et al.'s findings show that *Tess* reduces symptoms of depression and anxiety significantly (Fulmer, Joerin, et al. 2018). Stephens, Joerin, et al. integrated *Tess* in behavioral coaching for adolescents with obesity symptoms. For this study, *Tess* has been customized to the domain of pediatric weight management: rather than focusing solely on mental health, recommended behavioral counselling methods have been included in the system as well. The majority of conversations (73.6%) were initiated by Tess and users reported her to be helpful 96% of the time, suggesting the AI chat assistant approach being feasible for younger patients (Stephens, Joerin, et al. 2019).

[5] `https://www.x2ai.com`

CoachAI uses a chat bot as tool for delivering interventions by coaches, rather than the chat assistant being the main interventionist. The chat bot collects user data regarding nutrition, physical activity, and general state of well-being, creates user profiles and clusters the users into groups. The coaches can view their clients' data on a web interface and assign them activities and questionnaires, which the chat bot will deliver. The chat assistant automatically handles tasks such as collecting feedback on the activities the coach assigned (Fadhil, Schiavo, et al. 2019).

Tess and CoachAI are both not designed to replace a human professional, but rather relive personnel from tasks that can be automated. As described, chat assistants can deliver social support and support patients in changing their behaviors. Either by enhancing professional coaching sessions, by being reachable out of office hours, or delivering interventions fully automated, chat assistants provide a great opportunity to develop an efficient addition to current healthcare offers at a large scale.

However, there are some limitations in existing systems. Firstly, there are issues with the chat assistants themselves: on the one hand, NLP requires a large set of training data to develop a system that fully understands the users. Especially in proof-of-concept studies, this results in participants reporting poor understanding of intends by the chat assistants (Milne-Ives, De Cock, et al. 2020; Nadarzynski, Miles, et al. 2019). On the other hand, pre-defined answers for the users to select from, limit the freedom of input (Luo, Aguilera, et al. 2021). Secondly, there are concerns regarding the accuracy of the systems or security concerns when disclosing personal information (Nadarzynski, Miles, et al. 2019; Luo, Aguilera, et al. 2021).

Based on the showcased theoretical background to behavior change, diet tracking and self-monitoring, the limitations of current nutrition and diet apps, and the analyzed existing systems and related work, we developed a proof-of-concept chat bot. The *Real-time Artificial Intelligent Nutrition Assistant* (RAINA) implements the basic functionality to track a user's nutrition using the BZfE nutrition pyramid (Brüggemann 2018). The system addresses users in the *Preparation*, *Action* and *Maintenance* phase of the TTM and makes use of several BCTs: *Goal Setting and Planning*, *Feedback and Monitoring*, *Shaping Knowledge* and *Social Support*. RAINA is designed to support the actionable knowledge of integrating the nutrition pyramid in everyday dietary planning. She represents a smart, caring and motivating virtual dietary assistant. On the one hand, she provides credible sources to self-teach the nutrition pyramid, making her usable on her own. On the other hand, she can be used in combination with professional nutrition counselling to deepen the understanding of the nutrition pyramid and apply the learned knowledge. In the next section, the implementation of RAINA is described, followed by an evaluation of the system during a seven-day user-study.

4. Implementation

RAINA was built using the Rasa Framework[1] as back-end and the messenger app Telegram[2] as the channel for communication. Telegram allows easy and platform independent access to RAINA since it is widely supported by mobile and desktop operating systems.

When starting a new conversation with RAINA, she will send the user a link to the materials by the BZfE[3]. After the user finished reading the materials, RAINA will ask four questions to recall the learned knowledge. If a user did not understand a topic, RAINA will explain it briefly. At the end of the initial dialogue, RAINA will ask the user to add a portion of water and asks the user to retrieve the pyramid. We implemented those two messages as exercises for the user, since those two functions will be the most frequently used when interacting with RAINA to track the diet. RAINA generates an image, based on the original *BZfE Nutrition Pyramid* (Figure 2.2), to provide visual feedback on the consumed foods. After the initial dialogue, the user can text RAINA at any time to add consumed servings of food to the pyramid. In the background, the user is added to the database to be eligible for notifications, after the initial dialogue has been finished. This operation is performed after the dialogue to not interrupt the tutorial with a scheduled message.

In addition to the basic functionality of adding and subtracting portions from the pyramid, RAINA is also able to retrieve the pyramid for a specific date or to calculate the average pyramid over a given timeframe. Furthermore, RAINA reaches out to the users by herself twice a day. The goal is to motivate them and remind them to enter all of their consumed foods into the pyramid: Firstly, RAINA wishes the users a good morning at 8 a.m. and asks them how motivated they are to start the day. Secondly, RAINA sends the users their pyramid of the day in the evening and asks them to reflect on their nutrition up to this point of the day. Furthermore, she incentivizes them to cook themselves a dinner based on the remaining fields of the pyramid (Figure 4.4). Moreover, every Sunday, RAINA sends the users a weekly overview containing the average pyramid over the last seven days and asks them to reflect on their nutrition

[1] `https://rasa.com`

[2] `https://telegram.org`

[3] `https://www.bzfe.de/ernaehrung/die-ernaehrungspyramide/die-ernaehrungspyramide-eine-fuer-alle`

for the past week. Depending on the answers to the questions, RAINA replies with a message to either cheer up the user, or to praise the user's performance.

Furthermore, RAINA is able to answer simple questions about the food groups, such as which foods belong to a certain group or how big a serving of a specific food group is. General questions about the pyramid e.g., what names should be used to address the fields in the pyramid or on which sources RAINA is based, are implemented as well. To support the user experience, one can always ask the assistant directly which functions she supports and how to execute them, instead of having to open the PDF file with instructions which has been sent to the study participants.

The system architecture is shown in Figure 4.1 with the main components being the *Rasa Open Source*[4] machine learning framework and the *Rasa Action Server*[5]. Additionally, an *Amazon Web Services Simple Storage Service (S3)*[6] has been integrated to store images and send them via attachments with the assistant's messages.

4.1. Rasa Open Source

For the chat assistant to understand the intentions of the user, Rasa Open Source uses *Intents* which describe requests by a user and *Actions* which describe the assistants reactions to the requests. There are to components which determine the chat assistant's reaction to a given user input: the *Natural Language Understanding (NLU)* component as well as the *Dialogue Policies* (Rasa Technologies 2021b). As the name suggests the *NLU* component performs the task of transforming a user's messages to *Intents* which can be further processed by the system. The *Dialogue Policies* define the chat assistants behavior i.e., which *Actions* follow on the detected *Intent*.

In this section the configuration to train RAINA is being described, with the most important components being explained in detail. In Figure 4.1 this corresponds to the *Rasa Open Source* container.

4.1.1. Natural Language Understanding

The *NLU* component is being trained using a fully customizable pipeline[7]. In this pipeline, the user's message is passed through a *Tokenizer* and multiple *Featurizers* to match the message to an *Intent*. After the training is finished and the *NLU* unit receives a message, the message is passed through the pipeline like the training examples. The

[4]`https://rasa.com/docs/rasa`
[5]`https://rasa.com/docs/action-server`
[6]`https://aws.amazon.com/de/s3`
[7]`https://rasa.com/docs/rasa/model-configuration`

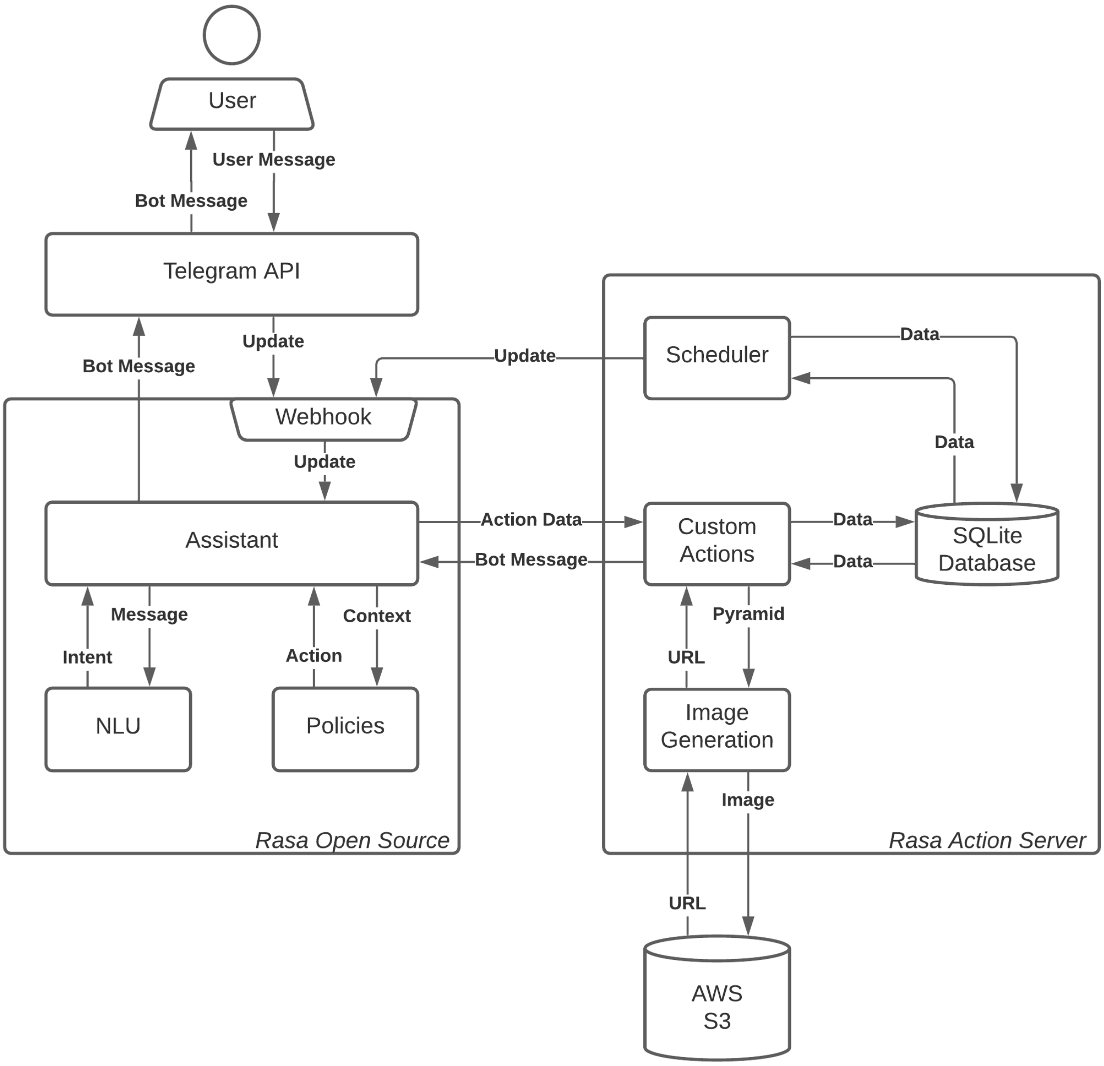


Figure 4.1.: System Architecture

NLU component then predicts an *Intent*, to which the message most likely matches and returns it together with the confidence of the prediction.

For the prototype of this thesis, the recommended default pipeline has been used containing the *WhitespaceTokenizer*[8], *RegexFeaturizer*[9], *LexicalSyntacticFeaturizer*[10] and the *CountVectorsFeaturizer*[11]. The *WhitespaceTokenizer* parses the initial message into tokens by separating the words at the whitespaces. The following chain of featurizers builds a vector containing features based on the tokens. Those feature vectors are then used by the *DIETClassifier*[12] to match then initial message to an *Intent*, which is known by the system. The more message examples exist for an *Intent*, the more feature vectors are known for this *Intent*. This increases the chance for a slightly new wording, describing a known *Intent*, to be recognized and matched by the *DIETClassifier*.

4.1.2. Conversation Policies

A Rasa assistant's predictions for the actions to perform next, based on an *Intent* and a conversation's context, are trained by providing so-called *Stories* and *Rules*. Both forms of training data are built by writing sequences of *Intents* and *Actions* to represent a conversation.

For loose dialogues, a user might have with the assistant, *Stories* are provided describing the expected behavior of the chat assistant in possible conversation scenarios. Based on those *Stories*, the model is being trained to predict the next action, taking the last *Intent* and the conversation's context into account. However, there are certain messages to which the system should always react the same way. For example: when users asks if they are talking to a bot, the bot should always reply with *"yes"*, independently from the context of the conversation.

To train the *Action* prediction of RAINA, the following *Conversation Policies* have been used: the *Memoization Policy*[13] to train the model based on *Stories* and the *RulePolicy*[14] to implement the given *Rules*.

4.1.3. Entity Extraction

After the assistant identifies an *Intent*, *Entities* are used for pieces of information which should be further processed. Rasa Open Source provides a large selection of *NLU*

[8] `https://rasa.com/docs/rasa/components#whitespacetokenizer`
[9] `https://rasa.com/docs/rasa/components#regexfeaturizer`
[10] `https://rasa.com/docs/rasa/components#lexicalsyntacticfeaturizer`
[11] `https://rasa.com/docs/rasa/components#countvectorsfeaturizer`
[12] `https://rasa.com/docs/rasa/components#dietclassifier`
[13] `https://rasa.com/docs/rasa/policies#memoization-policy`
[14] `https://rasa.com/docs/rasa/policies#rule-policy`

pipeline components to extract *Entities* from user messages, each with an individual use case.

For the system implemented in this thesis, three different types of *Entities* are extracted from the users' messages: *"number"* and *"pyramid_field"* are used to identify the amount and the food group the users are referring to. The *Entity "time"* is being used to fetch pyramids from the past as well as to calculate the average pyramid over a given period of time.

When a user message is passed through the Rasa *NLU* pipeline and a known *Entity* is being identified, it will be stored in the the conversation *Tracker*[15] and can be accessed in a custom *Action*, implemented in the *Rasa Action Server*. For the prototype the *Entity* extractors *DucklingEntityExtractor*[16] and *DIETClassifier*[17] have been added to the pipeline.

DucklingEntityExtractor

The *Entities "number"* and *"time"* are frequently used by chat assistants and are standard use cases for *Entity* extraction. Therefore, there is a dedicated *Entity* extractor for those types called *DucklingEntityExtractor*. Although it has been configured to use German as language, there are a few issues which had to be handled manually in the custom *Actions*.

On the one hand, *Duckling* struggled with the German grammar: when extracting numbers Duckling does not identify *"[...] eine Portion Obst [...]"* (*"[...] one serving of fruits [...]"*) as one serving but does not extract any number at all. This is due to the fact, that the number *"one* in German on its own is written as *"eins"*, which *Duckling* is able to understand. However, when used in a sentence to add one serving, the German grammar changes *"ein**s**"* to *"ein**e**"*, which Duckling does not identify as a number. Since this is only the case for a single number, this exception will be caught in the custom *Action* adding, respectively removing, serving. In this case the value will be set to one as default number of servings.

One the other hand, the German notation of decimal numbers uses a comma (*"[...] 2,5 Portionen Obst [...]"*) instead of a point (*"[...] 2.5 portions of fruits."*) as in the English notation. Using a point instead results in Duckling recognizing a date instead of a number and causing yet another exception. For the proof-of-concept implementation, this exception is also being caught and handled by defaulting to the value *"one"*.

[15] `https://rasa.com/docs/action-server/sdk-tracker`
[16] `https://rasa.com/docs/rasa/components#ducklingentityextractor`
[17] `https://rasa.com/docs/rasa/components#dietclassifier-1`

DIETClassifier

The *Dual Intent and Entity Transformer (DIET)* has been developed by Rasa Technologies and is being used for two areas of application. As the name suggests, the *DIETClassifier* is being used to identify *Intents* (subsection 4.1.1) and to extract *Entities* as well. Using the DIETClassifier enables developers to use custom *Entity* types and train the model accordingly. In this thesis, the *Entity* type *"pyramid_field"* is being extracted using the DIETClassifier. The model is trained so that the *Entity* type can adopt the value of the food groups represented in the pyramid. The value of the *Entity* will then be used to access the field of the pyramid in the database (subsection 4.2.1).

Synonyms

For the assistant to understand a variety of words describing the food groups themselves as well as some commonly used products in the food groups, *Synonyms*[18] are being used. The model has been trained to recognize 333 synonyms for the nine different food groups.

Firstly, the synonyms contain paraphrases and different spellings for the names of the food groups. Secondly, specific products and foods are also being handled such as different side dishes specified by the BZfE, which belong to the food group of carbohydrates like bread, rice, or potatoes. This allows users to add some specific foods directly, instead of using the title of the food group respectively pyramid field.

The *EntitySynonymMapper*[19] maps extracted *Synonyms* onto the specified *Entity* value specified in the training data, after extracting the *Synonyms* from the message. For example: If a user decides to add an apple, the assistant recognizes *"apple"* as *Synonym*, maps it to the value *"fruits"* and sets the value of the *Entity* type *"pyramid_field"* to *"fruits"*, which is then stored in the Tracker. The *Rasa Action Server* then adds a portion to the database for the field *"fruits"* in the corresponding column of the *pyramids* table.

4.1.4. Fallback Behavior

The Rasa NLU component is usually not being trained solely on sample messages written by the developers. Since it is impossible for a small group of people to brainstorm all different wordings for an *Intent*, real world examples from test conversations are needed to get a better result in regard to natural language understanding. The process of developing a conversational AI by repeatedly collecting user data, adding it to the training data, and retraining the model is called *Conversation Driven Development* (Rasa

[18] `https://rasa.com/docs/rasa/nlu-training-data#synonyms`
[19] `https://rasa.com/docs/rasa/components#entitysynonymmapper`

Technologies 2021a). It takes several cycles and continuous integration of training data to train the assistant to a point where it understands most expected user inputs for its field of application.

However, there will always be messages the *NLU* cannot to match to an *Intent*. If this is the case and the chat assistant is not confident to a certain level what the user means, it will ask the user to rephrase their message. This is being implemented using the *FallbackClassifier*[20] and setting a minimum threshold for the *NLU* confidence. If the NLU predicts an intent with a confidence lower than the provided threshold, the user will be asked to rephrase the message.

Similar to the *NLU* fallback, a fallback for the *Dialogue Policies* is being implemented in the *Rule Policy*: if the assistant's confidence for the predicted *Action* is lower than the given threshold, it will tell the user it is unsure what the user expects. The user will then be asked to make sure the request is supported by the system and rephrase the message.

4.2. Rasa Action Server

The Rasa framework provides the possibility to implement custom functionality by creating a so-called *Action Server*[21]. When the assistant predicts a custom *Action* defined in the domain, it sends a HTTP POST request to the *Action Server* endpoint. This request contains information including the name of the predicted custom action, the conversation ID and extracted *Entities*. *Action Servers* can be implemented in any programming language as long as they provide an endpoint which accepts POST requests. The Rasa SDK for Python provides the APIs required to implement a *Custom Action Server* and contains methods to get easy access to the information of the conversation as well as send messages to back to the conversation. The *Tracker*[22] represents the bot's memory and provides access to the conversation data, while the *Dispatcher*[23] implements functionality to send messages back to the user. In this chapter, the components of the *Action Server* used for RAINA are described in detail. In Figure 4.1 all components described are contained in the *Rasa Action Server* container. The *AWS S3* is depicted outside of the *Action Server* since it is an external service.

[20] `https://rasa.com/docs/rasa/components#fallbackclassifier`
[21] `https://rasa.com/docs/action-server`
[22] `https://rasa.com/docs/action-server/sdk-tracker`
[23] `https://rasa.com/docs/action-server/sdk-dispatcher`

4.2.1. Database

For storing the user data, we chose the lightweight solution *SQLite3*[24] as database system in our implementation. Unlike client-server based SQL databases such as *PostgreSQL* or *MySQL*, *SQLite* is a serverless solution resulting in a higher transaction speed. All data stored in the *SQLite* database is written to a single file directly in the file system, instead of being spread across multiple files as it is the case with client-server based databases (Hipp, Wyrick & Company 2021).

This single file containing the whole database integrates well with the approach to run the *Action Server* as a *Docker*[25] container with a focus on quick progress. The latest version of the database-structure is always integrated directly in the corresponding *Docker* image of the *Action Server*. This allowed fast development speed of the *Action Server* and an automatic adaption of the database in parallel.

A drawback of SQLite databases is the missing possibility to remotely connect to them, due to the serverless architecture. To fetch and analyze the user data post study, a *Docker* volume-mount has been implemented to be able to copy the database file from the remote server to the local machine.

There was no need to save whole conversations in the database, since Rasa integrated this in their deployment solution RasaX, which will be described in section 4.3. In the *SQLite* database only data used by the action server itself is being stored, which benefits strongly from the high transaction speed *SQlite* provides. The user data is being stored in the following tables:

1. **user**: user IDs and their registration dates
2. **pyramid**: nutrition pyramids by date and user ID
3. **reflection**: answers to the check-in questions by date and user ID

The pyramids are being represented by one column for each food group storing the number of consumed servings as a float with the default value zero. The columns are incremented/decremented by mapping the extracted *Entity* to an enum, which contains the column name. Answers to the check-in questions are stored with one column containing the question type and one column containing the score.

[24] `https://docs.python.org/3/library/sqlite3.html`

[25] `https://www.docker.com`

4.2.2. Image Generation

Visualizing the plain numbers stored in the database is the key element of providing visual feedback to the users. A first idea was to build a string out of food emojis and format it with indentation, so it represents a pyramid. This approach was connected to several issues: Firstly, the emoji-pyramid does not represent the original BZfE Nutrition Pyramid since the icons differ. Secondly, the formatting of the generated string might change when being used on devices with a smaller screen, shifting emojis in the next line breaking the triangular shape of the pyramid. This approach also eliminates the possibility to show more fields than intended in the original pyramid, if a user eats more portions than there are considered in the BZfE Nutrition Pyramid as well as only partially consumed portions of foods.

The closest representation of the BZfE pyramid is the original image itself. To indicate that a portion has been entered, one could simply cross it out by printing two lines over the field in question. With this method half of a portion could be displayed with only one crossing line but smaller fractions are impossible to insert into the pyramid. Based on the idea to print crosses in the original image, we developed a concept to represent the pyramid with easy visibility of the entered portions and taking fractions of portions into account as well. The original nutrition pyramid image was converted to a black and white image (see: Figure 4.2) and the colorful icons for the pyramid fields have been cut out and are now stored separately as individual images.

Figure 4.2.: Adapted Base Image

Figure 4.3.: Partially Filled Pyramid

The pyramid images are now being generated on runtime whenever a user requests to see a pyramid: the pyramid-data is fetched from the database and for each serving inserted, the corresponding colorful icon will be printed in the black and white image of the pyramid. This results in a fully grey pyramid with the entered fields in color.

For editing images in this way, the Python imaging library *Pillow*[26] has been used. The positions for the icons in the correct order (left to right) for each food group are saved in the script. Using *Pillow*, it is possible to crop images while printing them into the black and white image of the pyramid. If a user inserted a fraction of a portion into the pyramid, the icons width is being cropped to display the corresponding fraction of icon for the given amount. If a user has inserted 4.75 portions of water, the pyramid will contain four whole water icons and one icon covering only 75% of the next water icon. Additionally, portions added above the recommended amount are now displayed next to the pyramid at the level of the corresponding food group, which happened for the icon of sweets on the top of the pyramid (see: Figure 4.3). We have chosen the positions of the additional portions so that the distance between the pyramid and the icons is greater than the distances inside the pyramid, in order to easily see the inserted excess of portions.

4.2.3. Image Cloud Storage and Sending

After generating the image for the user, the file will be saved in the local file system of the *Action Server*. For keeping the memory usage as low as possible, the filename equals the user ID. This results in only one image being stored for each user since the last image with the filename will be overwritten. Since the number of files equals the number of users, is approach is reasonably memory effective. However, the amount of memory required scales linearly with the number of users. For this proof of concept implementation this is not an issue since there is only a limited amount of test users. If the chat assistant should be deployed and used on a larger scale, we recommend to implement a scheduled function deleting the images at the end of each day.

Rasa provides the possibility to send images to users by adding a link to the image parameter of the message. However, it is not possible to send files directly via the chat-bot. To get a public link to the image of the pyramid, the file will be uploaded to a *Amazon Web Services (AWS) Simple Storage Service (S3)*. The *S3* bucket is configured to be publicly accessible so the chat clients are able to download and display the image. After generating the image with *Pillow* and saving the file, the file is being uploaded to the *AWS S3* immediately. For this step the Python library *Boto3*[27] has been used, which is the official Amazon Web Services Python SDK.

Since the name of the image-file will not change, because it is the user's ID, subsequently the link to the pyramid file in the *S3* for a user will always be the same as well. Regardless of the name of the file, the link might point to a newly generated pyramid. This turned out to be an issue with Telegram. Telegram seems to save images for used

[26] https://pillow.readthedocs.io/en/stable/index.html

[27] https://boto3.amazonaws.com/v1/documentation/api/latest/index.html

links in its cache and uses those when the same link will be sent again, causing the users to always get the image of the first pyramid they requested. We bypass this issue by adding a hashed timestamp as variable to the link, which changes its appearance but not its functionality. The Telegram API does not find the link in its cache and therefore downloads the latest image from the *S3*.

4.2.4. Message Scheduling

For sending messages at specific times to each user, a scheduled function is needed. In this section the approach for reaching out to the users used in this thesis is described in detail, since we do not use a functionality provided by the Rasa framework, but a workaround using the applications Telegram webhook[28]. The reasoning for this decision is described in the following section.

Rasa ReminderScheduled Events

The Rasa framework provides the *ReminderScheduled*[29] event which can be returned via a custom *Action*. Those events are being executed at the specified time with the passed *Intent* and *Entities*. However custom *Actions* are only executed if the chat assistant predicts them as an follow-up *Action* i.e., if the user triggers it within a *Story* or *Rule*. This means those events are designed for user to schedule reminders and are only executed once at the given time, without the possibility to schedule multiple events at once. One could use those events in combination with rules implemented in the assistant to create events cascading: the first event is being scheduled when the user registers and when this event is being executed it schedules the next one. This "chain" of *ReminderScheduled* events would then be perceived by the user as scheduled messages. The problem with this workaround is that the *ReminderScheduled* events are not stored in a database but in the bot's memory. This means that all reminders will be lost, if the system needs to be restarted, updated, or shuts down due to a failure[30].

Telegram Webhook

To get reliable scheduled messages a scheduled function needs to be executed to send the message to all users at a given point of time. For this thesis the main channel for the bot, which has also been used for the evaluation, was Telegram. Rasa interacts with the Telegram bot API via a webhook[31], which is specified in the bot's configuration. A

[28] `https://rasa.com/docs/rasa/connectors/telegram`

[29] `https://rasa.com/docs/rasa/reaching-out-to-user/#scheduling-reminders`

[30] `https://rasa.com/docs/rasa/reaching-out-to-user/#cancelling-reminders`

[31] `https://core.telegram.org/bots/webhooks`

webhook is an exposed endpoint to enable server-to-server communication by sending requests to it. In the case of Telegram and the Rasa back-end in the following way: whenever a user sends a message to the Telegram bot, Telegram sends a message to the Rasa backen-end informing it that the conversation has been updated. This message is implemented as a POST request to the Rasa webhook. The request contains all necessary information for the chat assistant to know how and to which user to respond, such as the the message which has been sent, the chat which has been updated, and the message ID.

Simulating an Conversation-Update

Since the webhook is defined in the bot's configuration and is publicly accessible, it is possible to send requests to it by custom services. The POST requests can be designed like the update-requests the Telegram API sends, simulating an update in a conversation. The values of the parameters in the body of the POST-requests payload can be chosen freely, which means we can select the user and the message which has been sent to them. In this case the telegram chat ID equals the conversation ID which is being stored in the Rasa *Tracker*.

For the message, which will be injected into the conversation update, we can use another functionality of the Rasa framework, which is addressing *Intents* directly, bypassing the NLU. Instead of sending a real text message which will be interpreted by the NLU, it is possible to send *Intents* directly to the Rasa Core by sending /[*intent*] as a message. In the prototype three different scheduled check-in questions have been implemented. Each one is being triggered by an individual intent and a corresponding rule, to make the bot "reach out" to the user.

To get a list of users to send the check-in questions to, the user table in the SQlite database is being used. For each user the conversation ID is being stored as user ID paired with the date the user started to use the chat assistant. Since the explained approach uses the Telegram webhook, only valid telegram chats can be updated. The bot was deployed using RasaX which offers a webchat interface to interact with the current deployed version of the bot (section 4.3).

Conversation IDs created via the webchat are also being inserted into the user table, but are unable to receive the check-in questions. This is not an issue since the webchat is only being used for short interactions to test the assistant, rather than chatting over an extended period of time. However, the user IDs have to be filtered out of the list before the scheduled function starts to send the messages. This is not a problem since the Telegram conversation IDs are integers, while the RasaX conversation IDs are alphanumeric strings which are clearly distinguishable.

Scheduled Functions

This strategy has been used to implement functions scheduled at specific times during the day using the Python library *schedule*[32]. Those functions send a POST request from the *Action Server* to the Telegram webhook, which leads the chatbot to executing a custom *Action*, connected using a rule. Each *Intent* for a check-in question has a corresponding custom *Action* sending the message to the user. The user does not notice any of this, but just receives the message of the bot, thinking it reached out to the user by itself. Using the Telegram webhook to make a chatbot reach out to the user is a common practice, which has also been used by Ludwig Horner in his Bachelor's Thesis and Wenjian Li in his Master's Thesis (Horner 2019; Li 2020).

4.2.5. Storing Check-In Answers

There are three check-in questions implemented, each one with five possible answers. With the use of Telegram buttons answering the questions is a matter of seconds. The answers have been designed to appear appealing by using fitting emojis to support the meaning of the answer. As example, the *Evening Check-In Question* is shown in Figure 4.4.

In the *Action Server* the texts on the buttons are not being used, but the title of the button itself, to distinguish which answer a user selected. The five answers are mapped to scores on a 5-point Likert scale which are stored in the *reflection* table, paired with the type of the question as well the user ID and as the date it has been answered.

To display a button in the chat, the button is added to the bot's utterance with a text to be displayed on it. The button itself represents an *Intent* which will be processed by the chat assistant. Although the Rasa framework supports sending *Entities* as a payload with their buttons, this does not seem to work with Telegram's buttons. When using a custom chat interface, one could simply send the question type and the score as *Entities* with the buttons, using a single *Intent* for all buttons. Since for this thesis Telegram has been the main interface to interact with the chatbot, a workaround has been developed to be able to store the answers to the check-in question. Using a distinct button for every of the five answers for each of the three questions, we created 15 new intents using a naming-convention, which allowed us to parse the name of the buttons into the corresponding question-type and score to store it into the database.

When a button has been pressed by the user, a custom *Action* is being executed in the *Action Server*. The last *Intent* which has been received can be read from the *Trackers* memory. For example: if the user does not feel motivated in the morning the tracker will receive an intent called *BUTTON_morning_check_in_1*. This title will then be split

[32]`https://schedule.readthedocs.io/en/stable`

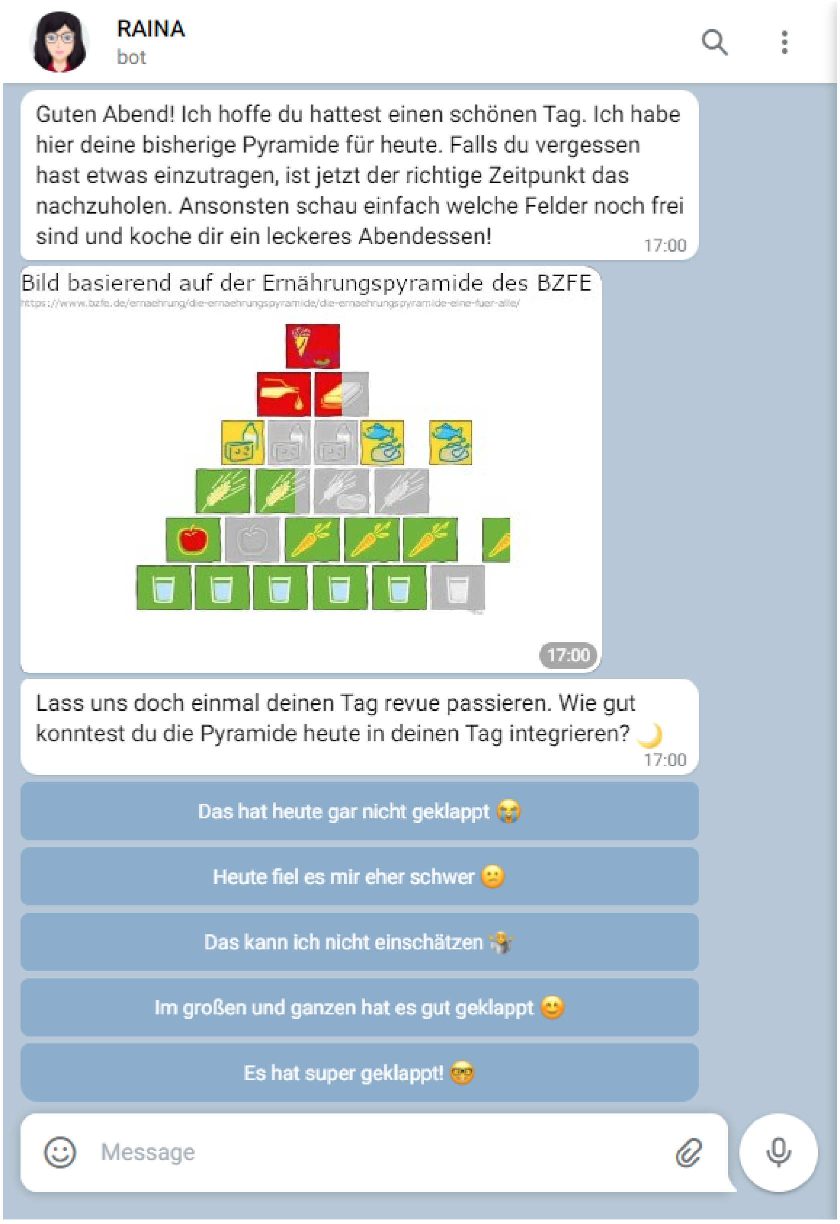


Figure 4.4.: Evening Check-In

into the title of the question (i.e., "*morning_check_in*") and the score (i.e., "*1*") the user selected. The title and the score will be stored in the database, with 1 being the lowest possible and 5 being the highest possible point on the Likert-Scale. Additionally, the bot sends the user a motivational message if the score is greater than three and a tries to cheer the user up otherwise.

4.3. RasaX Deployment

The challenge when developing a conversational AI chat assistant based on supervised learning is the amount of training data the system needs to learn from. Rasa Technologies provides a free tool for all Rasa developers to support *Conversation-Driven Development (CDD)* called *RasaX*[33]. *RasaX* is a self-hosted system via which a Rasa project can be deployed, shared, tested, and monitored. Once *RasaX* is deployed on a server, it can be connected to the git repository in which the training data for the bot is being stored.

A model based on the latest training data in the repository can be trained via *RasaX* at any time directly on the server. The virtual machine running on the server used for this thesis made use of eight vCPUs as well as 16 GB of RAM. The training and activation of a new model took about six to eight minutes. The *Custom Action Server* is being deployed on the server as well using Docker. The Docker image of the latest version of the *Action Server* has been pushed to a Dockerhub repository and was integrated in the docker-compose file of *RasaX*.

4.3.1. RasaX Webinterface

In addition to the usual channels which are supported by Rasa Open Source, *RasaX* provides a webchat to quickly share and test the currently running version of the assistant. The website, the sharing-link leads to, consists of a very short description of the assistant as well as a chat window. Each conversation between a tester and the assistant can be monitored and analyzed in the RasaX admin console. More than that, developers also have the possibility to either adapt and use real-world test conversations as training data or train the bot with interactive learning. The *Stories* and examples for *Intents* being generated with those methods can be directly pushed to the connected git repository. *RasaX* automatically creates a new branch with a merge request to be reviewed and adds the user generated training data to the repository.

[33] `https://rasa.com/docs/rasa-x`

4.3.2. Telegram Connection

For the user study, as well as the continuous testing the messenger app *Telegram* has been selected to interact with the assistant. The main reasons for Telegram are ease of use for the testers, fast integration into the Rasa project as well as the fact that the short time period of the thesis did not allow to build a custom front-end in addition to the chat bot. The creation of a Telegram bot is simple by texting a bot called *BotFather*[34] by Telegram. The *BotFather* provides an access token for the created bot to use the Telegram HTTP API and allows the developer to customize the bot by setting a description or profile picture. For the chat assistant in this thesis an avatar has been created using the online editor *Avatar Maker*[35].

The Layout of the texts as well as the emojis used in the messages the assistant sends to the users are fitted to Telegram. There are plugins for IDEs adding emoji support to YML-files, which are being used to define the assistant's utterances. Although emojis can be added using commands, the emojis used in the messages have been copied from the telegram desktop application into the files. The plugins do not support, all of Telegram's emojis at the time of implementing the assistant by command, but still displayed them if copied and pasted out of the Telegram application. When the messages have been sent, all emojis used are converted to Telegram emojis, making the messages more appealing.

4.3.3. Undefined Intents

As described earlier, it is possible to bypass the Rasa *NLU* by sending /[*intent*] as a message. This applies for all channels the assistant has been connected to. However, this seems to be a problem specifically with Telegram if the intent the user is trying to send to the assistant is not defined in its domain. When first testing the assistant on Telegram, users immediately tried to shut the bot down by sending intents to it such as */stop* or */shutdown*. Those commands resulted the Telegram bot to enter a crash loop in the conversation, the undefined intent has been sent to, causing it to be unreachable until the RasaX server has been restarted. Although this issue occurs when using Telegram, RasaX itself handles this issue internally, when using the provided Webchat. This indicates that there may be an error in the Rasa Open-Source framework interacting with the Telegram API. The issue itself could not be fixed, but some intents or commands users might send which are commonly used to control bots such as */stop*, */shutdown*, */end* and */ping* have been added to the assistants domain. Those intents are being handled with separate rules prompting the users to reformulate their intent.

[34] `https://core.telegram.org/bots#3-how-do-i-create-a-bot`
[35] `https://avatarmaker.com`

5. Evaluation

For evaluation of the chat assistant, we conducted a seven-day user-study. Firstly, we were interested in the usability of the system and the user interactions. Secondly, we wanted to evaluate if there is a change in the participants' view on their health, their self-efficacy regarding nutrition and their abilities in reflective thinking. Due to the short time frame of this bachelor's thesis, it was not possible to examine the effects of RAINA on participants' dietary behaviors and habits over a longer period of time.

5.1. Study Design

The study consisted of two online questionnaires and a seven-day trial period for RAINA on Telegram with the option to continue use after the user study was completed. The study was conducted fully remotely, sending the participants all necessary information via email. At the start of the study, the participants received the link to the first online questionnaire as well as a PDF with a summary of RAINA's abilities and instructions on how to communicate with her.

In the initial questionnaire we asked the participants for some general information consisting of their demographic data, if they have ever taken part in a professional nutritional counseling, and if they have used a nutrition app to monitor their diet before this study. Subsequently, there were four question groups for gathering information on the participants' *Health Consciousness*, *Reflective Thinking*, *Nutrition Self-Efficacy* and *Nutrition Self-Assessment*.

After seven days the link to the second online questionnaire has been sent, with the information that RAINA will be available for four more weeks. The question groups *Health Consciousness* and *Nutrition Self-Efficacy* have been asked again without any modifications compared to the first questionnaire. To the question group regarding *Reflective Thinking*, several questions have been added to evaluate the participants' reflection while the study took place. The question group of *Nutrition Self-Assessment* was replaced with *Questions about RAINA* and general questions about the study in general which should be answered with free text input.

The development of both, the pre- and the post-study questionnaire, will be described in in the next section. The complete questionnaires as well as the instructions, which have been provided, can be found in the appendix.

5.2. Measurements

We used question-groups and items specifically developed for the topics to be evaluated with the questionnaires. Some scales and questions were adapted to our study setting and context as described in this section. The question groups *Health Consciousness*, *Reflective Thinking* and *Nutrition Self-Efficacy* have been used in both questionnaires. *Demographics* and *Nutrition Self-Assessment* were only part of the first questionnaire, while *Questions about RAINA* were only asked in the second questionnaire. All items of the scales described below are rated on a five-point Likert-Scale, with *"totally disagree"* corresponding to a numeric score of 1, being the lowest and *"totally agree"* corresponding to a numeric score of 5 being the highest possible answer.

The participants' *Health Consciousness* in general has been evaluated using a scale developed by Hong. The original scale consisted of eleven items, of which five have been used in this study. This question-group aims to display the participants' assessment on their personal responsibility for living a healthy life (Hong 2009).

After asking the participants about their *Health Consciousness*, *Reflective Thinking* was evaluated using questions from a questionnaire created by Kember, Leung, et al. The original questionnaire was designed to be used for measurement of *Reflective Thinking* in the domain of university lectures (Kember, Leung, et al. 2000). Of the four core-constructs, the questions measuring *"Habitual Action"* and *"Reflection"* have been used in the pre-study questionnaire since they aimed on general reflection. In the post-study questionnaire, the construct *"Understanding"* has been added, in which questions referring to a lecture, in this case the use of RAINA, are being asked.

For measuring the participants' *Nutrition Self-Efficacy*, we used the entire *"Nutrition Self-Efficacy Scale"* developed by Schwarzer and Renner (Schwarzer and Renner 2009). As descried in chapter 3, *Self-Efficacy* describes a person's confidence to be able to handle high-risk situations without falling back into old behavioral patterns (Prochaska and Velicer 1997). The scale consists of five possible barriers, which one might face when trying to stick to healthy foods. The participants are asked to state how certain they are to overcome those barriers (Schwarzer and Renner 2009).

At the end of the pre-study questionnaire, the participants were asked on how they personally assess their diet in general. For this group, we used several questions from Shamsalinia, Ghadimi, et al., which fitted best to this study. The questionnaire by Shamsalinia, Ghadimi, et al. was designed for participants with a mean age around 60 years and contained some questions regarding an age-appropriate diet and medication (Shamsalinia, Ghadimi, et al. 2019). Those questions have not been included in this study.

For the evaluation of the system itself we included a set of questions specifically about RAINA and the *BZfE Nutrition Pyramid* in the post-study questionnaire. On the one hand, a set of items making use of the Likert-Scale, just like the scales above were included. On the other hand, questions with textual answers have been included to add a qualitative part to the study. Those questions are mostly self-devised, but are taking the fourth construct (*"Critical Reflection"*) of Kember, Leung, et al.'s questionnaire and interview questions from the system evaluation of the *"Reflection Companion"* by Kocielnik, Xiao, et al. into account (Kember, Leung, et al. 2000; Kocielnik, Xiao, et al. 2018).

5.3. Participants

The participants for the seven-day user-study were recruited by presenting the study on social media with a Google Form for self-enrollment. After 2 weeks of voluntary registration the study period started, taking place from the 3rd to the 9th of January 2022. A total of 21 participants completed the pre-study questionnaire, but only 18 conversations with RAINA have been started and 18 post-study questionnaires have been submitted. Since the study took a quantitative approach and we designed the interviews and questionnaires to be completely anonymous, there was no way to discard the three excess pre-study questionnaires after the study. This results in a different total number of responses for the pre-study (n=21) and post-study (n=18) questionnaire.

Of the 21 participants, the gender distribution was fairly balanced with 9 female and 12 male participants. The age varied from 19 to 57 (M=28.14, SD=12.69), with 13 (62%) of the participants being 22 or 23 years old. The level of education was high, with 12 (57%) participants currently enrolled as university students and four (19%) participants working as developers in the IT domain with a university degree. Four (19%) participants were working in different field, with three of them having graduated from a university and one having completed vocational training. One participant did not specify their level of education or current job.

Participants were mostly in the normal weight category[1] (mean BMI=21.86 kg/m^2, SD=2.64), two female participants were underweight (BMI=16.10 kg/m^2 and BMI=16.80 kg/m^2), one male participant was slightly in the pre-obesity category (BMI=25.25 kg/m^2). Only one (5%) male participant reported to have had nutritional counselling. However, 10 (48%) participants have used a nutrition or diet app before the study. When looking at female and male participants separated, seven out of nine (78%) women have used a nutrition app, but only three out of twelve (25%) men.

[1] `https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi`

6. Results

In this chapter, we present the results in the chronological order of the user-study: beginning with the results of the *Pre-Study Questionnaire*, followed by the users' engagement with the system and the results of the check-in questions in the *User Data* section. Finally, the results of the *Post-Study Questionnaire* will be described.

At this point it is important to note that, due to a bug in the initial conversation, eleven (61%) participants who started a conversation with RAINA did not get the check-in questions. The conversations have been monitored, but there was no active intervention from our side to not distort the results of the study. Although the participants have been informed that they will receive daily check-in questions from RAINA and should contact us, if something does not work as expected, only one participant reached out to us. She stated that she does not receive the check-in questions she was expecting. In this case we helped and provided technical support for the participant to be able to use RAINA as intended.

The questions about RAINA in the *Post-Study Questionnaire* have been extended within the study by one more question asking if the user received notifications, followed by two follow-up questions depending on the previous answer. In the *User Data* section, we present the data looking at users with and without notifications separated.

6.1. Pre-Study Questionnaire

Nutrition Assessment

When self-assessing their diet, the participants' mean score (M=3.19, SD=0.58), in was close to the middle of the Likert-scale. The distribution of mean scores is shown in Figure 6.1. Most participants reported to have no issues with maintaining a healthy weight for their body height and age (M=3.76, SD=1.34)[A1]. They know which nutrients a healthy diet consists of (M=3.76, SD=1.04)[A7] and recognize healthy foods when they are buying them (M=4.28, SD=0.78)[A9].

However, the knowledge self-assessment on how to cook foods without decreasing their nutritional value differs, resulting in a mean close to the median (3.00) of the Likert-scale with a high standard deviation (M=3.05, SD=1.14)[A8]. Only about half of the Participants feel confident about eating healthy when they are busy or stressed

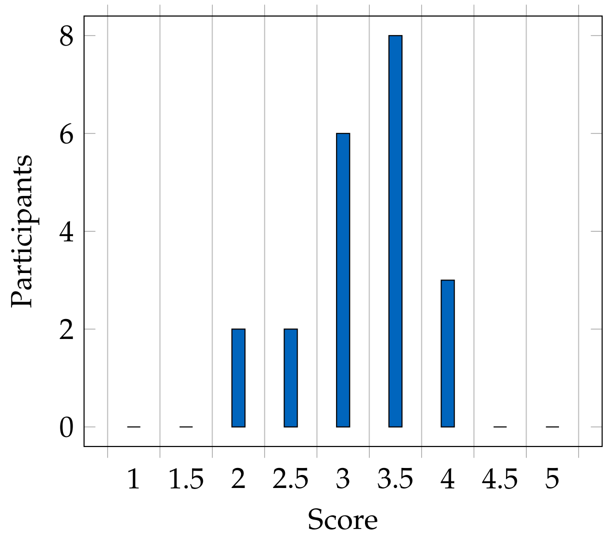


Figure 6.1.: Nutrition Assessment Mean Scores (rounded to nearest half)

(M=3.05, SD=1.07)[A2]. Similar results occurred on the assessment of participants' motivation to cook healthy meals when being alone (M=3.38, SD=1.36)[A3] or eating not at home (M=3.19, SD=1.03)[A4]. The majority of participants do not abandon fast food (M=1.95, SD=0.74)[A5] and do not decline unhealthy foods if it is offered to them (M=1.95, SD=0.92)[A6].

Furthermore, 17 participants (81%) reported to check supposed nutrition facts, they receive via family, friends, and media, with the help of the internet. Two (9.5%) use professional literature in the domain of nutrition and the remaining two (9.5%) do not validate facts about diet at all [A10].

Health Consciousness

Participants are reflecting on their health (M=3.86, SD=0.73)[H1], but are mostly not concerned about their health all the time (M=2.67, SD=0.91)[H2]. They are taking responsibility for their health (M=4.24, SD=0.62)[H3] and think that their health requires active participation from themselves (M=4.67, SD=0.58)[H4]. To the question if they only think about their health when they are sick, the participants mostly disagreed (M=2.43, SD=1.12)[H5]. The mean scores for all participants are shown in Figure 6.2 (M=3.59, SD=0.46).

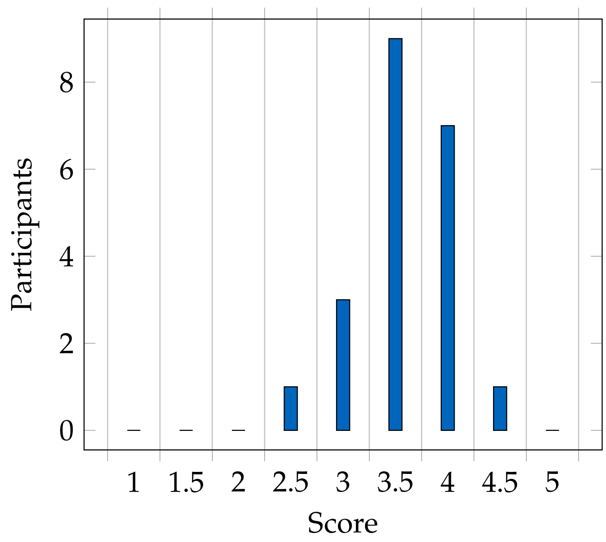


Figure 6.2.: Health Consciousness Mean Scores (Pre-Study, rounded to nearest half)

Reflective Thinking

In the context of *Habitual Action* (Kember, Leung, et al. 2000), the participants were divided in both, the execution of a task without thinking about it (M=3.43, SD=1.03)[R1] and remembering information which they learned once (M=3.38, SD=1.02)[R2]. This resulted in a high standard deviation greater than one. In the construct of *Reflection* ([R9], [R10], [R11], [R12]), the reported answers were grouped closer together with lower standard deviation (0.57 < SD < 0.96) and higher mean scores (3.71 < M < 3.90). This resulted in a high mean score with a low standard deviation for *Reflective Thinking* shown in Figure 6.3 (M=3.69, SD=0.40).

Nutrition Self-Efficacy

The participants are positive about overcoming the following challenges: taking a long time to develop the needed routines (M=3.90, SD=0.83)[S1], repeated attempts to change their diet (M=3.76, SD=0.77)[S2] and not getting support from others (M=3.61, SD=0.86)[S4]. Changing their whole eating behavior (M=3.33, SD=1.11)[S3] and creating a detailed plan to stick to healthy foods (M=3.33, SD=1.11)[S5] present larger challenges to some of the users, resulting in lower mean scores with higher standard deviations. The participants showed high mean scores in *Nutrition Self-Efficacy* (M=3.57, SD=0.75), displayed in Figure 6.4. However, the mean scores are spread more over the range of the scale than for *Health consciousness* and *Reflective Thinking*.

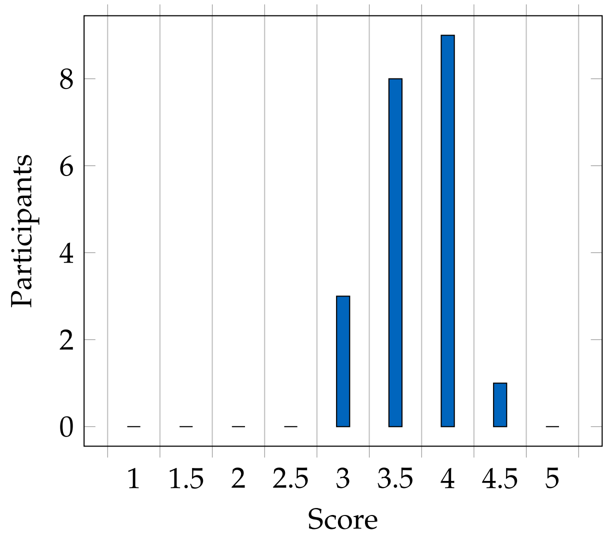


Figure 6.3.: Reflective Thinking Mean Scores (Pre-Study, rounded to nearest half)

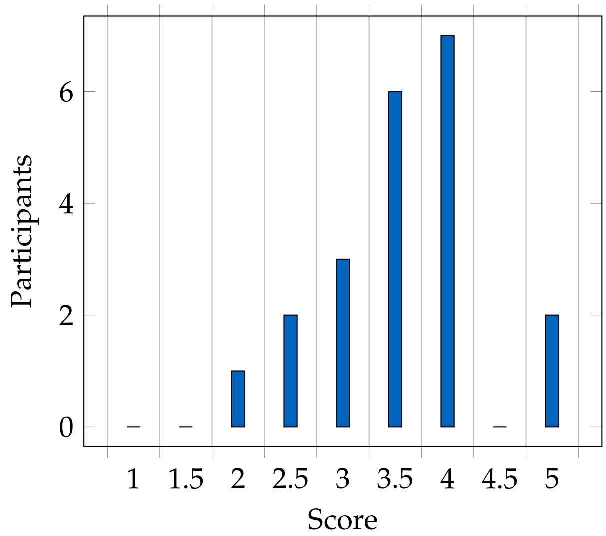


Figure 6.4.: Nutrition Self Efficacy Mean Scores (Pre-Study, rounded to nearest half)

Aggregated Scores

When looking at the mean scores aggregated for each user individually in the question groups *Health Consciousness*, *Reflective Thinking* and *Nutrition Self-Efficacy*, we could notice a pattern: the scores in each of the categories did not deviate strongly from one-another, which means that participants with a low score in one of the categories tend to have a low score in the others as well. The difference between the lowest and the highest score for a user was 0.50 in nine (43%) of the cases, 1.00 in eight (38%) of the cases and 1.50 in four (19%) of the cases. The mean scores for each participant in the first questionnaire are attached in the appendix in Table D.1.

An example for the high difference was on participant, who had a score of 4.00 in *Health Consciousness* and *Reflective Thinking*, but only a 2.50 in *Nutrition Self-Efficacy*. There were only two participants who reached the highest possible score of 5.00, both in the category *Nutrition Self-Efficacy*, with corresponding high scores (4.00, 4.00) and (3.50, 4.00) in the categories of *Health Consciousness* and *Reflective Thinking*. The lowest score measured was 2.00, also in the category *Nutrition Self-Efficacy*. The participant had also a low score in *Health Consciousness* (2.50) and average score in *Reflective Thinking* (3.50). Overall, the mean scores over all categories for each participant individually, were in the interval of 2.67 to 4.33 with standard deviations between 0.24 and 0.70.

6.2. User Data

User Engagement

We measured the engagement with the system in pyramids with at least one field value not equal to zero. The reasoning is, that the evening check-in question generates a empty pyramid, if the user has not inserted any foods up to this point, to send the user the image. We can not count those, since then the users who did get notification messages, would have been tracked as active every day, which was not the case as described in this section.

The overall engagement with the system is shown in Figure 6.5. Over the course of the seven-day trial period, a total of 18 participants reached out to RAINA and inserted at least one pyramid with an entry into the database. However, at no day of the study, all users have been active. The number of active days for each user summed up is shown in Figure 6.6, the individual days each user engaged with RAINA is displayed in Figure 6.7. In both figures, the users who did get check-in questions are visualized in blue, while the users who did not get any check-in messages from RAINA are plotted in orange.

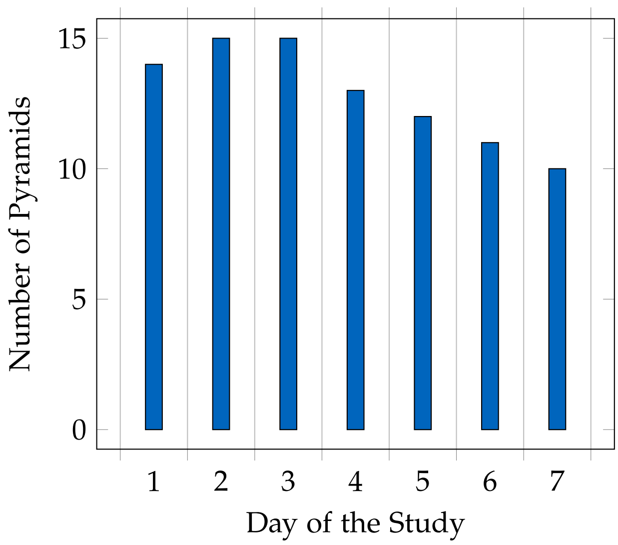


Figure 6.5.: Overall User Engagement

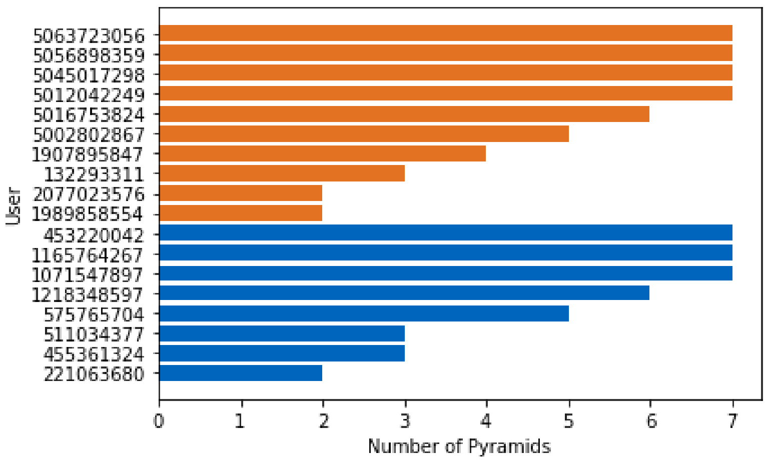


Figure 6.6.: Number of Active Days per User

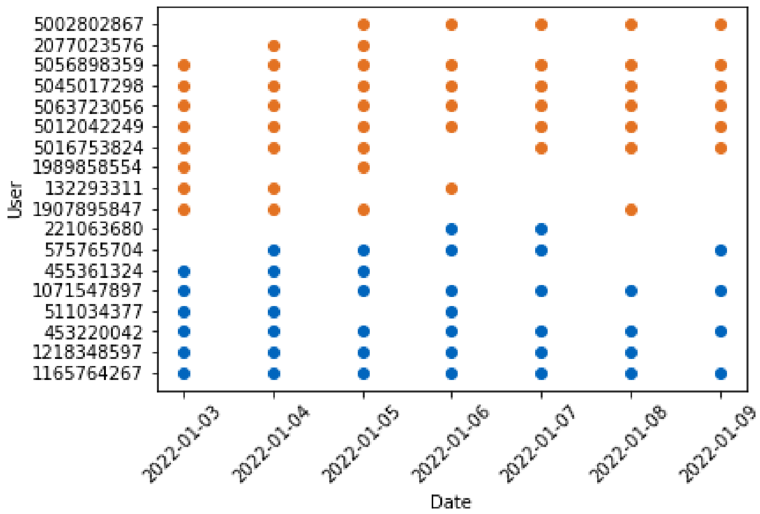


Figure 6.7.: Active Days per User

We could not identify a distinct pattern for quitting the usage of RAINA when observing on which days the users engaged with RAINA and when they stopped. While one user just started the study delayed by two days and continued using the chat assistant until the end, two others started late and used the system only for two days each. Some users skipped one day of usage, interacted with RAINA once more and the stopped interaction. Another user tested the system for three days and then quit (see: Figure 6.7). Particularly interesting is the fact, that both groups, with or without additional check-in messages from RAINA, inserted the exact same number of pyramids on average (5.00).

Check-In Data

Looking at the average number of answers to the check-in questions, they were lower than the average number of pyramids within the testing period, with an average of 4.00 answers for the morning and 4.13 for the evening check-in. In the morning check-in the users have been asked to report how motivated they feel to start the day. The weekly average of 4.00 within the interval of 3.60 to 4.25 remained stable at a high level of motivation (Figure 6.8).

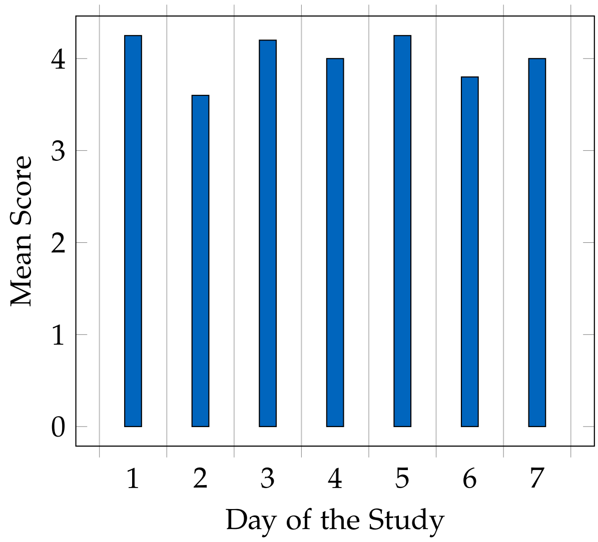


Figure 6.8.: Average Morning Check-In Scores

The scores of the evening check-in were much lower with an average of 2.70 within the interval of 1.50 to 3.67 (Figure 6.9). In the evening check-in the users have been asked how well they have been able to integrate the pyramid in their day so far. On the first day, the score (3.67) was the highest within the week, followed by the lowest score (1.50) on the second day, which is visualized in Figure 6.9. However, the score generally increased within the study as the user continued to use RAINA.

6.3. Post-Study Questionnaire

Changes in Health Consciousness

The mean score for *Health Consciousness* (M=3.42, SD=0.49) of all participants decreased slightly by 0.17 compared to the pre-study results (M=3.59, SD=0.46). When looking at the questions in detail, the participants' reflection on their health (M=4.00, SD=0.84)[H1] increased by 0.14, while the scores for concern about their health (M=2.39, SD=0.98)[H2] decreased by 0.28. The mean scores for the following questions decreased as well with a decrease of 0.24 for health responsibility (M=4.00, SD=0.59)[H3], 0.14 for active participation to a healthy life (M=4.53, SD=0.62)[H4] and 0.10 for the question if users only think about their health when they are sick (M=2.33, SD=1.02)[H5]. The distribution of mean scores for *Health Consciousness* for the post-study, compared to the pre-study survey is shown in Figure 6.10.

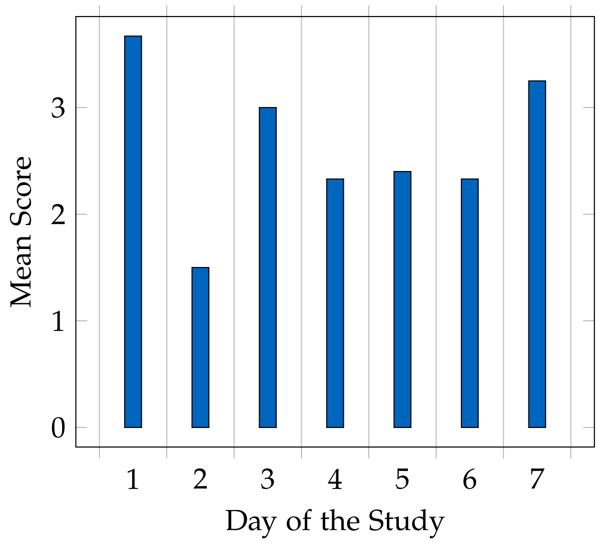


Figure 6.9.: Average Evening Check-In Scores

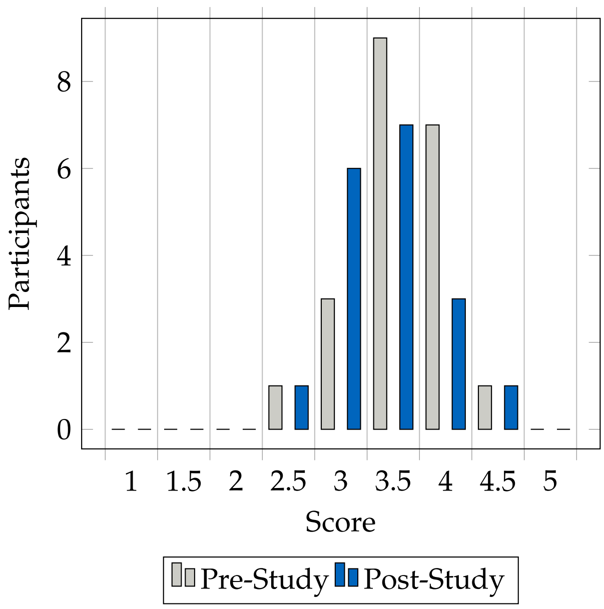


Figure 6.10.: Comparison of Health Consciousness Mean Scores (Pre- and Post-Study, rounded to nearest half)

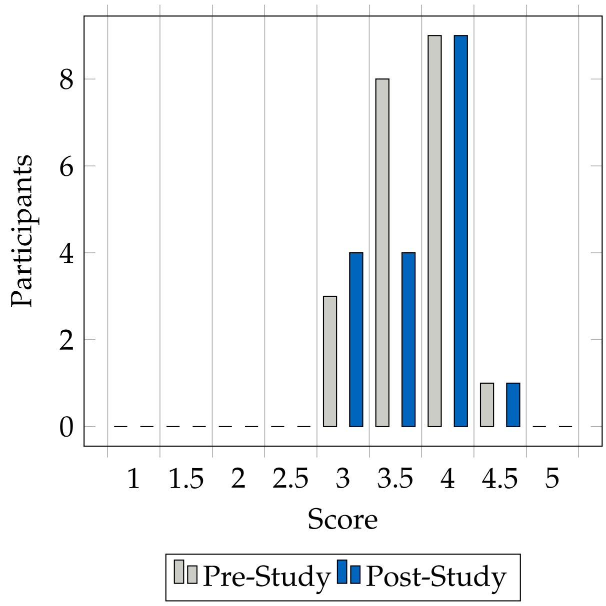


Figure 6.11.: Comparison of Reflective Thinking Mean Scores for Base Questions (Pre- and Post-Study, rounded to nearest half)

Changes in Reflective Thinking

As described in chapter 5, questions have been added to this group. To compare the results, we will first take a look at the "base questions" i.e., the questions which have been asked in both, the pre- and the post-study questionnaire. After this we will present the results for the newly added questions and the combined score.

When looking at the results for the base questions, the overall mean score (M=3.694, SD=0.46) only increased by 0.004, which represents almost no change compared to the pre-study results (M=3.690, SD=0.40). In Figure 6.11 the distribution of scores is visualized. The sub-questions did change similarly only by marginal values, with the highest change being 0.15 for the item, if the users question the way others do something and think of a better approach (M=4.05, SD=0.64)[R9].

Coming to the questions about *Reflective Thinking* in combination with the usage of RAINA, the first questions belong to the construct of *Habitual Action* (Kember, Leung, et al. 2000). The opinions were divided on both, resulting in high standard deviation. On the one hand, the participants slightly disagreed that while the study took place, they did things so often, that they started doing them without thinking about it (M=2.89, SD=1.13)[R3]. On the other hand, the participants slightly agreed, that if they were following RAINA's instructions, they did not have to think much (M=3.44,

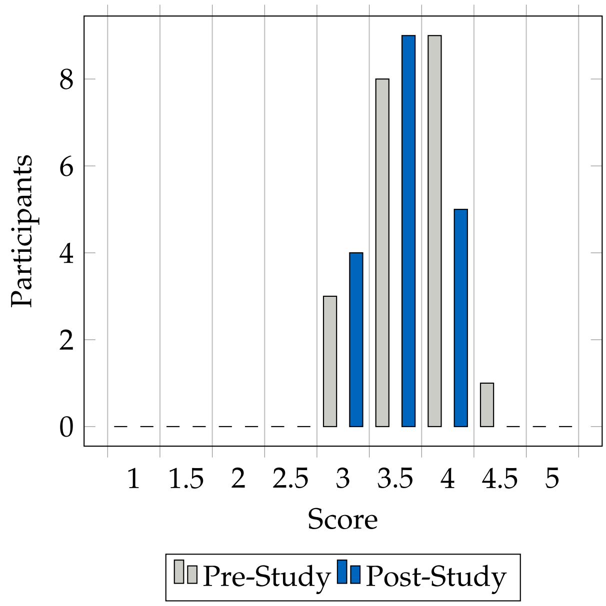


Figure 6.12.: Comparison of Reflective Thinking Mean Scores for Base Questions (Pre-Study) and Combined Questions (Post-Study) (rounded to nearest half)

SD=1.20)[R4]. The last four questions all belong to the construct of *Understanding* (Kember, Leung, et al. 2000). The participants generally agreed (3.56 < M < 3.83) that in order to successfully participate in the study they had to understand the topics RAINA taught them ([R5], [R6], [R7]).

However, the standard deviation was rather high (0.85 < SD < 1.19), since few participants disagreed. In contrast to the high mean scores in those questions, the participants disagreed (M=2.61, SD=0.85)[R8], that they had to continuously think about the newly learned information while the study took place.

This results in a lower mean score (M=3.33, SD=0.54) for the questions in regards to RAINA than for the base set of questions (M=3.694, SD=0.46). When combining all questions, asked in the group of *Reflective Thinking* in the post-study questionnaire (Figure 6.12), the resulting mean score (M=3.53, SD=0.36) is 0.16 lower than the score of the pre-study questionnaire (M=3.69, SD=0.40), where only the base questions were used.

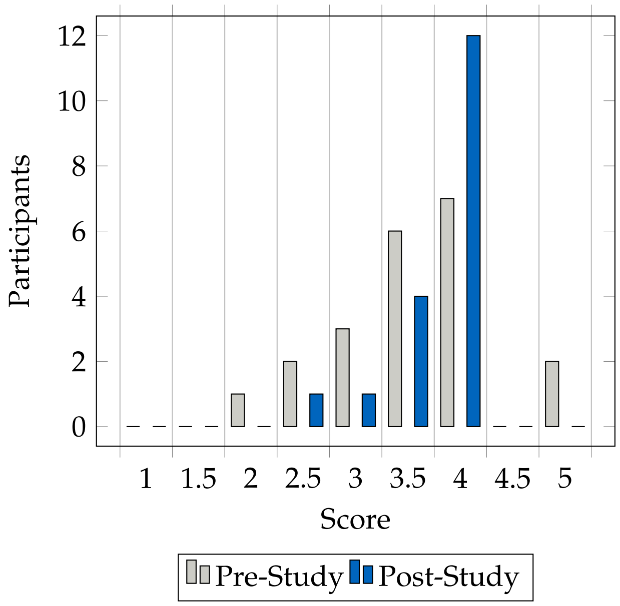


Figure 6.13.: Comparison of Nutrition Self-Efficacy Mean Scores (Pre- and Post-Study, rounded to nearest half)

Changes in Nutrition Self-Efficacy

The mean score for *Nutrition Self-Efficacy* (M=3.75, SD=0.43) increased by 0.18, with the standard deviation decreasing by 0.32, compare to the pre-study results (M=3.57, SD=0.75), which is displayed in Figure 6.13. The participants' confidence in overcoming the barriers of taking a long time to develop new routines (M=3.94, SD=0.54)[S1], having to reconsider their whole diet (M=3.56, SD=1.10)[S3] and needing to develop a detailed plan (M=3.56, SD=0.98)[S5] increased by 0.04, 0.23 and 0.22 respectively. Mean scores for overcoming the barriers of needing several attempts to stick to healthy foods (M=3.72, SD=0.75)[S2] and not receiving social support (M=3.59, SD=0.71)[S4] decreased by 0.04 and 0.03.

Changes in Aggregated Scores

For the aggregation of scores for each user, we selected the same questions for each group, which we have used in the pre-study questionnaire. The difference between the lowest and the highest score for a user was 0.00 in 3 (17%) of the cases, 0.05 in seven (39%) of the cases and 1.00 in eight (44%) of the cases. While in the pre-study questionnaire two participants reached the highest possible score in *Nutrition Self-Efficacy*, after the week of using RAINA, the highest sore reached in one group was

4.5, which was reached in *Health Consciousness* and *Reflective Thinking* once each. While those scores belong to two different users, the two occurrences of the lowest mean score calculated (2.5), belong both to the same participant in the categories of *Health Consciousness* and *Nutrition Self-Efficacy*.

Overall, the aggregated mean scores for the participants were in the interval of 2.67 to 4.17 with standard deviations between 0.00 and 0.47. While the range, in which the aggregated mean scores for each user are located, remained rather unchanged, the mean scores in general are now grouped even closer together. As described above, the observed changes in the individual question groups are not prominent when looking at the mean scores. However, the distribution of answers over the Likert-scale decreased. The mean scores for the post-study questionnaire are displayed in Table D.2.

Perception of RAINA

Of the 18 participants, who interacted with RAINA, eleven (61%) reported to have been familiar with the system of the BZfE and the Nutrition Pyramid [RA1]. The participants agreed, that the short introduction and repetition of the pyramid has been helpful (M=4.39, SD=0.78)[RA2], with no one disagreeing and only three participants being undecided. The majority had to adapt their diet in order to match the recommendations in the pyramid (M=3.65, SD=0.86)[RA3]. When asked if the person had already thought about the pyramid fields while shopping, the answers varied (M=3.0, SD=1.50)[RA4].

When asked about interacting with RAINA, the majority did not inform her immediately when they have eaten something (M=2.67, SD=1.08)[RA5]. Furthermore, half of the participants forgot to inform RAINA to update their pyramid sometimes (M=3.11, SD=1.37)[RA6], but most of the participants added the servings of the day subsequently in the evening (M=3.53, SD=1.37)[RA7].

The visualization of the pyramid was very well received by most of the participants. The majority reported that the pyramid helped them get an overview over their diet in general (M=4.22, SD=1.11)[RA8] and all of the participants agreed on the statement, that they were able the remember the graphic of the pyramid very well (M=4.67, SD=0.48)[RA9].

As reported at the beginning of chapter 6, unfortunately only seven (39%) of 18 users received check-in questions [C1]. Of those who did receive check-in messages from RAINA, the users assessed the messages to be helpful (M=4.43, SD=0.79)[CY1]. Although the check-in questions were perceived as helpful, only one person would have liked to receive more messages while all others disagreed (M=2.28, SD=1.25)[CY2]. The majority of users who used RAINA without the check-in questions would have liked to receive messages (M=4.13, M=0.99)[CN1] and stated the check-ins would have helped them to stay active in participation (M=4.0, SD=1.31)[CN2]. However, except for one participant, no one reached out to us, when they did not receive the check-in questions as announced.

General Experiences

Overall, the majority of users reported that RAINA affected their motivation to stick to a healthy diet in a positive way [F1]. One participant stated that their diet already has been close to the recommended amounts of foods in the pyramid and that the last bit of change to get it perfectly would be too much of an effort: *"According to the 80/20 principle, adjusting my diet to the pyramid would be an 80% effort for the last 20%"*. Having the visual representation of what has been consumed over the day in form of the pyramid, supported the participants to reflect on their food choices for the day and motivated them to correct their flaws for the next day: *"In the evening, you were always made aware of what you had missed during the day. As a result, you were motivated to eat more balanced meals the next day"*.

When asked about problems which occurred while the study took place [F2], some participants stated that the recognition of foods did not always work properly and RAINA did not know some foods they were trying to insert into their pyramid. However, when the participants used the names of the food groups or synonyms instead of specific products, there were little issues in recognizing foods: *"Initially, there were a few problems recognizing some food items. In these cases I used a known synonym for the respective field, which worked well"*. Furthermore, some participants stated, that RAINA removed portions when they wanted to add them: *"Once I wrote 'Add a portion of xy' and RAINA subtracted it"*. Usability-wise the chat assistant was easy to use and the system of the pyramid was convenient: *"The operation of the bot was intuitive and easy to use"*. One participant even used RAINA voice controlled: *"Via voice input, communication with the bot was very comfortable"*. Even though RAINA herself does not support voice recognition, the keyboards of smartphones allow the user to dictate their texts, transforming it to written text immediately, which made the "voice input" possible. The same participant also explicitly liked not having to count calories but getting a general understanding of their diet with the pyramid: *"It was nice that no calories were counted, but you got a good overview of food intake"*.

Monitoring and reflection, even though the study period has been only a week, showed impacts in the participants view on their diet [F3]. One the one hand, only two participants reported to have learned nothing new about their diet: *"Nothing new. I don't really eat "healthy" but I already knew that." / "I already ate relatively consciously before and did not have to change much"*. On the other hand, a participant who thought their nutrition was healthy, reconsidered their assessment when using RAINA: *"You think* *you're eating healthy, but you end up eating too many carbohydrates and sugar and far too few fruits and vegetables. You underestimate how many 'extras' [i.e sweets and snacks] you eat a day without logging"*. In general RAINA raised awareness to most of the users, that they did not consume enough fruits and vegetables, but too much of carbohydrates and extras.

Even though nine (50%) of the users reported that they would use the system after the study has ended [F4], only one user continued engaging with RAINA for additional 2.5 weeks. The reasons provided for discontinuing the usage varied. Several users stated to continue using RAINA, if she supports more functionality and the recognition of intents and entities will be improved: *"I would continue to use the system as long as further improvements and feature developments are made. The tool is excellent for getting a quick and rough overview of the current nutritional behavior"*. In contrast, other users reported that the system itself does not fit to their desired diet: firstly, there has been one vegan participant to whom the pyramid does not fit without major adjustments. Secondly, there was one competitive athlete, who stated to train about 20h per week. For this immense amount of physical activity, the recommended diet in the nutrition pyramid would not provide enough energy, since it is designed for a person with an average amount of physical activity. Thirdly, a user for who did not have to make any major modifications to their diet reported that RAINA will not be used, but the knowledge about the pyramid will remain as orientation: *"Due to the fact that not much has changed for me in terms of my diet, I will not continue to use the system itself, but the knowledge that was imparted to me"*.

When asked about ideas and functionality, the users would like in future versions of RAINA, several participants wished for functionality to entry pyramid fields faster. One idea was to be able to insert multiple foods in the same message, while another participant suggested to implement custom recipes: *"From my point of view, the following function would be great: entering "recipes" - this could be used to pre-store which nutrients a certain dish contains with my preparation, so that I don't have to add the individual nutrients manually each time"*. Another feature which has been mentioned was to rate the pyramid at the end of each day, to provide additional feedback to the picture in form of a score: *"an automatic daily summary of the pyramid, with a nutrition score (S, A, B, C, D, etc.)"*.

7. Discussion

The goal of this thesis has been to develop a conversational AI chatbot that implements an easy-to-understand theory for a healthy and balanced diet. Moreover, the assistant should further support the users to integrate and implement the theories in their everyday life. Our proof-of-concept chat assistant RAINA teaches the users the BZfE Nutrition Pyramid and is their daily contact to track their food intake. In this chapter we summarize and interpret the previously presented results.

7.1. General Findings

RAINA has been very well received by the users and has achieved her goal of educating and motivating the majority of the study participants to eat a balanced diet. The approach to track servings of certain food groups worked well and there has been positive feedback by participants, preferring this approach over calorie counting with conventional nutrition applications. Those findings are also in line with Casas, Mugellini, et al., where respondents positively emphasized the absence of precise nutrient tracking (Casas, Mugellini, et al. 2018). The difference between *Rupert le nutritionniste* by Casas, Mugellini, et al. and RAINA is the approach to set the goals: while *Rupert le nutritionniste* allowed the users to set their own goals, RAINA made use of the pre-set intake goals in form of the *BZfE Nutrition Pyramid*. Nonetheless, in both studies the participants reported an increase in awareness regarding their nutrition.

When investigating the interaction behavior of the participants with RAINA by analyzing the chat histories, we made interesting findings: at the beginning of the study, the majority of participants tested the limits of RAINA either by trying numerous synonyms for the food groups or using different wordings for their intents. Over the course of the study, the messages became shorter with less variations in wording, up to some cases in which participants just wrote the name of the food group and expected RAINA to add it to the pyramid. Even though RAINA is able to answer questions about portion sizes and the products which belong to the food groups, only few participants reached out to her, suggesting the system of the pyramid is easy to understand and remember.

To evaluate the impact, our approach has on the users' mindsets, we used two questionnaires to gather information based on the participants self-assessment. The participants we recruited to take part in our seven-day user study assessed their knowledge about healthy foods and the ability to follow a balanced diet as good. However, the majority did not pursue a healthy diet frantically, but also treated themselves with fast food occasionally. Compared with the provided body data, the self-reported answers match with the calculated BMIs for the participants: except for very few exceptions, all participants were in the normal weight category.

7.2. Impact on Health Consciousness

The group of participants was conscious about their health: with 3.0 being the median of the Likert-scale to measure the participants, only one person had a score lower than than the median in the pre-study questionnaire, with a mean score of 2.50. Most of the participants reached a score between 3.50 and 4.00 in the initial questionnaire. Within the course of the study, the average score of the participants' health consciousness decreased slightly. The lowest score remained at 2.50, which was achieved by only one participant.

The highest change in scores could be observed in the participants' concern about their health. While the participants are less concerned about their health, they also reported to reflect more about their health in general than before the study. Combining these two insights, you can see a fundamentally positive development: less concern, but more reflection about the participants health. However, with one score increasing and the other one decreasing, the change is not observable when looking at the overall mean score. With the little decrease in the other subsections, which might be caused by the three missing post-questionnaires, the overall score for *Health Consciousness* decreased.

7.3. Impact on Reflective Thinking

Although participants stated to be more reflective about their health, we could not find any correlation with the results in the question group of *Reflective Thinking*. Assuming that the *Reflective Thinking* abilities of the participants have improved, we should be able to see a clear increase in the average scores, but the values remained virtually the same for the base set of questions. After we included the scores of the questions that were asked exclusively in the second questionnaire, the average score dropped minimally. Looking at the results of the added questions in detail, the findings suggest that RAINA did not have a great impact on *Habitual Action*, which may be due to the short time of the trial period in which no new habits could develop. Nevertheless, the knowledge that RAINA had taught in the beginning of the study, was important to understand the system of the *BZfE Nutrition Pyramid* in general as well as to use the chat assistant.

When looking at Figure 6.11 one can observe that the bars at the score of 3.5 differ, while the others are on the same height. Having the same average score for the pre- and the post-study in mind, this might be confusing at first. However, the average scores both times were about 3.69, which means that the missing values close to the average value do not impact it as much as values further away from the mean.

7.4. Impact on Nutrition Self-Efficacy

Within the user study, the *Nutrition Self-Efficacy* average score increased slightly, with the standard deviation decreasing in parallel. While before the testing period, the scores were spread across the scale, after one week of using RAINA the average scores are grouped closer together with more than half of the participants reaching an average score of 4.0 in regard to *Nutrition Self-Efficacy*. This finding suggests that the *BZfE Nutrition Pyramid*, coupled with daily interaction with RAINA, may have a positive impact on confidence in overcoming barriers to healthy eating.

However, while two participants reached the perfect score of 5.0 in the pre-study questionnaire, this was not the case in the post-study survey. A possible explanation for this might be that those two participants had a different perception of a healthy diet than it is taught by the BZfE. This could lead to a decrease in self-efficacy since the barriers will appear harder to overcome if the goal is harder to achieve in general. Nevertheless, this also applies the other way around: participants who perceived a balanced and healthy diet harder to achieve before the study, than it actually was within the study, will assess the barriers as lower and therefore be more confident to overcome those.

7.5. Theoretical Contribution

This thesis contributes to the field of conversational AI chat assistants in the domain of health behavior change. More precisely speaking implementing and monitoring a healthy and balanced diet in daily life. Our findings suggest that the approach of tracking food intake in a coarser fashion, with servings as measurement instead of precise caloric tracking, can increase a person's self-reported awareness in regard to their diet. Providing visual feedback in form the *BZfE Nutrition Pyramid* was well received by the participants of our study and assessed as helpful for reflecting on their diet.

As described above we did not identify significant changes in the users' *Health Consciousness*, *Reflective Thinking* and *Nutrition Self-Efficacy* in the seven-day user-study. The results were in line with the theories described in chapter 2. The participants assessed their diet as good at the beginning of the study, were healthy and in the normal weight category. According to Prochaska and Velicer the changes of habits and behavior within the *Action* phase, have to be of an impact that "[...] is sufficient to reduce risks for disease." (Prochaska and Velicer 1997) (page 39). This requirement does not apply to the participants of our study, since none of them have been overweight or obese, thus the changes in diet would have been too small to be evaluated as valid for behavior change. Additionally, the process of actual behavior change takes several months, which was not possible to achieve within this thesis due to the short working period. This opens the possibility to conduct further work on the basis of this thesis, investigating if this approach is reasonable for long term behavior change.

In the following chapter we will describe the limitations of this thesis in more detail, followed by specific approaches for future work.

8. Limitations

While developing the chat assistant RAINA as well as evaluating the system, we faced several limitations. Those limitations and their impact on this thesis are described in this chapter.

The limitation with the most impact on the implementation was the short time of development, due to the time constraint of the thesis. With a short development time in order to be able to conduct a user-study, it has not been possible to integrate the approach of *Conversation Driven Development* as it usually being used. This means that there has not been enough time to fully test the system and collect user generated training data, once the first prototype of RAINA has been finished.

Rather than that, the user-study itself could be seen as the first opportunity to gather a large amount of user-generated input to train the bot. The data on which the system has been trained prior to the study was written manually by a single developer, resulting in insufficient coverage of examples for some intents. During the study, the users sent over a thousand messages to RAINA. Based on those messages it would be possible to retrain the model, to enhance the *Intent* and *Entity* recognition as well as add all synonyms and foods RAINA did not know while the study took place.

As described in subsection 4.1.3, the *Duckling Entity Extractor* did not perform as expected, due to the complexity of the German grammar. However, we have been dependent on this pipeline component since the functionality of extracting dates and numbers from plain text has been important for the prototype. Furthermore, especially at the end of the implementation phase, as more and more training examples have been added, we began to notice some hardware limitations: the time to train new models increased the more synonyms and stories were added, slowing down the development process considerably. For further development, more computational power might be needed, especially if implementing completely new functions as they require large amounts of new training data.

When looking at the study, we went for a rather quantitative approach and anonymized the questionnaires and conversations. However, it would have been better to recruit more participants since then the users who did not take part in all sections of the study would have less of an impact on the results. Another approach would be to match both of the questionnaires and the conversation to a user by an ID, to be able to remove incomplete data sets. This would also give us the opportunity to track the development of individual users and analyze it more precisely, leading to a more qualitative evaluation. Furthermore, the participants selected for the study were acquaintances of the author, resulting in similar demographic information: most of the participants being university students in their early twenties. Also, bias in answering the questions cannot be completely ruled out due to a personal relationship with the author.

In addition to having a greater group of subjects in the study, the duration of the study should also be extended to a few weeks up to several months. At this point we are not able to make statements on how RAINA actually affects behavior change over the long term, since as described in chapter 2 the process of behavior change is a process taking a long period of time (Prochaska and Velicer 1997). This was not possible to evaluate in the short timeframe of this thesis. Following up in this, the subjects for further studies should be selected in a way to get a group of participants who would not already reach high scores in *Health Consciousness*, *Reflective Thinking* and *Nutrition Self-Efficacy* before the study starts. If the scores at the beginning are lower, there is a higher chance in increasing them over the course of a study since there is more room for improvement. Kocielnik, Xiao, et al. reported a similar issue when evaluating the *Reflection Companion* and refer to it as "a known outcome or limitation of bounded scales" (Kocielnik, Xiao, et al. 2018) (page 13).

In the next chapter we describe our ideas and for the further development of RAINA and our approach for a possible follow-up user study.

9. Future Work

Based on the results of our study, there are two questions which are interesting to further investigate. Firstly, the timeframe of the study has been too short to evaluate if RAINA supports long term behavior change. As described in chapter 2 the process of changing behavior takes at least several months. Secondly, the initial scores of the participants in our user study have been rather high, leaving little room for improvement. For a follow-up study, the selection of participants should be refined to focus more on the impact on long-term behavior change than on usability of the system. An idea would be to recruit clients of nutritionists, who are interested to use RAINA in addition to their nutritional counselling. According to Prochaska and Velicer they would be either in the *Preparation* or *Action* phase (Prochaska and Velicer 1997). Future work could also even try to aim for persons in the *Contemplation* phase trying to convince them to give RAINA a try and thus move into the *Preparation* phase.

While the questionnaires may be re-used for a future study with little modifications, RAINA herself needs some adjustments and improvements. The model will be sharpened with the large amount of training data, we gathered while the study took place. Using RasaX, we can match user generated messages to the existing *Intents* efficiently and create new *Stories* based on fragments of the chat histories. Furthermore, we can investigate each individual conversation in detail and determine if RAINA needs to learn new *Intents* and with them new functionality and responses. The responses RAINA currently replies with could also be enhanced by multiple new messages or wordings for existing responses, making the conversations less repetitive. Furthermore, we will train RAINA to understand all additional synonyms for the food groups which the study participants used, but RAINA did not know at the time of the study.

Additionally, we will enhance existing functionality as well as add new functions, based on the detailed feedback we received. Many proposed improvements had to do with the retrospective insertion of the servings into the pyramid. While it has been a design decision to not be able to edit past pyramids in order to prevent cheating, some participants reported that they forgot to add servings and would have liked to add them the next day when they noticed. Pyramids which are not filled properly will negatively affect the weekly report, so we decided to add the functionality to edit the pyramid of the previous day, but not further back in time. This way one could add servings they forgot, but not manipulate the weekly overview.

Another enhancement is to teach RAINA to understand multiple inputs in one message, making the entry of servings into of the pyramid more convenient and faster. Based on this feature and the mentioned idea of integrating recipes, we will include the ability to create personal recipes to be able to add multiple servings of different food groups at once. However, those pre-defined dishes will not be publicly available to all users, since the method of preparation as well as the amount of ingredients may vary.

Last but not least, the weekly overview will be enhanced by a score describing the pyramid. This score will be calculated based on the weighted differences between the food groups: deviations in the green groups at the bottom of the pyramid will weigh less than the red groups on top of the pyramid, resulting in a rating from A (best) to F (worst) for a pyramid.

10. Conclusion

The goal of this thesis was to implement a conversational AI chat assistant for dietary monitoring using a coarser approach to nutrition tracking than conventional nutrition and diet applications. We provided information on underlying theories of behavior change in health: firstly, the *Transtheoretical Model* by Prochaska and Velicer describing the process of behavior change in seven stages (Prochaska and Velicer 1997). Secondly, we described some popular *Behavior Change Techniques* from the *Behavior Change Technique Taxonomy (v1)* by Michie, Richardson, et al. which support people changing their habits (Michie, Richardson, et al. 2013). Furthermore, we introduced the concept of *Nutrition Pyramids* for teaching and implementing a healthy diet and described the system by the *BZfE* in detail, which we also used to base our proof-of-concept implementation on. We described conventional nutrition applications and their drawbacks and presented related work in the domain of health behavior change supported by chat bots, focusing mainly on weight management and diet improvement.

The implementation of RAINA, our virtual dietary assistant, is based on the *Rasa Framework* and accessible through the *Telegram Messenger*. RAINA teaches the users the system of the *BZfE Nutrition Pyramid* and tracks a person's diet using servings of the food groups used in the *Nutrition Pyramid* as metric. Furthermore, she checks in on the users twice a day and is able to answer simple questions about the *Nutrition Pyramid*. Rather than providing information on the diet in plain numbers of consumed servings, RAINA generates an image of the pyramid on run-time, whenever a user want to take a look at their consumed foods. Every evening before dinner, RAINA reaches out to the users and shows them their current pyramid for the day to provide an overview to supports users to prepare dinner accordingly. Furthermore, users are asked how well they were able to integrate the pyramid for this day in the same check-in.

We evaluated our approach in a seven-day user-study with one questionnaire each before and after the testing period. RAINA was well received by the participants, providing new insights about their diet. The representation of pyramid was easy to understand and keep in mind. The approach of tracking the diet with servings and visualizing it with the image of the *Nutrition Pyramid* instead of tracking calories precisely was perceived positively. We did not identify significant changes in the participants *Health Consciousness*, *Reflective Thinking* and *Nutrition Self-Efficacy* within the seven days. While the *Health Consciousness* decreased minimally, the average *Reflective*

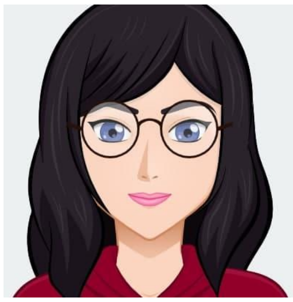
Thinking score remained at the same level when comparing the items used in both questionnaires and decreased slightly when adding the questions about RAINA to calculation of the average scores. The *Nutrition Self-Efficacy* increased slightly with the deviation of the scores decreasing, meaning the outlier values have moved closer to the average score, compared to the pre-study questionnaire.

Limitations for this thesis were the short development time for the prototype and the small time period in which the study took place. A bug at the start of the conversation with RAINA caused some of the users not getting the check-in messages. The missing cycles of *Conversation Driven Development* were noticeable when users tried to use synonyms which RAINA did not know at the time of the study. Furthermore, the results of the study were influenced by missing usage data for the chat assistant and post-study questionnaires of three users. To eliminate those limitations, future work can further enhance RAINA's exiting functionality as well as implement new functions based on user feedback. The study period can be chosen to be several weeks up to months to investigate RAINA's influence on long-term behavior change.

Overall, the approach we selected and implemented in our prototype constitutes a promising starting point for teaching a healthy and balanced diet and implementing in the daily life. It was well perceived by the study participants and might be suitable to support long-term behavior change, which has to be validated by further studies.

A. RAINA Instructions

RAINA – Anleitung



Einführung

1. Suche RAINA auf Telegram: @tum_raina_bot
2. Beginne eine Unterhaltung, indem du auf „Start“ klickst
3. RAINA wird dich dazu auffordern die Informationen des BZfE zur Ernährungspyramide auf den entsprechenden Webseiten zu lesen
4. Lies die Informationen durch und befolgen RAINAs Anweisungen genau

Die Einführung ist beendet so bald dir RAINA sagt, dass du sie anschreiben kannst, wenn du etwas eintragen möchtest. Danach kannst du RAINA jederzeit anschreiben, um etwas ein deine Pyramide einzufügen.

Wichtig: RAINA kennt die Namen der Felder der Pyramide und 333 Synonyme. Ganze Gerichte und spezifische Lebensmittel kann sie nicht in die zugehörigen Lebensmittelgruppen aufteilen: das ist deine Aufgabe. Sollte sie einmal nicht verstehen welches Feld/Lebensmittelgruppe gemeint ist, frag sie einfach wie die Felder der Pyramide heißen und verwende die genannten Bezeichnungen.

Hinzufügen und Abziehen von Portionen/Feldern

RAINA erkennt Zahlen, sowohl ausgeschrieben als auch als Nummern. Achte auf die deutsche Schreibweise von Dezimal-Zahlen: z.B. „0,5“ für eine halbe Portion. Solltest du einmal Portionen abziehen wollen, weil zum Beispiel eine falsche Anzahl eingetragen wurde, funktioniert dies analog zum Hinzufügen. Bitte achte darauf pro Nachricht nur eine Lebensmittelgruppe zu bearbeiten!

- *„Ich habe eine Portion Obst gegessen.“*
- *„Ziehe 0,5 Portionen Wasser ab.“*

Einen Überblick über die Pyramide behalten

RAINA kann dich über die aktuell eingetragene Anzahl der Portionen einer genannten Lebensmittelgruppe informieren:

- *„Wie viel Wasser habe ich heute getrunken?“*
- *„Wie viel Obst kann ich heute noch essen?“*

RAINA kann dir deine Pyramide für den aktuellen Tag als Bild senden:

- *„Wie sieht meine Pyramide heute aus?“*
- *„Was habe ich heute gegessen?“*

RAINA kann dir deine Pyramide für ein konkretes Datum zeigen und den Durchschnitt über einen Zeitraum bilden und dir schicken:

- *„Was habe ich gestern gegessen?“*
- *„Wie sah meine Pyramide am 31.12.2021 aus?“*
- *„Zeige mir meine Pyramide vom 24.12.2021 bis zum 31.12.2021.“*

Check-In Fragen

RAINA sendet dir jeden Morgen und Nachmittag eine kleine Check-In Frage. Um diese zu beantworten, klicke einfach auf den passenden Button. Sollte der Text auf einem Knopf abgeschnitten sein: die Antworten sind von schlecht (oberste Antwort) nach sehr gut (unterste Antwort) angeordnet.

FAQ

RAINA kann dir einfache Fragen zu den Lebensmittelgruppen und Portionsgrößen beantworten. Du kannst sie auch fragen, welche Funktionen RAINA aktuell unterstützt und dir jederzeit die Links zum BZfE schicken lassen. Sollte RAINA dein Anliegen einmal nicht verstehen, versuche es anders zu formulieren. Zur Not sieh auf den Websiten des BZfE nach, sollte sie dir nicht weiterhelfen können.

Anmerkungen

Alle Konversationen mit RAINA sowie die Antworten zu den Umfragen und Check-In Fragen werden anonym gespeichert, um sie für die Bachelor-Arbeit zu analysieren und auszuwerten. Solltest du RAINA **nicht** über den vollen Studien-Zeitraum nutzen, fülle bitte dennoch die zweite Umfrage aus und gib deine Gründe dafür an!

Sollte es Fragen oder Probleme geben kontaktiere mich gerne unter: anton.steuer@tum.de

Links zum BZfE

Ernährungspyramide:

https://www.bzfe.de/ernaehrung/die-ernaehrungspyramide/die-ernaehrungspyramide-eine-fuer-alle/

Lebensmittelgruppen:

https://www.bzfe.de/ernaehrung/die-ernaehrungspyramide/die-ernaehrungspyramide-eine-fuer-alle/ernaehrungspyramide-was-esse-ich/

Menge an Portionen:

https://www.bzfe.de/ernaehrung/die-ernaehrungspyramide/die-ernaehrungspyramide-eine-fuer-alle/ernaehrungspyramide-wie-viel-esse-ich/

Portionsgrößen:

https://www.bzfe.de/ernaehrung/die-ernaehrungspyramide/die-ernaehrungspyramide-eine-fuer-alle/ernaehrungspyramide-wie-gross-ist-eine-portion/

B. Pre-Study Questionnaire

Demographic Information

[D1] Alter
[D2] Körpergröße
[D3] Körpergewicht
[D4] Geschlecht
[D5] abgeschlossene Ausbildung
[D6] aktueller Beruf
[D7] Haben Sie bereits an einer professionellen Ernährungsberatung teilgenommen?
[D8] Haben Sie bereits eine App genutzt, um einen Überblick über Ihre Ernährung zu behalten und Kalorien zu zählen?

Health Consciousness

[H1] Ich reflektiere meine Gesundheit sehr oft.
[H2] Ich bin ständig über meine Gesundheit besorgt.
[H3] Ich übernehme Verantwortung für meine Gesundheit.
[H4] Eine gute Gesundheit erfordert aktive Teilnahme meinerseits.
[H5] Ich mache mir nur Gedanken über meine Gesundheit, wenn ich krank bin.

Reflective Thinking

- [R1] Wenn ich eine Tätigkeit ausübe, kann ich sie tun, ohne darüber nachzudenken, was ich tue.
- [R2] So lange ich mir die Informationen, die ich anfangs gelernt habe, merken kann, muss ich nicht viel nachdenken.
- [R9] Ich stelle manchmal die Art und Weise, wie andere etwas tun in Frage und versuche eine bessere Lösung zu finden.
- [R10] Ich denke oft darüber nach was ich getan habe und überlege mir alternative Wege wie ich es tun könnte.
- [R11] Ich denke oft über mein Handeln nach, um zu überlegen, ob ich es hätte verbessern könnte.
- [R12] Ich bewerte meine Erfahrungen oft neu, um daraus zu lernen und mich für meinen nächsten Auftritt zu verbessern.

Nutrition Self-Efficacy

Ich kann es schaffen, mich an gesunde Lebensmittel zu halten, ...

- [S1] ... auch wenn ich eine lange Zeit brauche die notwendigen Routinen zu entwickeln.
- [S2] ... auch wenn ich es mehrfach versuchen muss, bis es funktioniert.
- [S3] ... auch wenn ich meine gesamte Ernährungsweise überdenken muss.
- [S4] ... auch wenn ich bei meinen ersten Versuchen nicht viel Unterstützung von anderen erfahre.
- [S5] ... auch wenn ich einen detaillierten Plan erstellen muss.

Nutrition Assessment

[A1] Ich habe keine Schwierigkeiten ein gesundes Gewicht für meine Größe und mein Alter zu halten.

[A2] Ich verfolge eine ausgewogene Ernährung, auch wenn es stressig ist oder ich sehr beschäftigt bin.

[A3] Ich bin motiviert genug gesunde Gerichte zuzubereiten, wenn ich allein bin.

[A4] Ich achte auf eine ausgewogene Ernährung, auch wenn ich nicht zuhause esse.

[A5] Ich esse kein ungesundes Essen oder Junkfood (z.B. Fastfood, Chips, etc.).

[A6] Ich lehne ungesundes Essen ab, wenn es mir angeboten wird.

[A7] Ich weiß welche Nährstoffe eine gesunde Ernährung ausmachen.

[A8] Ich kann Gerichte so zubereiten, dass der Nährwert der Zutaten nicht verringert wird.

[A9] Ich erkenne gesunde Lebensmittel, wenn ich einkaufen gehe.

[A10] Ich überprüfe die Wahrheit und Genauigkeit von Ernährungs-Fakten, die ich über Familie, Freunde oder Medien erfahre...
... überhaupt nicht.
... selbst mit Hilfe des Internets.
... selbst mit Hilfe von Fachliteratur.
... in einer professionellen Ernährungsberatung.

C. Post-Study Questionnaire

Health Consciousness

[H1] Ich reflektiere meine Gesundheit sehr oft.
[H2] Ich bin ständig über meine Gesundheit besorgt.
[H3] Ich übernehme Verantwortung für meine Gesundheit.
[H4] Eine gute Gesundheit erfordert aktive Teilnahme meinerseits.
[H5] Ich mache mir nur Gedanken über meine Gesundheit, wenn ich krank bin.

Reflective Thinking

[R1] Wenn ich eine Tätigkeit ausübe, kann ich sie tun, ohne darüber nachzudenken, was ich tue.
[R2] So lange ich mir die Informationen, die ich anfangs gelernt habe, merken kann, muss ich nicht viel nachdenken.
[R3] Während der Studie habe ich Dinge so oft gemacht, dass ich begonnen habe sie zu tun, ohne darüber nachzudenken.
[R4] Wenn ich den Anweisungen von RAINA folge, muss ich nicht allzu viel nachdenken.
[R5] Für diese Studie musste ich die von RAINA vermittelten Konzepte verstehen.
[R6] Um diese Studie zu absolvieren, war es nötig den Inhalt zu verstehen.
[R7] Ich musste die von RAINA vermittelten Inhalte verstehen, um den Chat-Assistenten nutzen zu können.
[R8] Während der Studie musste ich beständig über den Inhalt nachdenken, der mir beigebracht wurde.
[R9] Ich stelle manchmal die Art und Weise, wie andere etwas tun in Frage und versuche eine bessere Lösung zu finden.
[R10] Ich denke oft darüber nach was ich getan habe und überlege mir alternative Wege wie ich es tun könnte.
[R11] Ich denke oft über mein Handeln nach, um zu überlegen, ob ich es hätte verbessern könnte.
[R12] Ich bewerte meine Erfahrungen oft neu, um daraus zu lernen und mich für meinen nächsten Auftritt zu verbessern.

Nutrition Self-Efficacy

Ich kann es schaffen, mich an gesunde Lebensmittel zu halten, ...

[S1] ... auch wenn ich eine lange Zeit brauche die notwendigen Routinen zu entwickeln.
[S2] ... auch wenn ich es mehrfach versuchen muss, bis es funktioniert.
[S3] ... auch wenn ich meine gesamte Ernährungsweise überdenken muss.
[S4] ... auch wenn ich bei meinen ersten Versuchen nicht viel Unterstützung von anderen erfahre.
[S5] ... auch wenn ich einen detaillierten Plan erstellen muss.

Questions about RAINA

[RA1] Die kurze Einführung von RAINA hat mir geholfen die Pyramide noch einmal zu wiederholen.
[RA2] Ich musste meine Ernährung anpassen, um die Pyramide zu treffen.
[RA3] Ich habe mir bereits vor den Mahlzeiten (beim Einkaufen oder beim Kochen) Gedanken über die Felder der Pyramide gemacht.
[RA4] Ich habe RAINA immer direkt informiert, wenn ich etwas gegessen/getrunken habe.
[RA5] Ich habe oft vergessen RAINA zu schreiben.
[RA6] Ich habe die vergessenen Portionen immer abends nachgetragen.
[RA7] Die Visualisierung der Pyramide hat mir geholfen meine Ernährung einzuschätzen.
[RA8] Die Darstellung der Pyramide ist mir gut im Gedächtnis geblieben.

Questions about the Check-Ins

[C1] Ich habe Check-In Fragen/Erinnerungen von RAINA erhalten.
[CY1] Die Nachrichten von RAINA haben mir geholfen dabei zu bleiben.
[CY2] Ich hätte gerne mehr Nachrichten/Erinnerungen von RAINA bekommen.
[CN1] Ich hätte gerne Erinnerungen bekommen.
[CN2] Durch Erinnerungen hätte ich öfter mit RAINA interagiert.

Questions with Free Textual Answers

[F1] Wie hat sich die Nutzung des Systems auf Ihre Motivation, sich ausgewogen zu ernähren, ausgewirkt?

[F2] Wie waren Ihre generellen Erfahrung bei der Nutzung von RAINA? Was hat gut funktioniert und wo gab es Probleme?

[F3] Welche Dinge haben Sie über Ihr Ernährungs-Verhalten gelernt?

[F4] Würden Sie das System nach der Studie privat weiter nutzen? Welche Gründe gibt es für Ihre Entscheidung?

[F5] Haben Sie sonstige Anmerkungen? (z.B. Gründe für ein vorzeitiges Beenden der Nutzung während der Studiendauer, gewünschte Funktionen etc.)

D. Aggregated Scores

	Health Consciousness	Reflective Thinking	Self-Efficacy	M	SD
1	2.5	3.5	2.0	2.67	0.62
2	3.0	3.5	3.5	3.33	0.24
3	3.0	4.0	3.5	3.5	0.41
4	3.0	4.0	4.0	3.67	0.47
5	3.5	3.0	3.0	3.17	0.24
6	3.5	3.0	4.0	3.5	0.41
7	3.5	3.5	3.0	3.33	0.24
8	3.5	3.5	4.0	3.67	0.24
9	3.5	3.5	4.0	3.67	0.24
10	3.5	4.0	2.5	3.33	0.62
11	3.5	4.0	3.5	3.67	0.24
12	3.5	4.0	5.0	4.17	0.62
13	3.5	4.5	4.0	4.0	0.41
14	4.0	3.0	4.0	3.67	0.47
15	4.0	3.5	3.5	3.67	0.24
16	4.0	3.5	4.0	3.83	0.24
17	4.0	4.0	2.5	3.5	0.71
18	4.0	4.0	3.0	3.67	0.47
19	4.0	4.0	3.5	3.83	0.24
20	4.0	4.0	5.0	4.33	0.47
21	4.5	3.5	3.5	3.83	0.47

Table D.1.: Aggregated Scores for each User (Pre-Study, ordered ascending by Health Assessment, Reflective Thinking, Self-Efficacy)

	Health Consciousness	Reflective Thinking	Self-Efficacy	M	SD
1	2.5	3.0	2.5	2.67	0.24
2	3.0	3.0	4.0	3.33	0.47
3	3.0	3.5	3.0	3.17	0.24
4	3.0	3.5	4.0	3.5	0.41
5	3.0	4.0	3.5	3.5	0.41
6	3.0	4.0	4.0	3.67	0.47
7	3.0	4.0	4.0	3.67	0.47
8	3.5	3.0	4.0	3.5	0.41
9	3.5	3.0	4.0	3.5	0.41
10	3.5	3.5	3.5	3.5	0.0
11	3.5	4.0	3.5	3.67	0.24
12	3.5	4.0	4.0	3.83	0.24
13	3.5	4.0	4.0	3.83	0.24
14	3.5	4.5	3.5	3.83	0.47
15	4.0	3.5	4.0	3.83	0.24
16	4.0	4.0	4.0	4.0	0.0
17	4.0	4.0	4.0	4.0	0.0
18	4.5	4.0	4.0	4.17	0.24

Table D.2.: Aggregated Scores for each User (Post-Study, ordered ascending by Health Assessment, Reflective Thinking, Self-Efficacy)

List of Figures

List of Tables

Bibliography

Achananuparp, P., E.-P. Lim, and V. Abhishek (2018). "Does Journaling Encourage Healthier Choices?" In: ACM. DOI: 10.1145/3194658.3194663.

Bandura, A. (1977). "Self-efficacy: Toward a unifying theory of behavioral change." In: *Psychological Review* 84.2, pp. 191–215. ISSN: 1939-1471(Electronic),0033-295X(Print). DOI: 10.1037/0033-295X.84.2.191.

Baumer, E. P., V. Khovanskaya, M. Matthews, L. Reynolds, V. S. Sosik, and G. Gay (2014). *Reviewing reflection: on the use of reflection in interactive system design*. Conference Paper. DOI: 10.1145/2598510.2598598.

Brüggemann, I. (2018). *Die Ernährungspyramide - Richtig essen lehren und lernen*. Pamphlet.

BZfE (2020). *Ernährungspyramide: Wie groß ist eine Portion?* URL: https://www.bzfe.de/ernaehrung/die-ernaehrungspyramide/die-ernaehrungspyramide-eine-fuer-alle/ernaehrungspyramide-wie-gross-ist-eine-portion/. (accessed: 09.01.2022).

– (2021). *Ernährungspyramide: Was esse ich?* URL: https://www.bzfe.de/ernaehrung/die-ernaehrungspyramide/die-ernaehrungspyramide-eine-fuer-alle/ernaehrungspyramide-was-esse-ich/. (accessed: 09.01.2022).

Calvaresi, D., J.-P. Calbimonte, F. Debusson, A. Najjar, and M. Schumacher (2019). "Social Network Chatbots for Smoking Cessation: Agent and Multi-Agent Frameworks." In: ACM. DOI: 10.1145/3350546.3352532.

Casas, J., E. Mugellini, and O. A. Khaled (2018). *Food Diary Coaching Chatbot*. Conference Paper. DOI: 10.1145/3267305.3274191.

Chan, A., R. Kow, and J. K. Cheng (2017). "Adolescents' Perceptions on Smartphone Applications (Apps) for Health Management." In: *Journal of Mobile Technology in Medicine*. DOI: doi:10.7309/jmtm.6.2.6.

Cordeiro, F., D. A. Epstein, E. Thomaz, E. Bales, A. K. Jagannathan, G. D. Abowd, and J. Fogarty (2015). "Barriers and Negative Nudges: Exploring Challenges in Food Journaling." In: ACM. DOI: 10.1145/2702123.2702155.

El Kamali, M., L. Angelini, M. Caon, G. Andreoni, O. A. Khaled, and E. Mugellini (2018). "Towards the NESTORE e-Coach." In: ACM. DOI: 10.1145/3267305.3274188.

Fadhil, A., G. Schiavo, and Y. Wang (2019). "CoachAI: A Conversational Agent Assisted Health Coaching Platform." In: *arXiv pre-print server*. DOI: arxiv:1904.11961.

Flothkötter, M. (2021). *Ernährungspyramide: Wie viel esse ich?* URL: https://www.bzfe.de/ernaehrung/die-ernaehrungspyramide/die-ernaehrungspyramide-eine-fuer-alle/ernaehrungspyramide-wie-viel-esse-ich/. (accessed: 08.01.2022).

Franco, R. Z., R. Fallaize, J. A. Lovegrove, and F. Hwang (2016). "Popular Nutrition-Related Mobile Apps: A Feature Assessment." In: *JMIR mHealth and uHealth* 4.3, e85. ISSN: 2291-5222. DOI: 10.2196/mhealth.5846.

Fulmer, R., A. Joerin, B. Gentile, L. Lakerink, and M. Rauws (2018). "Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial." In: *JMIR Mental Health* 5.4, e64. ISSN: 2368-7959. DOI: 10.2196/mental.9782.

Gentner, T., T. Neitzel, J. Schulze, and R. Buettner (2020). "A Systematic Literature Review of Medical Chatbot Research from a Behavior Change Perspective." In: IEEE. DOI: 10.1109/compsac48688.2020.0-172.

Hankonen, N., S. Sutton, A. T. Prevost, R. K. Simmons, S. J. Griffin, A. L. Kinmonth, and W. Hardeman (2015). "Which Behavior Change Techniques are Associated with Changes in Physical Activity, Diet and Body Mass Index in People with Recently Diagnosed Diabetes?" In: *Annals of Behavioral Medicine* 49.1, pp. 7–17. ISSN: 0883-6612. DOI: 10.1007/s12160-014-9624-9.

Hipp, Wyrick & Company (2021). *How SQLite Works*. URL: https://www.sqlite.org/howitworks.html. (accessed: 27.12.2021).

Ho, A., J. Hancock, and A. S. Miner (2018). "Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot." In: *Journal of Communication* 68.4, pp. 712–733. ISSN: 0021-9916. DOI: 10.1093/joc/jqy026.

Hong, H. (2009). "Scale development for measuring health consciousness: Re-conceptualization." In: *that Matters to the Practice*, p. 212.

Horner, L. (2019). "Dialogsystembasierte Ernährungserhebung für einen virtuellen Ernährungsberater." unpublished bachelor's thesis supervised by Prof. Dr. Georg Groh, advised by Monika Wintergerst, Technical University of Munich, Munich, Germany.

Kember, D., D. Y. P. Leung, A. Jones, A. Y. Loke, J. McKay, K. Sinclair, H. Tse, C. Webb, F. K. Yuet Wong, M. Wong, and E. Yeung (2000). "Development of a Questionnaire to Measure the Level of Reflective Thinking." In: *Assessment & Evaluation in Higher Education* 25.4, pp. 381–395. ISSN: 0260-2938. DOI: 10.1080/713611442.

Kocielnik, R., L. Xiao, D. Avrahami, and G. Hsieh (2018). "Reflection Companion." In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.2, pp. 1–26. ISSN: 2474-9567. DOI: 10.1145/3214273.

Latif, S., R. Rana, J. Qadir, A. Ali, M. A. Imran, and M. S. Younis (2017). "Mobile Health in the Developing World: Review of Literature and Lessons from a Case Study." In: *IEEE Access* 5, pp. 11540–11556. DOI: 10.1109/ACCESS.2017.2710800.

Li, W. (2020). "Detection and Dialog-Based Self-Reporting of Stress for Eating Behavior Prediction." unpublished master's thesis supervised by Prof. Dr. Georg Groh, advised by Monika Wintergerst, Technical University of Munich, Munich, Germany.

Luo, T. C., A. Aguilera, C. R. Lyles, and C. A. Figueroa (2021). "Promoting Physical Activity Through Conversational Agents: Mixed Methods Systematic Review." In: *Journal of Medical Internet Research* 23.9, e25486. ISSN: 1438-8871. DOI: `10.2196/25486`.

Maher, C. A., C. R. Davis, R. G. Curtis, C. E. Short, and K. J. Murphy (2020). "A Physical Activity and Diet Program Delivered by Artificially Intelligent Virtual Health Coach: Proof-of-Concept Study." In: *JMIR mHealth and uHealth* 8.7, e17558. ISSN: 2291-5222. DOI: `10.2196/17558`.

Michie, S., M. Richardson, M. Johnston, C. Abraham, J. Francis, W. Hardeman, M. P. Eccles, J. Cane, and C. E. Wood (2013). "The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions." In: *Annals of Behavioral Medicine* 46.1, pp. 81–95. ISSN: 0883-6612. DOI: `10.1007/s12160-013-9486-6`.

Milne-Ives, M., C. De Cock, E. Lim, M. H. Shehadeh, N. De Pennington, G. Mole, E. Normando, and E. Meinert (2020). "The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review." In: *Journal of Medical Internet Research* 22.10, e20346. ISSN: 1438-8871. DOI: `10.2196/20346`.

Mori, M., K. Macdorman, and N. Kageki (2012). "The Uncanny Valley [From the Field]." In: *IEEE Robotics & Automation Magazine* 19.2, pp. 98–100. ISSN: 1070-9932. DOI: `10.1109/mra.2012.2192811`.

Nadarzynski, T., O. Miles, A. Cowie, and D. Ridge (2019). "Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study." In: *Digital Health* 5. ISSN: 2055-2076. DOI: `10.1177/2055207619871808`.

Nass, C., J. Steuer, and E. R. Tauber (1994). "Computers are social actors." In: ACM Press. DOI: `10.1145/191666.191703`.

Nutrition Australia (2014). *Healthy Eating Pyramid*. URL: `https://nutritionaustralia.org/fact-sheets/healthy-eating-pyramid/`. (accessed: 08.01.2022).

Park, S., J. Choi, S. Lee, C. Oh, C. Kim, S. La, J. Lee, and B. Suh (2019). "Designing a Chatbot for a Brief Motivational Interview on Stress Management: Qualitative Case Study." In: *Journal of Medical Internet Research* 21.4, e12231. ISSN: 1438-8871. DOI: `10.2196/12231`.

Pereira, J. and Ó. Díaz (2019). "Using Health Chatbots for Behavior Change: A Mapping Study." In: *Journal of Medical Systems* 43.5. ISSN: 0148-5598. DOI: `10.1007/s10916-019-1237-1`.

Perski, O., D. Crane, E. Beard, and J. Brown (2019). "Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study." In: *DIGITAL HEALTH* 5, p. 205520761988067. ISSN: 2055-2076. DOI: 10.1177/2055207619880676.

Prasetyo, P. K., P. Achananuparp, and E.-P. Lim (2020). "Foodbot: A Goal-Oriented Just-in-Time Healthy Eating Interventions Chatbot." In: ACM. DOI: 10.1145/3421937.3421960.

Prochaska, J. O. (2008). "Decision Making in the Transtheoretical Model of Behavior Change." In: *Medical Decision Making* 28.6, pp. 845–849. ISSN: 0272-989X. DOI: 10.1177/0272989x08327068.

Prochaska, J. O. and W. F. Velicer (1997). "The Transtheoretical Model of Health Behavior Change." In: *American Journal of Health Promotion* 12(1), pp. 38–48.

Rasa Technologies (2021a). *Conversation-Driven Development*. URL: https://rasa.com/docs/rasa/conversation-driven-development/. (accessed: 29.12.2021).

– (2021b). *Rasa Architecture Overview*. URL: https://rasa.com/docs/rasa/arch-overview/. (accessed: 01.01.2022).

Schwarzer, R. and B. Renner (2009). "Health-Specific Self-Efficacy Scales." In:

Seitz, H. (2020). *Die Ernährungspyramide*. URL: https://www.bzfe.de/ernaehrung/die-ernaehrungspyramide/die-ernaehrungspyramide-eine-fuer-alle/. (accessed: 08.01.2022).

Shamsalinia, A., R. Ghadimi, R. T. Chafjiri, F. Norouzinejad, A. Pourhabib, and F. Ghaffari (2019). "Nutrition self-efficacy assessment: designing and psychometric evaluation in a community-dwelling elderly population." In: *Journal of Health, Population and Nutrition* 38.1. ISSN: 2072-1315. DOI: 10.1186/s41043-019-0203-3.

Sherman, S. J. (1980). "On the self-erasing nature of errors of prediction." In: *Journal of Personality and Social Psychology* 39.2, pp. 211–221. ISSN: 1939-1315(Electronic),0022-3514(Print). DOI: 10.1037/0022-3514.39.2.211.

Statista (2017). *Umfrage zur Nutzung von E-Health-Apps für ausgewählte Zwecke in Deutschland 2017*. URL: https://de.statista.com/statistik/daten/studie/698074/umfrage/umfrage-zur-nutzung-von-e-health-apps-fuer-ausgewaehlte-zwecke-in-deutschland/. (accessed: 08.01.2022).

Stephens, T. N., A. Joerin, M. Rauws, and L. N. Werk (2019). "Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot." In: *Translational Behavioral Medicine* 9.3, pp. 440–447. ISSN: 1869-6716. DOI: 10.1093/tbm/ibz043.

Strong, K. A., S. L. Parks, E. Anderson, R. Winett, and B. M. Davy (2008). "Weight Gain Prevention: Identifying Theory-Based Targets for Health Behavior Change in Young Adults." In: *Journal of the American Dietetic Association* 108.10, 1708–1715.e3. ISSN: 0002-8223. DOI: 10.1016/j.jada.2008.07.007.

Ta, V., C. Griffith, C. Boatfield, X. Wang, M. Civitello, H. Bader, E. Decero, and A. Loggarakis (2020). "User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis." In: *Journal of Medical Internet Research* 22.3, e16235. ISSN: 1438-8871. DOI: 10.2196/16235.

Vasiloglou, M. F., S. Christodoulidis, E. Reber, T. Stathopoulou, Y. Lu, Z. Stanga, and S. Mougiakakou (2021). "Perspectives and Preferences of Adult Smartphone Users Regarding Nutrition and Diet Apps: Web-Based Survey Study." In: *JMIR mHealth and uHealth* 9.7, e27885. ISSN: 2291-5222. DOI: 10.2196/27885.

West, J. H., L. M. Belvedere, R. Andreasen, C. Frandsen, P. C. Hall, and B. T. Crookston (2017). "Controlling Your "App"etite: How Diet and Nutrition-Related Mobile Apps Lead to Behavior Change." In: *JMIR mHealth and uHealth* 5.7, e95. ISSN: 2291-5222. DOI: 10.2196/mhealth.7410.

Willett Walter C.and Skerrett, P. J. (2005). *Healthy Eating Pyramid*. URL: https://www.hsph.harvard.edu/nutritionsource/healthy-eating-pyramid/. (accessed: 08.01.2022).

Williams, P., L. G. Block, and G. J. Fitzsimons (2004). "When Consumers Do Not Recognize "Benign" Intention Questions as Persuasion Attempts." In: *Journal of Consumer Research* 31.

– (2006). "Simply asking questions about health behaviors increases both healthy and unhealthy behaviors." In: *Social Influence* 1.2, pp. 117–127. ISSN: 1553-4510. DOI: 10.1080/15534510600630850.

World Health Organization (2013). *Obesity: Health consequences of being overweight*. URL: https://www.who.int/news-room/q-a-detail/obesity-health-consequences-of-being-overweight. (01.02.2022).

– (2021a). *Obesity*. URL: https://www.who.int/news-room/facts-in-pictures/detail/6-facts-on-obesity. (01.02.2022).

– (2021b). *Obesity and overweight*. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight. (01.02.2022).

World Obesity Federation (2021). *COVID-19 and Obesity: The 2021 Atlas*. Report. World Obesity Federation.

Zhang, J., Y. J. Oh, P. Lange, Z. Yu, and Y. Fukuoka (2020). "Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint." In: *Journal of Medical Internet Research* 22.9, e22845. ISSN: 1438-8871. DOI: 10.2196/22845.