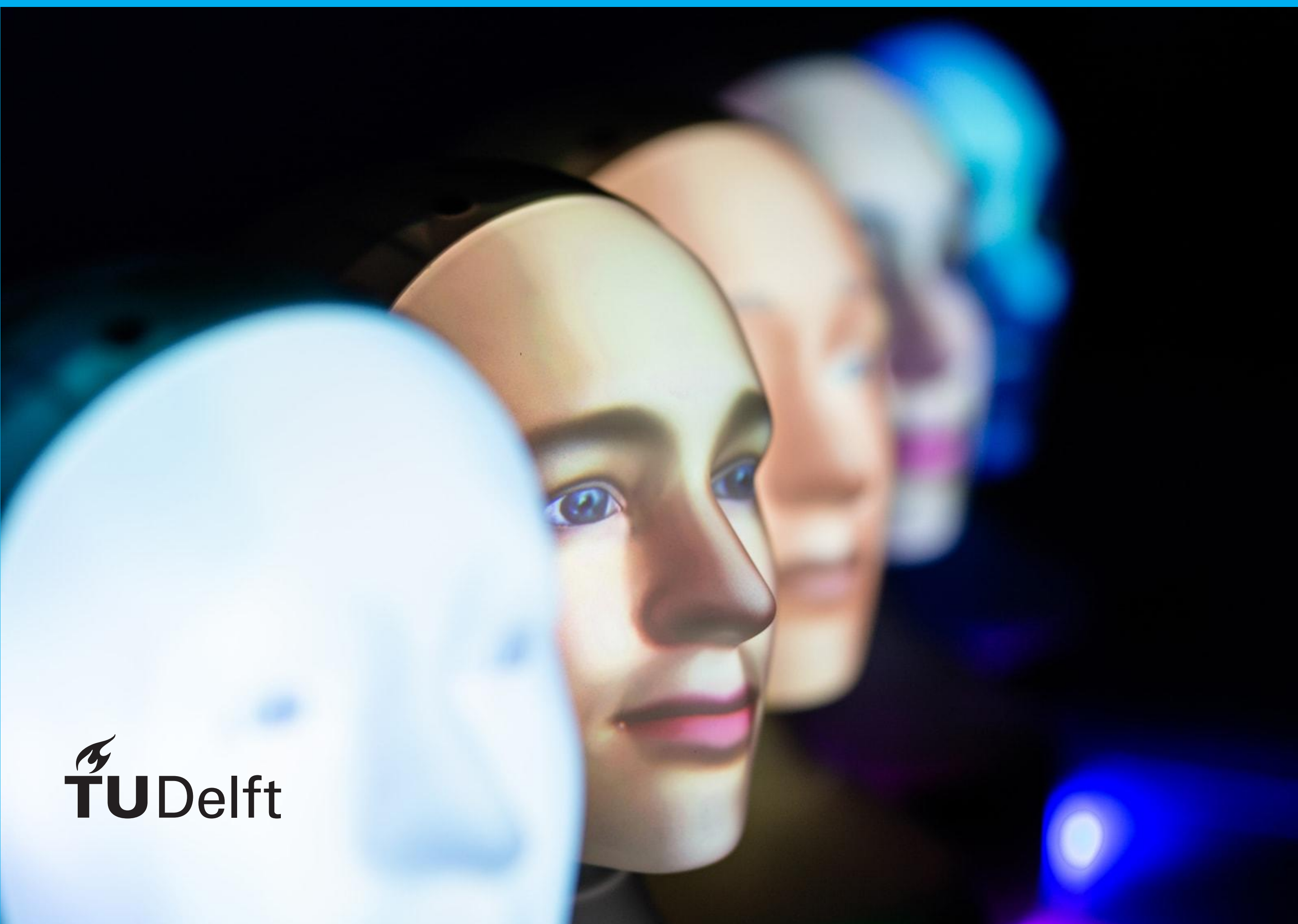
Experience Sharing Conversational Agent for Type II Diabetes

Avinash Saravanan

4993381



Experience Sharing Conversational Agent for Type II Diabetes

by

Avinash Saravanan

to obtain the degree of Master of Science Computer Science
at the Delft University of Technology,
to be defended publicly on Friday July 23, 2021 at 10:30 AM.

Student number:	4993381	
Project duration:	Nov 17, 2020 – July 23, 2021	
Thesis committee:	Prof. dr. M.A. (Mark) Neerincx,	TU Delft, supervisor
	Dr. Catharine Oertel,	TU Delft, co-supervisor
	Dr. Hayley Hung,	TU Delft, Computer Vision Lab

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Preface

Artificial intelligence and many of the newer technologies that utilize it have been the source of much hype and interest within daily news cycles and financial forecasts within the tech sector. However, despite the technology showing much promise, its visibility and presence in our daily lives is limited due to limitations of current methods, lack of human capital devoted to/capable of expanding the current state of the art, or simply a lack of resources.

Many uses of artificial intelligence are deemed as limited due to ambitious expectations and the aforementioned limitations, but even with some of the limitations, it is still capable of facilitating the work and efforts of humans; albeit in a limited capacity.

To explore the potential of the technology in its current state and how effectively it can be applied within a *visible* social context that can have high impact to humanity, an experience sharing robot for type II Diabetes prevention was considered.

Given the social context, and how diet can be a very personal aspect of one's daily routine, the need for a memory for personalization quickly became a focus for the research which led to the question of how memory should be used for creating a more personal connection and provide motivation. This led to the focus on shared memories between the robot and the user being used as a way to provide motivation using details provided by the users themselves. This work is quite unique in that it considers not only the scientific questions involved with such problems, but also the design approach that was taken which allows for a thorough and more holistic perspective that can inform those who wish to practically apply findings in their own work within research as well as outside of research a clear understanding of the practical and scientific considerations that should be taken into account. It is my hope that the results of this work can possibly lead to a future where robots and agents better understand the people they are interacting with and use that understanding to improve their interactions in the same way humans make relations with one another.

I would like to thank and acknowledge all those whose involvement aided in the completion of this project. To begin, I would like to thank my supervisors Mark Neerincx and Catharine Oertel whose guidance facilitated the development of the initial ideas into the final product you are reading now. I wish to thank Hayley Hung for agreeing to be part of the thesis committee.

With regards to technical and administrative support, I would like to express my appreciation for the assistance and advice provided by Bart Vastenhouw, Ruud de Jong, Anita Hoogmoed, Santosh Ilamparuthi, and Jose David Aguas Lopes. Finally, I wish to thank the many individuals I consulted with from the Interactive Intelligence group at TU Delft such as Maria Tsfasman, Levent Gungen, Eric Gu, and Merijn Bruines who provided guidance on different aspects of the prototype and experiment.

I would also like to acknowledge the 4TU Human Technology Research Center and Medical Delta for providing support and funding that allowed for this research to be completed.

Avinash Saravanan
Delft, July 2021

Abstract

In this work, a conversational agent for developing a healthy diet habit over multiple days framed within the context of type II diabetes is designed, developed, and tested. The *primary goal* of this work is to determine how motivational phrases that refer to past events with intrinsic motivational value can affect goal achievement and user experience. This is accomplished by utilizing a novel motivational memory model (m2model) and designing a system that can test and utilize such functionality for improved long term interactions. As such, the research contains both scientific and application based relevance.

Since the agent is used in long term multi-session interactions, a computational memory model is designed and implemented. This model facilitates the conversation over multiple sessions and stores past events for the creation and use of motivational phrases in the later sessions. To implement such functionality, a functioning memory capable of persisting information was needed.

A novel motivational memory was considered because such a memory was expected to better support the agent's primary purpose of encouraging individuals to make a diet change by sharing positive nutrition experiences. In many prior works, the use of a computational memory model in agents has been used more for ensuring topic adherence, difficulty scaling, personalization, or to reduce repetition, but designing a memory specifically for the purpose of motivational rephrasing is lacking in past work. Similarly, rephrasing of experiences with intrinsic value are rare and the combination of the two is lacking in previous literature.

Despite shared experiences and their use for motivational purposes being promising, since a motivational memory is a rarity, how the use of shared experiences rephrased into motivational statements can affect an interaction and how exactly such phrases should be used to optimize an interaction was unclear. This led to the following design questions which were considered during the construction of the agent:

1. How should shared experiences be modeled and organized within the memory?
2. How should shared experiences be modified to generate motivational phrases?
3. When can shared experiences be used?

It should be noted that the design questions cover how shared experiences should be modeled within a memory for motivational rephrasing, but does not ask whether they should be used, and what types of effects they may have. For this, to determine the effect of motivational rephrasing and whether they result in a benefit, the following scientific research questions were formulated:

1. Does referring to previous sessions improve goal attainment (performance)?
2. Does referring to previous sessions improve user experience?
3. Does a variety of references help more than making the same reference? Similarly, can a variety of references over a number of sessions be more useful than a single repeated reference?

In order to determine the answers to the scientific research questions, the following conditions were identified. There are three conditions that are studied to determine the effectiveness of references to shared memories in a between subjects study.

1. Interactions with no references shared experiences.
2. Interactions that only make references to the same shared experience.
3. Interactions that make references to various shared experiences.

To evaluate the differences between these three conditions, metrics relating to achievement of the chosen goal and metrics relating to the user experience are analyzed.

This was accomplished by carrying out a between subjects experiment with three groups of participants where one group was assigned to one condition. Collectively, 79 participants a majority of which was recruited through Prolific engaged with the agent across the three groups through a series of Zoom meetings. The participants completed a Godspeed questionnaire at the end of their three sessions with the conversational agent along with additional questions that provided greater clarity for the particular use case of the conversational agent.

The results, once analyzed, found that there was a significant difference between interactions that only make references to the same shared experience and interactions that make references to various shared experiences where various shared experiences resulted in a significant increase in motivation from prior to the experiment to during the experiment. There was, however, no significant difference in levels of goal achievement between the three conditions, but future works with more challenging goals may offer more clarity with regards to the effect of the memory model on goal achievement. In addition to this, there were some metrics that may have resulted in significant differences, but did not due to low power and as a result, any future works should consider a larger sample size for greater insights.

By completing the design, implementation, and evaluation of the conversational agent equipped with a motivational memory, this study offers the following contributions:

1. Provide greater clarity to how shared experiences should be used and not only whether they should be used.
2. A motivational memory utilizing shared experiences with intrinsic value
3. Determine the effect of motivational references on goal achievement and user experience.

Contents

1 Introduction

The primary purpose of this project is twofold: development and design of an agent and motivational memory model (m2model) capable of experience sharing and determining the effect of shared experiences with personal intrinsic value within the context of long term goal based interaction. Intrinsic value within the context of this work refers to intrinsic motivational value where the past event has some relevance to a user's inner motivations, desires, and concepts that are valued by the user with regards towards their activities. To facilitate this, a conversational agent was developed with a memory that was designed with the intention of encoding events with intrinsic value and rephrasing them into motivational phrases that can be later used in future sessions for praise or encouragement depending on the context.

The context that the conversational agent operates within is that of type II diabetes (T2D) prevention or treatment by choosing and reaching a diet related goal. Type II diabetes is also known as adult-onset diabetes. Type II diabetes is the most common form of diabetes [22] which makes up 90% of all diabetes cases. This constitutes more than 392 million people in the world [22] which is a significant percentage of the global population.

It should also be noted that during the writing of this report, the world embraced distance based technologies to combat the spread of the coronavirus. This makes the use of a conversational agent that can be used to support the efforts of human dietitians particularly compelling because it allows for patients to receive assistance without needing to place themselves at risk by leaving their residence.

The conversational agent that is considered is an agent to assist those interested in improving their diet or who are interested in prevention of T2D. Specifically, the scenario that will be considered is an agent that provides dietary advice to provide more positive outcomes to reduce the possibilities of T2D. The agent will work with the participant over a period of time in three sessions to promote the development of positive habits and utilize the technique of experience sharing to enforce and motivate the user by referring to past experiences. This approach stems from Cognitive Evaluation Theory (CET) [24] which is a sub-theory of the macro theory of self-determination theory (SDT) [93] where external events can influence intrinsic motivation and influence the perception of their levels of competence with regards to a particular activity [25]. By utilizing past experiences, it is possible to infer what a user's perception of their competence is and strengthen or tune their perception accordingly through the use of the externally applied usage of motivational phrases generated from such past events.

This work first considers the design of the agent with respect to this scenario. The project will take a human centered design approach which uses an approach to design as specified within Socio-Cognitive Engineering (SCE) [73]. This implies an approach that considers the different aspects of design in the following order.

1. Foundation (section 2)
2. Specification (section 4)
3. Ontology (section A)
4. Evaluation (section 5)

After the design is completed (as specified in sections 2, 3, and 4), it is implemented and evaluated (as specified in section 5). The evaluation is in the form of a study that is be completed by participants who wished to make an improvement on their diet or who wished to prevent T2D in the future. The participants were asked questions after completing the sessions regarding their experience as well as how they felt about the experience sharing. With regards to experience sharing, this work will evaluate whether referring to past shared experiences with intrinsic improves the overall experience with the conversational agent and achievement, and whether a references to different experiences over multiple sessions rather than the same experience leads to an improvement in the user experience and achievement.

In addition to this, the progress of the participant towards their dietary goal was also recorded to determine how effective the agent was at assisting the participant in achieving their goal. Source code used in this work can be found in the following repository: https://github.com/asarav/MSc-Thesis-Experience-Sharing-Conversational-Agent.

1.1. Background

1.1.1. Long Term Relationships and Experience Sharing

Within long term relationships and self-determination theory, relatedness or social relations are an important aspect of motivational needs. Self-determination theory is a behavioral change theory that has been successfully used in the past for the development of healthy lifestyle support. The theory is explained at further depth in paragraph 1.1.3.

One way of increasing relatedness is with self-disclosure. By sharing personal information between the agent and the user, it may be possible to increase relatedness. One study that examined this was Burger et al. [14] which utilized a disclosure module within the PAL project to foster feelings of relatedness between the children who were the patients and the virtual agents they interacted with.

Another way of increasing relatedness and maintaining long term relationships is discussed by Kasap et al. [48] which is to be able to remember and refer back to past memories in current conversation. These references to shared experiences or memories is the primary area of study within this work. This is an area of study that is still being actively explored. One work that looks into shared experiences is a masters thesis that utilized the PAL project to create an experiment that used an episodic memory to refer to shared experiences [95]. Goedheijft [95] found that the experiences did not result in any significant changes in affect or motivation, but the report did note that the full potential of episodic memory was not tested and that further work was required. Main differences between [95] and this proposal include, but are not limited to: the domain of type II diabetes instead of type I, the type of participant which is adults and the testing of the effect of a variety of shared memories. The type of memory that may be used for this proposal may also differ based on considerations that are brought up during the design phase.

Outside of conversational experiences and relationships, episodic memory has been shown to have useful benefits in intelligent agents [78]. Nuxoll [78] found that the use of episodic memory in pilot experiments such as maze navigation unveiled that an episodic memory or some modification of it can lead to several cognitive capabilities that are vital for general intelligence.

Existing Architectures for Long Term Relationships Many existing systems use architectures like SOAR [55] or ACT-R [2] which are cognitive architectures that use theory of cognition as inspiration. Although these architectures have been used successfully in practice, these are also generalized architectures and were developed before the 21st century and more modern architectures have been developed that build off of these two architectures. The primary purpose of such architectures is to store information before and during the conversation and use that information to determine how to act.

ACT-R splits knowledge into two types of categories: declarative knowledge and procedural knowledge. Declarative knowledge represents information in a discrete way whereas, procedural knowledge within ACT-R consists of processes as the name suggests. To support such types of knowledge, ACT-R uses a declarative memory which stores facts and a procedural memory which stores productions or knowledge regarding how to perform certain actions.

SOAR uses a different approach which tries to reach a goal by searching through a problem space to find appropriate solutions to reach a particular goal. Such solutions are composed of operators and whether or not to apply them given contexts and situations which utilize modules within the architecture. Of these modules, two are related to memory: procedural memory and working memory. The procedural memory is used with the working memory to apply certain procedures based on conditional rules to the current situation which is represented by the working memory.

Kasap et al. [48] in a continuation of the work in [49] presented a conversational agent that utilized an architecture that had three main components to maintain a long term relationship within the context of tutoring.

The three components used were a Belief-Desire-Intention (BDI) architecture, Hierarchical Task Network (HTN), and an episodic memory. The HTN allows for larger tasks to be divided into smaller tasks to allow for concepts to be taught in more manageable segments. The episodic memory is split between a short term and long term memory where the short term memory contains the entire conversation and the long term memory contains the processed data which holds only the most useful or relevant data that can be used in future conversations. In this project, the memory that will be used will be similar, but will be designed specifically for the purpose of diet management and will therefore be more specialized and stripped down in comparison.

Elvir et al. [30] describes and implement a conversational memory or a unified episodic memory architecture for remembering conversations. The paper describes how it encodes memories and provides implementation details such as how memories are retrieved and stored within a backend database. It also explains how to annotate statements made by the user to derive relevant features that can then be used in the prioritization and organization of individual memories within the database. The researchers found that an ECA with memory was able to answer questions more effectively than the same ECA without memory. This architecture does not implement a forgetting mechanism where memories may decay over time. In future work, they consider incorporation of multi-modal input and an ability to handle extreme loads as a possible avenue for improvement of their architecture.

1.1.2. Conversational Agents

Conversational agents have a variety of applications from general chit chat to task-specific purposes [45]. The types of conversational agents range from simple chatbots to fully embodied robots. Types of embodiments can include, but are not limited to chat bots, wearables, virtual agents, and robots.

Smartphone based conversational agents such as Siri, Google Assistant, S Voice, and Cortona have reached a level of ubiquity and use where some users feel comfortable enough to ask for health based information even when such conversational agents have not been equipped to fully handle such critical topics [66].

While all conversational agents have some form of dialog manager [41] which determines what responses to provide to the user based on the statements made by the user, the way in which the user may interact with a chatbot will depend on the embodiment. For example, a chatbot receives input from the user in the form of text. Conversely, a robot or system equipped with speakers and a microphone may be capable of communicating through speech.

The conversational agent that is considered in this work is that of a task based agent that is limited to the domain of diet and it utilizes a Furhat robot which converts to speech to text to process the user's statements before processing such statements in a dialog manager.

Memory in Conversational Agents Within conversational agents, determining how to model and design the memory of an agent is an active area of research. Many take the approach of utilizing a cognitive architecture which has been a subject of research for over 40 years [53]. These architectures draw inspiration from the workings of the human brain to determine whether to include a working memory, procedural memory, episodic memory, or some variation of the aforementioned memory modules [54]. Within paragraph 1.1.1, a more detailed explanation of the architectures or SOAR and Act-R are provided along with approaches to long term interactions. Since such architectures typically draw inspiration from human memory, such memory models have limitations since there are disagreements regarding how human memory operates and the area is still an active area of research [57]. It should also be noted that certain architectures prioritize certain aspects of an interaction similar to how the memory defined in this work prioritizes motivational. For example, for long term interactions, autobiographic memories may be used to avoid repetition within a conversational agent [43] and to ensure a sense of narrative consistency throughout an interaction with regards to a sense of the role the agent may play and the role that is held by the user. Such an approach does have some similarity to that of the motivational memory proposed in this work, but where the autobiographic memory may be used to construct a life story using past events, it does not intentionally bias the past event in one way or another which in the case of the motivational memory is needed to change the inclinations of the user towards being in favor of reaching a particular goal. Some approaches utilize the emotions associated with a past event to determine how to encode and decode the event for use. Such mechanisms take emotional tagging as inspiration to enhance life stories and narratives by allowing for more convincing transitions from one idea to the next [18] while others use emotional models to allow agents to display emotions which can be effective within certain contexts when emotions are consistent [98]. The while motivational memory proposed in this does take sentiment into account, it does not, however, perform emotional tagging in the same manner as the aforementioned works and instead utilizes sentiment to narrow down contexts in which certain past events are appropriate to retrieve. In contrast, the robot described within this work is designed to display consistently positive emotion for the purpose of motivation and negative emotions were deemed counterproductive within a goal based interaction where a positive interaction is one of the aspects being optimized by the agent.

1.1.3. Behavior Change Support Systems and Goal Achievement

A behavior change support system (BCSS) is a software or technical system with that is designed change or reinforce behaviors. The systems persuade the users and change their opinions or views on a certain subject, thereby changing the actions that user performs of their own will to reach the outcome that is desired by the designers of the system [80].

There are many types of changes such as compliance change, behavior change, and attitude change. In the case of this work, since the goal is to encourage long lasting habits, attitude change and behavior change will be the main types of changes that will be considered.

Within the theories considered within BCSS, there are certain areas that are most relevant when considering the design of the memory which is specified in section 1.2.4.

Goal Setting Theory of Motivation Goal setting theory in particular is relevant due to the obvious connection to the diet related goals that are being followed by users when interacting with the system described in this work. As specified by Locke and Latham [61], there are five elements that need to be considered when setting a goal:

1. Clarity
2. Challenge
3. Commitment
4. Feedback
5. Complexity

Within these five elements, the motivational memory applies most directly to feedback which is used to allow the patient to assess their progress and determine if they are on track. All other elements are covered in the education, question answering and goal setting modules in the logic that is executed by the conversational agent. The functionality of these modules are described in detail in section 3.

Trans Theoretical Model of Change Within the trans theoretical model of change [87], of the six steps of pre-contemplation, contemplation, preparation, action, maintenance, and termination, this falls within action and maintenance. The way the memory operates is to support existing actions towards the goal in the case of when a memory is used for the purpose of encouragement, and it is used for action to encourage actions towards the goals when milestones are not met. One thing to note, however, is because of the duration of the experiment, the six month timescale stated within [87] is not quite as applicable.

Self-Determination Theory Self-determination theory (SDT) is a that concerns motivation and how people motivate the choices they make without outside influence [93] which is directly connected to the intrinsic motivations of people. The theory considers factors such as competence, autonomy, and relatedness in determining how motivation is formed. Interventions and behavior changes have been a staple of disease prevention and treatment in within healthcare [96]. Ng et. al [76] analyzed the use of self determination theory applied to health contexts. Such health contexts were split into the two categories of mental health and physical health. Within mental health, desired outcomes included less depression, less anxiety, etc. Within physical health, desired outcomes included less smoking, exercise, weight loss, etc. Ng et. al found through a meta-analysis that there were positive relationships between factors of self-determination theory and beneficial health outcomes.

As a result, insight into how patients intrinsically motivate their behaviors, choices, and habits allow for more effective approaches to behavior change systems [86] such as the one implemented in this work, making SDT very relevant when providing a motivationally fulfilling experience to end users.

1.1.4. Socio-Cognitive Engineering and Human Centered Design

Using a design centered approach can ensure that the final product will be better suited to serve the needs of the users and stakeholders. Such approaches try to optimize the user experience by identifying their needs and desires within the context of the system being developed. Such approaches have been used in many applications from product design to web design to the development of user-friendly information retrieval systems [21] and have can be applied to a wide variety of areas that have a human need that needs to be optimized.

To apply human centered design principles to cross disciplinary fields and areas with a social dimension, the frameworks of Situated Cognitive Engineering and Socio-Cognitive Engineering were introduced. Situated Cognitive Engineering allows for rapid iteration of a design where designers or researchers from different backgrounds can collaboratively develop the system being considered in an iterative manner to tune and refine the many aspects of the system. The framework allows for consideration of different aspects and provides a series of steps where the design follows from human factors and operational demands in the scenario being considered to specify the core functionalities and requirements to create a prototype that can then be tested and refined [72].

Socio-Cognitive Engineering takes an approach that considers developments in new human-centered technologies such as artificial intelligence (AI), robotics, digital interfaces, and conversational agents which can have a social dimension [100] in the interaction between the human and the system that is being considered [73]. An example of a work that informed the design of the system described in this work is that of the PAL project which utilizes a robot to interact with children [101] with type 1 diabetes. The project attempts to improve the interaction between the robot and children and optimize the inclination of children to play with the robot and by extension, learn more about how to manage their type 1 diabetes. Similar to PAL, the system described in this work also tries to improve management of a form of diabetes through a goal based approach to diet management.

The individual steps and components of the process will be further explained in section 1.4 and its subcomponents.

1.1.5. Type II Diabetes

As mentioned at the beginning of chapter 1, a significant percentage of the global population has Type II Diabetes [22] . As a result, technologies that can assist in managing diabetes can have an impact on the lives of many individuals.

Type II diabetes is a type of diabetes mellitus which is found in adults. Diabetes mellitus of which type II diabetes is most common is rising and is considered to be one of the largest endocrinological drivers of mortality according to the Global Burden of Disease (GBD) [10] leading to a 92.7% rise in over 20 years as noted by the GBD which analyzes morbidity and mortality of diseases. While the prevalence of diabetes varies by country, the global rates by which diabetes is growing is rising every decade [110]. Because of the increasing rates of diagnosis and mortality, strategies and approaches to preventing and treating diabetes become a growing priority as time passes.

Symptoms and Effects Type II Diabetes is characterized by high insulin resistance, low insulin levels and high blood sugar levels. This is caused by the pancreas being unable to produce enough insulin due to high blood sugar levels [106]. Doctors typically diagnose their patients with this disease when they notice symptoms such as increased thirst, urination, and fatigue. This, combined with a family history of diabetes, high blood sugar levels, an unhealthy diet, or obesity can act as an indicator of type II diabetes. Type II Diabetes can be brought on by lifestyle, environmental factors and individuals can be made more susceptible to the disease based on their genetic makeup as well as preexisting medical conditions [82]. Such environmental factors include diet, exercise, sleep, and stress.

Treatment and Prevention To manage type II diabetes there are multiple aspects that are considered. Since diabetes is heavily influenced by diet and exercise and correlates with obesity, improving such aspects of patients' lifestyle habits is a major part of treatment [22]. Such methods are useful for prevention as well [90]. Since improving such lifestyle habits can help with other diseases that result from an unhealthy diet such as heart disease and obesity [4]. As a result, for those who already have type II diabetes, maintaining blood sugar levels within what would be considered a normal range can help patients avoid more serious problems that can result from the diabetes and maintaining the proper weight can increase longevity in many patients [35].

It is possible to use eHealth applications to guide patients through improving such lifestyle habits. For more clinical or prescription based approaches, dosages of insulin are applied to properly maintain blood sugar levels. Such management is typically done by the patient, however there are a growing number of guidance systems that are available for automated insulin dose guidance [9].

Psychology of Diabetes Patients In addition to the physical impacts of type II diabetes, there is also a psychological impact brought on by the disease [46]. Being diagnosed with diabetes can cause emotional distress which can lead to depressed mood or more serious mental disorders such as major depressive disorder (MDD) which is commonly known as depression. Lifestyle changes become more difficult to achieve when mental health acts as an obstacle to enacting such change. Since stress can worsen obesity and the effect of diabetes, properly managing mental health is essential for diabetes treatment. In serious cases where behavioral factors are preventing progress towards improving a patient's type 2 diabetes, the use of a psychological intervention has been explored as a way to induce change in a patient's habits [20]. Although this work does not perform psychological interventions, an understanding of the pressures faced by patients can better allow for systems to be developed to better suit their needs and nudge patients towards the appropriate behavior change.

1.1.6. Diet Management and Current Practices

To provide proper diet advice, it is important to refer to what is currently being done in practice by dietitians and nutritionists. The literature falls into two styles of research one is more regarding nutrition science and the other is regarding practice and interaction with patients in the form of practice papers.

With regards to type II diabetes, studies and experiences by dietitians in treating the condition showed that dietetic consultation improved clinical outcomes for patients with type II diabetes [102]. Some of the factors that dietitians found can affect the experience and impact of dietetic are appointment time, affordability, psychological barriers for physical activity, and prevention of prediabetes [28]. In the case of the system proposed, appointment time, wait times and affordability are not issues, which gives it an advantage over human dietitians. This could perhaps offer a certain level of convenience over human dietitians, but conversely due to limitations of current technology, there are also some some aspects that will be are convenient when compared to a human.

To address some of the psychological factors associated with type II diabetes, the use of cognitive behavioral therapy (CBT) has been found to be effective [99]. CBT focuses on changing the way people think and how they react to those thoughts. This can be particularly helpful for those who lapsed into negative eating habits due to a troubled history that can result in a larger cognitive load. This has similarities to some techniques in motivational interviewing and behavior change such as self determination theory, making it quite relevant to determining the approach used by the system to motivate potential users.

With regards to diabetes patients and their motivation to attend appointments, certain factors such as beliefs about the effectiveness of treatment, or simply not prioritizing the appointment have been found to lead to absences [105]. A lack of confidence in one's own abilities combined with beliefs about the treatment can be detrimental towards properly making the appropriate changes needed to properly manage type 2 diabetes. Hurley et. al [44] found that improving the self-efficacy of patients by involving them in their own education regarding their diabetes improved their ability to manage their diabetes. By involving fewer lectures and more practical lessons, the self-efficacy and by extension, the self management skills of patients improved, making self-efficacy an important aspect of diabetes management.

1.1.7. eHealth

eHealth is a recent trend in healthcare that leverages new electronic technologies to improve the performance and accessibility of healthcare [13] [31]. The field includes different categories such electronic health based recording, telemedicine, telesurgery, etc. Although there is some debate regarding the definition of eHealth [84], the introduction of eHealth and modern technologies into healthcare is a promising area that can improve the experience of both healthcare professionals and patients who are using the healthcare system for their needs. At the same time, there is a necessity to provide such systems in a manner that is easy to grasp [75] so as to allow for those with lower amounts of familiarity with such technology or technological literacy to not be excluded from any crucial functionalities provided by eHealth systems [74].

With regards to eHealth, this work will fall within the ”tele” category. Specifically, it will fall within the tele-dietetics category which looks at the use of diet management for disease prevention or treatment as well as lifestyle improvements which is still a relatively new area of study within eHealth [19]. Like with the other eHealth categories, this too has some of the same concerns regarding technological literacy, but this is also compounded by the use of conversational agents within a health context which is currently still quite unfamiliar territory for younger demographics who have higher eHealth literacy [74].

Current Works in eHealth Addressing Diabetes The application of eHealth towards diabetes can be quite promising and a conversational approach could be possibly effective due to such approaches taking into account the social dimensions involved in diabetes management [107]. Such applications can allow for longer term approaches that can potentially expand the efforts of a doctor or dietitian that patients are seeing to manage their form of diabetes.

Neerincx et al. [73] present the PAL (Personal Assistant for a healthy Lifestyle) project which is a system that is used for type I diabetes (T1D or T1DM) in children. This is a system that utilized the SCE methodology in its approach. The system utilizes a Nao robot and a virtual avatar to help children improve their self-management skills. As the name suggests, the agent acts as a friend or ”pal” for the child to provide a more engaging and motivating experience.

Baptista et al. [7] evaluates the acceptability of the use of an embodied conversational agent in type II diabetes self-management apps. Specifically, it uses an embodied conversational agent named Laura to deliver diabetes management education within the app known as My Diabetes Coach which is an app that uses gamification and voice skills to quantify and advise patients. While the app and the embodied conversational agent does operate within the same domain as this work, Baptista et al. do not utilize shared experiences, and the memory of the agent is based more on the diabetes related statistics of the patient collected through the app rather than conversations with the patient or focus on optimizing or improving motivation. Management targets are set beforehand rather than decided in conversation with the agent which implies an approach that is less reliant on self-determination theory, and the conversational interface uses a set of choices rather than a speech driven conversation.

1.2. Motivation and Novelty

Here, the question of why this system is being developed and what is novel about it is answered.

1.2.1. Impact of Motivational Approaches using Shared Events

Applications of a motivational approach to memory extend beyond that of diabetes, because it can be utilized for any type of behavior change when considered as a system that is not only motivational, but also persuasive as it is able to convince the user or direct the user to a certain outcome desired by both the user and the agent. Similar approaches can be applied to education, other health related outcomes, and even marketing approaches where the intention is to provide a dynamic and adaptable system that leads to a particular purchasing decision or series of purchasing decisions. As a result, the creation of a motivational memory that is able to utilize shared events of intrinsic value act as the first steps towards more general behavioral change based approaches within intelligent agents.

1.2.2. Application of a Conversational Agent to Tele-Dietetics

Within tele-dietetics, the use of conversational agents is highly limited with embodied conversational agents in the area being used in more generalized applications that are used more for healthcare support and monitoring than specifically diet management. By answering questions regarding the design of such an agent as described in this project, this work provides one of the only applications of tele-dietetics that utilizes a conversational agent and can offer insight into what are the requirements to create and automate the work of a traditional nutritionist or dietitian.

1.2.3. Impact on Health of Individuals

A conversational agent that offers nutrition advice for those who have diabetes or are looking to prevent diabetes can potentially help a large number of people that is only growing [22].

By utilizing a conversational agent that is specifically designed for diet management, it is possible to target and improve an important aspect of lifestyle better than a health management app, or a conversational agent that is used for multiple aspects of health.

The potential impacts can result in longer life expectancy, peace of mind, and a higher quality of life. In addition to this, individuals will be more educated regarding type II diabetes, and can therefore be in a position to help others who experience the same problems.

1.2.4. Novelty

In addition to the novelty of the conversational agent within tele-dietetics, the main area of research within this project looks into how the use of shared memories with intrinsic value can affect achievement and user experience.

Previous studies have tested whether shared memories can improve a certain aspect of the interaction [95] [48]. In such cases, the memory has been used more for ensuring topic adherence, difficulty scaling, personalization, or to reduce repetition. In other words, these studies examined the use of shared memories as a way to improve relatability and reduce the loss of novelty over an extended interaction to ensure that the interaction stays engaging.

Rephrasing and choosing of experiences with intrinsic value are rare and the combination of the two is lacking in literature despite reflection on the past being a staple of psychology from the inception of the field [88].

In addition to this, these studies do not, however, examine *how* within their experimental setup which means that they primarily test whether a memory should be used rather than how it should be used.

By testing how a variety of shared experiences can compare with simply using the same shared experience multiple times, exactly how shared experiences should be used to increase effectiveness can be further clarified. This applies not only to experiences with conversational agents, but also has potential applications within psychiatry and other fields that utilize long term interactions.

With regards to the specific areas of conversational agents and human robot interaction this research operates within, it should be noted that this research involves long term interaction as well as user adaptation for goal setting. Although there are many works that study long term interaction, these are in the minority when compared to single conversations within the literature. Similarly, use cases that utilize user adaptation for goal setting within a conversational approach over the long term are also in the minority.

Novelty of Memory Design Here, an overview and explanation of the consideration and purposes of the memory used are discussed. Further description of the techniques and structure used in the memory itself can be found in Design of Memory for Experience. Sharing

When considering the design of the memory, it should be asked why a memory may be needed.

To begin, to persist information across multiple sessions, a longer term memory is needed beyond data that is persisted within the a single conversation. In this case, since a diet related goal and its achievement is the primary use case, there is a need for a memory, not only for referring to past events, but also to simply remember the user and the progress that was made. Goals are not achieved in one day, as a result, multiple sessions are needed.

In addition to goal achievement, it is necessary to shape the progression of the dialogue and interaction to best fit the needs of the user so as to improve the chances of achieving a particular goal. Using a memory can allow for reduced repetition which results in a better experience that does not become boring over time. To improve achievement, an approach that takes into account motivation is considered. Shared experiences can be used to make a personal connection with users and reflect on past progress to improve motivation. Motivation can be a limited resource and reminders of why a goal is worth working towards can help reduce the rate of decreasing motivation making it a promising function that can potentially help with higher achievement as well as a better interaction.

There are multiple parts of the memory with each part serving a specific purpose:

1. Diet
2. User Metadata
3. Experience Sharing

The diet and metadata is used to facilitate progression towards a goal and identify the user. This includes information such as the name of the user, an identifier to recognize them with for future sessions, some physical data to help with the calculation of the milestones and goals to be met by the user, and finally data that records the progression of the user towards their goal.

Many memory models use progress as a way of personalization to infer if something is easy or not without an explicit statement from the user [51] , [48] , [59] .

The experience sharing is used for two purposes: The first is that of realigning the motivations of the towards the chosen goal throughout all of the sessions. The second is that of using shared experience in context to support praise and criticism of progress. In this way, it is able to strengthen and clarify sources of intrinsic motivation and use them as extrinsic motivation through praise and statements in the conversation. This is unique in that it is designed to encode shared experiences (statements made during the conversation instead of task related metadata) with the purpose of using them for motivation in the future (motivational memory). Many memory models that have multiple sessions or work with user data use progression or numerical data to potentially share an experience [48] , [95] , but do not necessarily use the user's own statement (e.g., their own words) as a way of sharing that experience for the purpose of encouragement or praise and such memory models typically don't discern a difference in the intrinsic value of different past events. Doing so implies understanding of what the user said and in what context and requires more information to be stored and processed in the encoding of the memory to determine the proper context to mention topics that were brought up in previous conversations. For memory models that are limited to one session, the memory is used more to keep track of the context and topics. The way in which the memory is used in literature when working towards a skill is difficulty scaling [48] , a personal connection, or greater relatability, but very rarely for encouragement or criticism within the context of longer term reflection and motivation (specifically reflection that extends beyond just the previous session or beyond the current interaction and considers intrinsically valuable events). For example, Goedheijt [95] does use this to allow children with type 1 diabetes to form a better relationship with a NAO robot for greater affection and relatability, but does not take into account whether past events are intrinsically significant and instead uses topic based relevance to enhance conversations. In Kasap et. al [48], this is used more for difficulty scaling, reflection on progress, and identification of problem areas within a tutoring context rather than advice or motivation over a longer period of time.

Expected Outcomes:

By utilizing the aforementioned memory with the three distinct parts of diet, user metadata, and experience sharing through motivational rephrasing, the following expected outcomes become more apparent:

1. There should be higher levels of relatability and motivation through the use of the shared memory references.
   (a) By extension, higher levels of relatability and motivation should lead to higher levels of achievement due to what could result in a higher emotional investment as well as more clarity regarding the patient's own intrinsic motivations which results from the use of the patient's answers to compose the shared memory reference.
2. There is personalized advice or feedback given based on the user's chosen goal and progression.
3. The user treats the conversational agent as though it is a trustworthy advisor that understands his or her motivations and struggles due to the way in which the conversational agent is able to point to specific topics covered in the conversation that do not necessarily directly relate to recorded metrics or numerical progress.

Relationship to the Temporal Properties of Motivation:

Motivation is affected by a variety of variables of which future perspective and outlook [3] can affect the quality of motivation. One aspect that can affect the future time perspective is past events due to the way in which people use past events to extrapolate into the future. Much of the rephrasing uses shared experiences to connect past behavior and mindsets to future behavior– particularly in the second session. These use sentiment and keywords that represent ideas and concepts that are important to the patient and ties those to the patient's current behavior and goal achievement in the future. As a result, it is able to tie the past events to a desirable future event which can ensure a greater clarity of how the intrinsic desires they have held since the beginning of the experiment relates to the end goal. The expectation is that by doing so, the success of the patient towards their goal will increase, and successes beyond the duration of the experiment will increase as well due to the forward-thinking outlook presented within the motivational phrases.

1.3. Formulation of Focus and Research Questions

The intention of the system developed is to encourage starting and solidifying a diet based habit that can reduce the impact of type II diabetes or reduce the chances of being diagnosed with type II diabetes.

The system is a conversational agent that instructs and guides the patient towards success in a diet related goal of their choice that can help them with regards to type II diabetes. Through dialogue in multiple sessions, the user will be able to learn what changes to make in their lifestyle and how to maintain those changes for a positive long term outcome.

To better ensure that patients actually succeed in their goals and have a positive user experience, references to shared experiences are leveraged to create higher levels of relatability, increased intrinsic motivation, and a slower decrease in novelty factor. A shared experience is an episode that was experienced by both parties of a conversation. In this case, that refers to the patient, and the conversational agent.

To refer to shared experiences, shared experiences must be stored within some form of episodic memory. When designing and implementing the the memory and dialog, the following design and engineering questions were considered:

1. How should shared experiences be modeled and organized within the memory?
2. How should shared experiences be modified to generate motivational phrases?
3. When can shared experiences be used?

To determine the effectiveness of the use of shared experiences in terms of user experience and goal achievement, the following research questions are asked to determine whether shared experiences should be used at all and how shared experiences should be used if they do lead to a better experience.

1. Does referring to previous sessions improve goal attainment (performance)?
2. Does referring to previous sessions improve user experience?
3. Does a variety of references help more than making the same reference? Similarly, can a variety of references over a number of sessions be more useful than a single repeated reference?

To test this, there are 3 conditions that are considered:

1. No references to previous events
2. References to only one event environment.
3. Multiple references to different events (variety) where no events are ever brought up twice

1.4. Design Process

The design is an integral part of the development of the conversational agent, and will determine what to prioritize and what to implement to improve the user experience.

It consists of the following components which are further described in their respective sections:

- FOUNDATION (Chapter 2)
- SPECIFICATION (Chapter 4)
- MODELS (Chapter A)
- EVALUATION (Chapter 5)

2

Foundation

The foundation is the component of the design that considers the reality and conditions in which the system will operate.

The foundation has the following components:

- Operational Demands (Section 2.1)
- Human Factors (Section 2.2
- Technology (Section 2.3)

In addition to these components, based on the use case, the design and formulation of the memory (Section 2.4) is described which takes into account the content of sessions conducted by the conversational agent.

2.1. Operational Demands

The operational demands defines the situation and conditions in which the system being designed and implemented will operate in. The operational demands are composed of the following sub-components:

- Environments (Section 2.1.1)
  - The setting in which the system and users will operate in.
- Stakeholders (Section 2.1.2)
  - Is composed of direct and indirect stakeholders. The individuals who will directly interact with the system are direct stakeholders. The individuals who do not interact with the system, but have an interest in the development and functioning of the system are indirect stakeholders.
- Tasks (Section 2.1.3)
  - The type of actions that are executed by users and stakeholders that the system will potentially support.
- Personas (Section 2.1.4)
  - A realization of users and stakeholders that tries to provide a realistic outline of the type of people the system is involved with. This is typically used to ground the design and implementation within a realistic setting [64].
- Problem Scenario (Section 2.1.5)
  - A description of how the previously mentioned sub-components result in a problem that can be solved by the system that is being considered in the design.

2.1.1. Environments

The envisioned diet based education and goal achievement is situated within the homes of the typical user. The user will be able to conveniently access the session from the comfort of their own home through a modern computer or a mobile device.

Users can also hold sessions with the conversational agent outside of their homes as long as privacy is ensured for the purposes of a uninterrupted session.

Although the home is different for each user, the home is where the patient lives, eats, and sleeps and it encompasses part of the user's beliefs regarding their own identity [68]. It is expected to an environment where they are most comfortable. It is also an environment where they can directly analyze their eating habits.

It is expected that such an environment can allow for a greater level of comfort and personal investment when compared to the office of a dietitian or nutrition due to the less impersonal nature of the user's own place of residence.

One disadvantage to note, however, is that the home is also filled with distractions and due to the large amount of variation between the homes of patients, there is a possibility that such an environment can be counterproductive towards reaching one's diet related goals.

2.1.2. Stakeholders

Here, stakeholders are presented. DSH refers to direct stakeholders and ISH refers to indirect stakeholders. For some of the stakeholders, actualized examples can be found in the Personas section (section 2.1.4). Stakeholders are accompanied by value acquisition which where the values held by the stakeholder originates from. The values are the particular values of the patient that are relevant within the context of diabetes treatment or prevention.

Patients *DSH01: Patient With Type II Diabetes 2.1.2* and *DSH02: Patient at risk for Type II Diabetes 2.1.2* are both collectively referred to as "the patient" and represent two different types of patients who can benefit from the system.

DSH01: Patient With Type II Diabetes

This refers to individuals who have already been diagnosed with type II diabetes.

Value Acquisition:

DSH01: Patient With Type II Diabetes will have the same values as DSH02: Patient at risk for Type II Diabetes except they may be more concerned with concerned with present livelihood when compared with DSH02: Patient at risk for Type II Diabetes [37]. This implies that their requirements are more urgent and that they are looking for short term improvements to reverse any immediate concerns and long term plans to prevent any regression in improvements made in the short term.

Values:

1. Wellbeing/Health
2. Peace of Mind
3. Privacy
4. Confidence in Self
5. Sense of Achievement
6. Autonomy
7. Improved Understanding

A persona can be found in section 2.1.4.

DSH02: Patient at risk for Type II Diabetes

This refers to individuals who are at risk of being diagnosed with type II diabetes based on their current habits or traits that lead to obesity. This can also include individuals who are at risk due to a family history of diabetes which may lead to a higher risk of acquiring diabetes in the future.

Value Acquisition:
Due to the large number of people that fall into this category, the values may vary significantly, but those who are obese generally have reduced physical fitness and a number of symptoms that impair their ability to operate effectively within society and in daily life. The values that they acquire are partially based on the aspects of life that they are deprived of because of their obesity [46]. In the case of diabetes prevention, it can also be a matter of maintaining peace of mind in the future.

Values:

1. Wellbeing/Health
2. Peace of Mind
3. Privacy
4. Confidence in Self
5. Sense of Achievement
6. Autonomy

A persona can be found in section 2.1.4.

DSH03: Doctor/Primary Care Physician The primary care physician of the patient will be aware of any complications that may occur as a result of changes in diet, so the input of the physician is important for avoiding any possible medical complications [23]. Because the doctor can be involved with the goals of the patient, they are listed as a direct stakeholder.

Value Acquisition:

The purpose of the primary care physician is to ensure the comfort and health of the DSH01 and DSH02. To do this, he or she typically diagnoses current ailments or tries to look for markers that would lead to an ailment in the future and treats them accordingly. Values are also informed by the hippocratic oath and other ethical considerations commonly taken up by medical professionals [27]. Unlike a dietitian, a doctor can take a more holistic approach that takes into account factors beyond diet to inform recommendations that can improve the health of the patient.

Values

1. Safety of the patient
2. Wellbeing of the patient
3. Optimizing the longevity of the patient
4. Optimizing the patient's wishes
5. Privacy of Patient Data/Patient Confidentiality
6. Reducing Liability
7. Following of Proper Protocols

A persona can be found in section 2.1.4.

ISH01: Family Members Family members who are associated with the patient will have an interest in the wellbeing of the patient. Typically, this would refer to a spouse, children, siblings and parents. In addition to this, since a family history of diabetes can be an influencing factor in diagnosis [36], having parents or relatives set a good example can be vital for the rest of the family in preventing type II diabetes themselves. Similarly, the health and wellbeing of the patient can have either a positive or negative impact on future outcomes of the family members themselves, which means that the progress of the patient can, in turn, also affect the family members which suggests that the patient's actions can affect more than his or her own wellbeing.

Value Acquisition:

The family member thinks of the patient with respect to the patient's impact on the rest of the family.

Values:

1. Longevity of the Patient
2. Wellbeing of the Patient
3. Mental health of all family members.

ISH02: Researcher The researcher refers to those involved with the development of the system. The researcher is concerned with the outcomes the system is trying to optimize which is a positive experience for the patient or user of the system along with an effective management of diet to facilitate prevention or treatment of type II diabetes. Value Acquisition

The researcher wishes to optimize the values of the primary stakeholders that are the subject of the study.

Values

1. Optimizing the values of all other stakeholders.
2. Reliability
3. Ethical Design
4. Privacy and security of all stakeholder data

2.1.3. Tasks

Since the activities will be carried out within the home of patients, it is expected that patients will be comfortable within the environment. This is the location that patients, eat, sleep and, with the availability of the internet, this is also a place in which the patient can access information and interact with others.

As a result, the home can give a sense of peace and comfort and allow for an innumerable number of activities to be completed [39]. With respect to the purposes of diabetes, the activities completed by the patient or user of the system will be that of 3 short sessions completed over a period of 3 days with the conversational agent as shown in figure 2.1. In the first session, the patient will choose a manageable diet related goal or habit to achieve that is related to type II diabetes or obesity, and be educated on what actions to take to achieve that goal. The goal should be something that can be achieved within 3 sessions. After the first session, the patient will try to carry out the recommended actions. In the second session, the progress of the patient will be checked. In the third session, a judgement will be made regarding whether the goal was achieved or not, and a retrospective will be carried out to talk about what went right and what went wrong.

While goal achievement and habits are typically developed over a time period greater than 3 days, this 3 day period holds the minimum that is needed to set a goal, pivot and adjust to align progress with the final goal, and finally achieve the goal and reflect on the progress made. This allows for a greater number of experiments to be held when compared to experiments containing, for example, a week long duration.

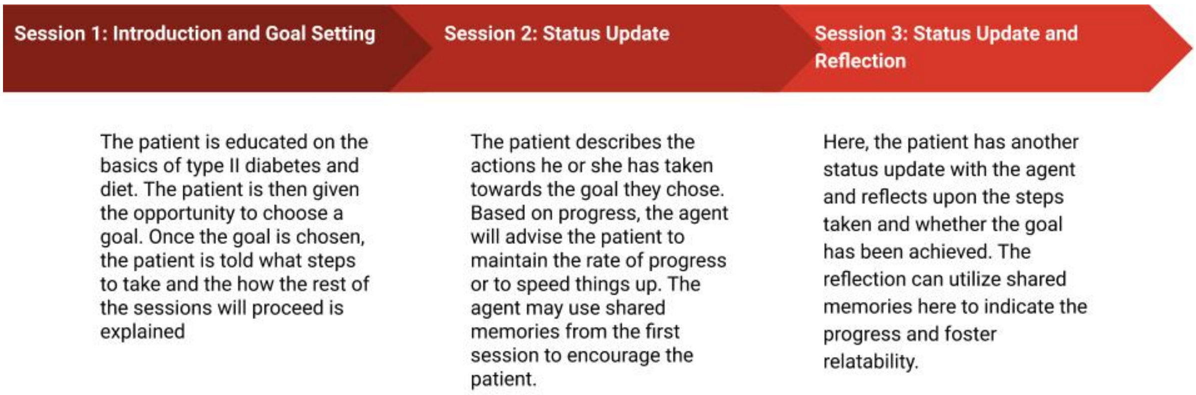


Figure 2.1: Diagram illustrating the 3 sessions the patient will have with the conversational agent.

The activities that a patient can work on include the following:

1. Calorie Restriction
2. Sugar Intake Reduction

While all of these activities have some overlap [58], the choices allow for the patient to have better control over the goal they want to achieve and more clarity regarding what it is that they are trying to achieve. Similarly, for those who have a preference, this can allow for a higher level of investment, because they can choose what they believe can be more meaningful or helpful for their own personal development and growth.

A third goal of diet composition management was considered as a possibility, but during implementation, it was realized that to determine diet composition, an in depth conversation needs to be held where all the diet items consumed by the user needs to be provided to the robot. Since providing such items during a conversation can result in a significantly different experience from the other two activities, this was removed as a possible activity for patients to work on.

2.1.4. Personas

DSH01: Dave (40) Dave (40) is a worker in the tech industry who has been diagnosed with type II diabetes. He is obese, and does not exercise. He spends most of his days working at a desk in an office. He is a divorced father of two children ISH01. He wishes to gain better habits so that he can ensure a better future for his children and avoid a premature death.

DSH02: Amy (30) Amy (30) is an office worker. She does not typically exercise and is usually sedentary. Diet involves what is offered at her work cafeteria, or fast food. She works at a desk during the day and watches television after work. She is concerned that her lifestyle will lead to consequences in the future and is hurting her social life and social opportunities.

DSH03: John (30) John is a primary care physician. When he sees that a patient has a trait that is associated with obesity such as a high weight for a particular height, hypertension, high cholesterol, etc. he considers all of the possible consequences such as heart disease, arthritis and type II diabetes. Based on the diagnosis and possible future consequences, he recommends changes that the patient DSH01 and DSH02 should make. In some cases, he may recommend that the patients see a dietitian for more frequent and personalized help based on the specific needs that they may have.

2.1.5. Problem Scenario

DSH03 John's primary desire is to improve the lifestyles of DSH01 Dave and Amy DSH02 so as to avoid serious consequences of their current lifestyle such as death. Depending on the levels of obesity, certain types of exercise can be harmful; especially at extremely high body weights. Because of this, diet change or surgery can be more feasible options. Since lifestyle change can be more manageable, diet change can be a more viable option for many.

Dave DSH01 and Amy DSH02 hope that they can change their habits to ensure a better future and improved ability to function within society [82].

They have tried to do so before, but due to factors such as the cost of healthcare and a lack of motivation and time, they tried a self guided approach which resulted in an initial attempt to make a change that did not solidify into a daily habit. Although they would like to try again, they would prefer an approach that would inform them of what they are doing wrong, provide them with appropriate advice that is tailored to their needs and provide additional motivation.

2.2. Human Factors

Sub-components of the human factors typically varies based on the scenario, but will contain Measuring Instruments (section 2.2.5 at the very least. This specifies what values and considerations must be taken into account to properly address human needs when considering the specific stakeholders who are involved. Those who are diagnosed with type II diabetes typically have the same concerns as those of patients with obesity, but those concerns may be more immediate in nature due to the diagnosis.

2.2.1. Autonomy

Autonomy refers to a person's ability to make decisions and take actions based on their own volition. When patients are allowed an option of choosing among a selection of goals, the goal becomes more meaningful to the patient and there is a greater sense of ownership [52]. As a result, there is a higher level of intrinsic motivation. Allowing patients to have more control of their goals will allow for patients to take a self directed approach which can improve self efficacy as well due to the greater level of involvement the individual may have in their own outcomes [104].

2.2.2. Diet Management

Diet is an aspect of lifestyle that is completely under the control of the patient and has a large impact on obesity and by extension type II diabetes [82].

Diet management refers to control of one's dietary intake by regulating the level of intake as well as what the diet composed of. Such management can be used to maintain a particular level of health by following a particular diet or reaching a goal like preventing or treating type II diabetes through a diet change that results in lower calories, sugar, or selection of conventionally healthier food items [79].

2.2.3. Ethics

Patients would like for those who are involved with their healthcare to follow ethical standards and respect the privacy of the patient. Information should not be recorded without the patient's consent, and the patient should be made clear of how the information that is recorded will be used. The patient should also not be intentionally harmed in any form during interactions with the proposed system. The set of standards that will be followed are similar to that of what may be found in the international code of medical ethics [103] which specifies what the doctor can and cannot do with relation to their patients. With regards to this work, the privacy and confidentiality of patient data will be protected and all data will be anonymized to follow General Data Protection Regulation (GDPR) regulations [38]. To ensure that all collected data was used properly, a consultation with a data steward at TU Delft was also carried out to ensure best practices were followed. In addition to this, approval was sought and given by an ethics committee to minimize any harms that can possibly result by carrying out experiments utilizing the system specified in this work. The use of ethics approval not only allows for the research to avoid ethical missteps, but it also can provide piece of mind to to participants of experiments that involve the specified system [56].

2.2.4. Goal Achievement

By achieving goals, the patient can have a sense of progress and build a sense of competency and a sense of confidence as they complete goals and subtasks [94]. As mentioned in section 1.1.3, this can provide a focus for one's actions to result in a set of directed actions that leads to a particular desirable outcome. This can be particularly useful when such an outcome is not achievable within a short term period and, as a result, requires careful planning and a non-trivial level of effort. For such outcomes, goal based approaches can be promising and sub-goals or milestones can be useful for managing the cognitive load [47] when working towards an outcome that is desired.

2.2.5. Measuring Instruments

The measuring instruments refers to tools that are used to record data. Such data can later be used to run statistical analyses and draw conclusions. While the instruments are described in this section, the actual metrics and questions being used are concretely specified in section 5.2.

MI1: Questionnaire for Experience To determine the user experience of the interaction, a questionnaire will be used to determine whether the user had a positive experience or a negative one. The type of questionnaire will be more similar to a godspeed questionnaire [108] than a social presence questionnaire [34]. The reason for this is that while the social presence of the robot can be useful, a Godspeed questionnaire is used quite often in other human robot interaction works that use socio-cognitive engineering, and can be used to gauge aspects not covered by the social presence of the robot such as safety, intelligence, etc. This questionnaire will utilize a Likert scale questions that are grouped by different categories [77] that relate to the user experience and perception of the conversational agent.

In addition to this, questions that are not covered by the Godspeed questionnaire will offer additional insight into the the results. Topics such as levels of motivation, self efficacy, country of origin, a direct question asking of whether the interaction was positive or not and many others that are specifically tailored to determine the nuances of the interaction with the specific system designed in this work can offer a more refined perspective that is not necessarily offered by the Godspeed questionnaire or the progress and actions of the user during their interaction with the conversational agent.

MI2: Achievement and Conversation Metrics To determine the level of achievement, the progress of the patient over all three of the sessions will be recorded.

This includes, but is not limited to, the intake of the patient during each session, whether the patient agrees with a shared memory reference, as well as the time taken for each session and a transcript of their statements during the interactions.

2.3. Technology

Sub-components of the technology will vary based on the technological demands brought on by the context being considered and the capabilities of current technology and their limitations.

To develop a conversational agent that is capable of conversing with patients and motivating them using shared experiences, technology that can support such functions are required.

2.3.1. Furhat Robot

An embodied conversational agent has many components which are leveraged to provide a better user experience. For this project, the Furhat conversational agent will be used. Furhat was developed by Furhat Robotics which is a Stockholm based startup.

Virtual Embodiment Different types of embodiments serve different purposes. With a virtual embodiment, the convenience of a traditional speech based conversational agent without a physical embodiment is available, but there is an additional level of relatability offered by providing a virtual embodiment [16]. Although Furhat does not have a torso, it does offer the ability for finer controls over facial gestures and has the option for a virtual embodiment which allows for it to fulfill the minimum conditions mentioned with regards to the advantages of a virtual embodiment [69].

Speech Recognition One advantage of a conversational agent that is capable of speech recognition over a text based chatbot is that it is possible for users who are incapable of using a keyboard to still communicate with the agent and for more realistic conversation to be simulated. In other words, users will type differently than they speak in a conversation. Furhat offers the ability to listen for the responses of users and can perform different actions based on the words that are spoken by the user.

Speech Generation Natural sounding speech generation in a conversational agent that is capable of adjusting its prosody can allow for the agent to add nuance to replies that it provides during its turns in a dialog. With Furhat, although the voice can be recognizably robotic, it has human qualities which allow for its voice to be acceptable to users. In the case of this work, the Amazon Polly voice has been

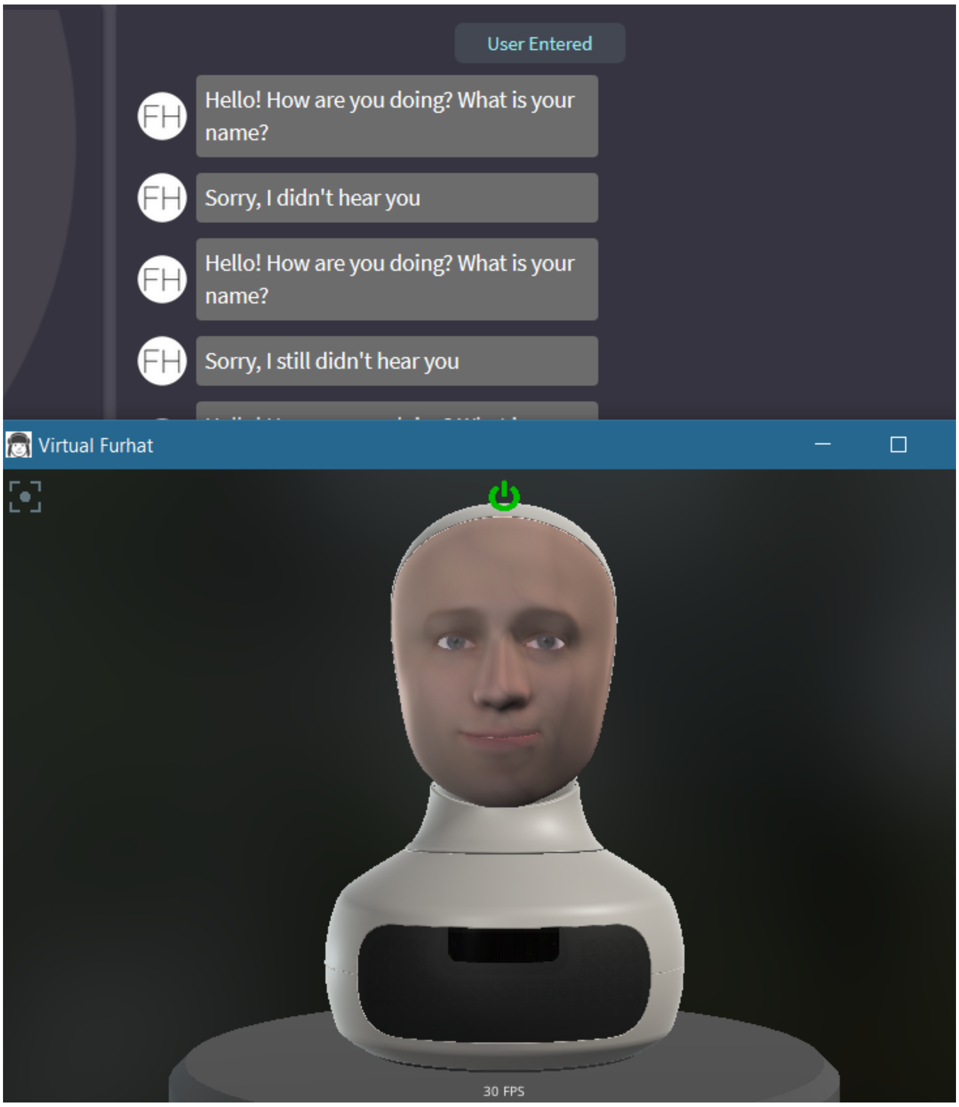


Figure 2.2: Image of a virtual Furhat robot and the user interface that accompanies the virtual Furhat robot.

used to minimize the robotic qualities of the voice that the robot used. In addition to this, stress and emphasis can be manually added to statements made by the Furhat agent to improve the human-like qualities of its speech.

Development Furhat utilizes Kotlin as its programming language and can be developed on using an IDE such as IntelliJ IDEA. It organizes dialog and responses in a state machine. In addition to this, the physical movements and positioning of the head and facial features can be programmed and modified. Furhat can also send and receive information from other servers, which allows for other languages and tools to be used to extend the functionality of Furhat. Furhat does not come with a way to persist data, so other methods should be used in combination with Furhat to store memories.

2.3.2. Natural Language Processing for Improved Understanding and Flexibility

Utilizing Natural Language Processing (NLP) techniques to decompose sentences can allow for more detailed information to be extracted beyond the textual content of a dialog. Aspects such as sentiment, context, and meaning can be determined and allow for a conversational agent to operate beyond certain responding to certain preprogramed responses from the user [70].

Within the system that is developed for experience sharing, keyword/phrase extraction, sentiment detection, and TF-IDF is used.

For keyword extraction, the Rake algorithm [92] within NLTK (Natural Language Toolkit) was used. Alternatives that were tested include pre-trained models as well as an implementation of TextRank [109] which is available within the Gensim library. Given the primary use case within this work, such alternatives were found to be less effective due to their primary focus being that of summarizing large amounts of text (e.g., articles, novels, etc.). In addition to this, such methods focus more on keywords instead of key phrases. Since keyword and keyphrase extraction in this work are used to paraphrase a user's answer to a motivational question, key phrases which can be extracted quickly from a small corpus can be useful if they describe what is considered to be the main subject of a user's answer to a question.

For sentiment detection, this is used to determine whether a user is confident or not to provide encouragement or praise. For this, NLTK's naive bayes implementation trained on twitter data is used. Twitter data is used, because the length of such data is similar to that of user answers when compared to alternatives such as novels, movie reviews, and novels. The simple naive bayes model was found to train quite quickly and resulted in consistently high accuracies of 95% or higher when tested with 10-fold cross validation which was deemed to be sufficiently accurate during qualitative evaluations of the model in pilots and test runs of the system. Other models such as decision trees, and VADER [29] were found to have lower accuracy. In the case of VADER, which is often compared with SVM's and Naive Bayes, and is said to perform better on shorter length documents, it was found during preliminary tests to decide which method to use for the system described in this work that VADER was more likely to classify negative tweets (which resulted in total accuracies of between 84 to 87%) as positive which could be problematic when trying to motivate someone. Simply put, if someone is lacking confidence and needs encouragement, incorrectly classifying their statements as confident or positive can deprive users of much needed encouragement that can lead to higher rates of achievement.

2.3.3. Data Storage/Persistence for Shared Memories

To properly remember conversations and share memories, some form of data persistence is needed. This can be in the form of storing data within a file, or something as sophisticated as a backend that communicates with a relational database.

For this work, a simple JSON based storage is used where each user's data is stored to its own JSON file. This form of storage is chosen primarily because of its simplicity, and the lack of a need for an additional server to run to enable persistence of data. Since the system will run on a modern computer, ensuring that cpu usage is minimized to avoid unexpected slowdowns during the conversation is a priority. In addition to this, limiting the cpu load can help to limit the thermal loads which by extension means that the fans of the computer can operate below a noise level that can possibly interfere with speech recognition and audio detection when using the virtual Furhat robot.

2.3.4. Video Telephony

Due to COVID-19, an in person trial will not be possible, but it will be possible to handle interactions online with patients by utilizing a video calling application such as Zoom or Skype.

In an ideal situation, the conversational agent would be made available on a website that can be visited at the user's convenience, but due to time restraints and a lack of availability of embodied conversational agents that can be hosted within a web browser, a video calling application is the best alternative. With the Furhat robot, since the virtual version of the Furhat robot was never intended to be used over a video calling application, ensuring that user responses can be properly detected by the robot is a priority.

To allow for this, a combination of sound engineering was used along with a delay to allow for audio from Zoom to be played from speakers into a microphone so that the virtual Furhat can detect responses. In an ideal scenario, directly piping the audio output from Zoom to the virtual Furhat with a virtual audio cable would avoid any problems with degradation of the audio, but in this work outputting audio into microphones was used.

2.4. Design of Memory for Experience Sharing

In this section, the components of the memory and how memories are encoded and decoded are discussed in detail.

2.4.1. Specifics of Reuse and Motivational Memory

To reuse a shared experience (a statement made by the user) and reflect on it, some understanding of the statement and the question needs to be encoded in the memory.

For the encoding, the following need to be understood:

1. The question asked
2. The answer given to the question
3. The context in which the answer was given.
4. The session in which the answer was given.

To ensure that statements relating to intrinsic motivation are considered, answers to questions that ask the user about their reasons for working towards a goal, or their feelings about a goal are utilized. This is similar to strategies and techniques utilized in motivational interviewing [81] to encourage goal oriented mindsets. Motivational interviewing is a form of counseling for eliciting behavior change where change talk is elicited by asking patients questions.

For the decoding, the saved data is used to reword the original statement into a reference that can be used as motivation or encouragement that can remind the patient of why they are working towards their goal, and to realign their motivations accordingly. To do so, the following is needed in the decoding:

1. Parts of speech (SPACY)
2. Sentiment Detection (NLTK Naive Bayes)
3. Keyword/Keyphrase Extraction (NLTK RAKE algorithm)
4. Possible spelling and grammar correction in case the speech recognition resulted in a sentence being incorrectly recorded (LanguageToolPython)

The parts of speech tagging when used in combination with keyword extraction can be used to determine if the extracted phrases actually contain useful information to summarize or paraphrase a user's past statements. The detected sentiment can be used to gauge the level of confidence a user has, so as to provide appropriate levels of encouragement or criticism. When these techniques are used in combination with the progress of the user, they can be used to constrain the different possible statements that can be given to advise and motivate a user.

Once these are made available, they can be reworded into phrases that can be used for encouragement. With regards to rewording, there are some variants that are used to allow for some flexibility in how an answer to a question can be later used as shared experiences:

1. Simple recitation of an answer
   (a) This involves simply replacing pronouns like ”I” and ”me” with ”you” to reference a past event.
2. Focusing on keywords/keyphrases
   (a) For keywords that include nouns (excluding pronouns), these can be used to talk about the main points of an answer without necessarily reciting it verbatim.
3. Ignoring content and using sentiment
   (a) Based on how someone answered a question, this can be used to simply talk about the feelings that the user had during the answering of the question rather than the answer. By using this in combination with a reference to the question that was asked, it can allow for an interpretation of the user’s mindset that goes beyond what is explicitly stated by the user.

One thing to note is that some combination of the the three can also be used. For example, in cases where a simple recitation of an answer is not enough for deep introspection or not convincing enough for the user to consider the conversational agent as intelligent, using sentiment to expand on the recitation to provide the appearance of comprehension can be used in tandem with the other techniques.

The motivational phrases generated are composed of the following parts:

1. The question asked in previous sessions.
2. The statement, keyword or sentiment of the answer provided.
3. A handcrafted statement that takes into account the sentiment of the answer and progress of the user to either encourage or praise.

An example of a case where a motivational phrase is constructed and used is as follows:

- Question Asked in Earlier Session: Are you feeling excited to start? Nervous? What feelings are you having right now?
- Answer Provided to Question: I think I feel a bit nervous.
- Rephrasing in the case of a milestone being met.
  1. **In our first session, I asked you how you were feeling before we started.** *You said bit nervous.* Given your ability to meet your milestone, there was nothing to worry about. Keep up the great work.
- Rephrasing in the case of a milestone not met.
  1. **In our first session, I asked you how you were feeling before we started.** *You said bit nervous. It seems that you may have been a bit nervous or worried, because you did not manage to reach your milestone.* A more optimistic and focused outlook may have served you better. Keep at it. I’m sure you will get there.

2.4.2. Types of Feedback

With regards to the purpose and contexts in which the rewording is used, these phrases are used to provide encouragement within the contexts of praise or constructive criticism.

Because the intention is to motivate, the type of feedback falls more into the categories of encouragement and praise as specified by Schunk and Lilly [97].

These also have the effect of reinforcing statements and advice that are provided to improve or maintain certain actions based on the rate of progress the user has towards their particular goal.

Because of this, the phrases used can have a corrective or confirmatory response based on whether the user has met their milestones or if they have not.

Although this is used within the contexts of encouragement or criticism, the references to the memories are used more as support of a particular piece of advice which means that they are not intended to be used as evaluative or descriptive feedback. Similarly, even though sentiment is extracted from answers given from the past, this is not used for mood matching, but rather to improve the relevance and appropriateness of a reworded phrase and determine in which context it would be appropriate to use a particular sentence.

3 Design Patterns

Design patterns refer to components that are utilized within the design that can be reused, or modified, much like an object within object oriented programming. The typical way in which design patterns are formulated is with the format where the core problem is stated and is then followed by the core solution which is expected to solve the problem.

3.1. DP01: Greeting

Greetings are a dialogue act that are typically used to start a conversation. It represents an acknowledgement of other parties in a conversation. Introductions are used in the case where parties in a conversation are meeting for the first time. To properly become acquainted with another person and address them within the conversation, details such as the other person's name are needed to allow for the relationship to grow beyond being strangers.

3.1.1. Core Problem

The conversational agent should introduce itself to the patient and allow for the patient to understand what its purpose is in the first session.

3.1.2. Core Solution

Build a self introduction state within the dialog. The agent introduces itself, explains its purpose and determines what it can call the patient. An illustration of the design can be found in the figure 3.1

3.2. DP02: Status Update

Given the nature of goal achievement, an incremental approach to make goal achievement more manageable requires some assessment of progress over incremental periods of time during the duration of time spent working towards the goal. To facilitate this, a status update that is used to determine progress, and compare that with the expected progress can be used to determine if the goal is likely to be achieved at the end of the duration or not. Within the dialogue, such status updates will occur in the second and third sessions where the intermediate milestone and the final goal and their achievement are evaluated respectively.

3.2.1. Core Problem

The conversational agent will need to have some way to determine the progress of the patient towards achieving the goal that they set for themselves.

3.2.2. Core Solution

Implement a series of states in the dialog that ask about what the patient accomplished with respect to their goal, and offer advice to handle any difficulties. Refer to the figure 3.2 for an example of how progress is requested and tied in with a memory reference within the context of a status update.

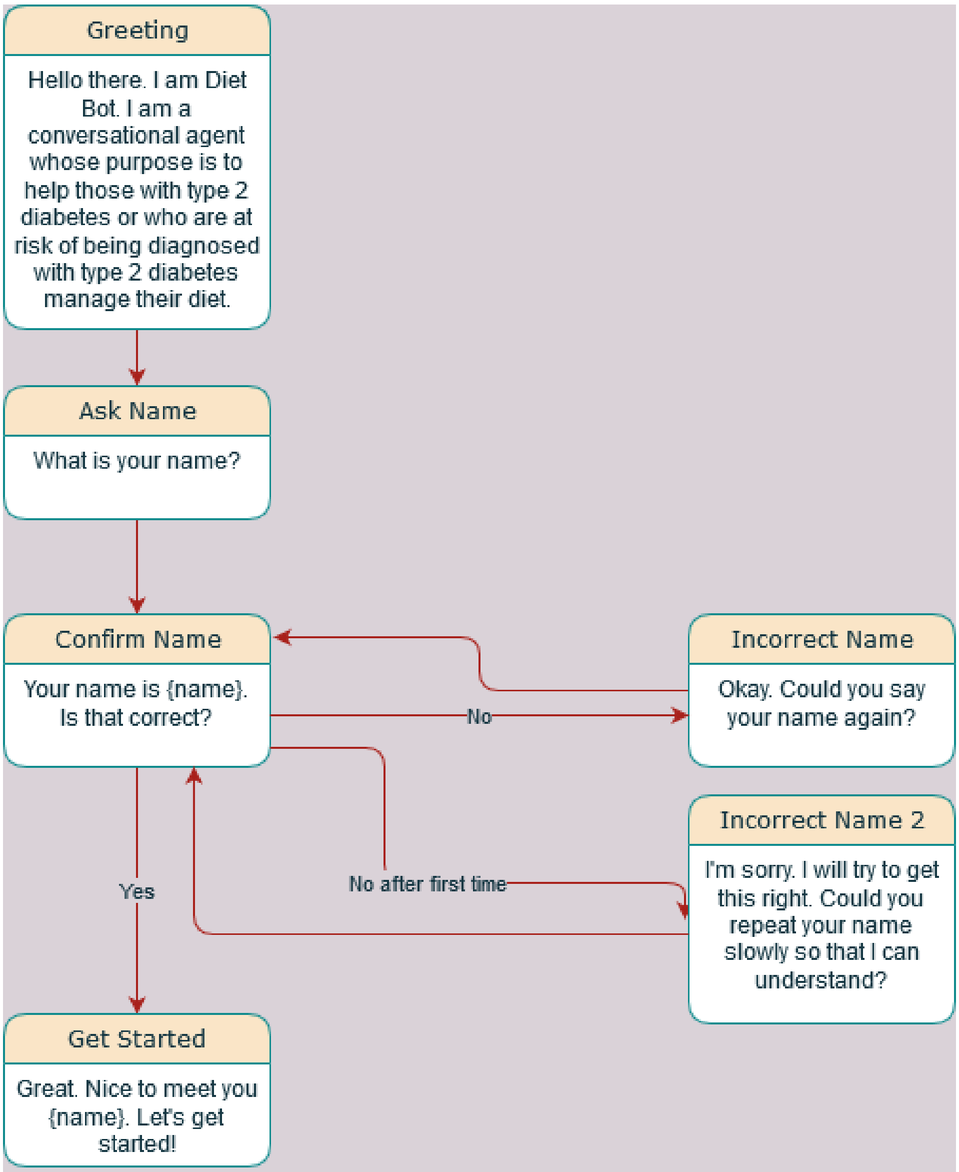


Figure 3.1: Dialogue states depicting the design of the design pattern DPO1: Greeting.

3.3. DP03: Motivationally Guided Shared Memory Reference

How should a memory be used? In this work, references to and reflection of past events that were shared between the user and the conversational agent will be used for improvement and reminding of motivations as well as realigning of motivations towards the chosen goal based on the progress made by a user. Reminiscence can allow for added motivation especially when tied to a form of feedback [32]. To do this, the memories stored in the DP06: Motivationally Guided Memory Storage Points will need to be reworded and used appropriately so that it can be used in a relevant manner that can have some meaning and significance for the user. This requires natural language processing to determine what the memory may mean to the user and an understanding of the context within which the memory will be referred to so that a past event is not used in a way that is incomprehensible or inappropriate.

With regards to the logistics of using a memory reference, these will be used in the second and third session after the first session has been completed to allow for a session in which memories are stored and processed in the long term memory to be completed before the memories are used for reflection. The two main contexts that will be considered are cases where a milestone or goal are met, or cases where they are not. This can constrain the use cases to particular contexts and thus allow for references to be manageable in scope and avoid the generation of any statements that are inappropriate or harmful to the user and their progress.

After a motivational rephrasing of a memory is provided to the user, the agent will ask the user if they agree or disagree with the statement. This can allow for the quality and relevance of the reference to a past event to be analyzed. This also allows for higher levels of user engagement and introspection since in addition to the robot speaking to the user, the user is able to offer their own opinion and consider whether the advice provided within the motivational rephrasing is truly relevant or not.

3.3.1. Core Problem

The agent should be able to remember shared memories that are relevant and present that memory in a way that is not jarring to the patient.

3.3.2. Core Solution

Build a module that can retrieve a memory based on context and reword the contents of the memory into an understandable sentence or series of sentences. In figure 3.2, a simplistic example of how a memory will be utilized for encouragement or constructive advice in a status update is presented at a high level.

3.4. DP04: Education and Exposition

The user will have a number of questions regarding type II diabetes, what type of diet they should follow, strategies for reaching their goal, and questions about how the goal was calculated. Answers to these questions can help the user understand how to reach their goal more effectively through increased clarity. Studies in the past have found that goal clarity can improve and affect motivation [12], making education and exposition a promising way to improve the general interaction. It can also help reduce the impact confounding variables such as the user's level of education and background

Anticipating every possible question the user may have and stating the answers within the dialogue flow before the user asks those questions is an option, but it is impractical and can lengthen the duration of the interaction and cause the user to lose patients. Instead, for general question answering the agent should be able to draw upon the most relevant answers when the user explicitly asks a particular question so as to allow for the user to have a level of autonomy in their own education, and to allow for the user to have greater control over the experience that they wish to have in the interaction. For this, an information retrieval based approach will be used to allow for greater flexibility and the removal of the need to explicitly program individual dialogue states for each particular question that can be asked.

3.4.1. Core Problem

The agent should be capable of drawing on knowledge about diabetes to provide relevant information about diabetes and obesity to properly inform the patient about why they are doing what they are to achieve a goal.

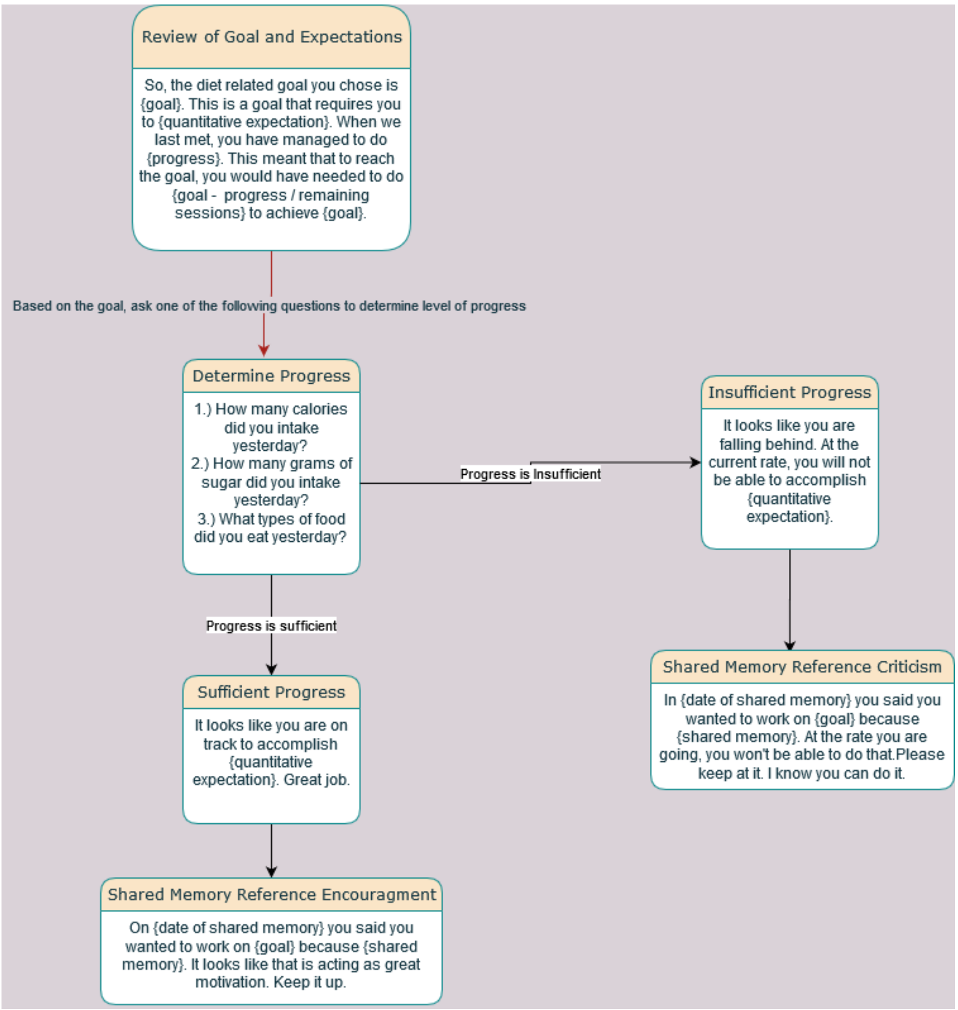


Figure 3.2: Dialogue states depicting the design of the design pattern DPO2 and DP03.

3.4.2. Core Solution

Build a rudimentary knowledge base or procure a database that has facts that are categorized by the types of contexts they can be used in. The knowledge base will utilize existing food databases for nutrition information and frequently asked questions. These frequently asked questions are compiled and organized into question-answer pairs which are used to determine which answer to provide to which question. Since the user can ask any question, the dialogue flow should constrain the user to ask about a particular topic. In addition to this, the information retrieval module will use a method like TF-IDF [8] to find the most similar question and answer if their question does not exactly match the stored list of questions.

3.5. DP05: Identity Recognition

For subsequent sessions, the agent should have some way of recognizing a previous user so that it is able to continue guiding the user towards their goal. To enable this, a unique identifier is required to recognize the user and distinguish the user from other users to avoid the possibility of overwriting the data of one user with another user's data. Under normal circumstances, a video based identification system with a camera would be more ideal, but because experiments will need to take place over Zoom, a user id is requested for identification purposes.

For this work, since Prolific [85] was utilized, it was initially considered that the user can supply their Prolific ID from the very beginning. In cases where speech recognition may be a problem such as an interaction over zoom, a shorter user ID could be supplied in place of a Prolific ID which can be as long as 24 alphanumeric characters. Alternatively, a user id can be provided by the experimenter prior to the interaction with the conversational agent in the case where some data is pre-filled to ensure a smoother interaction. In the case of this work, a shorter ID was provided by the experimenter. In other cases, what would be a typical way of conducting this design pattern would be to have the agent supply the ID to the user instead.

3.5.1. Core Problem

The agent should be able to distinguish the patient it is conversing with from other patients it has conversed with.

3.5.2. Core Solution

Implement a recognition module, or provide users with an identification number that they will need to remember and provide to the agent for each session they have with the agent.

Figure 3.4 shows an example of a scenario where the robot may interact with a patient. In the case of this work, since the users will not be actual patients and will be participants in an experiment, it is expected that they will come prepared with an id. As a result, only the left hand side of the flow would be relevant, and the robot would not consider providing or generating an id.

3.6. DP06: Motivationally Guided Memory Storage Points

The problem of determining what parts of previous conversations and sessions to refer to for DP03: Motivationally Guided Shared Memory Reference requires that the statements that are used to create references have some relevance to the underlying desires and motivations regarding why they are working towards a goal. This means that it is necessary to constrain the scope of possible statements to a specific topic so that the user is inclined to speak regarding the profound underlying desires that made them wish to work towards their particular goal. To this end, the statements to particular questions asking about their feelings and the reasoning behind their goals are stored to later be used for reference and reflection. The parts of the dialogue where such statements are made are referred to as storage points. Such techniques are similar to that which can be found within the ”focusing” and ”evoke” stages of motivational interviewing [91] where the struggles of the user as well as their reasons for working on their goal are explored. This has the advantage of directing the conversation towards a goal-oriented mindset, but it also reduces the need to programmatically choose what events may have intrinsic value when such lines of questioning directly target the intrinsic workings of the user. In addition to this, compared to a system that is lacking such questions, this leads to a higher level of introspection and clarity for the user regarding why they are working on a particular goal to begin with.

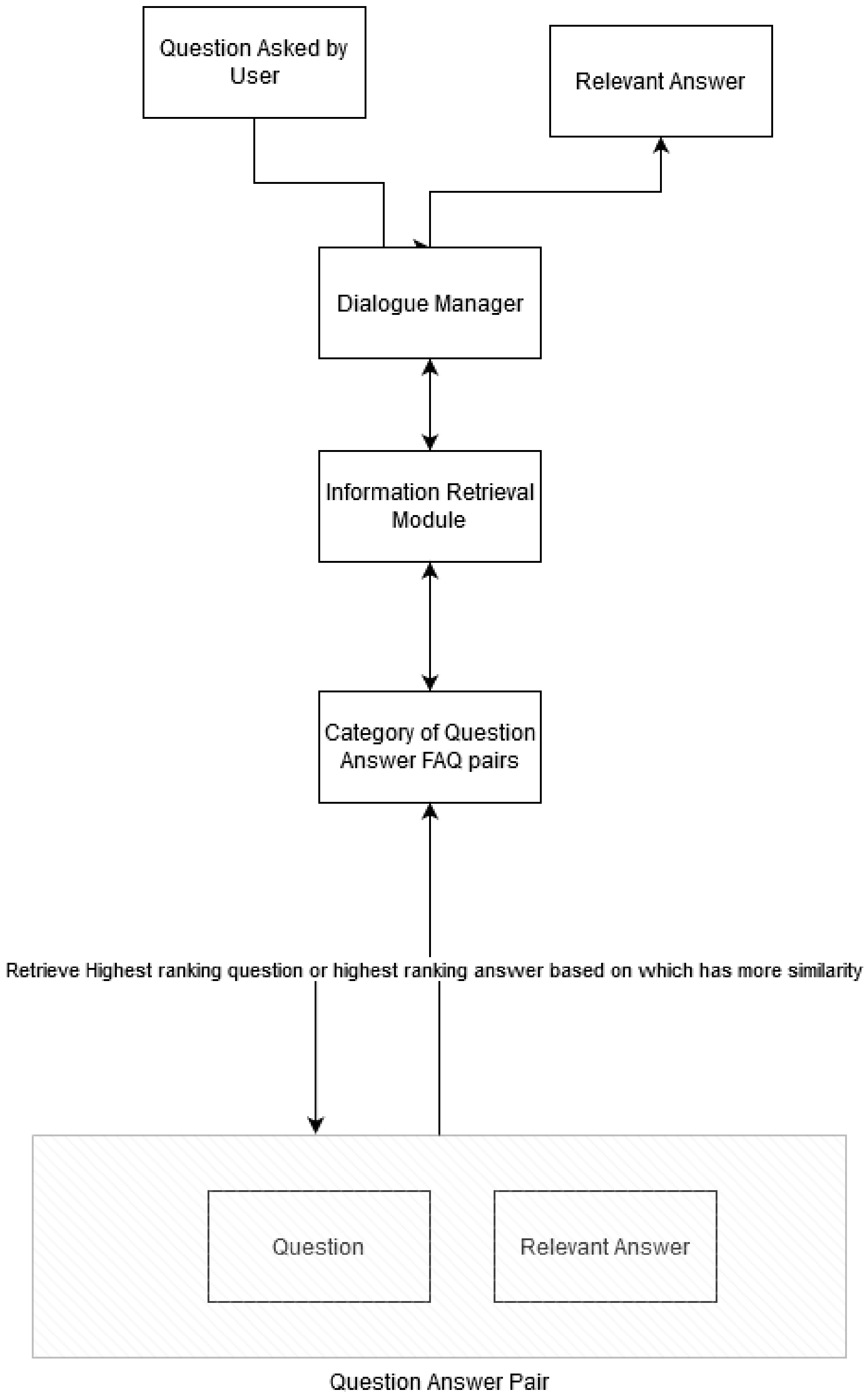


Figure 3.3: Model depicting the way in which information is retrieved for self-guided education within a conversation.

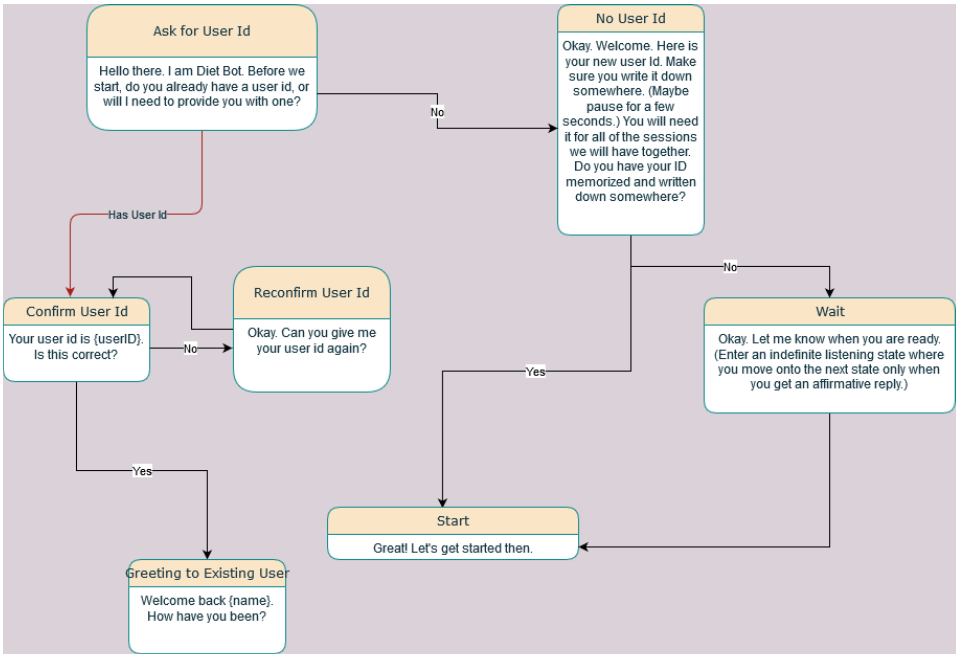


Figure 3.4: Dialog illustrating the way in which identity recognition would be possibly carried out in a clinical context. Within the experiment, only the left hand side of the flow would be considered

Figure 3.5 displays the types of questions that can be asked within such storage points and how such questions will vary between the first session and subsequent sessions as progress is made towards the goal.

3.6.1. Core Problem

The dialog should be able to guide the user to make certain meaningful and profound statements that can then be used as a shared memory.

3.6.2. Core Solution

Use questions that ask about the user's motivation and why they want to accomplish the goal. How they would feel if they accomplish the goal. What accomplishing the goal would allow for them to do. A example of how these questions can be asked in such a way that results in profound or meaningful memories is presented in following figure. One thing to note is that all of the questions asked have answers that can then be used to motivate the patient with their own words.

3.7. DP07: Context Appropriate Feedback

One aspect of a motivational memory and the primary goal of the system which is to encourage goal achievement is encouragement and criticism. To know when to use one over the other, context and the rate of progress must be known. Criticism when the user is meeting their milestone is counterproductive and criticism when the user has a low amount of confidence is also counterproductive. Similarly, not stating that the user's actions are lacking when they are not on track to meet the goal that they chose for themselves implies a lack of transparency. Within the implementation of the conversational agent, this is being handled in the second and third sessions by utilizing the progress of the user along with natural language processing techniques to recognize which of the two is most appropriate. In addition to this, to avoid hurting the feelings or the momentum of the user, strong language is avoided and encouraging phrases are used after stating that their actions are insufficient if they wish to reach their goal.

3.7.1. Core Problem

The agent should be able to encourage or criticize the patient in a way that will result in a significant enough impact on their affective state that it will help them find motivation to continue or change their

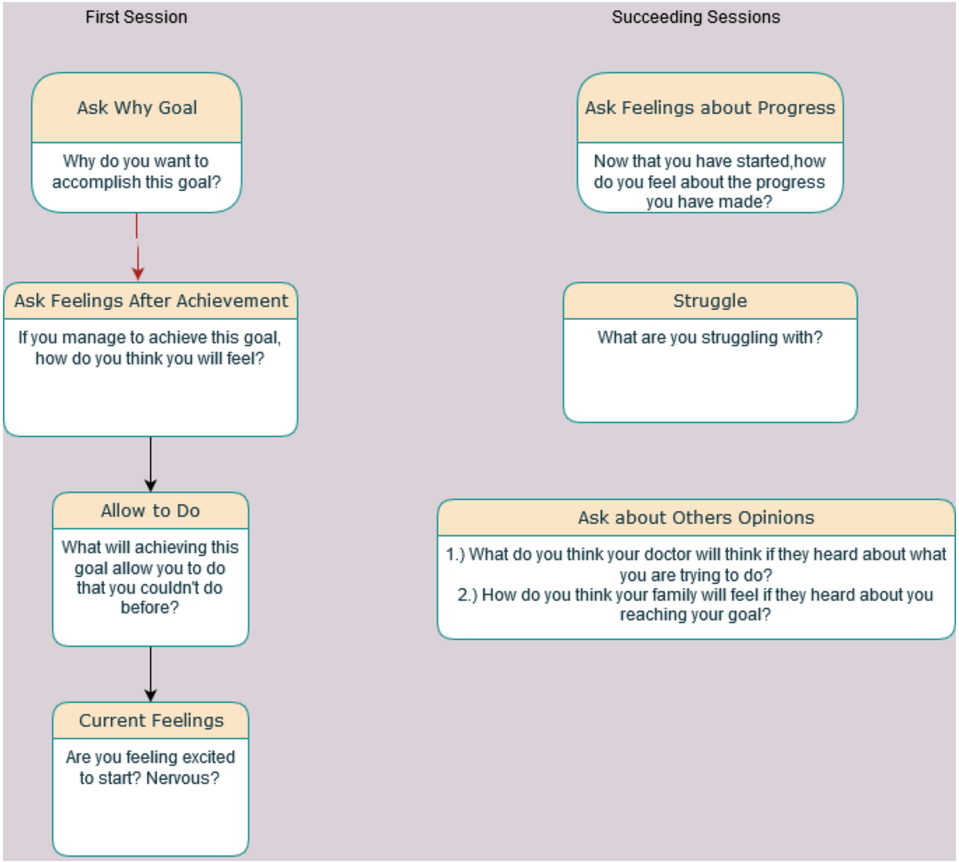


Figure 3.5: Dialog illustrating how questions may be asked to elicit answers with intrinsic value to the user

behavior accordingly. Similar techniques have been used successfully in multiple fields to provide intrinsic motivation such as education [62] and music performance [5]. As mentioned in section 1.2.4, to do this, the use of the feedback types of encouragement and praise will be used to provide motivational momentum based on the progress of the user during a status update as well as possibly incorporate diet related feedback.

3.7.2. Core Solution

Create a module of context appropriate phrases that can be used to encourage or praise based on the progress of the user during each successive session.

3.8. DP08: Requesting Progress

When determining progress, the way in which it is determined can affect the user's experience during the interaction. Having the user track their own progress over the multiple sessions is not only impractical, but also places additional stress on the user when their primary focus should be achieving a particular milestone or goal. To this end, the agent should be capable of storing information each session that is relevant to the patient's efforts towards their goal. In succeeding sessions, the agent can refer to this stored information to draw conclusions regarding whether progress is sufficient or insufficient. This means that when requesting progress, the agent should only need to ask the user what their current consumption is and determine the rest using information stored in memory to make the interaction as seamless as possible. The intended outcome is that users enjoy their interaction with the agent better and that their rates of achievement increase as well.

3.8.1. Core Problem

The agent should be able to determine what the current state of the patient's diet is, and how it may have changed compared to previous sessions.

3.8.2. Core Solution

Ask the patient about their diet. This could be its own reusable module since this will be used in multiple sessions.

Refer to the diagram in figure 3.2 for an example of how progress is requested and tied in with a memory reference within the context of a status update.

3.9. DP09: Milestone Calculation

Since one of the primary purposes of the conversational agent is incremental goal achievement, intermediate milestones are a very important part of that claim. To this end, milestones must be manageable and balanced and should take into account the initial learning curve and also be used to help invoke a feeling of familiarity as they find that similar increments in succeeding sessions are becoming easier to reach. In the case of the experiment being conducted in this study, within the three sessions, the milestone is expected to be completed by the second session and the final goal is expected to be completed by the final session. This implies that there are two main increments within the interaction and one increment is being worked toward at a time.

3.9.1. Core Problem

The conversational agent should be able to generate a reasonable goal and intermediate milestones that can be used to reach the goal.

3.9.2. Core Solution

Build a module that is capable of generating goals based on a number of thresholds for the individual goals so that the goals can be achieved within the three sessions.

Based on the final goal, the module should be capable of generating intermediate milestones using a linear curve and an exponential curve.

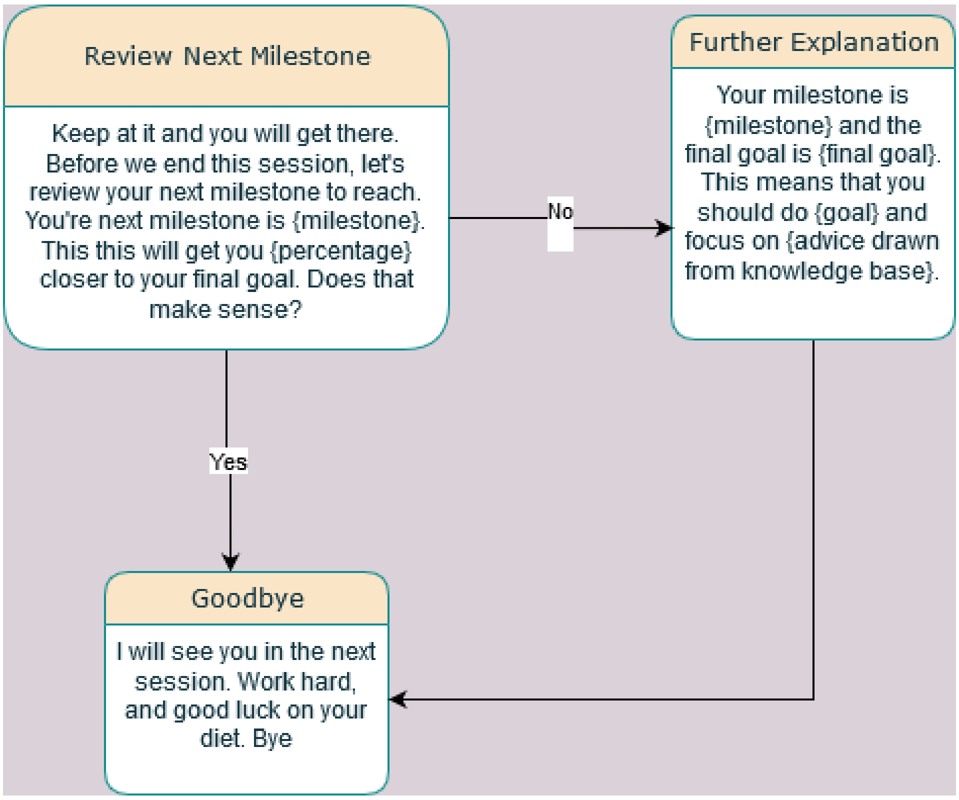


Figure 3.6: Dialog illustrating valediction which ends a conversation

3.10. DP10: Valediction

Similar to the greeting which begins a conversation, at the end of every session, a valediction is needed to end the session and leave on good terms with the user. In addition to this, the user must leave the session with an understanding of what is expected of them for the next session if there is one. Because of this, a thorough review of the milestone and goal should be provided. This implies that, advice specific to the goal should be provided as well so that vagueness and confusion is minimized at the end of the session.

3.10.1. Core Problem

The conversational agent should be able to end a conversation in a session.

3.10.2. Core Solution

Build a series of states that can be reused across sessions that can be used to end a conversation with the patient.

4

Specification

The specification is the component of the design that states what the system will actually do. The specification has the following components.

- Objectives (section 4.1)
  - The main goals of the system being developed
- Use Cases (section 4.2)
  - Realistic interactions that can occur between the users and the system and describe the process of events that occur from beginning to end.
- Requirements (section 4.3)
  - What is required by the system being developed to meet the objectives.
- Claims (section 4.4)
  - What the system claims to offer in terms of the services its provides and a documentation of the advantages, disadvantages and other relevant features relating to the claim.
- Value Stories (section 4.5)
  - Specifies the values of the stakeholders. Has a format that is as follows: ”As a <stakeholder> I want <function> to support <value>.”
- Design Scenario (section 4.6)
  - The design scenario provides a solution to the problem discussed in the problem scenario in section 2.1.5.

Requirements should show a clear connection or link with specific use cases. Similarly, there should be a clear connection between the design scenario and the problem scenario which is discussed in the operational demands component of the foundation.

4.1. Objectives

The main objective is to support DSH01 and DSH02 maintain a new habit by working towards a goal related to diet management. To this end, the conversational agent will have multiple sessions with the patient to allow for the patient to choose a goal and gradually work towards achieving it over the sessions.

4.1.1. OB01: Autonomy in Diet Related Goal Setting for the Patient

By allowing for the patient to have the ability to set their goals, they become more invested in the goal [26] and as a result, there is an added level of intrinsic motivation [60]. Within the context of the proposed conversational agent, this means choosing the type of goal which can be calorie restriction,or sugar reduction.

4.1.2. OB02: Gaining and Solidifying a Diet Related Habit for the Patient

To reach the goal, it is important that the patient is consistent in their efforts. As a result the patient should be motivated to start and be motivated to continue in their efforts to reach their respective goal. To improve the motivation, references to shared experiences will be used to increase relatability and introspection. By using shared experiences, it is more likely to provide a positive user experience that can induce strong emotions. By increasing the level of affective arousal, it is expected that behavior change will become more likely.

4.1.3. OB03: Improved Health and Sense of Achievement for the Patient

By completing a diet related goal, the impact of type II diabetes or risk of being diagnosed with diabetes should decrease which results in a positive outcome for the health of the patient. In addition to this, by accomplishing the goal, the patient will have a sense of achievement which can be very helpful when trying to achieve other goals in the future because of the confidence it builds.

4.1.4. OB04: Engaging Conversation Through Introspection

By utilizing shared experiences, the patient will be more engaged in the conversation, because they will be more likely to feel as though the conversational agent is actually listening to them and understanding who they are as a person. By referring to a statement the patient may have made in the past, the patient not only feels more related to the conversational agent, but they also begin to understand and appreciate their own motivations and desires.

4.1.5. OB05: Improved Understanding of Diet and T2D

During the sessions with the conversational agent, the patient should understand why they are performing certain actions to reach a goal. They should understand how the goal relates to type II diabetes. In addition to this, they should understand why certain staples of their diet may be harmful or not. To handle this appropriately, in initial sessions, the agent should explain the benefits of each type of goal by presenting facts and explaining the relationship between obesity and type II diabetes.

4.2. Use Cases

All use cases will generally follow the same steps, but will vary based on the goal that is chosen. Based on the goal, the intermediate milestones, advice, and dialogue will change slightly to be more appropriate for that specific goal. This ensures a more personalized interaction for the user. Use cases will assume that shared memories will be used, but it should be noted that for conditions that limit the use of shared memories, some steps may be skipped, or that the interactions may be shorter.

4.2.1. UC001: Struggling User

Objective	OB01: Autonomy in Diet Related Goal Setting for the Patient, OB02: Gaining and Solidifying a Diet Related Habit for the Patient, OB04: Engaging Conversation Through Introspection, OB05: Improved Understanding of Diet and T2D
Actors	DSH01: Patient With Type II, Diabetes, DSH02: Patient at risk for Type II Diabetes, Furhat Robot
Pre-condition	The user must be in front of the conversational agent so that it can enter the agent's field of view to begin the interaction. The user should be prepared with the following figures: caloric intake for the previous day, sugar intake for the previous day, and foods that were consumed the previous day.
Post-condition	The conversational agent educates the user regarding type II diabetes, and once the user chooses the goal of calorie restriction, the conversational agent chooses a set of realistic milestones that can be reached over the extended period of time covered by the 3 sessions with the conversational agent. Ideally, the user would have developed a habit of reduced caloric intake by the end of the 3 sessions.

Table 4.1: UC001 Objectives, Actors, Preconditions, and Post Conditions

Action sequence	Session 1:
	1. The first session begins. The conversational agent greets the patient and asks for their unique identifier. Since the user will not have a unique identifier, the agent then knows that this is the first session. In the case of a pregenerated scenario, the agent knows this is the first session, because no data is associated with the user.
	2. The conversational agent asks for the patient's name. It provides the patient with a unique identifier that will be used for future sessions in a non-pregenerated scenario and tells the patient to record the unique identifier so that it can be used for identification purposes in the next two sessions.
	3. The conversational agent explains to the user what its purpose is, and provides an introductory lesson on type II diabetes and the types of behaviors that can lead to prevention or management of type II diabetes.
	4. The conversational agent asks if the user understands, and if the user does not, the conversational agent can provide an explanation with certain parts rephrased.
	5. The conversational agent then informs the patient of the three goals it can work on with the patient and explains the goals in detail. These goals are calorie restriction and sugar reduction.
	6. The patient chooses one of the two goals. The conversational agent then asks questions regarding the patient's motivations, why they choose that goal, and other related questions. The answers to these questions will be stored and used for shared memory references in future sessions.
	7. The conversational agent asks the patient what their caloric intake or sugar intake was for the previous day based on the type of goal that they chose.
	8. Based on the current caloric intake or sugar intake, the conversational agent calculates a final goal between 80-90% of the current intake and calculates the intermediate milestone that needs to be achieved for the next session.
	9. The conversational agent provides advice and suggestions on how to reach the intermediary milestone and executes a valediction.

Table 4.2: UC001 Session 1 Action Sequence

Action Sequence	Session 2
	10. The second session begins. The conversational agent greets the patient and asks for their unique identifier.
	11. The conversational agent then asks the user's caloric intake or sugar intake for the previous day based on the type of goal that they chose.
	12. The patient's progress does not meet the milestone. This means that the conversational agent must provide encouragement to improve the patient's confidence.
	13. The conversational agent makes a reference to a memory from the previous session and uses a motivational phrase intended for encouragement in the context of not meeting the milestone that was generated from the user's past statement in the previous session. The conversational agent also asks if the user agrees or disagrees with the statement to encourage higher engagement and also increased retrospection.
	14. Since the user is not meeting the intermediate goal, the conversational agent offers the user the option of working towards a less ambitious goal to provide a certain amount of difficulty scaling to support the user's needs.
	15. The conversational agent then states what the next milestone is and provides tips on how to efficiently reach the next milestone.
	16. The conversational agent asks a new set of questions about the patient's motivation and desires to be used as shared memory references in the final session.
	17. The conversational agent executes a valediction and ends the conversation.

Table 4.3: UC001 Session 2 Action Sequence

Action Sequence	Session 3
	18. The third session begins. The conversational agent greets the patient and asks for their unique identifier.
	19. The conversational agent then asks the user's caloric intake or sugar intake for the previous day based on the type of goal that they chose.
	20. Depending on whether the user reached the final milestone or not, the conversational agent will praise or provide some advice.
	21. The conversational agent will then reflect on the progress that was made over all of the sessions, and make references to shared memories. For the possible scenarios, in this use case that can be considered, the conversational agent may say that the user struggled at first, but then achieved their goal or that the user did not manage to achieve their milestone or their goal and still needs to work to make a positive diet change.
	22. The conversational agent then provides suggestions for future goals the patient can carry out in a self directed manner. In the case of the user failing to meet their goal, the conversational agent will suggest working towards their goal, but over a longer period of time to ensure a higher level of ease with regards to taking the first step.
	23. The conversational agent executes a valediction that is appropriate for the final session and ends the conversation.

Table 4.4: UC001 Session 3 Action Sequence

UC Step	Requirements	Claims	Design Patterns
1	RQ001: Embodied Conversational Agent	CL005: A Friendly and Relatable Conversational Partner	DP01: Greeting
2	RQ004: User Identification	CL005: A Friendly and Relatable Conversational Partner	DP01: Greeting
3	RQ005: Knowledge Based Advice	CL003: Education	DP04: Education
4	RQ005: Knowledge Based Advice	CL003: Education	DP04: Education and Exposition
5	RQ002: Dialogue, RQ005: Knowledge Based Advice	CL001: Improved Autonomy through Goal Setting	DP04: Education and Exposition
6	RQ002: Dialogue, RQ003: Memory	CL002: An Engaging Conversation through Shared Experiences, CL005: A Friendly and Relatable Conversational Partner	DP06: Motivationally Guided Memory Storage Points
7	RQ002: Dialogue, RQ003: Memory	CL004: Incremental Goal Achievement	DP08: Requesting Progress
8	RQ005: Knowledge Based Advice	CL004: Incremental Goal Achievement	DP09: Milestone Calculation
9	RQ002: Dialogue, RQ005: Knowledge Based Advice	CL003: Education, CL005: A Friendly and Relatable Conversational Partner	DP04: Education and Exposition, DP10: Valediction
10	RQ001: Embodied Conversational Agent, RQ004: User Identification	CL005: A Friendly and Relatable Conversational Partner	DP01: Greeting
11	RQ002: Dialogue	CL004: Incremental Goal Achievement	DP02: Status Update
12	RQ002: Dialogue	CL004: Incremental Goal Achievement	DP07: Context Appropriate Feedback
13	RQ003: Memory	CL002: An Engaging Conversation through Shared Experiences	DP03: Motivationally Guided Shared Memory Reference
14	RQ005: Knowledge Based Advice	CL004: Incremental Goal Achievement	DP09: Milestone Calculation
15	RQ005: Knowledge Based Advice	CL003: Education	DP04: Education and Exposition
16	RQ002: Dialogue, RQ003: Memory	CL002: An Engaging Conversation through Shared Experiences, CL005: A Friendly and Relatable Conversational Partner	DP06: Motivationally Guided Memory Storage Points
17	RQ002: Dialogue	CL005: A Friendly and Relatable Conversational Partner	DP10: Valediction
18	RQ001: Embodied Conversational Agent, RQ004: User Identification	CL005: A Friendly and Relatable Conversational Partner	DP01: Greeting
19	RQ002: Dialogue	CL004: Incremental Goal Achievement	DP02: Status Update
20	RQ002: Dialogue	CL004: Incremental Goal Achievement	DP07: Context Appropriate Feedback
21	RQ003: Memory, RQ005: Knowledge Based Advice	CL004: Incremental Goal Achievement, CL003: Education	DP03: Motivationally Guided Shared Memory Reference
22	RQ005: Knowledge Based Advice	CL004: Incremental Goal Achievement, CL003: Education	DP04: Education and Exposition
23	RQ002: Dialogue	CL005: A Friendly and Relatable Conversational Partner	DP10: Valediction

Table 4.5: Related Requirements, Use cases, and Design Patterns for UC001

Objective	OB01: Autonomy in Diet Related Goal Setting for the Patient, OB02: Gaining and Solidifying a Diet Related Habit for the Patient, OB04: Engaging Conversation Through Introspection, OB05: Improved Understanding of Diet and T2D
Actors	DSH01: Patient With Type II Diabetes, DSH02: Patient at risk for Type II Diabetes, Furhat Robot
Pre-condition	The user must be in front of the conversational agent so that it can enter the agent's field of view to begin the interaction. The user should be prepared with the following figures: caloric intake for the previous day, sugar intake for the previous day, and foods that were consumed the previous day.
Post-condition	The conversational agent educates the user regarding type II diabetes, and once the user chooses the goal of sugar reduction, the conversational agent chooses a set of realistic milestones that can be reached over the extended period of time covered by the 3 sessions with the conversational agent. Ideally, the user would have developed a habit of reduced sugar intake by the end of the 3 sessions.

Table 4.6: UC002 Objectives, Actors, Preconditions, and Post Conditions

4.2.2. UC002: Overachieving User

Action sequence	Session 1:
	1. The first session begins. The conversational agent greets the patient and asks for their unique identifier. Since the user will not have a unique identifier, the agent then knows that this is the first session.
	2. The conversational agent asks for the patient's name. It provides the patient with a unique identifier that will be used for future sessions and tells the patient to record the unique identifier so that it can be used for identification purposes in the next two sessions.
	3. The conversational agent explains to the user what its purpose is, and provides an introductory lesson on type II diabetes and the types of behaviors that can lead to prevention or management of type II diabetes.
	4. The conversational agent asks if the user understands, and if the user does not, the conversational agent can provide an explanation with certain parts rephrased.
	5. The conversational agent then informs the patient of the three goals it can work on with the patient and explains the goals in detail. These goals are calorie restriction and sugar reduction.
	6. The patient chooses either calorie restriction or sugar reduction. The conversational agent then asks questions regarding the patient's motivations, why they choose that goal, and other related questions. The answers to these questions will be stored and used for shared memory references in future sessions.
	7. The conversational agent asks the patient what their caloric or sugar intake was for the previous day.
	8. Based on the current caloric or sugar intake, depending on the chosen goal, the conversational agent calculates a final goal between 80-90% of the current caloric or sugar intake and calculates the intermediate milestone that needs to be achieved for the next session.
	9. The conversational agent provides advice and suggestions on how to reach the intermediary milestone, recommended foods with less sugar, and executes a valediction.

Table 4.7: UC002 Session 1 Action Sequence

Action sequence	Session 2:
	10. The second session begins. The conversational agent greets the patient and asks for their unique identifier.
	11. The conversational agent then asks the user's caloric or sugar intake for the previous day depending on the goal that user chose in the previous session.
	12. The patient's progress not only meets the milestone, but it also exceeds it by a significant amount. Because of this, the conversational agent can provide praise since the patient is progressing in the correct direction and is also exceeding expectations.
	13. The conversational agent makes a reference to a memory from the previous session to support its praise and asks if the user agrees or disagrees with its statement to encourage higher engagement and retrospection.
	14. Since the user is significantly exceeding the intermediate goal, the conversational agent discusses a more ambitious goal which the user is free to follow if they prefer.
	15. The conversational agent then states what the next milestone is and provides tips on how to efficiently reach the next milestone. This is another opportunity to provide praise with a shared memory reference.
	16. The conversational agent asks a new set of questions about the patient's motivation and desires to be used as shared memory references in the final session.
	17. The conversational agent executes a valediction and ends the conversation.

Table 4.8: UC002 Session 2 Action Sequence

Action sequence	Session 3:
	18. The third session begins. The conversational agent greets the patient and asks for their unique identifier.
	19. The conversational agent then asks the user's caloric or sugar intake for the previous day.
	20. Depending on whether the user reached the final milestone or not, the conversational agent will praise or provide some advice. In this case, since the user is an overachieving patient, the robot will provide praise.
	21. The conversational agent will then reflect on the progress that was made over all of the sessions, and make references to shared memories.
	22. The conversational agent then provides suggestions for future goals the patient can carry out in a self directed manner. In this case, since the user did meet their final goal, the user is told that they can set goals on their own and work towards more ambitious goals. The conversational agent does, however, provide a warning stating that any succeeding goals should be reasonable, incremental, and if in doubt, to contact their general practitioner or dietitian/nutritionist.
	23. The conversational agent executes a valediction that is appropriate for the final session and ends the conversation.

Table 4.9: UC002 Session 3 Action Sequence

UC Step	Requirements	Claims	Design Patterns
1	RQ001: ECA, RQ004	CL005: A Friendly Partner	DP01: Greeting
2	RQ001, RQ004: User Identification	CL005: A Friendly Partner	DP01: Greeting
3	RQ005: Knowledge Based Advice	CL003: Education	DP04: Education and Exposition
4	RQ005: Knowledge Based Advice	CL003: Education	DP04: Education and Exposition
5	RQ002: Dialogue, RQ005: Knowledge Based Advice	CL001: Improved Autonomy through Goal Setting	DP04: Education and Exposition
6	RQ002: Dialogue, RQ003: Memory	CL002: An Engaging Conversation through Shared Experiences, CL005: A Friendly and Relatable Conversational Partner	DP06: Motivationally Guided Memory Storage Points
7	RQ002: Dialogue, RQ003: Memory	CL004: Incremental Goal Achievement	DP08: Requesting Progress
8	RQ005: Knowledge Based Advice	CL004: Incremental Goal Achievement	DP09: Milestone Calculation
9	RQ002: Dialogue, RQ005: Knowledge Based Advice	CL003: Education, CL005: A Friendly and Relatable Conversational Partner	DP04: Education and Exposition, DP10: Valediction
10	RQ001: Embodied Conversational Agent, RQ004: User Identification	CL005: A Friendly and Relatable Conversational Partner	DP01: Greeting
11	RQ002: Dialogue	CL004: Incremental Goal Achievement	DP02: Status Update
12	RQ002: Dialogue	CL004: Incremental Goal Achievement	DP07: Context Appropriate Feedback
13	RQ003: Memory	CL002: An Engaging Conversation through Shared Experiences	DP03: Motivationally Guided Shared Memory Reference
14	RQ005: Knowledge Based Advice	CL004: Incremental Goal Achievement	DP09: Milestone Calculation
15	RQ005: Knowledge Based Advice	CL003: Education	DP04: Education and Exposition
16	RQ002: Dialogue, RQ003: Memory	CL002: An Engaging Conversation through Shared Experiences, CL005: A Friendly and Relatable Conversational Partner	DP06: Motivationally Guided Memory Storage Points
17	RQ002: Dialogue	CL005	DP10: Valediction
18	RQ001: ECA, RQ004: User Identification	CL005: A Friendly and Relatable Conversational Partner	DP01: Greeting
19	RQ002: Dialogue	CL004: Incremental Goal Achievement	DP02: Status Update
20	RQ002: Dialogue	CL004: Incremental Goal Achievement	DP07: Context Appropriate Feedback
21	RQ003: Memory, RQ005: Knowledge Based Advice	CL004: Incremental Goal Achievement, CL003: Education	DP03: Motivationally Guided Shared Memory Reference
22	RQ005: Knowledge Based Advice	CL004: Incremental Goal Achievement, CL003: Education	DP04: Education and Exposition
23	RQ002: Dialogue	CL005: A Friendly and Relatable Conversational Partner	DP10: Valediction

Table 4.10: Related Requirements, Use cases, and Design Patterns for UC002

4.3. Requirements

Here, the requirements of the conversational agent are specified. All requirements apply to both use case UC001 (4.2.1) and UC002 (4.2.2)

4.3.1. RQ001: Embodied Conversational Agent

Specification The agent must be capable of displaying gestures, taking in input in the form of speech from the patient, and outputting responses using text to speech. In other words, it should be able to perform basic conversation with the patient such that it can provide information to the user through speech and take in and record information that is spoken by the patient.

4.3.2. RQ002: Dialogue

Specification The agent should be able to give appropriate responses to questions and answers given the context. For this, a state machine that is designed to handle branching dialogues and consider the most likely responses from the patient is needed. In addition to spoken responses, gestures like the nodding and shaking of the head can be useful for certain contexts and can help moderate the flow of the conversation. For question answering, an information retrieval based approach can be used to allow for a greater number of questions to be answered than can be feasible with a handcrafted state machine.

4.3.3. RQ003: Memory

Specification To share experiences, a long term memory is required to refer to shared experiences from other sessions and carry useful information across sessions. The agent must be able to remember the words spoken by the patient and the context that it was spoken in. It must then be able to determine sentiment so that it can determine whether to use the memory, and in which context it can use that experience. Once this is done it must be able to rephrase the experience into a motivational phrase such that it is not used in an inappropriate way within the context in which it is used.

4.3.4. RQ004: User Identification

Specification To allow for patients to have successive sessions, the agent should be capable of remembering individual patients so that it is capable of making a personal connection. To allow for this, patients will be provided with a unique identifier that they should record and provide to the conversational agent. In the case of this study, since participants will be recruited through Prolific, they will already have a unique identifier. This means that the agent should be capable of asking for the identifier, storing the id and associating any subsequent information provided by the patient with that identifier. If providing the Prolific ID through speech is too difficult, participants can be provided with a significantly shorter id.

4.3.5. RQ005: Knowledge Based Advice

Specification To properly advise the patient regarding the goals, their relationship to type II diabetes, as well as general diet related facts, the agent will need to be able to draw from a knowledge base that contains diet related data, diabetes related data, and apply that accordingly to the conversation when appropriate. The simplest form of such a knowledge base is that of an FAQ (frequently answered questions) bank that can address the majority of topics that a user may be concerned with.

For this, a search or information retrieval based approach will be used to retrieve the appropriate knowledge when the user has questions regarding a particular topic. In addition to this, some facts and words of advice that will apply to most patients will be provided during the interaction through the state machine that facilitates the dialogue for the three sessions.

4.3.6. RQ006: Goal Aligned Motivation

Specification The agent must be able to provide the proper advice to the patient to ensure that they are aware of their own motivations and are reminded of them through the use of praise and encouragement that is context specific to the patient's needs. This means to use praise to ensure continuation of certain actions when the progress is in line with what is expected to reach the goal and to use constructive criticism and encouragement when progress is insufficient. Since motivation can decrease over time, reminding the patient through use of shared experiences and framing them within the context of the goal that is currently pursued can increase the clarity of the goal and commitment to the goal itself.

4.4. Claims

Claims refer to what the system being developed claims to do. Claims are organized into upsides which represent the benefits of the claim, downsides which represent consequences, and finally the use cases and tests associated with the respective claim. Use cases refer to the use cases defined in section 4.2 and tests refer to the methods used to evaluate the claims which are defined in section 5.2.

4.4.1. CL001: Improved Autonomy through Goal Setting

Upside Patients are able to choose a manageable goal that would be most appropriate for them. This increases the chances of them achieving that goal. In addition to this, since patients are given a choice, they have a greater sense of ownership which can improve their intrinsic motivation. To measure levels of autonomy, MI1: Questionnaire for Experience (2.2.5) will be used and MI2: Achievement Metrics (2.2.5) will be used for tracking the type of goal that is chosen.

Downside The inclusion of a choice means that the dialogue will be slightly longer in the first session. For more impatient participants, this can be seen as a downside. Similarly, some participants may be dismayed by the posing of a choice as extra work rather than a benefit. Such downsides may be observed within MI1: Questionnaire for Experience (2.2.5).

Use Cases UC001: Calorie Restriction (4.2.1) and UC002: Overachieving User (4.2.2).

Tests Test001: Modified Godspeed Questionnaire (5.2.1).

4.4.2. CL002: An Engaging Conversation through Shared Experiences

Upside By utilizing shared experiences, there is an increased level of novelty in the interaction with the conversational agent. This means that the conversation itself becomes more engaging for the patient. In addition to having an improved experience, it may be possible for a more engaging experience to result in higher levels of investment in the goal achievement as well. To measure this, MI1: Questionnaire for Experience (2.2.5) and MI2: Achievement Metrics (2.2.5) will be used.

Downside In some cases, the use of shared experiences can result in a negative result if the experience has a negative connotation. What is key here is that the shared experience is interpreted and phrased in such a way that the user is not confused or taken aback by the reference. In addition to this, the use of shared experiences for motivation can slow down the dialogue. Such downsides may be observed within MI1: Questionnaire for Experience (2.2.5)).

Use Cases UC001: Calorie Restriction (4.2.1) and UC002: Overachieving User (4.2.2).

Tests Test001: Modified Godspeed Questionnaire (5.2.1).

4.4.3. CL003: Education

Upside The conversational agent informs the patient of why a goal is useful, how it relates to type II diabetes and obesity. It backs up all of the statements it makes with facts that are established in current research. As a result, the patient will have a greater understanding of type II diabetes and will be in a better position to prevent and treat type II diabetes after they finish all of the sessions with the conversational agent. To measure the level of knowledge gained, MI1: Questionnaire for Experience (2.2.5) will be used.

Downside For those who are more interested in simply achieving the goal, this can slow down the process. In addition to this, if the education provided is too high level for the patient, they may lose their interest in working towards their specific goal. With that said, all of these aspects will depend on the way in which the education is delivered, so allowing for patients to choose what they would like to have explained may be a better approach. Since the patient is able to ask the robot questions, the patient is given the option of a self-directed approach where they can explore the topics that they are interested in in greater detail. This can be detected within MI1: Questionnaire for Experience (2.2.5)

Use Cases UC001: Calorie Restriction (4.2.1) and UC002: Overachieving User (4.2.2).

Tests Test001: Modified Godspeed Questionnaire (5.2.1).

4.4.4. CL004: Incremental Goal Achievement

Upside By using intermediary milestones, it becomes easier to reach the final goal. This is true not only in terms of the actual work that is required, but also in terms of lessening the cognitive load by splitting the goal into easy to understand segments. To capture this, MI1: Questionnaire for Experience (2.2.5) and MI2: Achievement Metrics (2.2.5) are used.

Downside For some patients, the intermediary goals may slow down their progress in the case where their rate of achievement is faster than what was originally planned. Since the point is to create easily maintainable habits, this is not regarded as a problem, but it can be seen as a downside. In such cases, the user is given the ability to work towards a more ambitious goal as a way of difficulty scaling that keeps the interactions and the goal interesting. The effect on achievement will be captured in MI2: Achievement Metrics (2.2.5).

Use Cases UC001: Calorie Restriction (4.2.1) and UC002: Overachieving User (4.2.2).

Tests Test002: Achievement Metrics (5.2.2).

4.4.5. CL005: A Friendly and Relatable Conversational Partner

Upside By greeting the patient, providing advice, praise and executing human-like head movements, the conversational becomes an entity that can be trusted and related to. Such an approach should lead to a more positive experience and lead to a perception of the robot as something more than a stranger. It is expected that such relatability may also help with achievement rates and interest in the conversation. Questions asked in MI1: Questionnaire for Experience (2.2.5) can offer insight into such a claim.

Downside For patients who do not like exchanging pleasantries, or are less responsive to praise, this can be seen as an overhead that wastes time. For such users, the general approach of an embodied conversational agent may less effective in general compared to a text based chatbot or a completely self guided approach that is far more streamlined. Such opinions will be captured within MI1: Questionnaire for Experience (2.2.5).

Use Cases UC001: Calorie Restriction (4.2.1) and UC002: Overachieving User (4.2.2).

Tests Test001: Modified Godspeed Questionnaire (5.2.1).

4.5. Value Stories

Value stories, or user stories, specify how stakeholders would like the system to support their values and expectations. Each story specifies a relevant stakeholder, the function the stake holder wants the system to fulfill, and the values that are being supported by the function. Specification of such stories can allow for a clearer understanding of what the stake holders want so as to optimize for their highest priority values.

The format is as follows: As a <stakeholder> I want <function> to support <value>.

The value stories are listed in table 4.11.

Stakeholder	Function	Value
As a person with type 2 diabetes (DSH01: Patient With Type II Diabetes) or at risk of having type 2 diabetes (DSH02: Patient at risk for Type II Diabetes)	I want the agent to allow me to choose a diet related goal so that I have	more confidence, autonomy, and intrinsic motivation.
As a person with type 2 diabetes (DSH01: Patient With Type II Diabetes) or at risk of having type 2 diabetes (DSH02: Patient at risk for Type II Diabetes)	I want the agent to provide encouragement or criticism as needed so that I can have	an improved chance of achievement and clearer understanding of the goal.
As a person with type 2 diabetes (DSH01: Patient With Type II Diabetes) or at risk of having type 2 diabetes (DSH02: Patient at risk for Type II Diabetes)	I want the agent to remember aspects of our conversation so that I can have a more	satisfying experience and a deeper and more fulfilling interaction.
As a person with type 2 diabetes (DSH01: Patient With Type II Diabetes) or at risk of having type 2 diabetes (DSH02: Patient at risk for Type II Diabetes)	I would like the agent to provide advice specific to my needs based on the questions I ask so that I	can have a better understanding of why my goal is important and how I can achieve my goal
As a primary care physician (DSH03: Doctor/Primary Care Physician)	I want the agent to help patients to learn and solidify healthy diet related habits so that I can increase their	life expectancy and well-being.

Table 4.11: Value Stories

4.6. Design Scenario

The design scenario provides a solution to the problem scenario provided in Problem Scenario.

To help DSH01: Dave (40) (2.1.2) and DSH02: Amy (30) (2.1.2) gain positive habits to prevent or manage type II diabetes, DSH03: John (30) (2.1.2) introduces them to a conversational agent that can help them learn and maintain a diet related habit through three sessions. DSH01: Dave (40) (2.1.2) and DSH02: Amy (30) (2.1.2) will be referred to collectively as the patient (Patients) that will be helped by the conversational agent.

The patient begins the first session by introducing his or herself to the conversational agent. Once the patient and the agent have introduced themselves to one another, the agent explains to the patient what its purpose is. The agent informs the patient of the types of goals the patient can work towards and explains what the goal involves and the benefits of reaching that goal. The agent also informs the patient of how the goal is related to type II diabetes and why that goal can be useful. The agent asks the patient what type of goal the patient would like to work towards and what the patient's current diet looks like. The agent also asks why the patient wants to work on that goal and other questions regarding the patient's motivation and what outcomes they would like to see. This not only creates a shared experience, but it also increases the amount of introspection carried out by the patient which makes Dave DSH01 and Amy DSH02 have more clarity regarding their own motivations. Once the agent and the patient grasp the current state of the patient's diet, the agent calculates what would be a realistic milestone for the patient to reach by the final session and calculates intermediate milestones to reach by each succeeding session. The agent informs the patient of these milestones, and the patient is informed of exactly what is expected of them for the next session.

Between sessions DSH01: Dave (40) and DSH02: Amy (30) may mention what they are doing to their friends and ISH01: Family Members (2.1.2). Since interacting with a conversational agent can be considered an interesting, novel experience, the chances of DSH01: Dave (40) and DSH02: Amy (30) mentioning what they are doing in their first session is relatively high. Doing so can increase the level of accountability and ownership the patient has over accomplishing their diet related goal. Between sessions, the patient can either work towards their milestone or not reach their milestone.

A day later, the patient has their second session with the agent. Since the patient and the agent already know each other, they speak in more familiar terms during the greeting which makes the patient relate more to the agent. The agent then asks how much progress the patient made. If the patient reached the milestone, the agent praises the patient and then refers to a shared memory to use as encouragement. If the patient did not reach the milestone, the agent offers constructive advice and then refers to a shared memory as a way of holding the patient accountable to their own words. In the case the patient did not reach the milestone, whether to adjust the final milestone will be discussed. This allows for a certain level of personalization as well as difficulty scaling to allow for the patient's level of motivation to not be undercut by a final goal that is too easy or a final goal that is too hard. The agent then asks questions similar to the questions it asked in the first session to use as shared memories. The agent then reiterates the milestones and the actions the patient will need to take before concluding.

A day later, the patient has their final session with the agent. The final session proceeds in a fashion similar to the second session. However, when it is time to provide encouragement or advice, the agent instead reflects on both of the previous sessions and the rate of progress while making references to shared memories. The agent then offers a recommendation for future action. If the patient did not reach their goal, the agent could suggest working towards that goal over a longer period of time. If the patient did reach their goal, the agent could offer suggestions for subsequent goals they could work towards in the future. The agent then concludes and thanks the patient for the effort they expended regardless of the result.

Based on these suggestions, the patient can present their doctor DSH03: John (30) (2.1.2) with their results and discuss future actions to take or continue to maintain and improve upon their habits in a self-directed manner.

This scenario solves many problems. Regardless of whether the patient achieves their goal or not, the patient will have an improved understanding of type II diabetes and positive diet related habits. Also, because of the questions that are asked for the purpose of shared memories, the patient will have a better understanding of their own motivations and desires. If the patient actually makes an effort towards any of the milestones, this means that they are in a position to work towards a less ambitious goal. If the patient does achieve their goal, they will be able to move onto more challenging goals while also having increased confidence and a sense of achievement. For those who do not have the time or the funds to see an actual dietitian, this also has an added benefit of providing easily accessible diet advice for those who need it.

5

Evaluation

The evaluation determines the quality and performance of the system that is designed and implemented. The following are the components that make up the evaluation.

- Prototype (section 5.1)
  - This is the conversational agent or implementation of the system that is designed.
- Tests (section 5.2)
  - This specifies what questions will be asked, what will be recorded, and how that recorded data will be used to draw conclusions.
- Experimental Study Design (section 5.3)
  - This is the experiment that is carried out to evaluate the effectiveness of the prototype. This section covers the types of participants, how the test will be carried out, and any other considerations.
- Results (section 5.5)
  - This will contain the results of the test that was carried out.

Prior to the actual test, a pilot test with 3 individuals was carried out to fix any deficits and tune different aspects of the prototype as well as the study itself.

5.1. Prototype

The prototype refers to the actual system that is developed and used in the evaluation. Since this is a conversational agent with some novel functionalities, this consists of many different components that were needed to implement those functions which are specified in the following subsections.

5.1.1. Furhat

The Furhat robot comes with an SDK that allows for development using a virtual embodiment of the Furhat robot. An image of the virtual embodiment is provided in figure 2.2. This virtual embodiment is what is shown to participants over zoom calls.

For the prototype, only the speech recognition, the text to speech generation, and the embodiment with gestures are used.

With regards to important functionalities such as data persistence, user recognition, and NLP, the version of Furhat used is either lacking or limited which requires other modules to be used. For this reason, all other functionality is offloaded to a a flask server that is communicated with during the conversation until a termination signal is sent from the Flask server which managers the dialogue.

The version of the SDK that was used is Windows version 1.21.0. For code run by Furhat, Kotlin version 1.3.31 was used.

5.1.2. Flask Server

To allow for greater flexibility and additional features not available within Furhat, Furhat communicates with a Flask server [40] for all dialogue management. The Furhat robot sends and receives dialogue information using either GET or POST API requests. Since the server is run locally, there is no latency. By using a Flask server, it allows for the dialogue and more complex functionality to be written in function and for sophisticated Python libraries to be used if needed. This is particularly true for NLP related libraries that are used for more complex responses.

The version of Python used here is Python 3.7.

The version of Flask used is version 1.1.2.

NLP libraries used include the following:

- SPACY 2.1.3
- NLTK 3.4.5
- GENSIM 3.7.1

Libraries that are used for data processing and other tasks are the following:

- PANDAS 0.23.4
- NUMPY 1.19.4

5.1.3. Dialogue Manager

This is the dialogue manager which processes responses made by the user and determines how to make the agent respond. Any aspects that require training a model will usually load a model from a pickle file for quick retrieval to avoid unnecessarily training a model every time a conversation is held.

The dialogue manager has multiple components.

- Finite State Machine
  - A series of statements and responses that determines the flow of the dialogue. This is where the majority of the dialogue will be handled. This is a fairly standard way of organizing and managing dialogue [65].
- Information Retrieval
  - A module that is able to use a query to search for relevant information. This can be used for open ended question answering that can be less restrictive than a finite state machine, but lacks the direction or flow that can be offered by a finite state machine.
  - This utilizes techniques such as TF-IDF and Cosine Similarity [89] to find the best answer for a particular question using question and answer pairs. An alternative that was tried and considered was Chatterbot which is a conversational agent library that learns uses conversations as training data to determine what reply to give. Chatterbot, unfortunately, performed sub-optimally when compared to search based methods with regards to question answering, which made it impractical for the general question answering module compared to TF-IDF which had the best results.
- NLP
  - Sentiment detection
    - ◦ Naive Bayes trained on twitter data
  - Dialogue detection for detecting interrogative statements
    - ◦ Gradient Boosting Classifier trained on NPS chat data.
  - Case folding
- Data Storage
  - Storage of diet information

  - Storage of milestone progress
  - Storage of personal data
  - Storage of experiences
- Milestone Calculation
  - Using the chosen goal, it calculates milestones to reach for each session with the agent.
- Knowledge Search
  - A module that can search for diet based information. In this case, this is used for providing food related advice for memory references related to the user's favorite food to provide a healthier alternative that it more in line with their particular goal. The module consists of a CSV file containing food items with nutritional information that was provided by the USDA and a TF-IDF based method that looks up similar food items based on semantic similarity. This functions similarly to the information retrieval module, but unlike the question answering case, this formats diet related data into natural language that takes into account the context and chosen goal.

5.1.4. Prolific

Prolific is a crowd sourcing website that allows for researchers to recruit participants for studies [85]. Participants must be paid a minimum of $6.50 per hour of work. Participants within this work who were recruited from Prolific were paid a flat rate of £9 per hour for their time.

Prolific allows for researchers to filter participants based on certain criteria such as whether they have diabetes or not.

The filtration criteria used for participants within Prolific are as follows:

- Submission acceptance percentage
- Number of previous submissions
- No currently followed diet

This ensures that participants are at a particular level of quality, and that participants are not following another diet which can potentially interfere with the diet related habit that is being followed within the experiment they are participating in.

Prolific also provides participants with an ID that can be used within the interaction for user identification. This id, however, is 24 alphanumeric characters long, and as a result, it is impractical to have users provide it as their unique identifier. Because of this, a shorter id is generated for users to provide to the robot.

5.1.5. Video Based Interactions through Zoom

Due to COVID and the physical locations of participants varying, video calling applications will be used so that patients can interact with the virtual Furhat embodiment. For this to work properly, the audio from Furhat must be shared with the participant, and the audio from the participant's microphone should be sent to Furhat so that a conversation can be held remotely.

5.2. Tests

5.2.1. Test001: Modified Godspeed Questionnaire

To determine the experience of the interacting with the Furhat robot, a Godspeed questionnaire with additional questions pertaining to its effectiveness with regards goal achievement and habit building will be used. The following are standard questions in the Godspeed Questionnaire that utilize a 5 point Likert scale [77] . There has been much debate within literature regarding whether to use a 5 point Likert scale or some other alternative with 5 points representing a minimum for many types of data [83]. With the 5 point likert scale, Babacus and Boller find that the scale results in a lower level of frustration for participants [6]. Similarly, Sachdev and Verma found that a 5 point scale resulted in improved response quality and response rate [1].

Anthropomorphism						
Fake	1	2	3	4	5	Natural
Machinelike	1	2	3	4	5	Humanlike
Unconscious	1	2	3	4	5	Conscious
Artificial	1	2	3	4	5	Lifelike
Moving Rigidly	1	2	3	4	5	Moving Elegantly

Table 5.1: Questions asked for Anthropomorphism within the Godspeed Questionnaire

Animacy						
Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Inert	1	2	3	4	5	Interactive
Apathetic	1	2	3	4	5	Responsive

Table 5.2: Questions asked for Animacy within the Godspeed Questionnaire

The parts of the questionnaire that are traditionally part of the Godspeed questionnaire are defined in tables 5.1, 5.3, 5.4, and 5.5. For analyses using these categories, it should be noted that the individual items can be fused through summation to represent the category in full.

The following is an indication of the type of role the user felt the conversational agent had during the interaction. The user is expected to choose one of the following for this question.

- Brother or Sister
- Classmate/Colleague
- Stranger
- Relative
- Friend
- Parent
- Teacher/Coach
- Neighbor

Likeability						
Dislike	1	2	3	4	5	Like
Unfriendly	1	2	3	4	5	Friendly
Unkind	1	2	3	4	5	Kind
Unpleasant	1	2	3	4	5	Pleasant
Awful	1	2	3	4	5	Nice

Table 5.3: Questions asked for Animacy within the Godspeed Questionnaire

Percieved Intelligence						
Incompetent	1	2	3	4	5	Competent
Ignorant	1	2	3	4	5	Knowledgeable
Irresponsible	1	2	3	4	5	Responsible
Unintelligent	1	2	3	4	5	Intelligent
Foolish	1	2	3	4	5	Sensible

Table 5.4: Questions asked for Percieved Intelligence within the Godspeed Questionnaire

Perceived Safety						
Anxious	1	2	3	4	5	Relaxed
Agitated	1	2	3	4	5	Calm
Quiescent	1	2	3	4	5	Surprised

Table 5.5: The perceived safety questions are answered twice. They are answered with regards to how the user felt at the beginning and how they felt at the end.

Diet and Diabetes						
Is not useful for combating Type II Diabetes	1	2	3	4	5	Is useful for combating Type II Diabetes
Is not useful for combating obesity	1	2	3	4	5	Is useful for combating obesity
Is less convenient than a human dietitian	1	2	3	4	5	Is more convenient than a human dietitian
Interaction with a human is more preferable	1	2	3	4	5	Interaction with the robot is more preferable
Change in Diabetes Knowledge						
Knowledge Decreased	1	2	3	4	5	Knowledge Increased

Table 5.6: Questions asked for Diet and Diabetes within the Modified Godspeed Questionnaire

Motivation						
Was not Motivated at all	1	2	3	4	5	Was Highly Motivated

Table 5.7: Questions asked for Motivation within the Modified Godspeed Questionnaire

Additional Questions In addition to the standard questions that exist within the Godspeed questionnaire, additional questions were added to address the specific claims and concerns of this experiment to provide better insight into the end results. Questions that are an addition to the Godspeed questionnaire are defined in tables 5.8, 5.18, 5.7 which is asked before, during, and after the 3 sessions, and 5.9 which is asked before and after the experiment.

In addition to this, there are questions that take a yes or a no as an answer. These questions are the following:

- This is my first time interacting with this kind of robot.
- Have you been diagnosed with type 2 diabetes?
- Do you have a family history of type 2 diabetes?
- Would you like to use a similar system for working towards other health goals?
- Would you have preferred sessions with the system to have been longer or shorter? (Longer, Shorter, The duration was fine)
- Would you have preferred more sessions with the system or fewer sessions? (More sessions, less sessions, the number was fine)

Finally, there are questions relating to number of sessions and duration which take three possible answers.

- Would you have preferred sessions with the system to have been longer or shorter? (Longer, Shorter, The duration was fine)
- Would you have preferred more sessions with the system or fewer sessions? (More sessions, less sessions, the number was fine)

5.2.2. Test002: Achievement Metrics

To determine the how effective the system is at helping users to achieve their goals, a set of achievement metrics will be calculated based on the performance of patients during the three sessions.

Engagement						
Not Engaging	1	2	3	4	5	Engaging
Autonomy						
No Autonomy	1	2	3	4	5	High Autonomy
Negative or Positive Experience						
Negative	1	2	3	4	5	Positive

Table 5.8: Questions designed to answer claims presented in the design

Self-Efficacy						
Not Confident	1	2	3	4	5	Confident

Table 5.9: Self Efficacy Question that is asked before the experiment and after the experiment. (I am confident in my ability to make a diet change)

Milestone Adherence: This is calculated similarly to Mean Absolute Error [17] in regression problems. The difference between the MAE and the milestone adherence metric is that the metric will differ for the different goals.

The metric will simply calculate the differences between the progress made by the patient during each session and compares that with the milestones and sums these differences up.

$\Sigma|milestone - achievedValue|$

There are two variants of this metric. The first is as described while the second involves any instances where milestones that are exceeded to be treated as an absolute difference of zero. When comparing milestone adherence between patients who have different goals, the milestone adherence has been normalized so that they can be compared. This is done by using the initial intake as the normalizing constant since all succeeding consumption and goals are based off of the initial intake provided in the first session.

Milestone Difference: The difference between the milestone and the actual consumption in the second session.

Final Goal Difference: The difference between the final goal and actual consumption in the final session. There are two variants of this metric. The first is as described while the second is calculated under the assumption that the final goal was not changed in the second session.

Intention to Continue Diet: This is a binary metric that tells whether a participant will continue working on their diet after completing all three sessions with the conversational agent.

Milestone Achievement Count: This is a simple metric that checks whether the intermediate milestone was met or not.

Final Achievement: This is a simple binary metric that checks if the final goal was achieved or not. There are two variants of this metric. The first is as described while the second is calculated under the assumption that the final goal was not changed in the second session.

5.3. Experimental Study Design

The experiment that was conducted is a between subjects experiment that contained three equally sized groups of participants for each of the conditions covered by the research questions who will individually interact with the conversational agent over 3 separate sessions that will occur no less than within 24 hours of each others. In other words, there should be at least a day's worth of time that passes between consecutive sessions.

The majority of participants were recruited through Prolific. Participants were expected to fulfill the following criteria to participate in the study:

- Fluent in English
- Above the age of 18
  - Within Prolific, it is not necessary to filter based on age, because all participants must be 18 or older to use Prolific.
- Is not currently following a diet
  - In an ideal scenario, participants will have diabetes, obesity, or prediabetes, but due to ethical reasons, this is not a valid criteria to filter on. Instead, participants are filtered on whether

they are following a diet or not, because this can allow for a fair comparison to be made between participants from different conditions.

- Has a working computing device, stable internet connection, and headphones that can be used for the purpose of video meetings.
  - In the case that a participant does not have a working microphone, participants may be asked to switch to a smartphone microphone instead due to the higher quality hardware that can be found in phones.
- It would be preferable that participants do not know each other.
- Participants who are at risk of any health problems, has an excessively high BMI, or metabolic diseases may be excluded to avoid any ethical risks.
- Previous Submissions
  - Experienced participants are favored to avoid having the research spending time teaching participants how to use different features of Prolific.
- Acceptance Rate
  - Participants should be at a sufficiently high level of quality and should attend meetings on time. To ensure this, a high acceptance percentage of previous submissions is considered.

The participants were randomly split into three groups. These groups correspond to the conditions specified in the research questions and are as follows:

1. Experiences no references to shared memories
2. Experiences only references to a single shared memory
3. Experiences references to multiple shared memories

Although the references differed, the memories or rather motivational questions asked remained the same for all groups. This meant that the same questions were be used for all groups, but how participant's answers were used differed based on the group that the participant was assigned to.

To ensure that all participants were briefed and engaged with in the same way, a video was used to provide instructions to minimize any variations in the experiment that was outside of the conversations with the Furhat robot.

All groups were asked the same questions in the godspeed questionnaire at the end of third session and the same metrics were recorded for their goal achievement.

There are two forms that were completed by participants. The first form is a consent form that informs the participant of what the purpose of the study is, any ethical considerations and how any recorded data will be used. The second form contained the Godspeed questionnaire.

Qualtrics was leveraged for the creation of an online survey that was used for the consent form and the online survey that contained the Godspeed questionnaire.

Sessions were conducted with participants on an individual level through video calls where the Furhat robot's embodiment is seen by the participant through a shared screen and the Furhat robot's audio is heard by the participant through shared audio.

The following is an overview time needed for each of the sessions which resulted in the budget needed for processing all of the participants:

- approximately 10 minutes per interaction
- 5-7.5 minutes to set up and answer questions during each session.
- 10 minutes to complete Qualtrics surveys.
- 7.5 minutes to track personal diet.

This results in a total of approximately 70 minutes. To encourage participation, a payment classified by Prolific as "Great" was considered and total of 90 participants was considered as a ceiling.

In the case of this study, a base rate of £9 per hour is used. It should be noted that Prolific adds a service fee which adds 33% to all payments as an additional charge.

Study Design on Prolific: Within Prolific, this experiment is classified as a longitudinal study which means that it consists of multiple ”Prolific studies” over a longer period of time.

For this experiment, a participant will complete a total of 4 ”Prolific Studies”.

The first study involves the completion of the consent form, and scheduling of 3 zoom meetings through a web application called Calendly which allows for scheduling of appointments. Calendly automatically converts timezones to ensure that appointments are correctly scheduled. This first study includes the inclusion criteria mentioned previously, and excludes any participants who have already completed the experiment previously.

The second, third, and fourth studies involve the 3 zoom meetings respectively. The inclusion criteria for this is that of participants who have completed the ”Prolific Study” that precedes it within the longitudinal study.

Exclusion Criteria: It should be noted that depending on certain conditions and aspects, a participant's results can be rejected.

Typically, such participants who are rejected are rejected early in the process to avoid an inefficient use of time or the spending of money on subpar results.

- Does not have a sufficiently high quality audio setup
- Does not correctly follow the instructions in the consent form or fails honey pot questions/attention checks.
- Does not attend the first meeting or is not responsive on Prolific.
- Does not complete all 3 meetings.
- Does not have a high enough level of fluency in English.
- Does not consent to participation in the study

5.4. Participants and Demographic Information

Of a total of 93 participants, 79 had valid results. Of the valid participants, 36 were female, 42 were male, and 1 was non-binary. Participants are from many countries of origin with the majority being from the CET, EET, and BST timezones. This includes participants from Italy, Denmark, Mexico, Spain, South Africa, Canada, Poland, Hungary, United Kingdom, United States, Portugal, Ukraine, Brazil, Greece, Estonia, Zimbabwe, Iran, Turkey, Nigeria, China, Ireland, Slovenia, France, and Ecuador. Conditions 1 and 2 had 26 participants each, and condition 3 had 27 participants.

5.5. Results

5.5.1. Statistical Methods

There are 3 classes of data that are collected. The first is that of quantitative data which is collected during the conversation. These include figures such as intake on different days. The second is that of Likert scale data which is collected through the consent form and questionnaire at the end of the third session. Finally, there are binary questions that are asked during the conversation and within forms that are completed by the participant.

Milestone Data and Likert Scale Data Milestone data collected, such as intake, is continuous and as a result, it can be treated with parametric methods as long as assumptions are fulfilled. This makes a method like the student t-test or ANOVA appropriate for analyzing such data when the data is normally distributed. Since this experiment has three conditions, ANOVA is more appropriate since it is able to handle three or more groups unlike the t-test which is limited to two groups [71]. Within literature, Likert scales are often treated as continuous data despite being discrete in nature [42] due to the robust nature of the F-statistics that are used with ANOVA which has been cited as justification to ignore the assumption when sample size is sufficiently high [15]. In this case of this work, although ANOVA is relatively robust to its assumptions [11], verifying the assumptions will strengthen the use of the method, so a Shapiro-Wilk test will be used with the residuals [67] and used to determine whether ANOVA is normal and to test the homogeneity of variance, Levene's test will be used. Finally, post-hoc tests

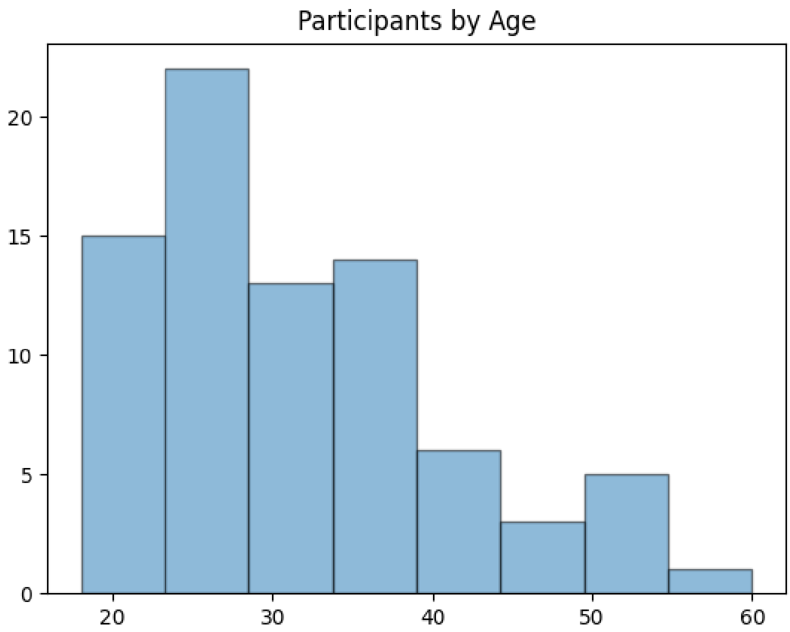


Figure 5.1: Histogram showing the age of participants in the study

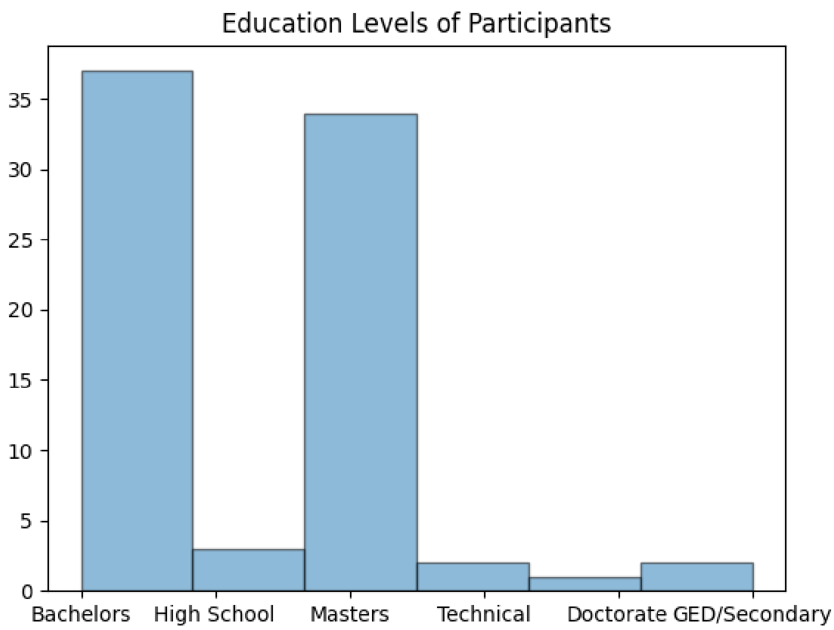


Figure 5.2: Histogram showing the education level of participants in the study

Anthropomorphism		**Fake/ Natural**	**Machine/ Human**	**Conscious/ Unconscious**	**Artificial/ Lifelike**	**Rigid/ Elegant**
All	***Mean***	3.316	2.899	3.671	2.899	3.291
	Std.	1.069	1.215	1.083	1.139	1.04
	Median	3	3	4	3	4
Condition 1	***Mean***	3.269	2.923	3.5	2.885	3.231
	Std.	1.151	1.197	1.03	1.107	0.863
	Median	3.5	3	4	3	4
Condition 2	***Mean***	3.154	2.846	3.462	2.808	3.231
	Std.	1.047	1.19	1.174	1.234	1.032
	Median	3	3	4	3	4
Condition 3	***Mean***	**3.519**	**2.926**	**4.037**	**3**	**3.407**
	Std.	1.014	1.299	0.98	1.109	1.217
	Median	**4**	3	4	3	4

Table 5.10: Descriptive Statistics for Anthropomorphism Questions. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

Animacy		**Dead/ Alive**	**Stagnant/ Lively**	**Mechanical/ Organic**	**Inert/ Interactive**	**Apathetic/ Responsive**
All	***Mean***	3.671	3.443	2.785	4.076	4.127
	Std.	1.118	1.071	1.129	0.958	0.925
	Median	4	4	3	4	4
Condition 1	***Mean***	**3.846**	**3.654**	2.5	4.038	4.038
	Std.	1.008	0.846	0.949	0.916	0.871
	Median	4	4	2.5	4	4
Condition 2	***Mean***	3.423	3.308	**3.154**	3.962	4.115
	Std.	1.102	1.192	1.190	1.076	0.993
	Median	4	4	3.5	4	4
Condition 3	***Mean***	3.741	3.370	2.704	**4.222**	**4.222**
	Std.	1.228	1.149	1.171	0.892	0.934
	Median	4	4	3	4	4

Table 5.11: Descriptive Statistics for Animacy Questions. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

will be used in the case of significant differences to determine what groups are different. For ANOVA Bonferroni will be utilized.

Binary Data To determine if binary results are significantly different from each other, Fisher's exact test [33] will be used to compare between the conditions due to the relatively small sample size per condition. The reason why Fisher's exact test will be used instead of chi square test is because Fisher's exact test can be used for smaller sample sizes, and unlike chi square, it is an exact test instead of an approximation [50]. For data that includes 3 choices such as preferred duration and preferred number of sessions, Pearson's chi squared will be used instead since Fischer's exact test is limited to 2 x 2 contingency tables.

5.5.2. Descriptive Statistics

To gain an understanding of the distribution of user data that were collected during the experiment, the mean, and standard deviation are provided. In addition to this, the median is also provided to provide a more clear understanding of the skew and distribution of the data.

Observations within Descriptive Statistics With the descriptive statistics collected, some observations can be made regarding the robot as a whole. It should be noted, that determining significant statistical differences between conditions can be found within the inferential statistics in subsection

Likeability		**Dislike/ Like**	**Unfriendly/ Friendly**	**Unkind/ Kind**	**Unpleasant/ Pleasant**	**Awful/ Nice**
All	***Mean***	4.215	4.519	4.531	4.380	4.430
	Std.	0.811	0.695	0.695	0.756	0.779
	Median	4	5	5	5	5
Condition 1	***Mean***	**4.308**	**4.654**	**4.615**	4.423	**4.5**
	Std.	0.549	0.629	0.697	0.758	0.762
	Median	4	5	5	5	5
Condition 2	***Mean***	4.038	4.5	**4.615**	**4.538**	4.423
	Std.	1.038	0.707	0.637	0.647	0.809
	Median	4	5	5	5	5
Condition 3	***Mean***	4.296	4.407	4.370	4.185	4.370
	Std.	0.775	0.747	0.742	0.834	0.792
	Median	4	5	4	4	5

Table 5.12: Descriptive Statistics for Likeability Questions. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

Perceived/ Intelligence		**Incompetent/ Competent**	**Ignorant/ Knowledgeable**	**Irresponsible/ Responsible**	**Unintelligent/ Intelligent**	**Foolish/ Sensible**
All	***Mean***	4.253	4.506	4.165	4.291	4.291
	Std.	0.759	0.677	0.854	0.803	0.787
	Median	4	5	4	4	4
Condition 1	***Mean***	4.269	**4.615**	**4.269**	4.346	4.115
	Std.	0.604	0.496	0.874	0.797	0.952
	Median	4	5	**4.5**	4.5	4
Condition 2	***Mean***	4.154	4.385	4.038	4.038	4.346
	Std.	0.881	0.898	0.916	0.958	0.745
	Median	4	5	4	4	**4.5**
Condition 3	***Mean***	**4.333**	4.519	4.185	**4.481**	**4.407**
	Std.	0.784	0.580	0.786	0.580	0.636
	Median	4	5	4	**5**	4

Table 5.13: Descriptive Statistics for Perceived Intelligence Questions. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

Safety Prior		**Anxious/ Relaxed**	**Agitated/ Calm**	**Quiescent/ Surprised**
All	***Mean***	3.646	4.050	3.253
	Std.	1.261	1.073	0.776
	Median	4	4	3
Condition 1	***Mean***	3.423	3.962	2.923
	Std.	1.301	1.148	0.744
	Median	3	4	3
Condition 2	***Mean***	**3.962**	**4.152**	**3.462**
	Std.	1.113	1.047	0.582
	Median	4	**4.5**	3
Condition 3	***Mean***	3.556	4.037	3.370
	Std.	1.340	1.055	0.884
	Median	4	4	3

Table 5.14: Descriptive Statistics for Safety Questions prior to the experiment. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

Safety After		**Anxious/ Relaxed**	**Agitated/ Calm**	**Quiescent/ Surprised**
All	***Mean***	4.241	4.418	3.544
	Std.	1.003	0.794	0.984
	Median	5	5	3
Condition 1	***Mean***	4.154	4.423	3.385
	Std.	1.120	0.809	0.941
	Median	5	5	3
Condition 2	***Mean***	4.269	4.346	3.5
	Std.	1.002	0.936	0.906
	Median	5	5	3
Condition 3	***Mean***	**4.296**	**4.481**	**3.741**
	Std.	0.912	0.643	1.095
	Median	5	5	**4**

Table 5.15: Descriptive Statistics for Safety Questions after the experiment. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

Change in Safety		**Anxious/ Relaxed**	**Agitated/ Calm**	**Quiescent/ Surprised**
All	***Mean***	0.595	0.367	0.291
	Std.	1.171	0.950	0.963
	Median	0	0	0
Condition 1	***Mean***	0.7317	**0.462**	**0.462**
	Std.	1.373	1.067	0.582
	Median	0	0	0
Condition 2	***Mean***	0.308	0.192	0.038
	Std.	0.788	0.849	0.958
	Median	0	0	0
Condition 3	***Mean***	**0.741**	0.444	0.370
	Std.	1.256	0.934	1.214
	Median	0	0	0

Table 5.16: Descriptive Statistics for Change in Safety Questions where Safety prior to the experiment and after the experiment are compared. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

Categories		**Anthropo-morphism**	**Animacy**	**Likeability**	**Perceived Intelligence**	**Safety Prior**	**Safety After**	**Change in Safety**
All	**Median**	16	19	23	22	12	13	1
	Mean	16.0759	18.1013	22.0759	21.5063	10.9494	12.2025	1.2532
	Std.	4.0849	3.9471	2.8455	2.8728	2.3254	1.8631	2.3009
Condition 1	**Median**	16	18	**23**	22	10	12.5	1
	Mean	15.8077	18.0769	**22.5**	21.6154	10.3077	11.9615	1.6538
	Std.	3.7632	3.0845	2.3195	2.3846	2.4457	1.7316	2.4485
Condition 2	**Median**	15.5	19	22.5	21.5	12	13	0
	Mean	15.5	17.9615	22.1154	20.9615	**11.5769**	**12.1154**	0.5385
	Std.	4.3841	4.6431	2.6882	3.6822	1.9631	2.1228	**1.7940**
Condition 3	**Median**	16	19	22	**23**	12	13	1
	Mean	**16.8889**	**18.2593**	21.6296	**21.9259**	10.9630	10.963	1.5556
	Std.	4.1075	4.1006	3.4323	2.4007	2.4413	1.7403	2.5013

Table 5.17: Descriptive Statistics for all Categories after the sub items for the categories have been summed up. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

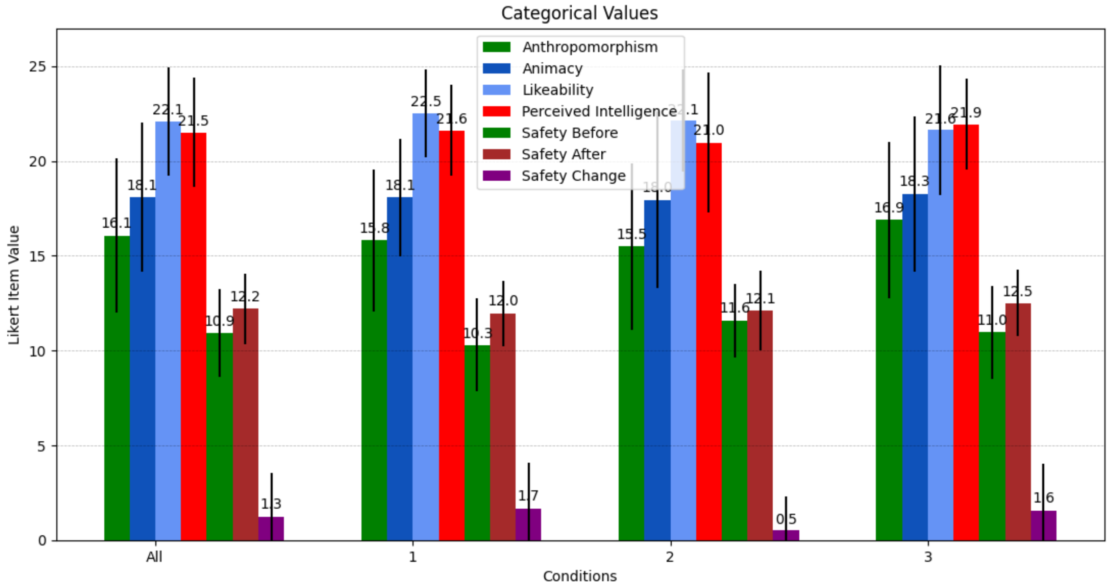


Figure 5.3: Histogram showing the values for category Likert scales after summation of sub items

Diet and Diabetes		**Useful/ Not Useful for Diabetes**	**Useful/ Not Useful for Obesity**	**Convenience**	**Preference**	**Change in Knowledge**
All	***Mean***	4.063	4.038	2.962	3.937	3.835
	Std.	0.852	0.912	1.224	1.042	0.758
	Median	4	4	3	4	4
Condition 1	***Mean***	3.885	3.885	2.846	4.038	**3.846**
	Std.	0.864	0.993	1.347	1.113	0.732
	Median	4	4	3	**4.5**	4
Condition 2	***Mean***	4.038	3.885	**3.077**	**4.077**	**3.846**
	Std.	0.958	1.033	1.383	0.977	0.732
	Median	4	4	3	4	4
Condition 3	***Mean***	**4.259**	**4.333**	2.963	3.704	3.815
	Std.	0.712	0.620	0.940	1.031	0.834
	Median	4	4	3	4	4

Table 5.18: Descriptive Statistics for Diet and Diabetes questions. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

Claims		Engagement	Autonomy	Negative/ Positive Experience
All	***Mean***	3.873	3.278	4.582
	Std.	1.017	1.176	0.672
	Median	4	3	5
Condition 1	***Mean***	3.731	3.269	4.538
	Std.	1.041	1.251	0.582
	Median	4	3	5
Condition 2	***Mean***	3.885	3.269	4.462
	Std.	1.071	1.151	0.859
	Median	4	3	5
Condition 3	***Mean***	**4**	**3.296**	**4.741**
	Std.	0.961	1.171	0.5257
	Median	4	3	5

Table 5.19: Descriptive Statistics for Claims questions. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

Self-Efficacy		Motivation Before	Motivation During	Motivation After
All	***Mean***	3.861	4.405	4.456
	Std.	1.009	0.707	0.828
	Median	4	5	5
Condition 1	***Mean***	**4.038**	**4.462**	4.5
	Std.	0.999	0.706	0.762
	Median	4	5	5
Condition 2	***Mean***	3.962	4.308	4.346
	Std.	0.871	0.788	0.892
	Median	4	4	5
Condition 3	***Mean***	3.593	4.444	**4.519**
	Std.	1.118	0.641	0.849
	Median	4	5	5

Table 5.20: Descriptive Statistics for motivation. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

Motivation Change		Motivation Before to During	Motivation During to After
All	***Mean***	0.544	0.051
	Std.	0.813	0.618
	Median	0	0
Condition 1	***Mean***	0.423	0.038
	Std.	0.945	0.445
	Median	0	0
Condition 2	***Mean***	0.346	0.0385
	Std.	0.629	0.662
	Median	0	0
Condition 3	***Mean***	**0.852**	**0.074**
	Std.	0.770	0.730
	Median	**1**	0

Table 5.21: Descriptive Statistics for change in motivation. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

Self-Efficacy		Efficacy Before	Efficacy After	Change in Efficacy
All	***Mean***	4.456	4.418	-0.03797
	Std.	0.694	0.727	0.93390
	Median	5	5	0
Condition 1	***Mean***	4.346	4.385	0.03846
	Std.	0.629	0.752	0.85398
	Median	4	4.5	0
Condition 2	***Mean***	4.5	4.269	-0.23077
	Std.	0.583	0.874	1.04909
	Median	5	4	0
Condition 3	***Mean***	**4.519**	**4.593**	**0.07407**
	Std.	0.849	0.501	0.85747
	Median	5	5	0

Table 5.22: Descriptive Statistics for self-efficacy. The maximum mean is bolded and underlined for each question item and if there is a maximum median, that is bolded and underlined.

First Time Interacting with Robot	Yes	No	Yes Frequency
All	71	8	0.8987
Condition 1	21	5	0.8076
Condition 2	26	0	1
Condition 3	24	3	0.8889

Table 5.23: Simple count/frequency for the binary question of whether this is the first time interacting with the robot.

Has Diabetes	Yes	No	Yes Frequency
All	2	77	0.0259
Condition 1	1	25	0.04
Condition 2	1	25	0.04
Condition 3	0	27	0

Table 5.24: Simple count/frequency for the binary question of whether a participant has type 2 diabetes or not.

Has Family History of Diabetes	Yes	No	Yes Frequency
All	29	50	0.3670
Condition 1	8	18	0.3077
Condition 2	13	13	0.5
Condition 3	8	19	0.2963

Table 5.25: Simple count/frequency for the binary question of whether a participant has a family history type 2 diabetes or not.

Is Willing to Use Similar System for Other Health Problems	Yes	No	Yes Frequency
All	70	9	0.8860
Condition 1	22	4	0.8462
Condition 2	22	4	0.8462
Condition 3	26	1	0.963

Table 5.26: Simple count/frequency for the binary question of whether a participant is willing to use a similar system for other health related goals.

Should Sessions have been Longer or Shorter?	Longer	Shorter	Duration was Appropriate	Appropriate Frequency
All	29	0	50	0.6329
Condition 1	13	0	13	0.5
Condition 2	6	0	20	0.7693
Condition 3	10	0	17	0.6296

Table 5.27: Simple count/frequency for the binary question of whether a participant is would have preferred longer or shorter sessions

More Sessions or Fewer Sessions	More Sessions	Fewer Sessions	Number of Sessions was Appropriate	Appropriate Frequency
All	37	0	42	0.5316
Condition 1	11	0	15	0.5768
Condition 2	12	0	14	0.5385
Condition 3	14	0	13	0.4815

Table 5.28: Simple count/frequency for the question of whether a participant is would have preferred more sessions or fewer sessions.

Type of Relationship	Teacher/ Coach	Stranger	Classmate/ Colleague	Friend
All	55	15	3	6
Condition 1	19	5	0	2
Condition 2	19	5	0	2
Condition 3	17	2	3	2

Table 5.29: Simple count for the question of the type of role or relationship the user had with the robot

Milestone Adherence		Milestone Adherence	Zeroed Milestone Adherence	Milestone Adherence without Goal Change	Zeroed Milestone Adherence without Goal Change
All	***Mean***	0.14690	0.02335	0.18149	0.02175
	Std.	0.17502	0.07559	0.18913	0.07460
	Median	0.08673	0	0.13	0
Condition 1	***Mean***	0.15438	0.03218	0.18036	0.03018
	Std.	0.21193	0.07317	0.21617	0.06936
	Median	0.07333	0	0.09659	0
Condition 2	***Mean***	**0.11520**	**0.00720**	**0.147353**	**0.00772**
	Std.	0.15829	0.01509	0.182650	0.01877
	Median	**0.05556**	0	**0.06261**	0
Condition 3	***Mean***	0.170225	0.03040	0.21545	0.02711
	Std.	0.151642	0.10669	0.16715	0.10692
	Median	0.14	0	0.21333	0

Table 5.30: Descriptive statistics for milestone adherence. The lowest mean and median values are bolded and underlined for each metric.

Milestone Differences		Milestone Difference	Goal Difference	Goal Difference without Goal Change
All	***Mean***	-0.0628	-0.0374	-0.0752
	Std.	0.14108	0.1031	0.1170
	Median	-0.03	-0.0143	-0.05
Condition 1	***Mean***	-0.0649	-0.0251	-0.0551
	Std.	0.1613	0.1040	0.1199
	Median	-0.0197	-0.0079	-0.0213
Condition 2	***Mean***	-0.0523	-0.04853	-0.0796
	Std.	0.0990	0.1297	0.1384
	Median	-0.0088	-0.0133	-0.0364
Condition 3	***Mean***	-0.0709	-0.0386	-0.0904
	Std.	0.1584	0.0707	0.0903
	Median	-0.0763	-0.0364	-0.0986

Table 5.31: Descriptive statistics for milestone difference.

Shared Experience Agreement		Agreement Percentage
Condition 2	***Mean***	0.904
	Std.	0.284
	Median	1
Condition 3	***Mean***	0.812
	Std.	0.226
	Median	0.75

Table 5.32: Descriptive statistics for shared experience agreement.

Milestone Achievement			
	Yes	***No***	***Yes Frequency***
All	69	10	0.873
Condition 1	22	4	0.846
Condition 2	23	3	0.885
Condition 3	24	3	**0.889**

Table 5.33: Descriptive statistics for milestone achievement metrics. Maximum values for each yes frequency are bolded.

Final Goal Achievement Metrics	Final Goal Achievement			Final Goal Achievement Without Goal Change		
	Yes	***No***	***Yes Frequency***	***Yes***	***No***	***Yes Frequency***
All	68	11	0.861	64	15	0.810
Condition 1	23	3	**0.885**	21	5	0.808
Condition 2	23	3	**0.885**	21	5	0.808
Condition 3	22	5	0.815	22	5	**0.815**

Table 5.34: Descriptive statistics for final goal achievement metrics. Maximum values for each yes frequency are bolded.

Will Continue Diet			
	Yes	***No***	***Yes Frequency***
All	73	6	0.924
Condition 1	23	3	0.885
Condition 2	24	2	0.923
Condition 3	26	1	**0.963**

Table 5.35: Descriptive statistics for whether participants will continue working on their diet after the experiment. Maximum value for yes frequency is bolded.

5.5.3 where p values and other statistical figures are provided.

With regards to the type of role, the majority of participants found the the robot to be like a teacher or coach with a smaller number seeing the robot as a stranger or friend. All instances of the robot being recognized as a classmate or a colleague were by participants who were part of group 3, but through a chi-square analysis, the differences were not found to be significant between the conditions.

For positive or negative experience, it appears that most participants had a highly positive experience regardless of the condition used. This is indicated by the median value of 5 for all conditions which is the highest possible value, and a mean value above 4.4.

For most aspects of the godspeed questionnaire, the robot scored generally well with likeability items and perceived intelligence scoring particularly high. Anthropomorphism resulted in the lowest results, but were still either at a neutral value or higher for most of the values.

For safety items, it should also be noted that all conditions resulted in an increase in safety which indicates that the robot had a calming effect.

Significant Results Non Binary						
	Shapiro-Wilk	**Levene**	**Anova**	**Anova F Statistic**	**Power**	**Power Sample Size**
Motivation Change From Before to During Experiment	0.000	0.435	0.0482	3.1562	0.9827	39.6819

Table 5.36: Inferential statistics for significant results of non-binary data collected.

Bonferroni Post-Hoc Tests				
	Group 1	**Group 2**	**pval**	**pval_corr**
Motivation Change From Before to During Experiment	1	2	0.7312	1.0
Motivation Change From Before to During Experiment	1	3	0.0756	0.2268
Motivation Change From Before to During Experiment	**2**	**3**	**0.0118**	**0.0353**

Table 5.37: Results of Bonferroni Post-Hoc Tests. Rejections of null hypotheses are underlined and bolded.

5.5.3. Inferential Statistics

For inferential statistics, only significant results will be presented. For determining significance, values of $p < 0.05$ are considered. It should be noted that for many of the insignificant results, the power was low and the recommended sample size based on power numbered between the hundreds to thousands based on the type of metric that was being analyzed.

In table 5.36 there is only one significant result that can be noted which is that of the change in motivation from prior to the conversation with the agent to during the conversation with the agent. For many unsignificant results the power indicated that a larger sample size may have helped with improving the result for ANOVA.

For binary data, all differences were found to be insignificant regardless of conditions that were compared. The same is true for data that required the need of chi squared tests.

To further analyze the significant results, the Bonferroni post-hoc test was used which is shown in table 5.37. This was utilized to determine which groups in particular are significantly different from each other.

As can be noted, groups 2 and 3 are significantly different according to the Bonferroni test run after ANOVA for motivation change.

5.5.4. Linear Mixed Models

To determine the effect of the different collected data on certain metrics, generalized linear mixed models (GLMM) are used to gain further insight into what affects certain metrics the most. For the following, only significant terms will be reported. Gender, age, education, and other variables that are defined prior to the experiment are treated as random effects and other variables are treated as fixed effects. For metrics that are part of a category within the godspeed questionnaire, these metrics were excluded, and the category metric that contains the likert scale data with summed values of their respective components were used instead.

For some metrics which were not used as a dependent variable, these metrics were excluded due to lack of convergence, or due to the distribution in the data which led to problems within the calculation of the model such as singular matrices, perfect separation, and hessian matrices that were not positive definite.

For final goal achievement which can be seen in table 5.38 it can be seen that there are two variables that can be said to affect the fit of the model significantly. The greatest value for the coefficient is provided by the intercept which suggests that all participants are predisposed to reaching the final goal. The value provided by the animacy category which is the second significant fixed effect does have a positive coefficient, but this coefficient is very small in comparison.

For final goal achievement where no changes to goals in the second session are considered, there are other metrics that were found to be significant. This can be seen in table 5.39. The greatest value for the coefficient is provided by the animacy category, but like case of standard final goal achievement, this positive correlation was quite small. Interestingly, the perceived understanding of diabetes after completing the experiment had a negative correlation where the higher the understanding was, the lower the final goal achievement may have been, but it should be noted that the p value for this was on the border of what may have been considered significant.

With regards to anthropomorphism and animacy, tables 5.40 and 5.41 show that there is a positive correlation between the two categories where the only significant fixed effect for anthropomorphism is that of animacy and the fixed effect with the lowest p value for animacy was that of anthropomorphism. This indicates that perceptions of the robot as animate resulted in it appearing more human-like and vice versa. This seems somewhat intuitive since most humans tend to be more animate than rigid.

For animacy, there was also a positive correllation with the perceived usefulness for diabetes, the perception of the interaction as positive or negative and the user's understanding of diabetes after the experiment. This could possibly indicate that perception of the robot as animate has a relationship with the conclusions and final thoughts or impressions a user had about the experience as a whole.

The likeability of the agent which can be seen in table 5.42, were influenced by the condition and perceived autonomy where the use of memories appeared to have a negative correlation with the likeability of the robot. Autonomy, on the other hand, had a far more understandable positive correlation.

Finally, GLMM was used for the perceived intelligence category which can be seen within table 5.43. For this metric, achievement of the intermediate milestone had a relatively high positive correlation with perceived intelligence which is supported by the relatively high negative correlation within milestone adherence where the further away a user is from achieving a particular milestone, the lower the perceived intelligence may be. What this possibly implies is that the perceived intelligence correlates with how closely an individual manages to keep up with a diet plan over the entirety of th experiment.

Final Goal Achievement					
	REML Criterion				
	152.5				
Scaled Residuals					
	Min	**1Q**	**Median**	**3Q**	**Max**
	-1.9588	-0.3436	0.1994	0.4822	1.5069
Random Effects					
	Intercept Variance	**Intercept Standard Deviation**	**Residual Variance**	**Residual Standard Deviation**	
	1.729e-01	0.415848	1.189e-01	0.344766	
Fixed Effects					
	Estimate	**Error**	**Df**	**t value**	**Pr(>|t|)**
Intercept	1.410329	0.655959	43.851087	2.150	0.0371
Animacy	0.059543	0.024737	40.777328	2.407	0.0207

Table 5.38: Significant results for GLMM for Final Goal Achievement Metric with Categorical Metrics used as Fixed Effects

Final Goal Achievement Without Goal Change					
	REML Criterion				
	159.1				
Scaled Residuals					
	Min	**1Q**	**Median**	**3Q**	**Max**
	-1.8371	-0.3537	0.1691	0.5562	1.4889
Random Effects					
	Intercept Variance	**Intercept Standard Deviation**	**Residual Variance**	**Residual Standard Deviation**	
	8.880e-02	0.297997	1.310e-01	0.361875	
Fixed Effects					
	Estimate	**Error**	**Df**	**t value**	**Pr(>|t|)**
Animacy	0.065174	0.026685	35.281972	2.442	0.0197
Likeability	-0.058934	0.024891	43.235349	-2.368	0.0224
Understanding of Diabetes	-0.159452	0.078144	34.142983	-2.040	0.0491

Table 5.39: Significant results for GLMM for Final Goal Achievement Metric Without Goal Change with Categorical Metrics used as Fixed Effects

Anthropomorphism					
	REML Criterion				
	347.4				
Scaled Residuals					
	Min	**1Q**	**Median**	**3Q**	**Max**
	-2.10507	-0.54577	0.02248	0.48450	1.65243
Random Effects					
	Intercept Variance	**Intercept Standard Deviation**	**Residual Variance**	**Residual Standard Deviation**	
	7.855649	2.80279	6.494864	2.54850	
Fixed Effects					
	Estimate	**Error**	**Df**	**t value**	**Pr(>|t|)**
Animacy	0.77948	0.14846	21.28719	5.251	3.19e-05

Table 5.40: Significant results for GLMM for Anthropomorphism with Categorical Metrics used as Fixed Effects

Animacy					
	REML Criterion				
	313.2				
Scaled Residuals					
	Min	**1Q**	**Median**	**3Q**	**Max**
	-2.52026	-0.45326	0.07103	0.56722	1.89986
Random Effects					
	Intercept Variance	**Intercept Standard Deviation**	**Residual Variance**	**Residual Standard Deviation**	
	8.171794	2.85864	3.085567	1.75658	
Fixed Effects					
	Estimate	**Error**	**Df**	**t value**	**Pr(>|t|)**
Anthropomorphism	0.38116	0.07822	45.64198	4.873	1.37e-05
Likeability	0.27492	0.11951	38.16355	2.300	0.02698
Useful Or Not Diabetes	1.14182	0.47781	45.35388	2.390	0.02108
Change in Motivation After	-0.90840	0.40493	44.65351	-2.243	0.02988
Positive or Negative Experience	1.07864	0.49686	36.06485	2.171	0.03660
Understanding Diabetes	1.06673	0.37722	43.23848	2.828	0.00707

Table 5.41: Significant results for GLMM for Animacy with Categorical Metrics used as Fixed Effects

Likeability					
	REML Criterion				
	316.7				
Scaled Residuals					
	Min	**1Q**	**Median**	**3Q**	**Max**
	-1.51065	-0.59081	0.05962	0.58778	2.01603
Random Effects					
	Intercept Variance	**Intercept Standard Deviation**	**Residual Variance**	**Residual Standard Deviation**	
	5.057e-01	0.711134	3.344e+00	1.828680	
Fixed Effects					
	Estimate	**Error**	**Df**	**t value**	**Pr(>|t|)**
Condition	-0.956570	0.442603	46.232138	-2.161	0.03589
Autonomy	0.794516	0.289728	45.821114	2.742	0.00867

Table 5.42: Significant results for GLMM for Likeability with Categorical Metrics used as Fixed Effects

Perceived Intelligence					
	REML Criterion				
	286.9				
Scaled Residuals					
	Min	**1Q**	**Median**	**3Q**	**Max**
	-2.51646	-0.37584	0.06731	0.40558	1.62470
Random Effects					
	Intercept Variance	**Intercept Standard Deviation**	**Residual Variance**	**Residual Standard Deviation**	
	0.5777896	0.76012	1.7591907	1.32634	
Fixed Effects					
	Estimate	**Error**	**Df**	**t value**	**Pr(>|t|)**
Milestone Achievement	1.496824	0.591430	47.071535	2.531	0.01478
Milestone Adherence	-2.759992	1.186990	44.356329	-2.325	0.02470
Engagement	1.046368	0.227005	43.830960	4.609	3.48e-05
Duration	0.679047	0.210495	44.548139	3.226	0.00235
Number of Sessions	-0.577273	0.218360	46.337916	-2.644	0.01115

Table 5.43: Significant results for GLMM for Perceived Intelligence with Categorical Metrics used as Fixed Effects

5.5.5. Qualitative Findings

In addition to quantitative data that was collected, participants were allowed to type additional comments regarding their interaction with the robot and the experiment and sometimes voiced their impressions verbally after conversations.

No Judgement and Bias For some of the participants, they found certain advantages the robot had over a human which was not fully captured by the questionnaire. Three of the participants noted that unlike a human, the robot is not judgemental, or has any prejudices that could result in a negative experience. In this sense, the robot may be preferable especially for those who have had negative experiences in health related contexts in the past.

Compatibility of the Experiment Structure with Personal Schedule and Activity Level For a small number of participants, the three day period needed for the experiment was somewhat problematic due to personal circumstances. Some participants may cite a hard day at work which resulted in no time to eat. Similarly, some participants may cite being more active on a particular day which resulted in a higher intake for that particular day. Such problems are unavoidable since the level of activity is not taken into account in the design in terms of the conversational agent itself as well as the experimental design that requires that the experiment be conducted over a three day period.

Uniqueness of the Study compared to other Studies on Prolific Prior to carrying out the study, one concern that was held was that since participants were being paid for their time, that they would not take the experiment seriously or be interested in the content of the experiment itself. This was surprisingly not the case due to the unique nature of this experiment compared to other studies carried out on Prolific. Many experiments on Prolific are survey based or hosted online. This means that Zoom meetings or video calls are rare. In addition to that, such Zoom calls are usually held between people, which places this experiment in a category of its own in terms of the subject area and the way in which it is carried out compared to most studies on Prolific. Side benefits involving improving one's diet also seemed to encourage participants who were interested in the contents of the study rather than participants who were only interested in the reward.

Thoughts on Programmatically Generated Vocal Gestures Although vocal gestures were added to make the robot feel more natural or human, some of the participants found the procedurally generated vocal gestures to be distracting or awkward. This is because the robotic vocal gestures through Amazon's neural Polly voice lack the prosodic flexibility to be nuanced enough to utilize the most appropriate inflections when generating the speech for a vocal gesture. This can result in vocal gestures sometimes sounding abrupt rather than natural. For the neural Polly voice in particular, many of the Speech Synthesis Markup language (SSML) tags are only partially supported which results in a sub-optimal experience.

Significance of Beginning Efforts Towards Diet For many, the impact of simply counting calories and sugar has been stated to make a large difference in terms of how participants viewed their diet. Having an awareness of what one is eating and how much, resulted in participants immediately having a clear idea of what food items contributed the most to their caloric and sugar intake which may have improved their chances of success before even speaking with the conversational agent.

Anthropomorphism The visual aspects of the robot had a mixed response upon further discussion of the experiment after completion of the three sessions. Some participants did not mind the appearance while some found the human-like face to be unnerving. One participant made a comparison to Robocop which has a similar appearance without its signature helmet. With regards to the voice, this had a mixed response as well, but a majority of the participants found the Polly voice to be surprisingly human with the voice sounding less rigid than other voices. According to many participants, the voice did not sound robotic, and the main flaw that the voice had was with prosody and pacing which is a problem that participants noted is common in other robotic voices as well.

6
Conclusion

6.1. Discussion

6.1.1. Implications of Significant Results

Within the results collected, there was found to be a significant difference for the change in motivation from prior to the experiment to during the experiment. For motivation change, this was found to be significant for Anova. In addition to this, motivation change had a high power and the sample size that was used was high enough. The Post-Hoc test shows that when correcting with Bonferroni for the ANOVA test, only groups 2 and 3 within motivation change were significantly different. What this shows with respect to motivation change is that the third condition's change in motivation was significantly higher than the second condition. This indicates that simply having a reference to a shared experience is not enough to create a significant or noticeable change in the perception of a robot as a source of motivation. What this implies is that there is a minimum number of references to shared experiences needed to receive the full motivational benefit of those past events. This is further supported by the lack of a significant difference between groups 1 and 2 which were the groups that had no shared experiences and the group that was limited to only one shared experience to refer to respectively. With regards to why this is, it could simply be the case that referring to the same shared experience would result in a repetitive experience rather than a more fulfilling one. Of course, the data does not necessarily confirm nor deny whether users were able to remember the differences between the references made in sessions two and three and future work can help to determine the level of recall that a participant had regarding their experience. It could also be the case that referring to a variety of memories creates a more memorable experience that leaves a deeper impression. This indicates that a variety of shared experiences provides a greater motivation and that the motivational memory is actually capable of motivation when used correctly. In addition to this, the simple presence of a motivational memory is not enough to significantly motivate participants.

6.1.2. Research Questions

Does referring to previous sessions improve goal attainment (performance?) Based on the significant results, goal attainment, specifically final goal achievement as well as other performance related metrics, was not different between the three conditions. As a result, the null hypothesis cannot be rejected. Future experiments should consider testing with a higher sample size. An interesting aspect of the results to note is that although participants from group 3 were significantly more motivated, this did not necessarily translate to improved goal achievement. The absence of motivation related metrics in the GLMM results for final goal achievement also supports this statement.

Does referring to previous sessions improve user experience? This is true, but only for motivation change from prior to the experiment to during the experiment for the third condition. This implies that simply using references to previous sessions does not necessarily result in a better experience.

Does a variety of references help more than making the reference? Similarly, can a variety of references over a number of sessions be more useful than a single repeated reference? This has been proven to be true as can be shown by the improved motivation exhibited by participants within the third group. It should be noted, however that only motivation increased in a significant manner, and other aspects of the experience, while promising are not significant, and a larger sample size that results in greater power can help to resolve some of these issues.

6.1.3. Payment for Study

For payment, participants were paid a flat rate of 9 British Pounds per hour for their time. This is a rate that was recommended by Prolific for video calls which implies that the rate that was used was not overly high and was at an appropriate level where it was not underpaying or overpaying. This implies that the payment itself and its effect on the participant's actions, behavior, and data collected can be considered to be less than if participants were overpaid for their time.

6.1.4. Conversations through Zoom

While Zoom was not an optimal way to carry out the experiments, it was possible to carry out the experiment through Zoom which shows promise for the future of research containing conversational agents. It should be noted that the virtual Furhat is different in terms of presence when compared to the actual Furhat robot due to the way in which sound is delivered through Zoom as well as the simple fact that the 2d virtual Furhat does not directly match the translucent appearance appearance of the physical Furhat robot. This difference can be improved upon by using more realistic 3d models for the virtual Furhat robot. Similarly, based on the size of a participant's screen, the size of the robot may appear small or large, making the participant's perception of the robot vary based on the type of hardware they are using.

Delays and Lag The physical location of the participant as well as the status of their internet connection can result in a certain level of delay or lag. To allow for enough time to provided to participants, a relatively long listening time is set for the robot. However, in terms of perception, this can result in a participant saying something and then spending one to two seconds staring at the robot before the robot replies due to the delay. There is a possibility that such a delay can affect the participant's opinion of the robot, but in most cases, when there was a delay, it was quite obvious to participants that such delays are caused by Zoom and not the robot itself.

Varying Levels of Internet Connection Based on the internet connection, video can appear to flicker and sound can cut off. Such problems are difficult to detect within the experiment, and it should be noted that while this is similar to delays and lag where the problem is caused by Zoom, the way in which sound is cut off can create the appearance of either a staggered conversation or simply create the appearance of sentences being cut short. In the first case, the participant can easily recognize the problem being attributed to Zoom or the internet connection. In the second case, however, the participant may not recognize that there was even a problem.

Varying Levels of Audio Quality All participants are required to have a relatively high quality microphone so that their voice can be properly captured and conveyed to allow for the speech recognition to work with minimal complications. Typically, the best audio setup is that of a head mounted microphone, but not all participants may have necessarily had such a microphone. During the pilot studies, it was found that mobile phones had a high level of suitability for the experiment when compared to the built in microphones typically found in laptops or desktop computers. By allowing participants to join the Zoom meeting with their phone, the number of eligible participants increased significantly while maintaining acceptable levels of audio quality.

6.1.5. Levels of English Proficiency

The majority of participants are from countries where English is not the native language. While this was a concern prior to the experiment, the use of the first study as a way to filter out participants who were incapable of understanding instructions written in English was quite successful in selecting participants who had a sufficient level of English for the experiment. In addition to this, since the explanations and questions asked by the robot are intended to be understood by the general layperson, the requirements with regards to proficiency in English were not particularly high. This meant that the level of English proficiency by the participants who completed all three sessions was not a concern.

6.1.6. Technological Proficiency of Participants from Prolific

About half of the participants were inexperienced with regards to using Zoom. Despite that, in all cases, participants were able to determine how to use Zoom and connect to audio without the need for a high amount of assistance. With regards to technological proficiency within Prolific, this was less of a problem, because participants who were given access to the initial study were filtered based on their prior experience and the the percentage of completed submissions they had that were approved. This ensured that participants were not completely inexperienced with regards to using Prolific.

6.1.7. Strong Accents and Variants of English

There were some cases where pronunciation of words affected the way in which the word was recognized by the robot. Examples of such instances include Italian accents which attach a vowel to the end of words, and thick Scottish accents. It should be noted that in such cases where even the experimenter had difficulty understanding the speech, the robot's speech recognition was not considered to be at fault, and the audio quality was also not considered to be problematic.

6.1.8. Timescales for Goal Achievement

As mentioned in prior sections, goal achievement typically takes place over longer time periods. As a result, while the study that was carried out can be used to make statements regarding short term goal achievement, whether such results can be extrapolated to long term goal achievement is questionable.

6.1.9. Lack of Visual Aid

Within didactic approaches, visual aid can be quite effective at conveying a message. Rather than having a droning voice present information, doing so in a visual manner may be more effective in terms of comprehension as well as time due to the way in which visual methods can provide information more succinctly. This is true not only within teaching contexts, but also within clinical contexts where conveying dietary information in an easy to understand way can make the difference between a successful diet change and a unsuccessful diet change.

6.2. Addressing of Claims and Contributions

Since design was a significant part of this this work, whether the claims that were proposed for the designed system were properly delivered regardless of groups considered are analyzed in this section.

6.2.1. CL001: Improved Autonomy through Goal Setting

The robot allows the user to choose a goal, and also allows the user to change the goal to be more or less ambitious based on the progress that was made as a form of difficulty setting. In this sense, the robot allows for the user to have autonomy with regards to the type of changes that they can make towards their goal. In addition to this, users are expected to determine how they want to work towards their goal by changing their diet. They have the option of choosing healthier options, or eating less. That said, the answers provided for level of autonomy appeared to be slightly above a neutral value which indicates that the level of autonomy can be increased and that it may be considered to be the bare minimum. So, to conclude, while the system allows a certain level of autonomy, based on the opinions of participants, a greater level of choice on the part of the participants would have been preferred.

6.2.2. CL002: An Engaging Conversation through Shared Experiences

With regards to engagement, all conditions had an engagement value where the median was a 4 and the mean was above a neutral value of 3, which indicates that the conversation was somewhat engaging. In this sense, the system was partially successful, but a greater level of personalization as well as a higher level of interaction and input from the user may have lead to a higher appraisal regarding the level of engagement that the user had.

6.2.3. CL003: Education

With regards to the change in knowledge regarding diet and diabetes, the median is that of a 4 and the mean is above the neutral value of 3 which indicates that users did learn about how to manage their diet as well as about type II diabetes, but the level of change in their knowledge can be improved further. With regards to how such improvements can be made, further investigation will need to be conducted to determine the most effective way to deliver information.

6.2.4. CL004: Incremental Goal Achievement

More than 86% of all participants managed to reach their final goal after changing their goal in the second session to be either more or less ambitious. Without a change in the goal in the second session, slightly more than 80% of participants achieved their goal. This indicates that the proposed goals were realistic and reasonable, but not so easy that all participants were able to achieve them.

6.2.5. CL005: A Friendly and Relatable Conversational Partner

With regards to likability within the godspeed questionnaire, the robot scored medians and means of 4 or higher for all subcategories of likeability. With regards to friendliness which is a sub item, the robot was scored a median of 5 for all conditions, and means of 4.4 or higher. This indicates that the system was successful in being a friendly conversational partner and that any improvements will be marginal in value with respect to perceived friendliness.

6.3. Future Work

6.3.1. Exclusions of those with Metabolic Diseases

Although the designed system is a system for type II diabetes treatment and prevention, it can only be stated that the prevention aspect of the system was tested due to ethical limitations excluding many people with diabetes from the study. As a result, it could be argued that the conversational agent is useful for prevention, but the same argument would be more difficult to make for treatment of current conditions. For future studies, inclusion of participants with type II diabetes with a detailed observation of sugar levels to avoid any negative health effects is recommended.

6.3.2. Controlled Environment

As mentioned before, this system is intended to be used at the convenience of the participant and was intended to highlight a use case of a conversational agent that would be difficult to have a human complete due to lack of availability. Despite this

6.3.3. Longer Interactions

When analyzing the duration of the interactions, less than 64% of participants thought that the duration of the individual sessions was appropriate with the rest of the participants stating that the interactions should have been longer. This was true even for the third condition which had longer sessions due to the usage of shared experiences and motivational phrases. Future experiments with varying session lengths can help to provide greater insight into how long a session should be to achieve participant satisfaction as well as to maximize achievement and other metrics. In addition to this, even in human conversations, it is difficult for individuals to end a conversation when they want it to be ended [63], so such future work can help provide clarity on how to best design for a satisfying ending to a conversation. as well

6.3.4. More Sessions

As mentioned in prior sections, the experiment that was completed can be used to make statements regarding short term goal achievement over the period of days, but the same cannot be said for long term goal achievement over a period of months which is more common for behavior change scenarios [87]. In addition to this, less than 54% of participants thought that the number of sessions was appropriate which indicates that even for short term goal achievement, the number of sessions may have not been sufficient. Future experiments that test additional sessions can test how shared experiences and goal achievement are affected with a higher number of sessions. Naturally, the difficulty of goals should be scaled in proportion to the increase in sessions.

6.3.5. Effectiveness and Automation

Can this be used today to help individuals with their diet? Although this has been shown to have helped individuals reach diet related goals and most individuals stated that they would continue working on their diet, whether this can be used for partial automation has yet to be determined. To determine whether this can be used as a tool to assist nutritionists and dietitians, further work will need to be completed. A future work that includes nutritionists and dietitians in the design process as well as actual patients can help to shed light on how effectively a diet related system such as the one described in this work can partially automate the work of nutritionists and dietitians. Inclusion of dietitians within the development of the foundation, specification and evaluation as specified withing SCE [73] can allow for the robot to be tested in a realistic setting with immediate practical applications.

6.3.6. Reducing Manually Designed Aspects for Greater Flexibility and Personalization

Currently, the way in which motivational phrases are generated is by using a combination of natural language processing techniques and manually crafted phrases. While this is fine for a controlled and restricted use case like the one in this work, if one wishes to apply the system designed in this work to a broader use case, or allow for greater flexibility in terms of the possible advice given by the robot, reduction of handcrafted parts will be required and future works should take this into account. In addition to this mixed approaches that use the techniques presented in this work along with the personality types presented in [48] may allow for greater levels of personalization that can possibly used to improve the impact of the memory used by the conversational agent.

6.3.7. Determining Intrinsically Significant Values and Concepts Programmatically

Currently, the conversational agent sidesteps the problem of determining what is of intrinsic value to the participant due to the way in which it uses motivational interviewing techniques and questions to elicit answers that immediately get to the heart of why a person is working towards a goal and what their motivations are. This is quite effective within a period like the one described in this work, but if one were to extend the number of sessions, such methods may become quite repetitive over time. For future studies that have use cases where motivational interviewing are not valid, a programmatic method to deduce what has intrinsic value to the participant can possibly allow for less repetition and a better experience over a longer timescale.

6.3.8. Acceptable Levels of Repetition

In groups 2 and 3, there are two extremes with regards to how references to shared memories are used. In group 2, references to only one past event are made and in group 3, references to the same past event cannot be made which means that all events that are referred to must be different. The results show that the change in motivation from before the experiment to during the conversation is highest with group 3, but this does not state whether some level of repetition is acceptable. In previous works such as [48] and [95], it was made clear that there was some repetition in terms of the usages of the computational memory model across multiple sessions, however the level of repetition was not specified and as a result, this is an unexplored area within the literature. For this, additional groups that have varying levels of repetition and variety in the past events that are referred to can offer some insight towards what acceptable levels of repetition are with respect to references to past events.

6.4. Conclusion

In this work, a conversational agent was designed and evaluated to determine the effect of references to shared experiences or the use of past events with intrinsic motivational value phrased into motivational statements within the context of acquiring a diet related habit for type II diabetes prevention or treatment. To support the use of motivational statements that utilize past events with intrinsic value, a novel motivational memory model was designed and implemented. The system that was designed to test the effects of such references was that of a conversational agent that is able to walk patients or users through the process of acquiring a diet related habit of their choosing by helping them work towards a manageable goal based on their current diet and intake.

This system took the following design considerations into account during design and implementation:

1. How should shared experiences be modeled and organized within the memory?
2. How should shared experiences be modified to generate motivational phrases?
3. When can shared experiences be used?

These considerations took the context in which the experience was used and sentiment into account when determining how a experience should be rephrased so that it can be used appropriately in a particular context.

To determine the effect of shared references on goal achievement and user experience, three conditions were considered to determine the impact of shared references and to provide clarity regarding how shared references should be used. These conditions were as follows:

1. No references to previous events
2. References to only one event environment.
3. Multiple references to different events (variety) where no events are ever brought up twice

An experiment was conducted over Zoom meetings with participants who were recruited through Prolific to evaluate the the three conditions. It was found that the use of the motivational memory resulted in a significant self reported increase in motivation from before the experiment to during the experiment for the last condition where the same past event was never mentioned twice when compared to the second condition where only the same past event can be mentioned. The implications of such a finding is that simply using references to past events is not enough for motivation, and it shows that references to past events must be more deliberately chosen. Future work can further narrow down what the most important aspects of references to the past are to receive the greatest positive impact. While a relationship between references and goal achievement was not established, future studies can take the findings of this work into account to design and run experiments with a higher number of participants and more granular control of confounding conditions to better test the relationship between goal achievement, motivation, and references to past events. Areas to explore include, but are not limited to, the following:

1. In person interactions with the robot
2. Inclusion of individuals with type II diabetes and metabolic diseases with careful surveillance and monitoring
3. Varying levels of repetition of shared experiences.
4. Additional sessions and varying durations for individual sessions.
5. Designing and testing with the input of health professionals.
6. Development of systems to quantify levels of intrinsic motivational value for shared experiences to allow for the development of a motivational memory without handcrafted elements.

A

Models

This section includes ontologies and models that are relevant to the interaction. The ontology incorporates all previously mentioned components which include the foundation, specification and evaluation and provides a mapping or knowledge base that describes the relationship between those components.

In addition to the ontology, other models that specify the structure of different components and the architecture of the conversational agent are included as well.

A.1. Class Interaction

The class interaction diagram shows how different stakeholders and entities interact with one another to accomplish and carry out the functions specified in the Claims.

Unlike other models which may focus mainly on the interaction between the user and the conversational agent, this model illustrates the bigger picture from the perspective of the user and those the user interacts with outside of the interactions with the conversational agent to paint a picture of exactly how the patient will act within sessions and outside sessions as well.

Entities

- Patient
  - This refers to the patient who either has type II diabetes or is at risk of being diagnosed with type II diabetes.
- Embodiment
  - This refers to the virtual Furhat robot.
- Speech Recognition
  - This refers to the conversational agent's module that can convert spoken sound to text. This is built into Furhat.
- Speech Generator
  - This refers to the conversational agent's module that converts text to speech. This is built into Furhat
- Dialogue
  - This refers to the series of spoken phrases, questions, and responses that will be carried out by the conversational agent.
- Memory
  - This refers to data regarding the patient, their progress and any shared experiences.

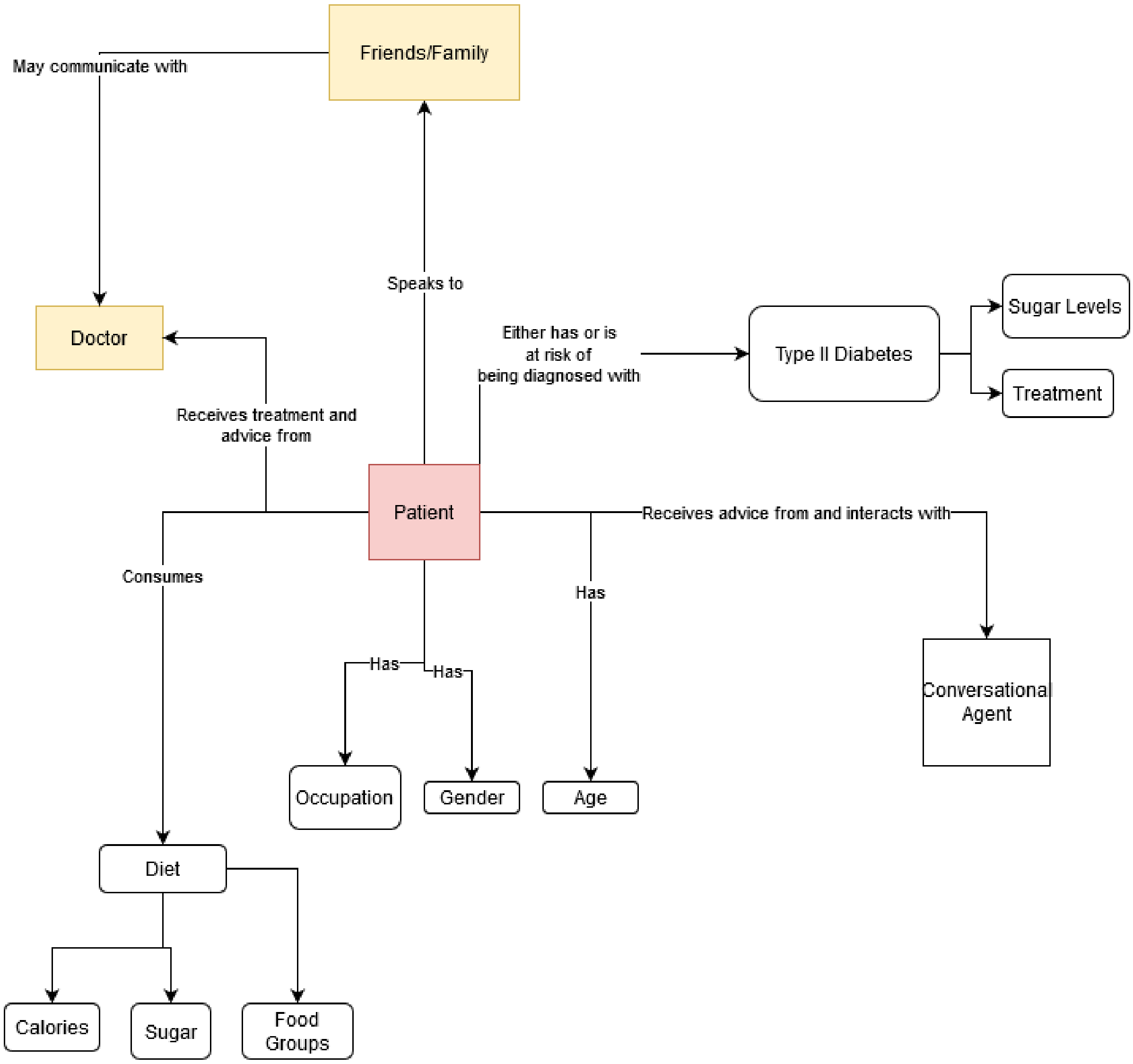


Figure A.1: Illustrates the interactions between all parties involved in the experiment

- Primary Care Physician
  - This refers to the primary care physician who may have directed the patient to interact with the conversational agent.
- Family/Friends
  - This refers to individuals who are within the social circle of the patient.

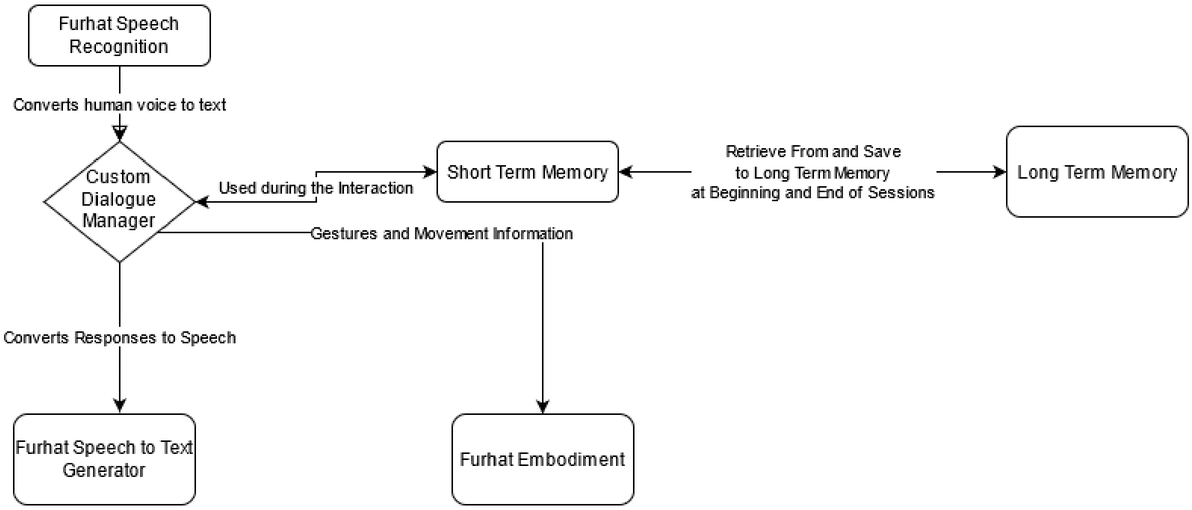


Figure A.2: Illustrates how the different components of the conversational agent work together to deliver an interaction

A.2. Conversational Agent Architecture

The conversational agent uses Furhat in tandem with a custom dialogue manager to allow it to perform functions that were not possible or difficult to implement with the Furhat SDK (software development kit) alone. In this conversational agent, the speech recognition is used to convert human responses to text. This text is sent to a flask server that is programmed in Python which contains the dialogue manager. The flask server processes the textual information from the response to determine which part of the dialogue flow it is in. If needed, the manager will access the short term memory during the interaction.

If the manager needs information regarding a user it has interacted with in the past, it will read in the information from long term memory relating to that user at the beginning of the interaction and process that into the short term memory for quick and easy access. Using this information, the manager can offer personalized responses that are relevant to the situation of the particular user that it is interacting with.

These responses are sent back to Furhat so that the embodiment can output gestures and the text to speech generator can output spoken language.

A.3. Dialogues

The dialogues determine what is said by the conversational agent and when.

It specifies what the expected responses by the participant are, as well as the flow of the conversation carried out between the conversational agent and the participant.

This section is split between the individual sessions that will be carried out between the conversational agent and the participant.

A.3.1. Session 1 Dialogue

This specifies the dialogue for the first session carried out between the conversational agent and the user.

In session 1, the primary goal is to lay the foundation for the sessions that will follow. This means that the conversational agent should learn all relevant information about the user to set an appropriate goal for the user to follow. The user should understand what the purpose of the conversational agent is, and the work that he or she will complete with it. It is here that the user is acquainted with the conversational agent for the first time.

If the user has any questions about diabetes, or about the goal and how it is calculated, these are answered in the search based answer sections that allow for general question answering.

By the end of the session, the user should have a better understanding of their goal and how it relates to type 2 diabetes. They may also have a clearer understanding of their own motivations due to the questions that are answered regarding their feelings and thoughts towards the goal which will be used as shared experiences in future sessions.

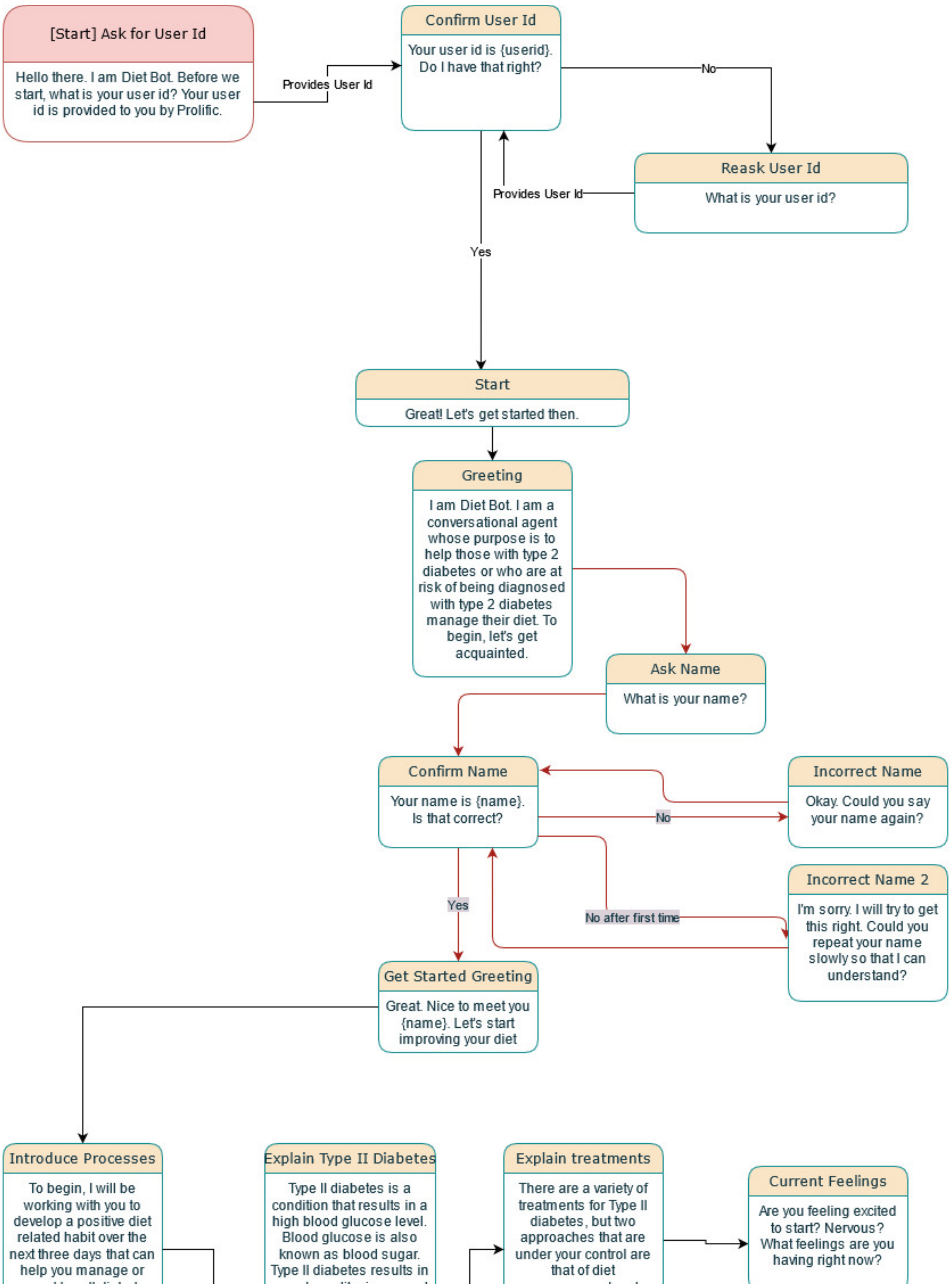


Figure A.3: Shows the flow of states used for the first session's interaction (Part 1)

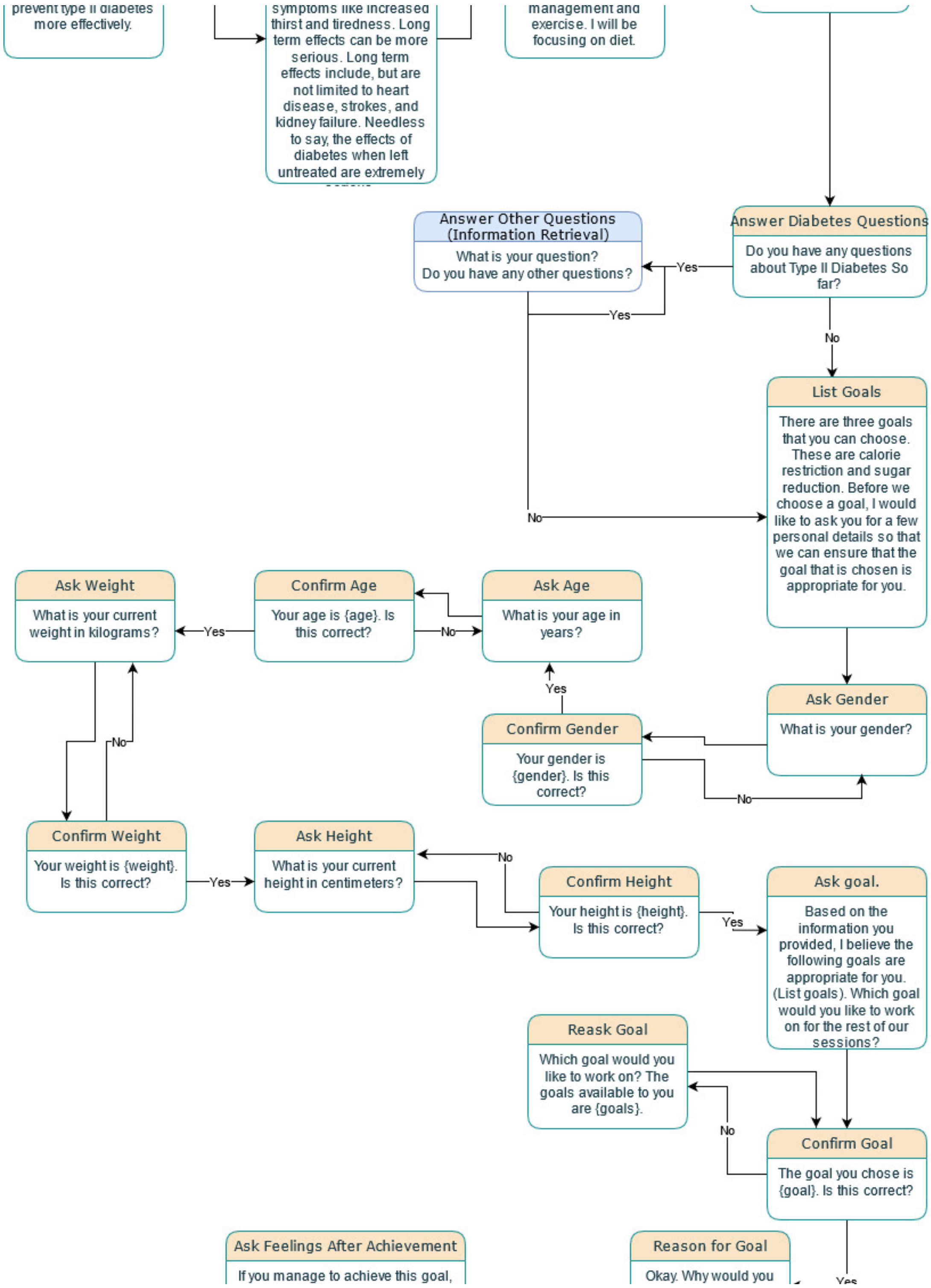


Figure A.4: Shows the flow of states used for the first session's interaction (Part 2)

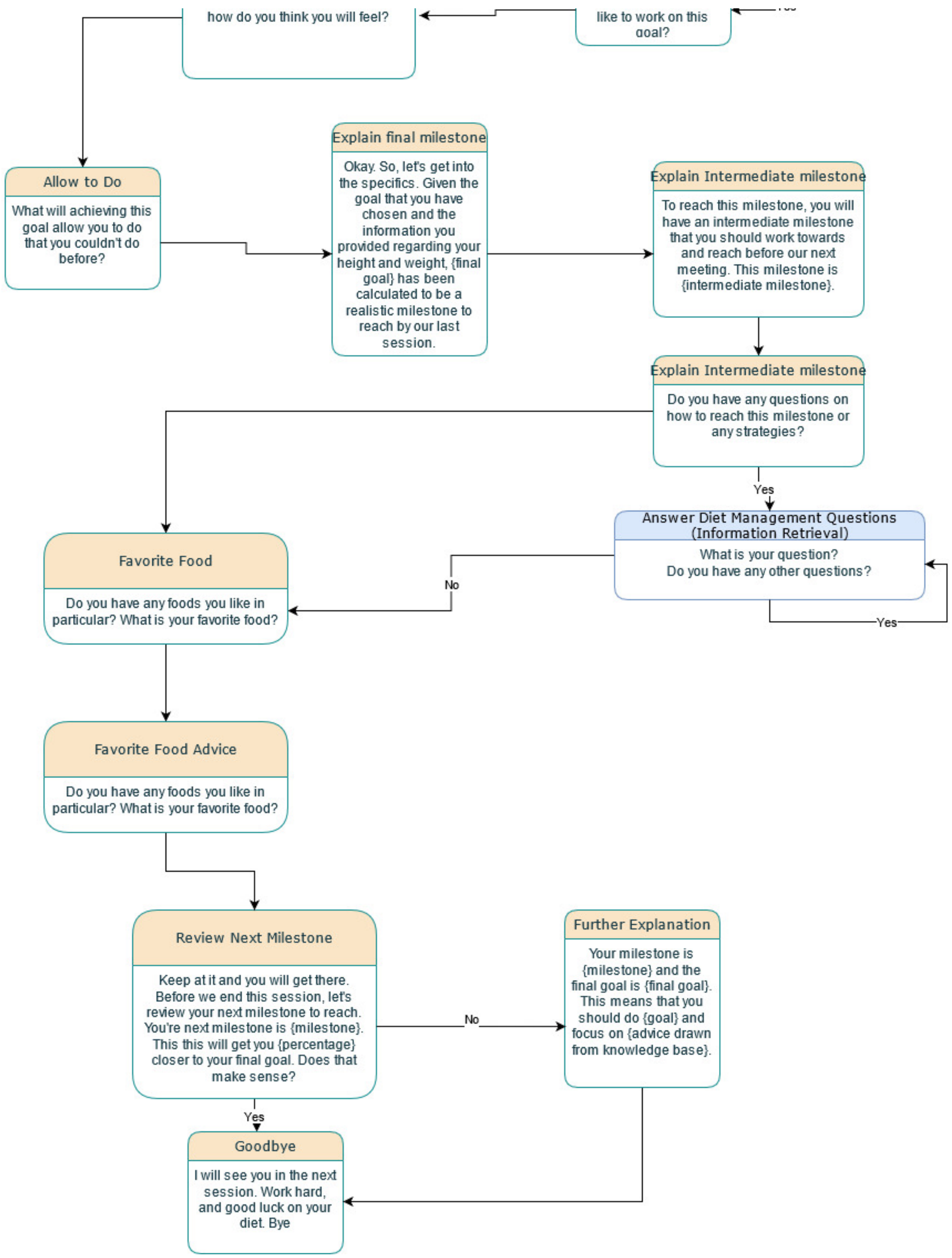


Figure A.5: Shows the flow of states used for the first session's interaction (Part 3)

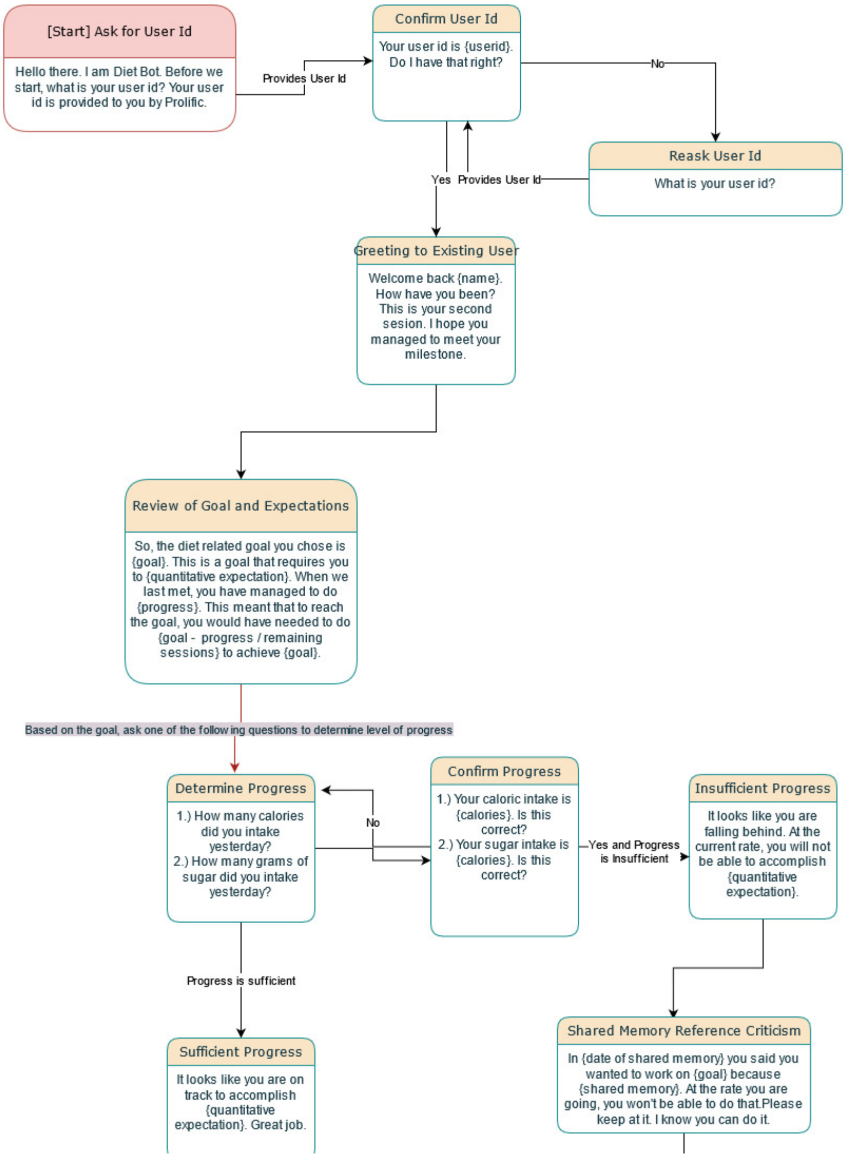


Figure A.6: Shows the flow of states used for the second session's interaction (Part 1)

A.3.2. Session 2 Dialogue

By the second session, the user and the conversational agent are already acquainted, and the primary goal is to determine progress and achievement of the milestone. Depending on whether the milestone is met or not, the user can continue towards their goal, or adjust it accordingly.

The point of this session is to offer advice, encouragement, and constructive criticism. Here, we can also identify what the user is specifically struggling with and try to provide tips. It is here that we see the first uses of shared experiences as a way to realign the motives of the user.

Because the user is already familiar with the conversational agent, this is expected to be a shorter interaction with the primary purpose of realigning user with their goal.

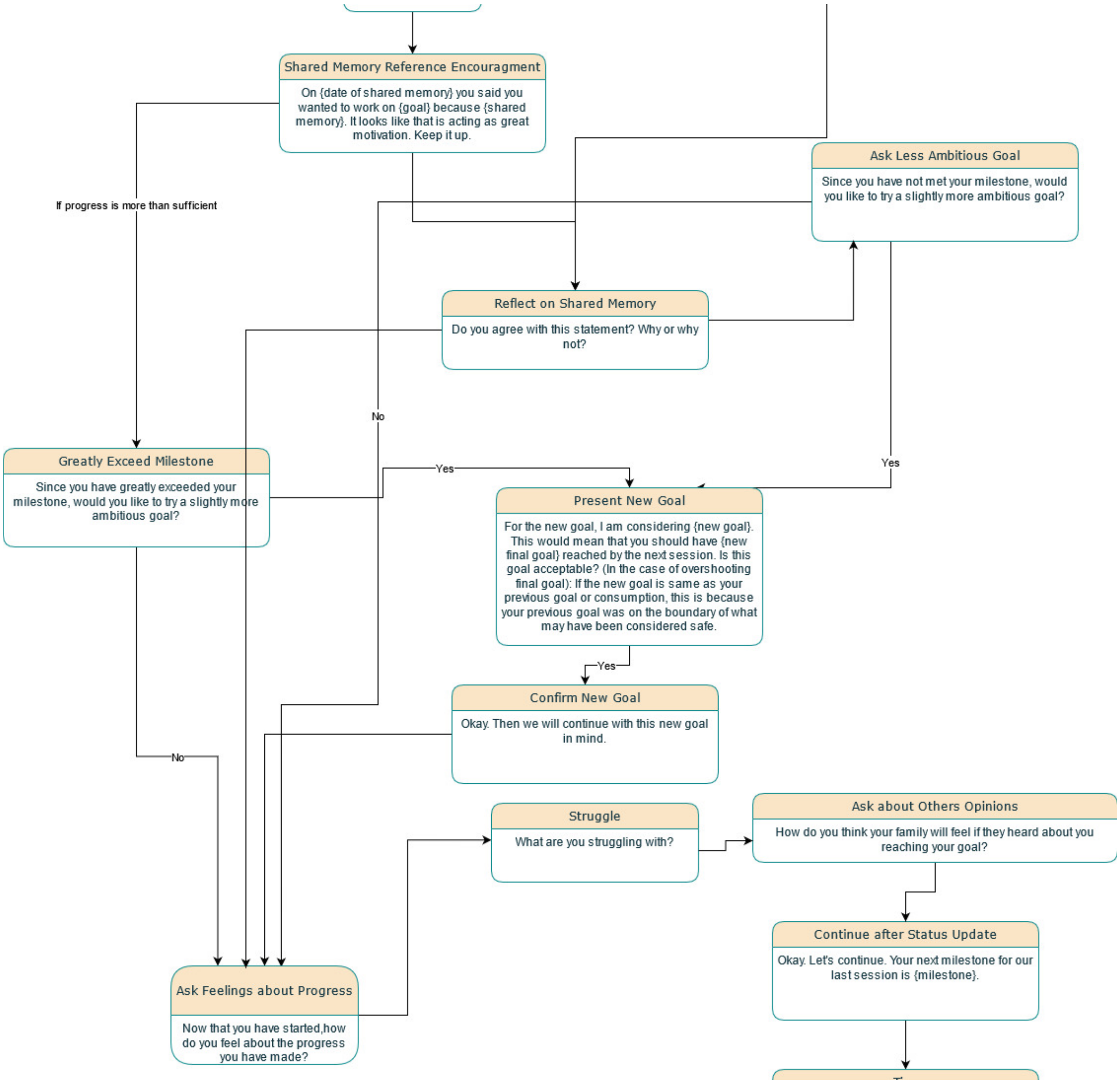


Figure A.7: Shows the flow of states used for the second session's interaction (Part 2)

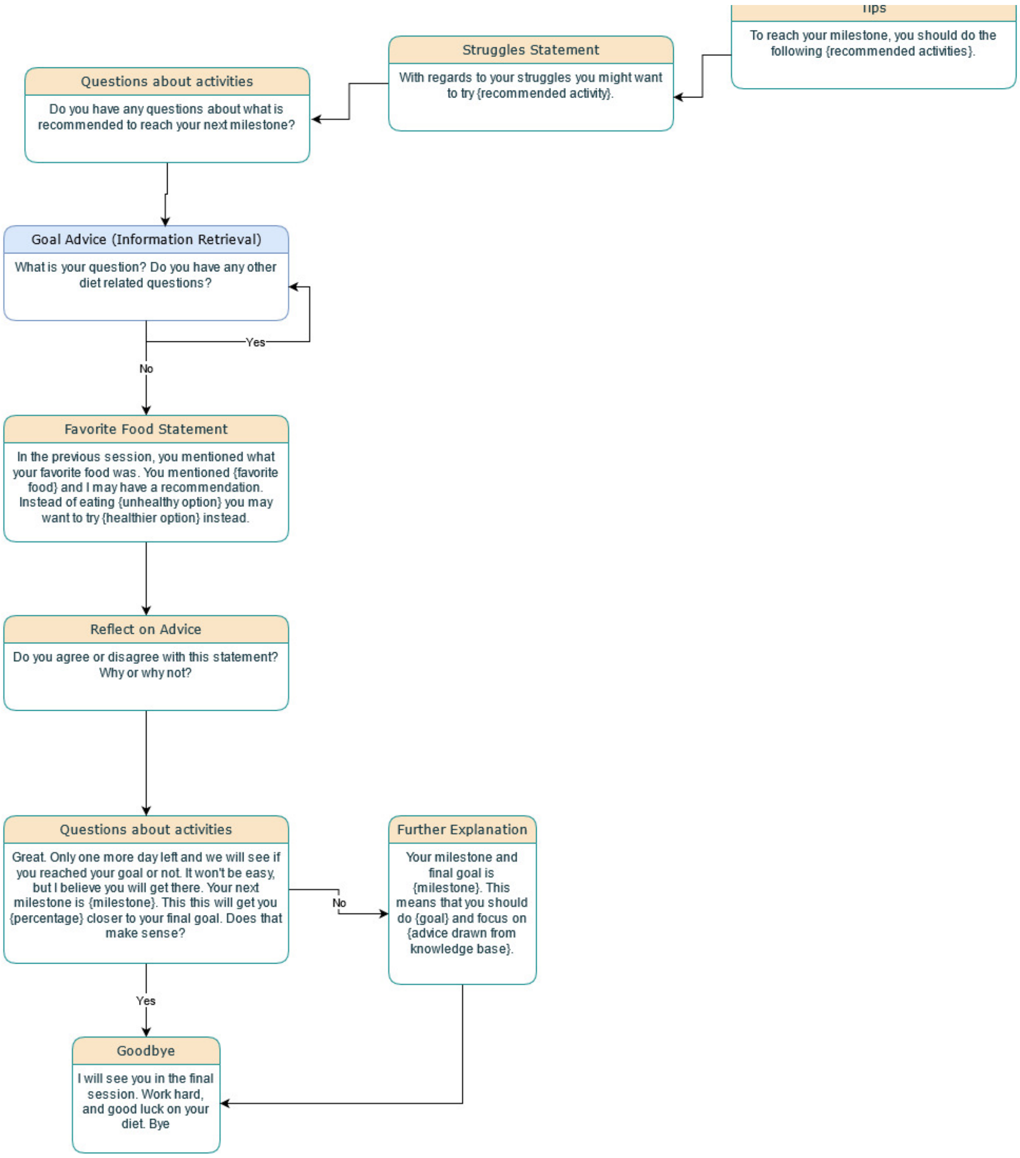


Figure A.8: Shows the flow of states used for the second session's interaction (Part 3)

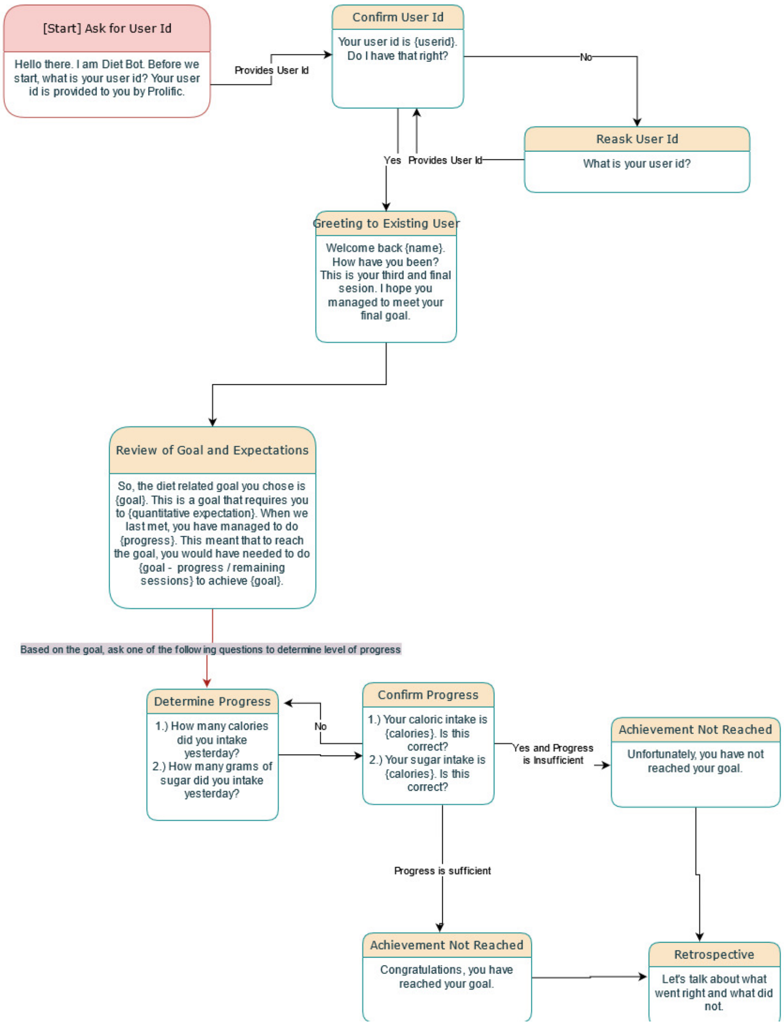


Figure A.9: Shows the flow of states used for the third session's interaction (Part 1)

A.3.3. Session 3 Dialogue

The purpose of the final session is to perform a final status update with the user with the purpose of determining whether they met their final goal or not.

Once this is determined, the conversational agent is expected to reflect on the progress over the past sessions in chronological order. It is here that experiences are referred to to frame the progress within the context of the user's motivations to provide them with perspective on what could be considered to be the bigger picture with regards to their personal desires.

This means that the user is allowed to look back on their accomplishments and see where they succeeded and where they struggled. The user is expected to leave this session with a clear idea of how to improve and how to continue working towards more ambitious habits related to their diet.

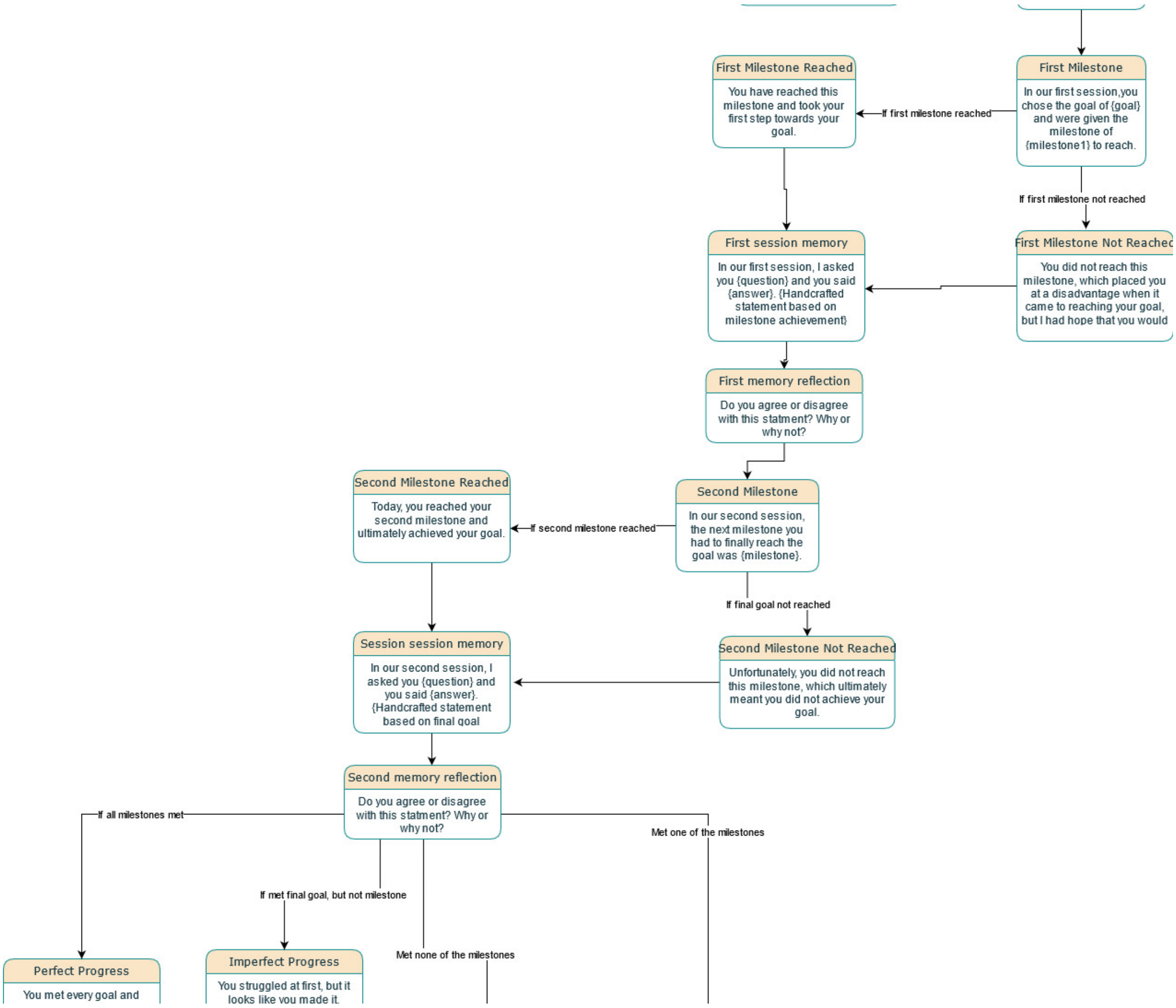


Figure A.10: Shows the flow of states used for the third session's interaction (Part 2)

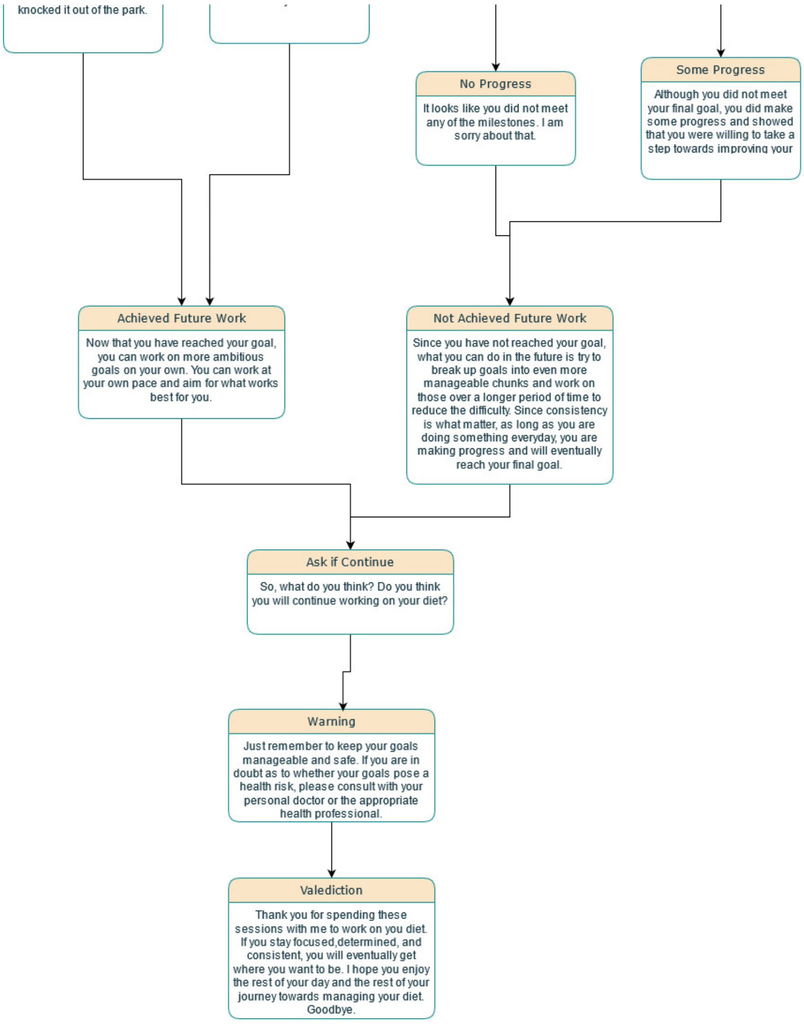


Figure A.11: Shows the flow of states used for the third session's interaction (Part 3)

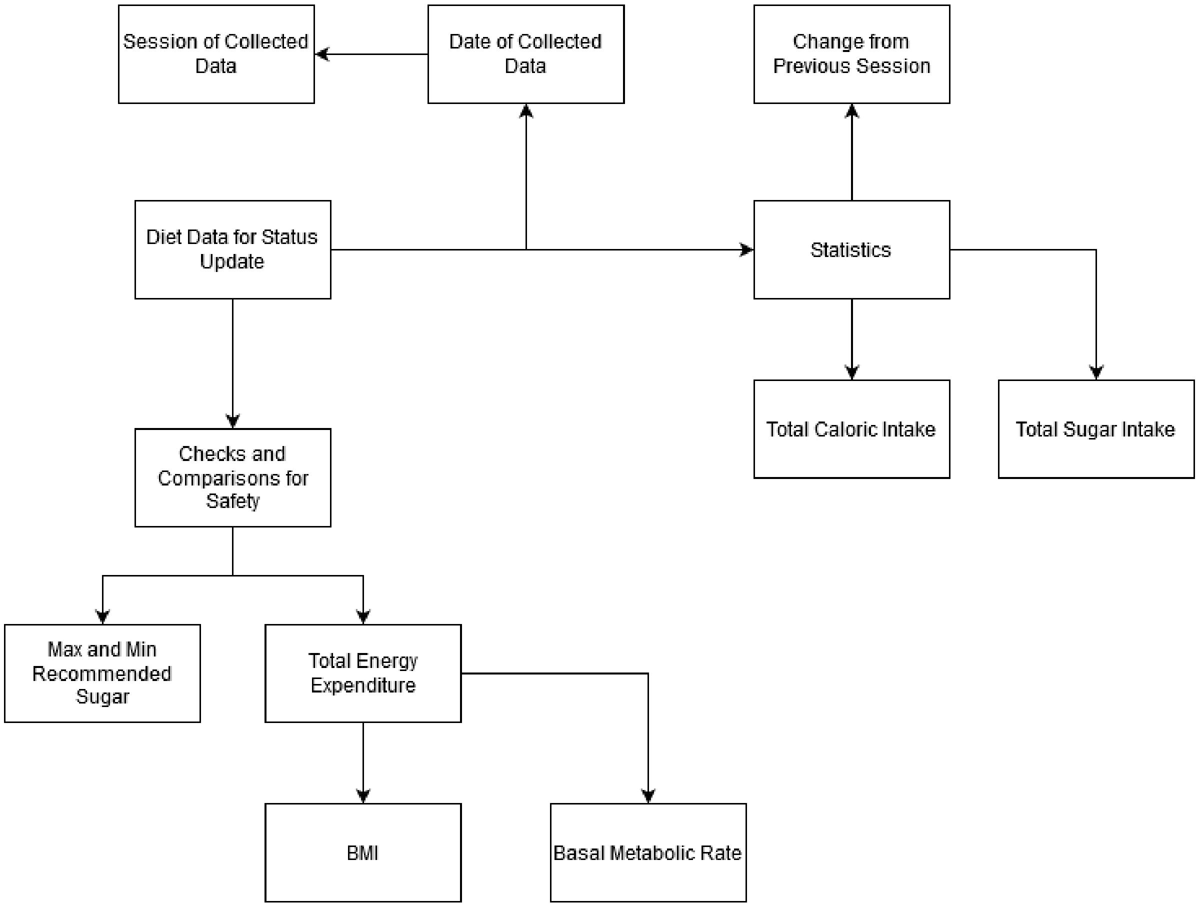


Figure A.12: Displays the organization of diet related data that is stored

A.4. Diet Related Data

The data that is collected and processed in each session that is related to diet is shown here.

In the initial session, some physiological data such as gender, age, height, and weight are collected to calculate some initial thresholds that will be used to avoid any diet changes or actions that can result in harm as specified in User Data. This data is used in future sessions to calculate some useful figures to determine progress and adherence to those thresholds.

Each session update usually focuses on a particular goal which is calorie restriction or sugar reduction. The progress of these goals are monitored by asking the user about their caloric intake or their sugar intake respectively.

By comparing the current intake with the intake from the previous session, a change can be recorded and then compared with whatever milestones were calculated in the initial session to determine if the user is on track or not. Similarly, to ensure that the participant is not going too far or engaging in any reductions that can be potentially harmful, comparisons and checks for safety are used which utilizes maximum and minimum recommended sugar values and total energy expenditure. Total energy expenditure is composed of BMI and basal metabolic rate (BMR). In the case of BMI, this can also be used to determine if the user is underweight or not in the first session and avoid any harmful reductions.

The corresponding model can be found in figure A.12.

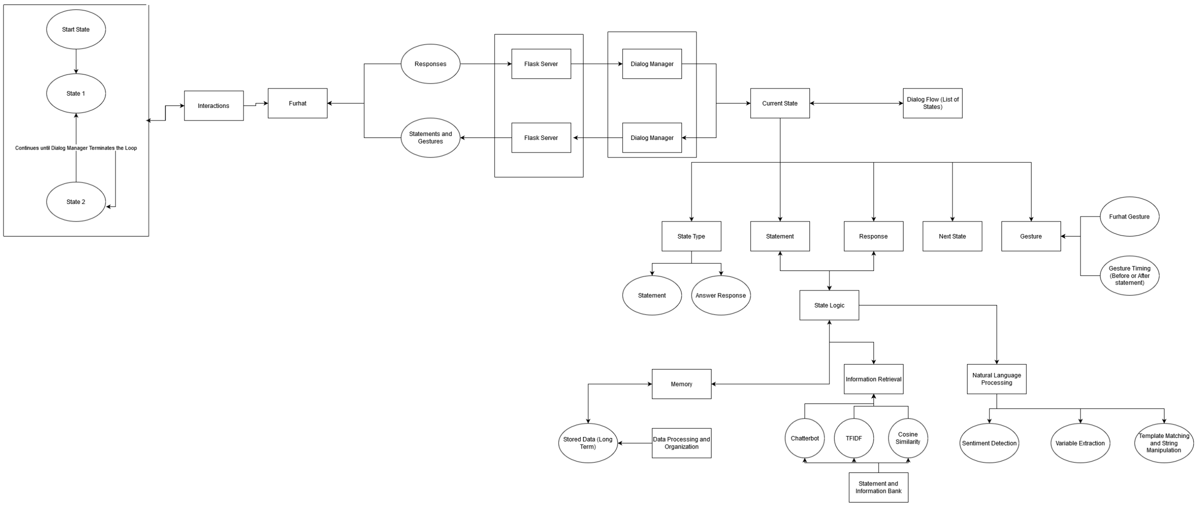


Figure A.13: Displays the organization of diet related data that is stored

A.5. Information Flow within Conversational Agent

In this model, a detailed display of all of the components used within the conversational agent are provided.

This not only includes the components themselves, but also the modules that were implemented that perform the specific logic for the module.

Starting from the right hand side, since Furhat is only used for its speech recognition, embodiment, gestures, and text to speech, the dialogue management functionality and states are used minimally. Specifically, Furhat only has two states which are used in a loop and a third state that is only run when a terminating command is sent from the actual dialogue manager.

Within these states, calls are made to the Flask server which contains the actual dialogue manager and these calls receive statements and gestures that are to be outputted by Furhat, or send responses by the user to the dialogue manager.

The dialogue manager utilizes a state machine in combination with NLP techniques and information retrieval techniques. In addition to this, task related and interaction related data is stored in a short term and long term memory to help facilitate the interaction, to personalize the interaction, and provide specific advice for a particular user.

The state machine utilizes three basic types of states which is that of a statement where the robot simply says a statement and moves on, a response answer where the robot expects a response, and finally an end state to inform the Furhat robot of when the conversation has been terminated.

Within the memory, there is a short term memory and a long term memory. The short term memory is only maintained during the interaction and is then transferred to the long term memory so that information regarding the user can be used across multiple sessions. Between sessions, some data processing and organization will be conducted that could not be conducted during the interaction due to computational loads or simply due to the fact that such processes were not needed during the session in which it was recorded.

Within the information retrieval module, question and answer pairs are stored and used for general question answering. Using libraries and NLP techniques such as chatterbot, TF-IDF, and cosine similarity, the module is able to retrieve an answer that best matches a question posed by the user.

Finally, as part of the NLP module which is used for processing long term data to reword into shared and parsing responses, functions which detect sentiment, and can be used for finding a number or particular answer within a response from the user are provided.

The corresponding model can be found within figure A.13.

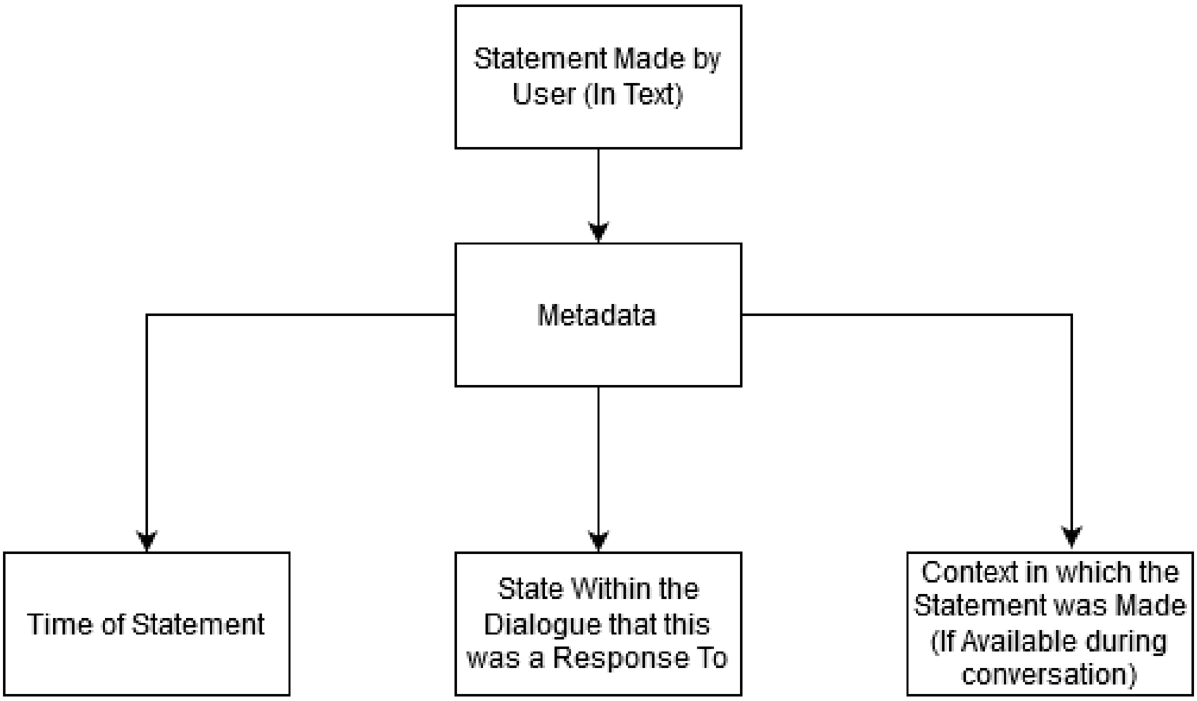


Figure A.14: Model showing what data is stored in the short term.

A.6. Processing of Experiential Data in Short Term Memory

The short term memory contains diet based information, metadata regarding the user, and statements made by the user during the interaction that may be used in the future as a shared memory reference.

The model presented here shows how a statement made is then converted into a memory that is stored in the short term memory in preparation for further processing in the long term memory so that it can be used in a future session.

During a conversation, the dialogue manager is aware of the state it is in within a flow, and can use that to provide some temporal information regarding the statement. Since it also knows what the statement was made in response to, a context can be stored during the interaction itself without any need for computationally intensive natural language processing methods.

Finally, the actual statement itself is stored in the short term memory so that it can later be processed by more computationally intensive methods in the long term memory.

The corresponding model can be found in figure A.14.

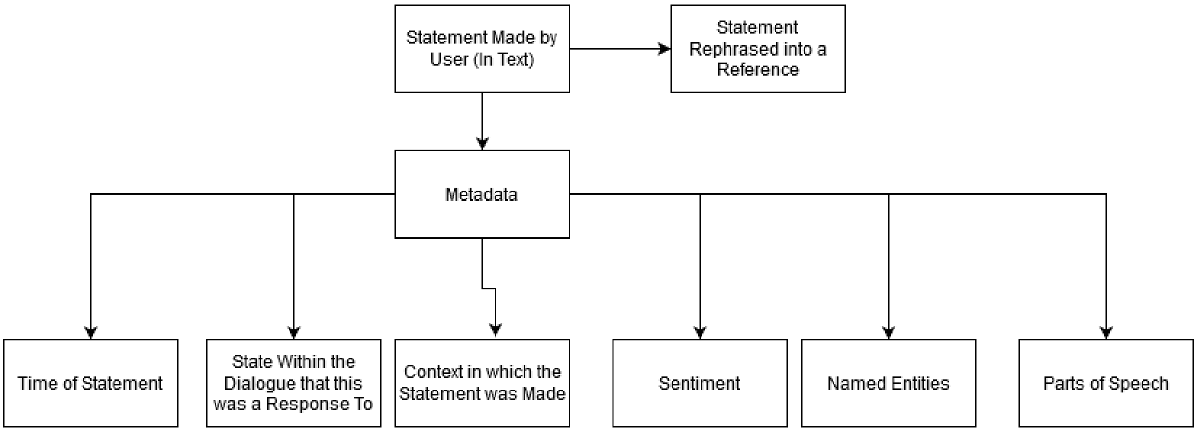


Figure A.15: Model showing what data is stored in the long term. When compared with short term data, it can be seen how short term data is processed to reach the state seen in the long term memory.

A.7. Processing of Experiential Data in Long Term Memory

As mentioned within Processing of Experiential Data in Short Term Memory, the long term memory has the advantage of additional time and computational resources which allows for certain functions such as natural language processing to be used to extract information such as sentiment, named entities and parts of speech. These can then be used to rephrase the statement into a reference, or to use key words to create a statement that can act as a reference from scratch.

The sentiment can be utilized to determine level of confidence that the user has when they provided a response to a motivationally guided question in the dialogue. This combined with a summarization of the user's statement can help to determine whether to praise or encourage in addition to the possible contexts of usage.

The later can be useful in the case where the statement made by the user is very long, or when Furhat's speech recognition incorrectly recognizes some words in a sentence.

Named entities are not necessarily used to generate motivational phrased, but are used for anonymization to protect the identity of the user in the case that they provide information that can be used to identify them. In addition to this, multiple statements may need to be prepared to handle the possibility of usage within different contexts within status updates in the later sessions as either encouragement, praise or constructive criticism as mentioned within the design pattern of DP07: Context Appropriate Feedback (3.7).

The corresponding model can be found in figure A.15.

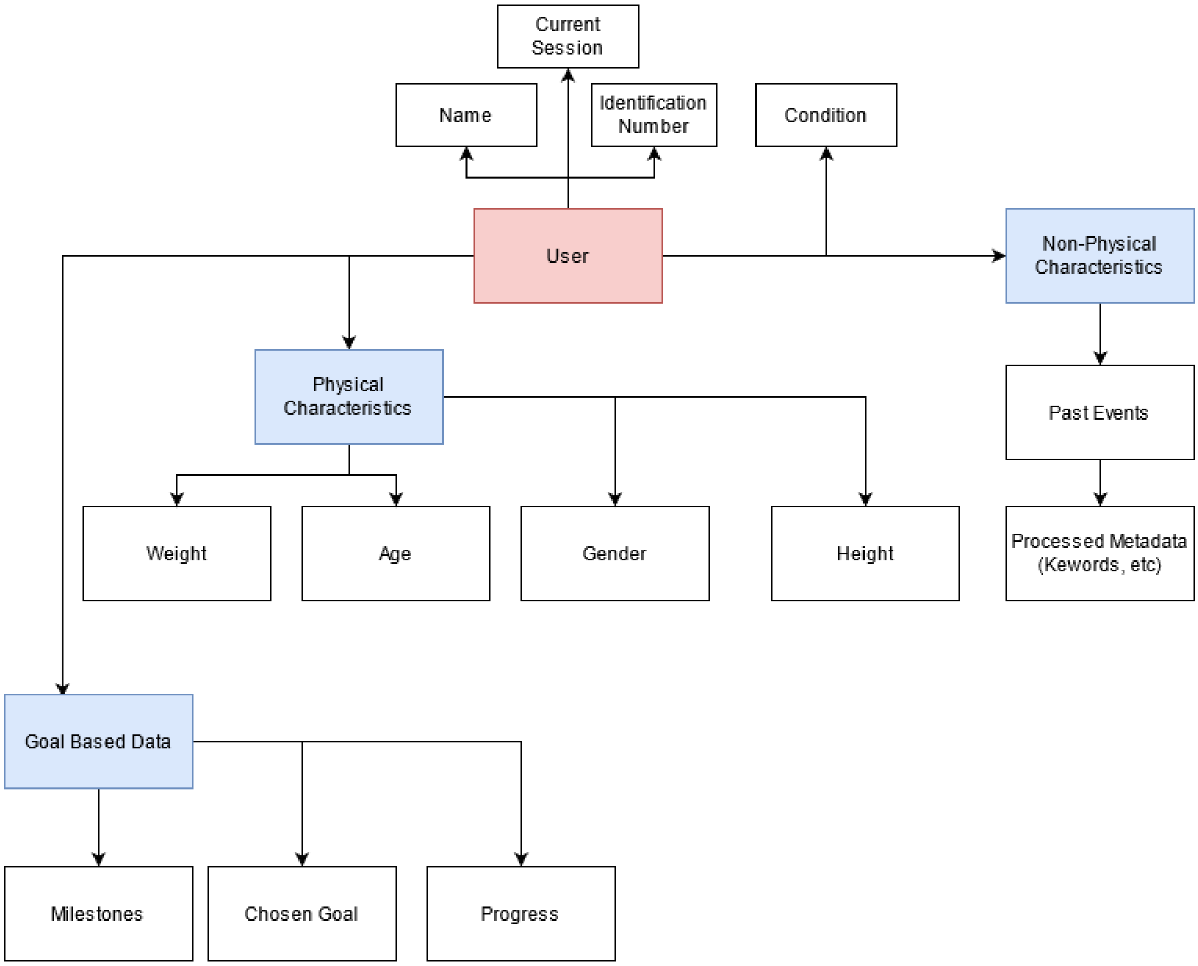


Figure A.16: Model showing what metadata is stored to represent an individual user.

A.8. User Data

The user data refers to metadata relating to the user that is used for the interaction. This contains a high level representation of the diet and memory data which are described in further detail in the respective models for diet and memory. The information used here has three primary purposes. The first is that of facilitating dialogue such that the agent can recognize and address the user. The second is that of facilitating progression through goals by maintaining milestones and experiences. This data is typically associated with a session or time so that it a chronological record can be utilized to reflect on progress in the third session. The third is that of recording physical data so as to facilitate proper diet advice and goal creation through calculation of metrics such as BMI and Total energy expenditure.

The corresponding model can be found within figure A.16.

Bibliography

[1] Gloria K.Q Agyapong. "The Effect of Service Quality on Customer Satisfaction in the Utility Industry – A Case of Vodafone (Ghana)". In: *International Journal of Business and Management* 6.5 (May 2011). DOI: 10.5539/ijbm.v6n5p203. URL: https://doi.org/10.5539/ijbm.v6n5p203.

[2] John R. Anderson, Michael Matessa, and Christian Lebiere. "ACT-R: A Theory of Higher Level Cognition and Its Relation to Visual Attention". In: *Human–Computer Interaction* 12.4 (Dec. 1997), pp. 439–462. DOI: 10.1207/s15327051hci1204_5. URL: https://doi.org/10.1207/s15327051hci1204_5.

[3] Lucija Andre et al. "Motivational power of future time perspective: Meta-analyses in education, work, and health". In: *PLOS ONE* 13.1 (Jan. 2018). Ed. by Andre Scherag, e0190492. DOI: 10.1371/journal.pone.0190492. URL: https://doi.org/10.1371/journal.pone.0190492.

[4] A. Astrup and N. Finer. "Redefining Type 2 diabetes: 'Diabesity' or 'Obesity Dependent Diabetes Mellitus'?" In: *Obesity Reviews* 1.2 (Oct. 2000), pp. 57–59. DOI: 10.1046/j.1467-789x.2000.00013.x. URL: https://doi.org/10.1046/j.1467-789x.2000.00013.x.

[5] Gordon D. Atlas, Terry Taggart, and Debra J. Goodell. "The effects of sensitivity to criticism on motivation and performance in music students". In: *British Journal of Music Education* 21.1 (Mar. 2004), pp. 81–87. DOI: 10.1017/s0265051703005540. URL: https://doi.org/10.1017/s0265051703005540.

[6] Emin Babakus and Gregory W. Boller. "An empirical assessment of the SERVQUAL scale". In: *Journal of Business Research* 24.3 (May 1992), pp. 253–268. DOI: 10.1016/0148-2963(92)90022-4. URL: https://doi.org/10.1016/0148-2963(92)90022-4.

[7] Shaira Baptista et al. "Acceptability of an Embodied Conversational Agent for Type 2 Diabetes Self-Management Education and Support via a Smartphone App: Mixed Methods Study". In: *JMIR mHealth and uHealth* 8.7 (July 2020), e17038. DOI: 10.2196/17038. URL: https://doi.org/10.2196/17038.

[8] Joeran Beel et al. "Research-paper recommender systems: a literature survey". In: *International Journal on Digital Libraries* 17.4 (July 2015), pp. 305–338. DOI: 10.1007/s00799-015-0156-0. URL: https://doi.org/10.1007/s00799-015-0156-0.

[9] Richard M Bergenstal et al. "Automated insulin dosing guidance to optimise insulin management in patients with type 2 diabetes: a multicentre, randomised controlled trial". In: *The Lancet* 393.10176 (Mar. 2019), pp. 1138–1148. DOI: 10.1016/s0140-6736(19)30368-x. URL: https://doi.org/10.1016/s0140-6736(19)30368-x.

[10] Jaikrit Bhutani and Sukriti Bhutani. "Worldwide burden of diabetes". In: *Indian Journal of Endocrinology and Metabolism* 18.6 (2014), p. 868. DOI: 10.4103/2230-8210.141388. URL: https://doi.org/10.4103/2230-8210.141388.

[11] María Blanca et al. "Non-normal data: Is ANOVA still a valid option?" In: *Psicothema* 29 (Oct. 2017), pp. 552–557. DOI: 10.7334/psicothema2016.383.

[12] San Bolkan, Alan K. Goodboy, and Dawn M. Kelsey. "Instructor Clarity and Student Motivation: Academic Performance as A Product of Students' Ability and Motivation to Process Instructional Material". In: *Communication Education* 65.2 (Oct. 2015), pp. 129–148. DOI: 10.1080/03634523.2015.1079329. URL: https://doi.org/10.1080/03634523.2015.1079329.

[13] Emiel A Boogerd et al. ""What Is eHealth": Time for An Update?" In: *JMIR Research Protocols* 4.1 (Mar. 2015), e29. DOI: 10.2196/resprot.4065. URL: https://doi.org/10.2196/resprot.4065.

[14] F. Burger, J. Broekens, and M. Neerincx. “Fostering Relatedness Between Children and Virtual Agents Through Reciprocal Self-disclosure”. In: *BNCAI*. 2016.

[15] James Carifio and R. Perla. “Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes”. In: *Journal of Social Sciences* 3 (2007), pp. 106–116.

[16] J. Cassell et al. “Embodiment in conversational interfaces”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI ’99*. ACM Press, 1999. DOI: 10.1145/302979.303150. URL: https://doi.org/10.1145/302979.303150.

[17] T. Chai and R. R. Draxler. “Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature”. In: *Geoscientific Model Development* 7.3 (June 2014), pp. 1247–1250. DOI: 10.5194/gmd-7-1247-2014. URL: https://doi.org/10.5194/gmd-7-1247-2014.

[18] Sung June Chang and In Ho Lee. “An emotion model using emotional memory and consciousness occupancy ratio”. In: *Proceedings of the 2005 international conference on Augmented tele-existence - ICAT ’05*. ACM Press, 2005. DOI: 10.1145/1152399.1152462. URL: https://doi.org/10.1145/1152399.1152462.

[19] Lousia Ming Yan Chung and Joanne Wai Yee Chung. “Tele-Dietetics with Food Images as Dietary Intake Record in Nutrition Assessment”. In: *Telemedicine and e-Health* 16.6 (July 2010), pp. 691–698. DOI: 10.1089/tmj.2009.0174. URL: https://doi.org/10.1089/tmj.2009.0174.

[20] Marie Clark and Sarah E Hampson. “Implementing a psychological intervention to improve lifestyle self-management in patients with Type 2 diabetes”. In: *Patient Education and Counseling* 42.3 (Mar. 2001), pp. 247–256. DOI: 10.1016/s0738-3991(00)00128-2. URL: https://doi.org/10.1016/s0738-3991(00)00128-2.

[21] Maria Costabile and Paolo Buono. “Principles for Human-Centred Design of IR Interfaces”. In: Jan. 2013, pp. 28–47. ISBN: 978-3-642-36414-3. DOI: 10.1007/978-3-642-36415-0_2.

[22] Goodarz Danaei et al. “National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2·7 million participants”. In: *The Lancet* 378.9785 (July 2011), pp. 31–40. DOI: 10.1016/s0140-6736(11)60679-x. URL: https://doi.org/10.1016/s0140-6736(11)60679-x.

[23] Jaime A. Davidson. “The Increasing Role of Primary Care Physicians in Caring for Patients With Type 2 Diabetes Mellitus”. In: *Mayo Clinic Proceedings* 85.12 (Dec. 2010), S3–S4. DOI: 10.4065/mcp.2010.0466. URL: https://doi.org/10.4065/mcp.2010.0466.

[24] Edward L. Deci. *Intrinsic Motivation*. Springer US, 1975. DOI: 10.1007/978-1-4613-4446-9. URL: https://doi.org/10.1007/978-1-4613-4446-9.

[25] Edward L. Deci and Richard M. Ryan. *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer US, 1985. DOI: 10.1007/978-1-4899-2271-7. URL: https://doi.org/10.1007/978-1-4899-2271-7.

[26] Leslie Dickinson. “Autonomy and motivation a literature review”. In: *System* 23.2 (May 1995), pp. 165–174. DOI: 10.1016/0346-251x(95)00005-5. URL: https://doi.org/10.1016/0346-251x(95)00005-5.

[27] E. Dziuk et al. “[Lung scanning performed with the use of microspheres labeled with radioactive isotopes in the diagnosis and evaluation of therapeutic results in malignant neoplasms]”. In: *Pol Arch Med Wewn* 53.3 (Mar. 1975), pp. 241–253.

[28] Kathaleen Briggs Early and Kathleen Stanley. “Position of the Academy of Nutrition and Dietetics: The Role of Medical Nutrition Therapy and Registered Dietitian Nutritionists in the Prevention and Treatment of Prediabetes and Type 2 Diabetes”. In: *Journal of the Academy of Nutrition and Dietetics* 118.2 (Feb. 2018), pp. 343–353. DOI: 10.1016/j.jand.2017.11.021. URL: https://doi.org/10.1016/j.jand.2017.11.021.

[29] Shihab Elbagir and Jing Yang. “Sentiment Analysis on Twitter with Python’s Natural Language Toolkit and VADER Sentiment Analyzer”. In: *IAENG Transactions on Engineering Sciences*. WORLD SCIENTIFIC, Jan. 2020. DOI: 10.1142/9789811215094_0005. URL: https://doi.org/10.1142/9789811215094_0005.

[30] Miguel Elvir et al. “Remembering a Conversation – A Conversational Memory Architecture for Embodied Conversational Agents”. In: *Journal of Intelligent Systems* 26.1 (Jan. 2017), pp. 1–21. DOI: 10.1515/jisys-2015-0094. URL: https://doi.org/10.1515/jisys-2015-0094.

[31] G Eysenbach. “What is e-health?” In: *Journal of Medical Internet Research* 3.2 (June 2001), e20. DOI: 10.2196/jmir.3.2.e20. URL: https://doi.org/10.2196/jmir.3.2.e20.

[32] H. J. Eysenck and C. D. Frith. “Reminiscence and Motivation”. In: *Reminiscence, Motivation, and Personality*. Springer US, 1977, pp. 131–164. DOI: 10.1007/978-1-4684-2244-3_5. URL: https://doi.org/10.1007/978-1-4684-2244-3_5.

[33] R.A. Fisher. *Statistical methods for research workers*. Edinburgh Oliver & Boyd, 1925.

[34] Siska Fitrianie et al. “What are We Measuring Anyway?” In: *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. ACM, July 2019. DOI: 10.1145/3308532.3329421. URL: https://doi.org/10.1145/3308532.3329421.

[35] Luigi Fontana and Frank B. Hu. “Optimal body weight for health and longevity: bridging basic, clinical, and population research”. In: *Aging Cell* 13.3 (Mar. 2014), pp. 391–400. DOI: 10.1111/acel.12207. URL: https://doi.org/10.1111/acel.12207.

[36] P. W. Franks. “Diabetes Family History: A Metabolic Storm You Should Not Sit Out”. In: *Diabetes* 59.11 (Oct. 2010), pp. 2732–2734. DOI: 10.2337/db10-0768. URL: https://doi.org/10.2337/db10-0768.

[37] Joyce P. Yi-Frazier et al. “The Impact of Positive Psychology on Diabetes Outcomes: A Review”. In: *Psychology* 03.12 (2012), pp. 1116–1124. DOI: 10.4236/psych.2012.312a165. URL: https://doi.org/10.4236/psych.2012.312a165.

[38] Michelle Goddard. “The EU General Data Protection Regulation (GDPR): European Regulation that has a Global Impact”. In: *International Journal of Market Research* 59.6 (Nov. 2017), pp. 703–705. DOI: 10.2501/ijmr-2017-050. URL: https://doi.org/10.2501/ijmr-2017-050.

[39] Lindsay T. Graham, Samuel D. Gosling, and Christopher K. Travis. “The Psychology of Home Environments”. In: *Perspectives on Psychological Science* 10.3 (May 2015), pp. 346–356. DOI: 10.1177/1745691615576761. URL: https://doi.org/10.1177/1745691615576761.

[40] Miguel Grinberg. *Flask Web Development: Developing Web Applications with Python*. 1st. O’Reilly Media, Inc., 2014. ISBN: 1449372627.

[41] Jan-Gerrit Harms et al. “Approaches for Dialog Management in Conversational Agents”. In: *IEEE Internet Computing* 23.2 (Mar. 2019), pp. 13–22. DOI: 10.1109/mic.2018.2881519. URL: https://doi.org/10.1109/mic.2018.2881519.

[42] Spencer E Harpe. “How to analyze Likert and other rating scale data”. In: *Currents in Pharmacy Teaching and Learning* 7 (2015), pp. 836–850.

[43] Wan Ho and Kerstin Dautenhahn. “Towards a Narrative Mind: The Creation of Coherent Life Stories for Believable Virtual Agents”. In: vol. 5208. Sept. 2008, pp. 59–72. ISBN: 978-3-540-85482-1. DOI: 10.1007/978-3-540-85483-8_6.

[44] Ann C. Hurley and Carole A. Shea. “Self-Efficacy: Strategy for Enhancing Diabetes Self-Care”. In: *The Diabetes Educator* 18.2 (Apr. 1992), pp. 146–150. DOI: 10.1177/014572179201800208. URL: https://doi.org/10.1177/014572179201800208.

[45] Shafquat Hussain, Omid Sianaki, and Nedal Ababneh. “A Survey on Conversational Agents/Chatbots Classification and Design Techniques”. In: Mar. 2019, pp. 946–956. ISBN: 978-3-319-98284-7. DOI: 10.1007/978-3-030-15035-8_93.

[46] BiranchiNarayan Jena, Sanjay Kalra, and Rajiv Yeravdekar. "Emotional and psychological needs of people with diabetes". In: *Indian Journal of Endocrinology and Metabolism* 22.5 (2018), p. 696. DOI: 10.4103/ijem.ijem_579_17. URL: https://doi.org/10.4103/ijem.ijem_579_17.

[47] Slava Kalyuga and Anne-Marie Singh. "Rethinking the Boundaries of Cognitive Load Theory in Complex Learning". In: *Educational Psychology Review* 28.4 (Dec. 2015), pp. 831–852. DOI: 10.1007/s10648-015-9352-0. URL: https://doi.org/10.1007/s10648-015-9352-0.

[48] Zerrin Kasap and Nadia Magnenat-Thalmann. "Building long-term relationships with virtual and robotic characters: the role of remembering". In: *The Visual Computer* 28.1 (Sept. 2011), pp. 87–97. DOI: 10.1007/s00371-011-0630-7. URL: https://doi.org/10.1007/s00371-011-0630-7.

[49] Zerrin Kasap and Nadia Magnenat-Thalmann. "Towards episodic memory-based long-term affective interaction with a human-like robot". In: *19th International Symposium in Robot and Human Interactive Communication*. IEEE, Sept. 2010. DOI: 10.1109/roman.2010.5598644. URL: https://doi.org/10.1109/roman.2010.5598644.

[50] Hae-Young Kim. "Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test". In: *Restorative Dentistry Endodontics* 42 (May 2017), pp. 152–155. DOI: 10.5395/rde.2017.42.2.152.

[51] Yonghee Kim et al. "Acquisition and Use of Long-Term Memory for Personalized Dialog Systems". In: Sept. 2014, pp. 78–87. DOI: 10.1007/978-3-319-15557-9_8.

[52] Anna Klimas. "A Goal-Setting Logbook as an Instrument Fostering Learner Autonomy". In: *Second Language Learning and Teaching*. Springer International Publishing, Dec. 2016, pp. 21–33. DOI: 10.1007/978-3-319-07764-2_2. URL: https://doi.org/10.1007/978-3-319-07764-2_2.

[53] Iuliia Kotseruba and John K. Tsotsos. "40 years of cognitive architectures: core cognitive abilities and practical applications". In: *Artificial Intelligence Review* 53.1 (July 2018), pp. 17–94. DOI: 10.1007/s10462-018-9646-y. URL: https://doi.org/10.1007/s10462-018-9646-y.

[54] John E. Laird, Christian Lebiere, and Paul S. Rosenbloom. "A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics". In: *AI Magazine* 38.4 (Dec. 2017), pp. 13–26. DOI: 10.1609/aimag.v38i4.2744. URL: https://doi.org/10.1609/aimag.v38i4.2744.

[55] John E. Laird, Allen Newell, and Paul S. Rosenbloom. "SOAR: An architecture for general intelligence". In: *Artificial Intelligence* 33.1 (Sept. 1987), pp. 1–64. DOI: 10.1016/0004-3702(87)90050-6. URL: https://doi.org/10.1016/0004-3702(87)90050-6.

[56] François Lemaire. "Do All Types of Human Research Need Ethics Committee Approval?" In: *American Journal of Respiratory and Critical Care Medicine* 174.4 (Aug. 2006), pp. 363–364. DOI: 10.1164/rccm.2603003. URL: https://doi.org/10.1164/rccm.2603003.

[57] J. M. Levenson and J. D. Sweatt. "Memory". In: *Cellular and Molecular Life Sciences* 63.9 (Apr. 2006), pp. 1009–1016. DOI: 10.1007/s00018-006-6026-6. URL: https://doi.org/10.1007/s00018-006-6026-6.

[58] Sylvia H Ley et al. "Prevention and management of type 2 diabetes: dietary components and nutritional strategies". In: *The Lancet* 383.9933 (June 2014), pp. 1999–2007. DOI: 10.1016/s0140-6736(14)60613-9. URL: https://doi.org/10.1016/s0140-6736(14)60613-9.

[59] Mei Yii Lim. "Memory Models for Intelligent Social Companions". In: *Studies in Computational Intelligence* 396 (Jan. 2012). DOI: 10.1007/978-3-642-25691-2_10.

[60] Edwin A. Locke. "Motivation through conscious goal setting". In: *Applied and Preventive Psychology* 5.2 (Mar. 1996), pp. 117–124. DOI: 10.1016/s0962-1849(96)80005-9. URL: https://doi.org/10.1016/s0962-1849(96)80005-9.

[61] Edwin A. Locke and Gary P. Latham. “Building a practically useful theory of goal setting and task motivation: A 35-year odyssey.” In: *American Psychologist* 57.9 (Sept. 2002), pp. 705–717. DOI: 10.1037/0003-066x.57.9.705. URL: https://doi.org/10.1037/0003-066x.57.9.705.

[62] Effie Maclellan. “Academic achievement”. In: *Active Learning in Higher Education* 6.3 (Nov. 2005), pp. 194–206. DOI: 10.1177/1469787405057750. URL: https://doi.org/10.1177/1469787405057750.

[63] Adam M. Mastroianni et al. “Do conversations end when people want them to?” In: *Proceedings of the National Academy of Sciences* 118.10 (Mar. 2021), e2011809118. DOI: 10.1073/pnas.2011809118. URL: https://doi.org/10.1073/pnas.2011809118.

[64] Tara Matthews, Tejinder Judge, and Steve Whittaker. “How do designers and user experience professionals actually perceive and use personas?” In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI ’12*. ACM Press, 2012. DOI: 10.1145/2207676.2208573. URL: https://doi.org/10.1145/2207676.2208573.

[65] Michael F. McTear. “Spoken dialogue technology”. In: *ACM Computing Surveys* 34.1 (Mar. 2002), pp. 90–169. DOI: 10.1145/505282.505285. URL: https://doi.org/10.1145/505282.505285.

[66] Adam S. Miner et al. “Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health”. In: *JAMA Internal Medicine* 176.5 (May 2016), p. 619. DOI: 10.1001/jamainternmed.2016.0400. URL: https://doi.org/10.1001/jamainternmed.2016.0400.

[67] Prabhakar Mishra et al. “Descriptive Statistics and Normality Tests for Statistical Data”. In: *Annals of Cardiac Anaesthesia* 22 (Jan. 2019), pp. 67–72. DOI: 10.4103/aca.ACA_157_18.

[68] JEANNE MOORE. “PLACING HOME IN CONTEXT”. In: *Journal of Environmental Psychology* 20.3 (Sept. 2000), pp. 207–217. DOI: 10.1006/jevp.2000.0178. URL: https://doi.org/10.1006/jevp.2000.0178.

[69] Samer Al Moubayed et al. “Furhat: A Back-Projected Human-Like Robot Head for Multiparty Human-Machine Interaction”. In: *Cognitive Behavioural Systems*. Springer Berlin Heidelberg, 2012, pp. 114–130. DOI: 10.1007/978-3-642-34584-5_9. URL: https://doi.org/10.1007/978-3-642-34584-5_9.

[70] Prakash M Nadkarni, Lucila Ohno-Machado, and Wendy W Chapman. “Natural language processing: an introduction”. In: *Journal of the American Medical Informatics Association* 18.5 (Sept. 2011), pp. 544–551. DOI: 10.1136/amiajnl-2011-000464. URL: https://doi.org/10.1136/amiajnl-2011-000464.

[71] B. Nayak and A. Hazra. “How to choose the right statistical test?” In: *Indian Journal of Ophthalmology* 59 (2011), pp. 85–86.

[72] Mark Neerincx and Jasper Lindenberg. “Situated cognitive engineering for complex task environments”. In: *Naturalistic Decision Making and Macrocognition* (Jan. 2012), pp. 373–389.

[73] Mark A. Neerincx et al. “Socio-Cognitive Engineering of a Robotic Partner for Child’s Diabetes Self-Management”. In: *Frontiers in Robotics and AI* 6 (Nov. 2019). DOI: 10.3389/frobt.2019.00118. URL: https://doi.org/10.3389/frobt.2019.00118.

[74] Efrat Neter and Esther Brainin. “eHealth Literacy: Extending the Digital Divide to the Realm of Health Information”. In: *Journal of Medical Internet Research* 14.1 (Jan. 2012), e19. DOI: 10.2196/jmir.1619. URL: https://doi.org/10.2196/jmir.1619.

[75] Linda Neuhauser and Gary L. Kreps. “Rethinking Communication in the E-health Era”. In: *Journal of Health Psychology* 8.1 (Jan. 2003), pp. 7–23. DOI: 10.1177/1359105303008001426. URL: https://doi.org/10.1177/1359105303008001426.

[76] Johan Y. Y. Ng et al. “Self-Determination Theory Applied to Health Contexts”. In: *Perspectives on Psychological Science* 7.4 (June 2012), pp. 325–340. DOI: 10.1177/1745691612447309. URL: https://doi.org/10.1177/1745691612447309.

[77] Geoff Norman. “Likert scales, levels of measurement and the “laws” of statistics”. In: *Advances in Health Sciences Education* 15.5 (Feb. 2010), pp. 625–632. DOI: 10.1007/s10459-010-9222-y. URL: https://doi.org/10.1007/s10459-010-9222-y.

[78] Andrew M. Nuxoll and John E. Laird. “Enhancing intelligent agents with episodic memory”. In: *Cognitive Systems Research* 17-18 (July 2012), pp. 34–48. DOI: 10.1016/j.cogsys.2011.10.002. URL: https://doi.org/10.1016/j.cogsys.2011.10.002.

[79] Bjørg Oftedal, Edvin Bru, and Bjørg Karlsen. “Motivation for diet and exercise management among adults with type 2 diabetes”. In: *Scandinavian Journal of Caring Sciences* 25.4 (Mar. 2011), pp. 735–744. DOI: 10.1111/j.1471-6712.2011.00884.x. URL: https://doi.org/10.1111/j.1471-6712.2011.00884.x.

[80] Harri Oinas-Kukkonen. “A foundation for the study of behavior change support systems”. In: *Personal and Ubiquitous Computing* 17.6 (July 2012), pp. 1223–1235. DOI: 10.1007/s00779-012-0591-5. URL: https://doi.org/10.1007/s00779-012-0591-5.

[81] Stefan Olafsson, Teresa O’Leary, and Timothy Bickmore. “Coerced Change-talk with Conversational Agents Promotes Confidence in Behavior Change”. In: *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM, May 2019. DOI: 10.1145/3329189.3329202. URL: https://doi.org/10.1145/3329189.3329202.

[82] Ben van Ommen et al. “From Diabetes Care to Diabetes Cure—The Integration of Systems Biology, eHealth, and Behavioral Change”. In: *Frontiers in Endocrinology* 8 (Jan. 2018). DOI: 10.3389/fendo.2017.00381. URL: https://doi.org/10.3389/fendo.2017.00381.

[83] Nina Østerås et al. “A randomised comparison of a four- and a five-point scale version of the Norwegian Function Assessment Scale”. In: *Health and Quality of Life Outcomes* 6.1 (2008), p. 14. DOI: 10.1186/1477-7525-6-14. URL: https://doi.org/10.1186/1477-7525-6-14.

[84] Claudia Pagliari et al. “What Is eHealth (4): A Scoping Exercise to Map the Field”. In: *Journal of Medical Internet Research* 7.1 (Mar. 2005), e9. DOI: 10.2196/jmir.7.1.e9. URL: https://doi.org/10.2196/jmir.7.1.e9.

[85] Stefan Palan and Christian Schitter. “Prolific.ac—A subject pool for online experiments”. In: *Journal of Behavioral and Experimental Finance* 17 (Mar. 2018), pp. 22–27. DOI: 10.1016/j.jbef.2017.12.004. URL: https://doi.org/10.1016/j.jbef.2017.12.004.

[86] Heather Patrick and Geoffrey C Williams. “Self-determination theory: its application to health behavior and complementarity with motivational interviewing”. In: *International Journal of Behavioral Nutrition and Physical Activity* 9.1 (2012), p. 18. DOI: 10.1186/1479-5868-9-18. URL: https://doi.org/10.1186/1479-5868-9-18.

[87] James O. Prochaska. “Decision Making in the Transtheoretical Model of Behavior Change”. In: *Medical Decision Making* 28.6 (Nov. 2008), pp. 845–849. DOI: 10.1177/0272989x08327068. URL: https://doi.org/10.1177/0272989x08327068.

[88] Susannah Radstone and Bill Schwarz. *Memory histories, theories, debates*. Fordham University Press, 2010.

[89] Anand Rajaraman and Jeffrey David Ullman. “Data Mining”. In: *Mining of Massive Datasets*. Cambridge University Press, pp. 1–17. DOI: 10.1017/cbo9781139058452.002. URL: https://doi.org/10.1017/cbo9781139058452.002.

[90] Cynthia Ripsin, Helen Kang, and Randall Urban. “Management of Blood Glucose in Type 2 Diabetes Mellitus”. In: *American family physician* 79 (Feb. 2009), pp. 29–36.

[91] Stephen Rollnick and William R. Miller. “What is Motivational Interviewing?” In: *Behavioural and Cognitive Psychotherapy* 23.4 (Oct. 1995), pp. 325–334. DOI: 10.1017/s135246580001643x. URL: https://doi.org/10.1017/s135246580001643x.

[92] Stuart Rose et al. “Automatic Keyword Extraction from Individual Documents”. In: *Text Mining*. John Wiley & Sons, Ltd, Mar. 2010, pp. 1–20. DOI: 10.1002/9780470689646.ch1. URL: https://doi.org/10.1002/9780470689646.ch1.

[93] Richard M. Ryan and Edward L. Deci. "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." In: *American Psychologist* 55.1 (2000), pp. 68–78. DOI: 10.1037/0003-066x.55.1.68. URL: https://doi.org/10.1037/0003-066x.55.1.68.

[94] Michael Scheier and Charles Carver. "Goals and confidence as self-regulatory elements underlying health and illness behavior". In: *The self-regulation of health and illness behaviour* (Jan. 2003).

[95] Bart Schreuder Goedheijt. *Recalling shared memories in an embodied conversational agent : personalized robot support for children with diabetes in the PAL project*. 2017. URL: http://essay.utwente.nl/80062/.

[96] Steven A. Schroeder. "We Can Do Better — Improving the Health of the American People". In: *New England Journal of Medicine* 357.12 (Sept. 2007), pp. 1221–1228. DOI: 10.1056/nejmsa073350. URL: https://doi.org/10.1056/nejmsa073350.

[97] Dale H. Schunk and Marsha W. Lilly. "Sex Differences in Self-Efficacy and Attributions: Influence of Performance Feedback". In: *The Journal of Early Adolescence* 4.3 (Sept. 1984), pp. 203–213. DOI: 10.1177/0272431684043004. URL: https://doi.org/10.1177/0272431684043004.

[98] Phoebe Sengers. "Schizophrenia and Narrative in Artificial Agents". In: *Leonardo* 35.4 (Aug. 2002), pp. 427–431. DOI: 10.1162/002409402760181240. URL: https://doi.org/10.1162/002409402760181240.

[99] Carole Sergeant and Pamela A. Dyson. "Diabetes and Diet: A Patient and Dietitian's Perspective". In: *Diabetes Therapy* 9.5 (Sept. 2018), pp. 1733–1739. DOI: 10.1007/s13300-018-0500-7. URL: https://doi.org/10.1007/s13300-018-0500-7.

[100] Mike Sharples. "Socio-cognitive Engineering". In: *Encyclopedia of Human Computer Interaction* (Jan. 2006). DOI: 10.4018/978-1-59140-562-7.ch080.

[101] Claudia Sinoo et al. "Friendship with a robot: Children's perception of similarity between a robot's physical and virtual embodiment that supports diabetes self-management". In: *Patient Education and Counseling* 101.7 (July 2018), pp. 1248–1255. DOI: 10.1016/j.pec.2018.02.008. URL: https://doi.org/10.1016/j.pec.2018.02.008.

[102] George Siopis, Stephen Colagiuri, and Margaret Allman-Farinelli. "Dietitians' experiences and perspectives regarding access to and delivery of dietetic services for people with type 2 diabetes mellitus". In: *Heliyon* 6.2 (Feb. 2020), e03344. DOI: 10.1016/j.heliyon.2020.e03344. URL: https://doi.org/10.1016/j.heliyon.2020.e03344.

[103] P. Sohl and H.A. Bassford. "Codes of medical ethics: Traditional foundations and contemporary practice". In: *Social Science & Medicine* 22.11 (1986), pp. 1175–1179. DOI: 10.1016/0277-9536(86)90184-x. URL: https://doi.org/10.1016/0277-9536(86)90184-x.

[104] Carlos M. P. Sousa, Filipe Coelho, and Encarna Guillamon-Saorin. "Personal Values, Autonomy, and Self-efficacy: Evidence from frontline service employees". In: *International Journal of Selection and Assessment* 20.2 (May 2012), pp. 159–170. DOI: 10.1111/j.1468-2389.2012.00589.x. URL: https://doi.org/10.1111/j.1468-2389.2012.00589.x.

[105] F. J. M. Spikmans et al. "Why do diabetic patients not attend appointments with their dietitian?" In: *Journal of Human Nutrition and Dietetics* 16.3 (June 2003), pp. 151–158. DOI: 10.1046/j.1365-277x.2003.00435.x. URL: https://doi.org/10.1046/j.1365-277x.2003.00435.x.

[106] Sandeep Vijan. "Type 2 Diabetes". In: *Annals of Internal Medicine* 152.5 (Mar. 2010), ITC3–1. DOI: 10.7326/0003-4819-152-5-201003020-01003. URL: https://doi.org/10.7326/0003-4819-152-5-201003020-01003.

[107] Allison Vorderstrasse et al. "Social Support for Diabetes Self-Management via eHealth Interventions". In: *Current Diabetes Reports* 16.7 (May 2016). DOI: 10.1007/s11892-016-0756-0. URL: https://doi.org/10.1007/s11892-016-0756-0.

[108] Astrid Weiss and Christoph Bartneck. “Meta analysis of the usage of the Godspeed Questionnaire Series”. In: *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Aug. 2015. DOI: 10.1109/roman.2015.7333568. URL: https://doi.org/10.1109/roman.2015.7333568.

[109] Yujun Wen, Hui Yuan, and Pengzhou Zhang. “Research on keyword extraction based on Word2Vec weighted TextRank”. In: *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*. IEEE, Oct. 2016. DOI: 10.1109/compcomm.2016.7925072. URL: https://doi.org/10.1109/compcomm.2016.7925072.

[110] B Zhou et al. “Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4·4 million participants”. In: *The Lancet* 387.10027 (Apr. 2016), pp. 1513–1530. DOI: 10.1016/s0140-6736(16)00618-8. URL: https://doi.org/10.1016/s0140-6736(16)00618-8.

List of Figures

List of Tables