Developing an Empathetic Chatbot Based on Conversational AI to Support Post-Stroke Anxiety Patients

Vanessa Neeff

Master of Science
Design Informatics
School of Informatics
University of Edinburgh
2021

Abstract

Post-stroke anxiety affects about one in four stroke survivors. Despite the need for psychological care after hospital discharge, the access to treatment is insufficient. Not only do conventional face-to-face therapies require patients to be mobile, but there is also a shortage of clinical professionals. To facilitate access to immediate psychological care, a chatbot was developed to guide patients through their therapy. First, user research interviews were conducted to explore how stroke survivors use technology and which expectations they have towards a mental health chatbot. Based on elaborated design considerations, a chatbot prototype guiding patients through their Cognitive Behavioural Therapy exercises was implemented. Subsequently, the chatbot was evaluated by two stroke clinicians and two stroke survivors. It was found that the prototype was perceived as efficient and helpful for patients with mild anxiety symptoms. Crucial points for improvement are the robustness and personalisation of chatbot characteristics.

Acknowledgements

I would like to express my gratitude towards my supervisor Dr. Maria Wolters for her generous help and constructive feedback throughout the year and especially during my final semester. Moreover, I would like to thank Dr. Yvonne Chun for her assistance and support without whom the project would not have come about. Many thanks also to my personal tutor Rabia Yasa Kostas who encouraged me throughout the project and provided invaluable feedback. I would also like to thank my flatmates, friends and family who have supported me all along the way and shown me unconditional compassion and love. Thank you for your selfless support and thank you for believing in me.

Table of Contents

Chapter 1

Introduction

1.1 Motivation

With 6 million fatalities annually, stroke remains one of the leading causes for death, physical and cognitive impairment [47]. Alone in the UK, there are more than 100,000 new stroke cases per year [66], with case numbers likely to increase by 60% annually [42]. Besides functional limitations in fine or gross motor skills such as coordination or walking [27], post-stroke recovery is often accompanied by social stigma, cognitive impairments and behavioural changes [35]. In particular, there is a high prevalence of mood disorders such as post-stroke depression and anxiety among stroke survivors [11] which can significantly reduce patients' quality of life after their discharge [25].

About 25% of stroke survivors have experienced significant levels of anxiety after stroke [14]. There is a clear demand for facilitated mental healthcare access which to date has not yet been met in the UK [85]. Moreover, conventional psychotherapy requires patients to meet their therapist in person at a clinic, which is often not feasible due to physical or psychological conditions [17]. Finally, the anxiety therapies currently offered are not specifically tailored to post-stroke anxiety patients and therefore often fail to achieve the intended treatment outcomes [19]. To overcome the above mentioned obstacles, Chun [18] has developed a telemedical treatment protocol aimed at delivering therapy specifically for stroke survivors with anxiety (TASK-CBT). The psychological intervention enables patients to manage therapy exercises by themselves using a website, followed by weekly telemedical therapy sessions with therapists. To provide further virtual assistance and empathetic feedback throughout the therapy, this project aims to develop a chatbot that guides patients through therapy exercises. Additionally, the development of a chatbot will facilitate current research activities aimed at validating the concept of a telemedical mental health intervention.

1.2 Objectives

The aim of this project is to investigate the key requirements of a conversational agent through user interviews with stroke survivors and to translate these findings into an interactive chatbot prototype. The prototype is expected to provide empathetic, useful feedback to users during their virtual therapy, thereby promoting user engagement with the chatbot. The TASK-CBT protocol developed by Chun et al. [18] lays the foundation for the chatbot dialogue to comply with clinical guidelines. To accomplish these objectives, patients with post-stroke anxiety will be interviewed to determine their attitudes towards technology, their relationship with digital applications and their expectations of a therapeutic chatbot. Based on identified user needs and previous TASK-CBT exercise datasets, design suggestions for the chatbot are made. These requirements finally inform the design of the chatbot conversation flow and are tested in subsequent user evaluations.

1.3 Contributions

The contributions outlined in this dissertation are specified as follows:

- **Chapter 2** summarises relevant literature findings and reviews previous work addressing chatbots for PSA.
- **Chapter 3** describes the user research activities taken to investigate stroke survivor's attitude towards technology and chatbots and the limitations of the study.
- **Chapter 4** outlines the different development phases of the chatbot implementation and illustrates the dialogue design.
- **Chapter 5** describes the evaluation of the chatbot conducted by stroke survivors and stroke physicians.
- **Chapter 6** discussed evaluation results, ethical and conceptual considerations and outlines implications for future research.
- **Chapter 7** summarises the project work, relevant findings and opportunities for improvement.

Chapter 2

Background

2.1 Quality of Life Post-Stroke

During a stroke, the central nervous system suffers an infarct caused by cell death due to an abnormal decrease of blood flow [68]. Transient Ischemic Attacks (TIA) on the other side can display the same causes and symptoms as a stroke but only last for a short period of time [26]. Given the heterogeneity of stroke and TIA symptoms, the severity of impairments differ greatly [82]. While some patients recover rapidly within minutes, others might never fully recover and suffer from long-lasting physical and cognitive impairments [82].

Regardless of the severity of the stroke, this event can lead to significant lifestyle changes that often also have a negative impact on the future prospects of those affected [5]. Stroke researchers have observed a decline in quality of life (QoL) in individuals post-stroke based on patients' own perceptions of their well-being [60]. In total, three domains are taken into account in health-related QoL: the perception of social, psychological and physical well-being [60]. The deterioration in QoL may be due to the fact that rehabilitation interventions focus mainly on physical recovery and less on psychological support to overcome the burdens of stroke-related lifestyle changes [21]. Especially after hospital discharge, the lack of ongoing out-patient services contribute to feelings of isolation and difficulties adapting to identity changes [21].

2.2 Post-Stroke Anxiety

One key factor responsible for the decline of QoL is post-stroke depression and anxiety [4]. Although there has been plenty of research regarding post-stroke depression in past decades, studies investigating post-stroke anxiety (PSA) on the other hand are still very limited [70]. Anxiety disorders are defined by persistent, excessive and disproportionate anxiety that interferes with patients' everyday lives [28].

Patients with PSA suffer primarily from excessive fear of having another stroke [17]. Research addressing PSA has shown that the chronic illness can be mainly attributed to two anxiety disorder subtypes: Generalised Anxiety Disorder (GAD) and Agoraphobia, a condition classified as a phobic disorder [19]. The accurate identification and classification of symptoms is important for clinical diagnosis, as GAD and Agoraphobia warrant different treatment approaches [19]. In particular, phobic disorders were found to be more dominant among stroke survivors (10%) than GAD (7%) [19].

Agoraphobia is characterised by the excessive fear of going to or being in specific places such as public transportation or enclosed indoor venues [28]. For stroke survivors, these fears usually stem from the belief that they will not be able to access help in the event of a clinical emergency, such as a recurrent stroke. If left untreated, Agoraphobia can interfere with everyday tasks such as grocery shopping and cause patients to stay at home permanently [28].

GAD on the other hand is defined as a chronic illness where individuals suffer from various types of anxious thoughts that are not limited to specific events, objects or outcomes [28]. Other symptoms include constant uneasiness and fatigue. People suffering from GAD often find it hard to manage their worries and distorted thoughts, whereas the extent of fear and worries are out-of-proportion to the actual risk of the incident. Common anxious thoughts often relate to everyday tasks, health conditions and the future quality of life [28].

2.3 Treatment of Post-Stroke Anxiety

2.3.1 Cognitive Behavioural Therapy

In general practice, Cognitive Behavioural Therapy (CBT) has been widely accepted by clinical guidelines as an evidence-based gold standard for treating various anxiety disorders [57]. Patients are equipped with the necessary skills to manage their anxiety in different scenarios and adapt to stress situations accordingly [31].

Especially for panic disorders associated with Agoraphobia, clinical trials conducted by Beurs et al. [23] have shown that CBT leads to longer-lasting and more efficient outcomes than pharmacotherapies. CBT is based on the theory that mood disorders are maintained by cognitive distortions and that through therapy, healthier behaviours can be acquired to cope with anxious thoughts [8]. Maladaptive thoughts can range from general beliefs and perceptions about life [8] to more specific thoughts about the patient's future well-being. Moreover, maladaptive thoughts influence people's emotions and actions when exposed to their fears [36]. During therapy, these anxious beliefs are identified and challenged by patients to reflect on their thought patterns [37]. During conventional face-to-face CBT, therapists guide their patients through daily or weekly exercises to help them recognise their cognitive distortions and adapt to healthier coping strategies [31].

2.3.2 Barriers and Accessibility

The lack of clinical resources is not the only burden towards accessible psychological care for stroke survivors in the UK [33]. Additionally, conventional CBT treatments require PSA patients to commute and meet their therapists on a weekly basis for a period of up to several months [17]. However, this is often not feasible due to the limited physical mobility of stroke survivors or the agoraphobic fear of commuting in general. Moreover, the prevalence of social anxiety among stroke survivors are further hindrances to psychological care. Due to these persistent barriers to mental health care access, a telemedical CBT treatment protocol specifically for stroke survivors was developed by Chun et al. [18].

2.3.3 A Telemedical Intervention: TASK-CBT

The advantages of the 'Treating Anxiety after Stroke' (TASK) CBT intervention developed by Chun et al. [18] can be briefly summarised as following: First, the CBT exercises are targeted specifically at the most prevalent anxiety disorder subtype among PSA patients, Agoraphobia. Therefore, treatment exercises incorporate exposure prevention elements to gradually habituate patients to their feared situations. Second, treatment exercises do not require patients to commute on a weekly basis. All tasks can be done remotely via a dedicated TASK website. Weekly therapy sessions via telephone accompany the web-based intervention to provide further assistance and ensure therapeutic alliance between the patient and the clinician [18]. Notably, a randomised clinical trial including 40 stroke and TIA patients has already indicated that TASK-CBT is superior to relaxation techniques [18] which are usually recommended to treat

GAD in non-stroke cohorts. There are further indications that web-based interventions might even be equally effective in treating anxiety disorders as conventional CBT treatments among non-stroke populations [6]. The web-based TASK-CBT incorporates different media including psychoeducational images and videos, patient stories and further links to knowledge bases about post-stroke care. The centerpiece of the website is made up of 6 different therapy exercises which can be completed by the patients on their own. The following exercises can be carried out and submitted online on a daily or weekly basis [18]:

1. Identification of anxious moments: Hereby, patients reflect about recent anxious situations and break their moments down into thoughts, feelings and actions.
2. Analysis of individuals' thinking patterns: Thinking styles such as catastrophising or black-and-white thinking are identified.
3. Challenging cognitive distortions: Negative thoughts are observed and rationally questioned by patients.
4. Exposure to feared situations: Patients are gradually exposed to feared situations or places without reacting with unhelpful coping mechanisms.
5. Active relaxation: On a daily basis, patients are reminded to actively relax for a dedicated period of time
6. Engagement in enjoyable activities: Patients are asked to engage in hobbies or leisure activities once a day.

2.4 Chatbots in Mental Health Care

2.4.1 Advantages

Automated conversational agents have been studied in recent years to investigate psychological treatment efficiency [58, 62]. These conversational agents, also called chatbots [41] are software programs that can engage in conversation with users and provide informative feedback to a variety of user input [56]. Artificial Intelligence (AI), Natural Language Processing (NLP) and rule-based decision trees provide the foundation for the goal-oriented human-computer interaction between chatbot and users [3].

The efficiency of psychological counseling through chatbots has been investigated in use cases such as drug abuse, phobic disorders and depression [79]. Study outcomes have shown that chatbots can successfully reduce symptoms through online mediated CBT [59]. In comparison to reading a psychoeducational paperback only, a mobile chatbot application developed by Oh [58] produced more promising results in treating panic disorder symptoms.

One major advantage of mental health chatbots is the facilitated accessibility of psychological care. To combat the shortage of trained clinicians [79], chatbot applications can even support patients long after their formal clinical treatment. Moreover, studies have found that webform-based CBT models face issues regarding poor patient adherence [69]. Therefore, the instant feedback and imitation of human conversation provided by a chatbot might cover for the lack of emotional support in static web-based treatments. Moreover, in comparison to mental health clinicians, chatbots were perceived as more trustworthy in a research study involving military members with mental health issues [52]. This might be due to societal stigma [52] and the fact that chatbots are less likely to project human bias regarding age, gender or ethnicity during virtual treatment sessions [62]. Additionally, chatbots offer the advantage of anonymity [52].

2.4.2 Features and Personality

In order to design a chatbot that specifically addresses the needs of PSA patients, previous research conducted by Lohse [51] aimed to identify and test such user requirements. Her systematic literature review [51] explored how stroke sufferers use technology in their everyday lives and how this translates into chatbot requirements. Subsequently, semi-structured interviews were conducted with non-stroke participants aged between 26 and 33 years to explore their perceptions on technical applications and mental health chatbots [51]. The following section describes Lohse's [51] principle research findings regarding chatbot features and how these relate to other study outcomes.

1. According to Lohse [51, p.20], an important prerequisite for sufficient acceptance of chatbots is the close alignment with clinical guidelines and transparency about user benefits. This is in line with other reviews on mental health chatbots that concluded that the absence of a clearly stated purpose is associated with lower chatbot acceptability [2]. Additionally, other studies have also found that

clinical relevancy is crucial to ensure that users disclose their personal information in a trustworthy environment [16]. Notably, a bonding relationship based on trust between the chatbot and patients has been reported as crucial for ongoing engagement [45].

2. Lohse [51] has further suggested that chatbot instructions should be 'straight-forward to understand' [51, p.20]. In addition, buttons should be the predominant response option [51, p.20] to facilitate the chatbot-human interaction. However, the extensive use of buttons and straight-forward instructions might come of as too robotic, thereby highlighting the fact that chatbot responses are all machine-generated. One study reported that users who were led to believe that chatbot responses were machine-generated consistently rated the responses with lower satisfaction [56]. However, regardless of the degree of realism required, the frequency of instructions are relevant because a proactive, dominant chatbot attitude has been reported to enhance users' trust towards mental health chatbots [45].

3. Additionally, Lohse [51] reported that a PSA chatbot should mimic real clinicians and introduce itself as a therapist. In consideration of the expectations of individual users, this could have significant consequences. Studies have found that users tend to perceive virtual agents as real people [53], which is also known as the computers as social actors (CASA) effect [69]. As a result, the limitations of the mental health chatbot may be unclear and the virtual agent may be mistaken for a human therapist. Therefore, other researchers in that field advise that chatbots should be presented as compassionate advisers rather than clinicians [29].

4. Finally, Lohse suggested 'empathy' and 'reflection' as important chatbot features [51, p.20] to increase user acceptance. Studies suggest that empathy expressed by chatbots towards humans may have similar effects as empathy between exclusively humans [24], thereby strengthening the bond which patients develop towards the chatbot. Interestingly, an experimental study assessing adolescents' perceptions towards counseling chatbots have found that clinical terminology was perceived less helpful than a friendly, empathetic language [54]. Empathy can be simulated through images or messages sent by the chatbot matching the emotional state of the user [22]. Another study researching patients' self-disclosure behaviours has found that human and empathetic interactions such as

small talk can enhance users' trust in chatbots as well [49].

2.5 Design Framework

To ensure that the efforts of following design activities can be effectively evaluated and that the chatbot prototype adds value to stroke survivors with PSA, the design framework 'Worth-Centred Design' (WCD) by Cockton [20] was adopted. The overall goal of the WCD framework is to produce digital prototypes that create 'collective value' [20, p.170] through an iterative approach between design arenas displayed in figure 2.1. Notably, the WCD framework was adopted and modified to the goals and research activities of this project. To start with, a design opportunity is identified by exploring users' needs through user interviews, literature research and data analysis. These activities shape the design intervention through user requirements and design suggestions. This in turn informs the design practice and the evaluation strategy (figure 2.1). Notably, evaluation activities do not only focus on the final implemented chatbot prototype but also assess early developments [20]. Finally, the iterative approach ensures that each development process, not just the prototype design, is carefully reviewed and modified as needed [20].

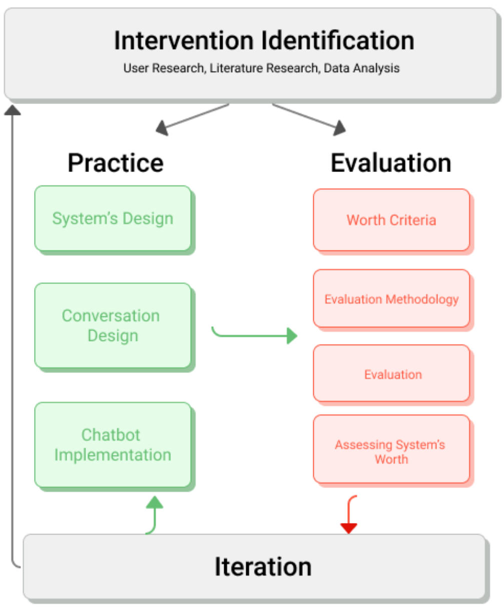

Figure 2.1: Illustration of The Worth-Centered Design adopted by Cockton [20] and Fatouma [15]. The figure shows the iterative approach to designing a chatbot prototype for stroke survivors.

Chapter 3

User Research

Numerous studies have evaluated clinicians' and patients' perception towards conversational agents in the past decade [76]. Researchers have assessed factors such as usefulness, characteristics and risks to gain understanding of the functional requirements of chatbots, whereby virtual interventions mostly targeted depression symptoms, autism [1] and other stroke unrelated anxiety disorders [69, 2]. To date and to our knowledge, there have not been any studies undertaken to explore stroke survivors' perception and requirements of a mental health chatbot to treat post-stroke anxiety. Therefore, user interviews with potential users of a TASK-CBT chatbot were conducted to understand how they use and perceive technology. Additionally, user research activities aimed to explore stroke survivors' attitude and expectations towards a therapist chatbot.

3.1 Methodology

The user research of this project followed a mixed-methods approach and consisted of two parts to inform the initial design of the TASK-CBT chatbot. First, a fully structured questionnaire was completed through telephone calls to quantify users' technology usage and their attitude towards technology and digital devices. Subsequently, a semi-structured interview was conducted to explore how stroke survivors use technology to cope with anxiety and which expectations they have regarding a post-stroke anxiety chatbot.

Recruitment of patients was carried out by chief investigator Dr. Yvonne Chun [18] of this TASK-CBT study who reached out to stroke and TIA patients that had already participated in the web-based TASK-CBT trials [18]. In total, 6 participants, 4 female and 2 male, were interviewed for this study. The participants were between 57 and 75 years old, the average age was 67. All participants were either retired or unemployed at the time of the interview. One participant held an academic degree, all other participants obtained professional qualifications. The interviews lasted between 60 and 80 minutes. The telephone calls were audio-recorded and transcribed in Microsoft Word [84] using Microsoft Dictate. Transcripts were then coded using the qualitative analysis software NVivo 1.4.1 [64] to conduct thematic analysis. This resulted in a codebook and main findings as summarised in section 3.2.3. Questionnaire data was analysed using Python [81] in the Jupyter Notebook Environment [43].

3.1.1 Ethics

All user research activities were approved by the Health Research Authority and reviewed by the NHS Research Ethics Committee (REC) due to the involvement of NHS patients (see Appendix chapter A). Audio, transcription and questionnaire data were fully anonymised and kept confidential as outlined in the Data Protection Information Sheet (see Appendix chapter A). In April 2021, recruited NHS patients were informed through the participant information form under the project name 'Co-designing a chatbot for people after a stroke (TASK-chatbot)' and consented to the secure online form 'Co-designing a Chatbot for Anxiety after Stroke' (see Appendix chapter A) via Redcap.

3.1.2 Data Collection

The fully structured questionnaire was adopted from the media and technology usage and attitude scale (MTUAS) [65] and modified accordingly to this project (see Appendix chapter B). Following the questionnaire, a semi-structured interview designed by Dr. Maria Wolters (see Appendix chapter B) was conducted via telephone. The goal was to quantify stroke patients' usage of different technologies and investigate their perception on technology and innovations such as a mental health chatbot. All user research activities were conducted remotely via telephone. After completion of all interviews, the results of the questionnaire were converted into a csv format to calculate descriptive statistics using Python Pandas, NumPy [55] and Seaborn [9] for data visualisation. The semi-structured user interviews were audio-recorded and notes were taken throughout the interview to document the interviewer's thoughts and observations.

3.1.3 Thematic Analysis

To investigate relevant topics concerning technology usage and attitude towards technological applications for health, thematic analysis [71] was performed on six transcribed interviews. During the pilot phase, two interviews were independently second-coded by two postgraduate researchers to check for researcher biases and to align coding schemes. Findings of the initial coding round were then discussed to discover coding differences by comparing quotes, themes and researchers' interpretations. The discussion ended in an agreement on a multidimensional coding frame.

During the second coding round, identified codes were re-organised and put into relation to each other to identify main themes and subcategories. By selecting additional interview material, the coding frame was further developed and modified accordingly to new emerging subcategories. Following the approach of Grounded Theory [78], the organisation of codes shed light on relevant topics and how they relate to each other given a specific user context. Notably, an empathetic stance was taken to interpret quotes and themes. The interview data was therefore predominantly interpreted closely to its literal meanings. Qualitative data was explored from different angles and perspectives, always considering the interviewees' situation and circumstances to draw connections between statements [83]. A third round of coding codes was performed to include codes extracted from the remaining interviews. The final codebook consists of a set of main categories and respective subcategories. Notably, subcategories within one main category are mutually exclusive so that codes only exist once within a main category [71]. All steps performed followed the guidelines of the SAGE Handbook of Qualitative Data Analysis [30].

3.2 Results

3.2.1 Quantitative Results

A descriptive analysis of the modified MTUAS questionnaire was conducted to quantify user behaviour for different technologies. The analysis is split in three main parts: The most frequently used technical devices among respondents (table 3.1 and figure 3.1), usage subscales of different activities performed and attitude subscales addressing participants' perception towards technology in general (table 3.2 and figure 3.2).

Device	Daily Users[%]	Total Users [%]
Smartphone	100.0	100.0
Digital TV	100.0	100.0
Tablet	66.7	83.3
Computer	16.7	83.3
Activity Tracker	50.0	66.7
Video Recorder	16.7	66.7
eReader	16.7	33.3
Smartwatch	0.0	0.0
Games Console	0.0	0.0

Table 3.1: Usage of Devices. Percentage of participants using listed devices on a daily basis and the total share of device owners.

As illustrated in figure 3.1, the most widely adopted technical devices among participants are smartphones and digital TVs. All six participants reported to use these devices on a daily basis. Notably, more participants used tablets on a daily basis than computers (67% and 17%, respectively) and only 83% of participants owned or used a computer at all (table 3.1). Moreover, both smartwatches and game consoles were not used by any participants. These findings suggest that the interviewed cohort exhibits a certain level of digital literacy, however they are less likely to be familiar with desktop apps or computer programs.

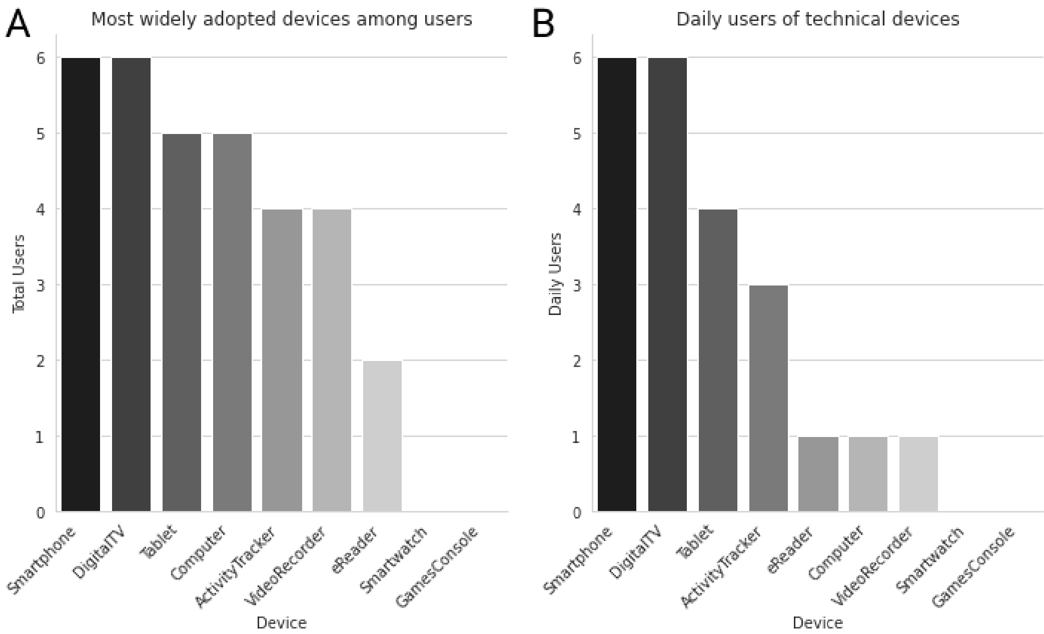

Figure 3.1: Frequently Used Technical Devices. Graph A displays the total number of daily users. Graph B shows the total number of users who reported to use listed devices either daily, weekly or rarely.

The usage subscales (figure 3.2) show that internet searching, emailing, text messaging and phone calling were among the most frequently performed activities with interquartile (IQR) values ranging between 4 and 5 (table 3.2). The broad variation of smartphone usage can be explained by the wide range of questions targeting the different activities performed with the device. While most participants reported a frequent use of apps on their smartphones, the majority of interviewees shared that they rarely use the video recording functionality. Moreover, only half of the participants used social media platforms, in particular, Facebook (figure 3.2). These participants reported that they use Facebook on a daily basis. Notably, usage values for media sharing activities were low among all participants. These findings suggest that overall, participants are quite active technology users and cover a broad range of activities. However, they have less tendencies to send and receive media files using their computers or tablets.

Usage^A	Median	IQR
Smartphone	4.0	2.0-5.0
Facebook Usage^B	2.5	1.0-4.0
Internet Searching	4.5	4.0-5.0
E-mailing	4.0	4.0-4.75
Media Sharing	2.0	1.0-3.0
Text Messaging	5.0	4.75-5.0
Phone Calling	4.5	4.0-5.0
Television Viewing	5.0	3.25-5.0

Attitude^C	Median	IQR
Multitasking	3.0	2.0-4.0
Anxious Dependency	4.0	3.25-5.0
Positive Attitude	4.0	4.0-5.0
Negative Attitude	3.0	2.0-4.0

Table 3.2: Usage and Attitude Subscales. Usage^A median values and interquartile ranges are indicated on a scale from 'never' (1) to 'several times a day' (6) except for Facebook Usage^B which ranges from 'never' (1) to 'hourly' (5). Attitude^C median values and interquartile ranges are indicated on a scale from 'strongly disagree' (1) to 'strongly agree' (5).

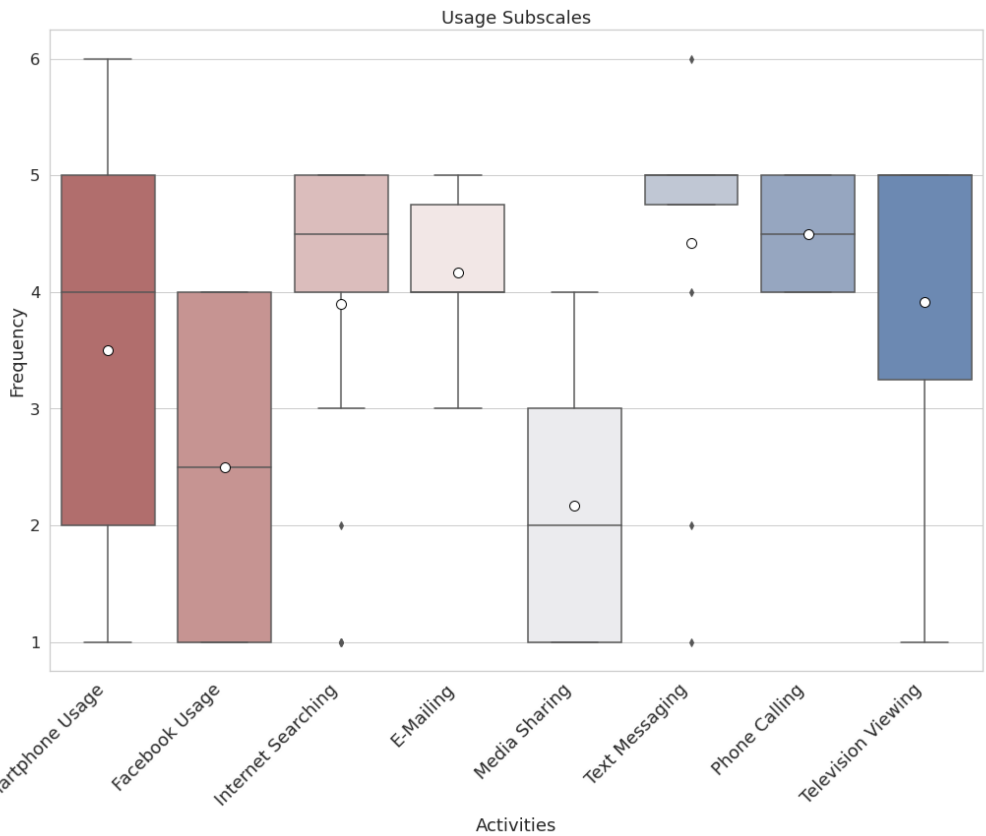

Figure 3.2: MTUAS Usage Subscales Boxplot. The frequency of usage was measured on a 6-scale ranging from 'never' (1) to 'several times a day'(6) except for Facebook Usage which ranges from 'never' (1) to 'hourly' (5). White dots indicate mean values, black lines indicate median values. The borders of the boxes mark the interquartile range.

As illustrated in figure 3.3, preference for multitasking behaviour was found to be relatively balanced among study respondents (IQR= 2 - 4). There are clear indications that participants exhibit strong technical dependencies (mean= 4.0) as half the participants reported to heavily rely on technology (IQR= 3.25 - 5.0), as seen in table 3.2. Moreover, all participants demonstrated a positive attitude towards technology. In contrast, the respondents were divided regarding negative statements dealing with technology (IQR= 2 - 4). These findings imply that most participants depend on their devices and applications while at the same time recognise the benefits and downfalls of technology.

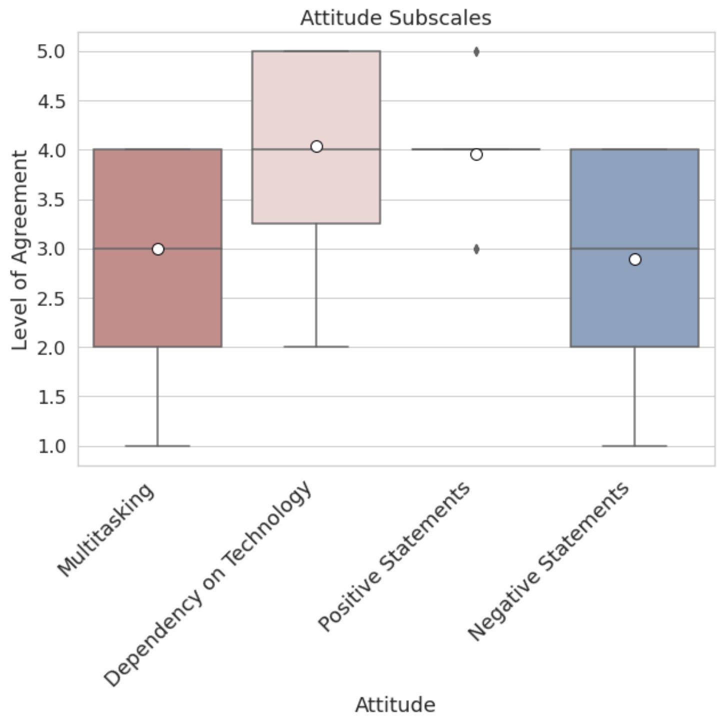

Figure 3.3: MTUAS Attitude Subscales Boxplot. Participants agreed to various statements on a 5-scale ranging from 'strongly disagree' (1) to 'strongly agree' (5). White dots indicate mean values, black lines indicate median values. The borders of the boxes mark the interquartile range.

3.2.2 Qualitative Results

The main categories identified from thematic analysis are highlighted below.

3.2.2.1 Technology's Role in Coping with Post-Stroke Anxiety

Technology and access to the internet play a big role in reassuring patients that fear a recurrent stroke. Five out of six users mentioned that access to online information was crucial to them to cope with post-stroke anxiety. Moreover, four out of six patients mentioned that the NHS website was their first point of reference to search for post-stroke information due to its perceived reliability. Three participants mentioned that access to their smartphone was essential to them in case they face an emergency situation such as a recurrent stroke. Another factor that speaks in favour of users' preference of using technology to face their anxiety symptoms is participants' reliance on technology to feel less isolated. Three interviewees mentioned that access to the internet and technology devices made them feel less isolated and two participants explicitly emphasised that turning to technology to cope with personal health issues gave them more safety due to a degree of anonymity.

"There's a safety vault there, a safety net. There's anonymity almost which I think is good." (P22)

3.2.2.2 Applications

Five out of six participants have used fitness tracking apps or wearables. The usage of health apps suggests that the interviewed stroke survivors are concerned with their own health and therefore rely on self-tracking applications. The majority of participants used fitnes self-tracking apps and devices primarily due to the broad coverage of metrics such as heart rate, pulse and calorie burn, to feel better about themselves due to a sense of accomplishment and to maintain or increase their health status. This suggests that reward systems can increase the acceptance of health applications.

'I do like knowing how many miles I walk in a day, how many steps, that sort of stuff. Just to maintain the fitness level.' (P19)

3.2.2.3 Preferred Representational Systems

Five out of six participants expressed positive attitudes towards audio-assisted applications in different contexts such as music, meditation, conversational agents or audio books. In particular, participants argued that applications working with speech recognition or audio allow them to lay back, relax and listen. On the other hand, three out of six participants also stated that in some instances they could process information better if it was presented visually, for example through videos or images.

3.2.2.4 Technological Curiosity vs Risk Avoiding Behaviour

All participants used their smartphones on a daily basis and were digitally literate concerning their frequent activities such as online researching, downloading apps and video calling. The majority of interviewees used the online search engine Google [34] on a regular basis to find information or to look for advice when needed. Importantly, these participants also perceived themselves as confident in dealing with everyday tasks on their smartphones.

On the other hand, three participants also displayed less risk tolerant behaviours towards new technology devices. In particular, these participants knew their technical limitations well. Some reported that they are usually keen to follow instructions which implies a less explorative approach to certain technologies, others relied on help from others to solve tasks.

'I always follow the instructions. You know, when the instructions are written out, I can follow them really well' (P22)

3.2.2.5 Negative Experiences with Technology

All participants have experienced frustrating incidents related to technology. While most situations related to circumstances outside of their control such as loss of internet connection or defective devices, three participants described how technology complicated their daily lives by lowering their self-esteem. For example, due to their worsened memory post-stroke, technological tasks were constantly disrupted or hindered. One participant explained that self-tracking results would lead to feelings of failure instead of success. In addition to above mentioned technology frustrations, inaccurate predictive text and frequent software updates made participants' lives more complicated. The latter might imply that stroke patients, especially those with memory difficulties, find it stressful to adapt to technological changes.

'When they change the app, I don't like that. I don't know why they need to change that, I got used to that. But if I just take my time, I'll pick it up. But I have to take time to pick it up.' (P21)

3.2.2.6 Anxious Situations

When participants were asked about how they used technology to cope with post-stroke anxiety, the majority mentioned specific instances that deal with anxiety or panic before going to sleep. Other anxious situations mentioned relate to commuting or meeting other people. One patient, on the other hand, noted that the deterioration of memory after the stroke often triggered the fear of forgetting things.

Interestingly, when asked about TASK exercises that were perceived as particularly helpful at the time of the web-based TASK trials [18], without exception, participants could only recall the relaxation exercise. This might be due to the fact that relaxation is a practical coping method that requires less steps and does not require any knowledge about its theoretical background. These findings also highlight the importance of facilitating the memorisation process so that relevant coping strategies can be recalled from memory after the therapy has ended.

'I can just remember deep breathing and remember to go through what I have been taught with the TASK treatment. But unfortunately, my memory is bad and I forget it and that's why I get frustrated.' (P21)

3.2.2.7 Post-Stroke Digital Literacy

Five out of six participants reported that they did not have major issues using their smartphones post-stroke and four participants reported the same about their laptops or tablets. The most common change observed after the strokes and TIAs was more time spent on their technical devices. However, two participants reported that they had greater difficulty solving problems that required higher cognitive effort.

'I used to have to break things down a bit. You know, I'm not as good as you know, before, I could just quickly glance through something and I could pick up all the relevant points that I was having to start off.' (P22)

3.2.2.8 Experience with Chatbots

The majority of participants possessed a good understanding of what a chatbot is and which functional limitations it has.

'It's all about asking the right questions, the reason why it works, to get the answer you need.' (P22)

The majority of participants was familiar with Siri and Alexa. Generally, they were satisfied with their experiences, only one participant did not like the virtual agents due to the high error rate in voice recognition when dealing with different British Accents.

Other participants described similar negative experiences when dealing with chatbots. In particular, incorrect recognition of user input was perceived as frustrating because it forced users to reword their inquiry multiple times before the chatbot provided the correct answers. Moreover, it was reported that the interaction with chatbots caused anxiety if the chatbot response speed was inappropriate or perceived as too fast so that the participant felt under pressure to come up with a user input as quickly as possible.

'I could see that they replied, and I would have to get back to them and then maybe the doorbell would go and you know, so it can cause anxiety.' (P25)

3.2.2.9 User Needs

All of the interviewed participants showed interest and willingness to test the TASK-CBT chatbot. Participants also described the perceived benefits of such a chatbot. The accessibility, time saving aspects, efficiency and straightforwardness of such a chatbot were factors that were perceived particularly beneficial in comparison to traditional therapy. When asked about chatbot features that a post-stroke anxiety chatbot should exhibit, two participants mentioned that a certain degree of personalisation is required. By personalisation, interviewees referred to a database with each patients' medical background and history that would inform the chatbot response. This indicates that value is created when chatbots are able to tailor their responses and direct the conversation accordingly to the specific user needs.

'Providing there's a backup from each person's personal record, you know, the questions and answers that you're going to get.' (P20)

In addition, participants expected the chatbot to provide enough context when going through the CBT exercises, to be robust to a variety of user inputs and to be reassuring. They expected the chatbot to make patients feel less isolated with their anxieties. One participant mentioned that the chatbot should be validated by doctors, suggesting that its trustworthiness could be increased if it is clearly stated that it is a medically approved intervention. When asked about the chatbot's personality, five out of six participants stated that the chatbot should possess humane and empathetic features. Participants gave examples of anthropomorphic characteristics such as short 'typing' breaks before the response is displayed on screen, addressing the user with their names and offering the user to take a break in between. The simulated caring nature of such a virtual agent is expected to act in consideration of users' interests.

Finally, three participants expressed their expectations towards the chatbot's identity. All of them preferred a clinical, professional stance which might be due to the above mentioned desire of reassurance and trustworthiness during the therapy. Moreover, all participants have already had therapy sessions with real clinicians in the past and therefore might favour a medical perspective for online interventions, too.

'I think if I was using something like that in a post TIA situation, I'd be looking for a clinician as opposed to a friend.' (P23)

3.2.3 Design Considerations

The main goal of this user research was to identify the perception and attitude of former post-stroke anxiety patients towards technology and their expectations towards a psychotherapy supporting chatbot. The summarised findings from the user research and literature review were then used to draft the initial user requirements (UR) for the chatbot prototype.

- **UR1**: The chatbot needs to provide context about TASK-CBT (2.3.3) and the application itself (2.4.2). Furthermore, it needs to give clear instructions on what to expect and how the chatbot will help users to cope with their anxiety. To position itself as an adviser with a professional stance (3.2.2.9), the chatbot should clearly refer to itself as a chatbot rather than a medical person (2.4.2) and deliver psychoeducational content in the onboarding phase (3.2.2.9).

- **UR2**: The conversation flow requires a good balance between user and chatbot input. Generally, the chatbot should dominate the dialogue (2.4.2) and lead the conversation. However, the user should retain flexibility over the pace of the conversation. They should be able to decide how fast or slow the conversation progresses (3.2.2.8).

- **UR3**: The chatbot has to be easily accessible (3.2.2.1) and understandable (3.2.2.7). Wherever possible, the chatbot should convey information step by step and reduce cognitive effort as much as possible (3.2.2.7). Therefore, longer passages of pure text should be avoided. Visual media should be introduced to make information easier to process and to facilitate the learning progress (3.2.2.3).

- **UR4**: The language of the chatbot should be empathetic and demonstrate a caring, interested personality (3.2.2.9). Empathetic feedback should be tailored specifically to users' situations to increase perceived trustworthiness and authenticity (3.2.2.9, 2.4.2).
- **UR5**: The chatbot should summarise essential steps of the TASK-CBT at the end of each session so that users can memorise new learnings better (3.2.2.6). To evoke a sense of achievement after an exercise has been completed, retrospective feedback should be implemented (3.2.2.2).
- **UR6**: To mitigate the risks of not correctly understanding users' input, buttons instead of free text options will be implemented at critical points in the conversation flow to reduce the chatbot complexity (2.4.2). Fallback response options to unexpected user input is required to cope with errors and provide transparency about the chatbot's limitations (3.2.2.8).
- **UR7**: The chatbot should provide some degree of reassurance regarding users' mental health (3.2.2.9). A confident and reassuring stance is expected to help users feel more at ease during the virtual therapy (3.2.2.9). The degree of medical reassurance will not be covered due to ethical considerations and will be further elucidated in chapter 6.
- **UR8**: The conversation flow has to cover all exercise questions from the original TASK-CBT protocol (2.3.3). This ensures that the chatbot is coherent with clinical therapy guidelines.

3.3 Discussion

3.3.1 Limitations and Risks

All participants displayed sufficient levels of digital literacy to utilise mobile and web applications. Furthermore, all stroke survivors involved welcomed the idea of a chatbot guiding patients through TASK-CBT. Notably, all participants were older than 55 years and retired or not employed. They were also familiar with the web-based TASK-CBT intervention beforehand and had participated in the TASK-CBT randomised controlled trial in the past [18].

Biased Participant Cohort. All patients gave the impression that they had a rather positive outlook on life despite their strokes. Because younger participants were not recruited for this study, these findings do not necessarily reflect the opinions of younger stroke survivors. Young patients are expected to face different social stigmas and life circumstances [77]. Notably, post-stroke anxiety is more prevalent among young patients [12] and it is reported that older people tend to have a more stoic view on their future than younger generations [61]. Moreover, the identified perceptions towards technology are unlikely to be fully representative of stroke sufferers in general, as the recruited participants were expected to be more digitally literate than the average population in their age group. Additionally, recruited participants did not suffer from recent strokes. On average, the last stroke or TIA was more than 3.5 years ago which has significant implications for the perceived urgency of PSA treatment. To increase the reliability of user research outcomes, it is advised to conduct both qualitative and quantitative studies with a larger cohort of patients across all age groups [18-65 and 65+] and stroke survivors who are not yet familiar with TASK-CBT.

Interview Quality Issues. To facilitate research activities, all interviews were conducted remotely via telephone calls. The quality of telephone networks and sound varied greatly, hence certain interviews yielded more information than others, depending on whether the interviewer was able to follow up on certain answers. In addition, a large proportion of the audio recordings was transcribed manually, as Microsoft Dictate has a limited ability to recognise different English and Scottish accents. For future research, especially in patients with post-stroke Aphasia, it is suggested to conduct qualitative interviews face-to-face or via online calls for improved audio quality.

Validity of Data Analysis. Moreover, due to time limitations, only two of the six qualitative interviews were second-coded by another researcher. This might have effected the overall coding quality as the final code book was not validated by an independent researcher. To increase the reliability of the coding results, a larger proportion of interviews should be second-coded before a coding frame is established. The quantitative questionnaires were conducted with a cohort size of 6 participants as well. Notably, only median and IQR values were outlined in table 3.2 to show the entire distribution. However, due to the small sample size, these results are only of limited significance. In order to achieve reproducible and reliable results, further studies should recruit more participants to validate the variability of subscales (figure 3.2).

3.3.2 Research Implications

Designing for Older Stroke Survivors. The aim of this study was to understand how stroke survivors use technology in their daily lives and which factors have to be taken into account when designing a PSA chatbot. The quantitative study revealed that all participants used their smartphones on a daily basis and possessed sufficient levels of digital literacy to engage with novelties and mobile applications. However, the limited use of computers indicates that older stroke sufferers may be less exposed to complex software applications than the younger non-stroke cohort interviewed by Lohse [51]. Furthermore, social aspects might play a minor role in stroke survivors beyond the age of 55, as only half of the participants used social media channels.

The qualitative findings on the other hand suggest that the majority of participants favoured Apple instead of Android smartphones due to perceived ease of use. In line with quantitative findings, participants reported an infrequent use of computers and preferred to use their smartphones or tablets instead. Notably, all participants have suffered from PSA and found the TASK-CBT exercises helpful during the time of intervention [18]. However, no specific coping mechanisms other than breathing exercises were recalled. These findings suggest that older stroke survivors or stroke patients with impaired memory require additional support in learning CBT techniques in comparison to younger, non-stroke populations who reported to have 'used strategies similar to CBT' [51, p.36]. Another factor that distinguishes older stroke survivors from younger non-stroke populations is the source of technology frustration. None of the participants described any issues concerning 'addiction and distraction' [51, p.33]. Participants in this study did not feel overwhelmed by notifications and social obligations, but reported frustrations related to software updates and high cognitive load.

Examining the Perception towards Empathetic Chatbot Language. Generally, participants welcomed the idea of talking to a mental health chatbot. The biggest challenge they saw was in mimicking human interactions with characteristics such as empathy and care. All participants have mentioned reassurance in the context of coping with PSA. Although it was not specified whether they meant reassurance regarding their physical or mental health, participants reported that they have used technology frequently to search for more information regarding their stroke and symptoms. Moreover, most participants explained that the chatbot should tailor its responses accordingly to users' medical history and personal needs.

Future research activities should explore stroke survivors' understanding of empathetic language in different social and cultural contexts across all age groups. Future research aiming at the development of a holistic application for PSA patients should consider native mobile applications that work with older operating systems, too. Moreover, the chatbot should be linked to patients' existing TASK-CBT records to keep track of their progresses. Additionally, future user research activities should include stroke survivors dealing with significant cognitive impairments to discover further barriers in technology.

Chapter 4

Design Practice

This chapter summarises the design practices taken to build an interactive chatbot prototype. The main goal of the design practices is to build a web-based chatbot prototype that guides users through an initial onboarding of the TASK-CBT intervention followed by the first post-stroke anxiety exercise. The underlying chatbot requirements are based on the findings generated from the preceding user research activities and knowledge acquired through background research on this topic. The conversation flow was designed on the basis of the TASK-CBT protocol developed by Chun et al. [18] and modified accordingly to a range of common user scenarios.

4.1 Methodology

4.1.1 Dialogue Design

Based on the findings and user requirements identified in section 3.2.2, the conversation flow was outlined prior to the chatbot implementation. This was necessary to test whether all dialogue pathways are covered by the intervention and how different user inputs will inform the next step during the conversation. Sketching and drafting the initial conversation flow (figure 4.1) helped to inform the final chatbot decision tree and underlying logic which was then implemented as described in section 4.2. Moreover, a user journey was visualised (figure 4.2) to help researchers and collaborators involved in this project to understand how PSA patients go on with their lives after discharge.

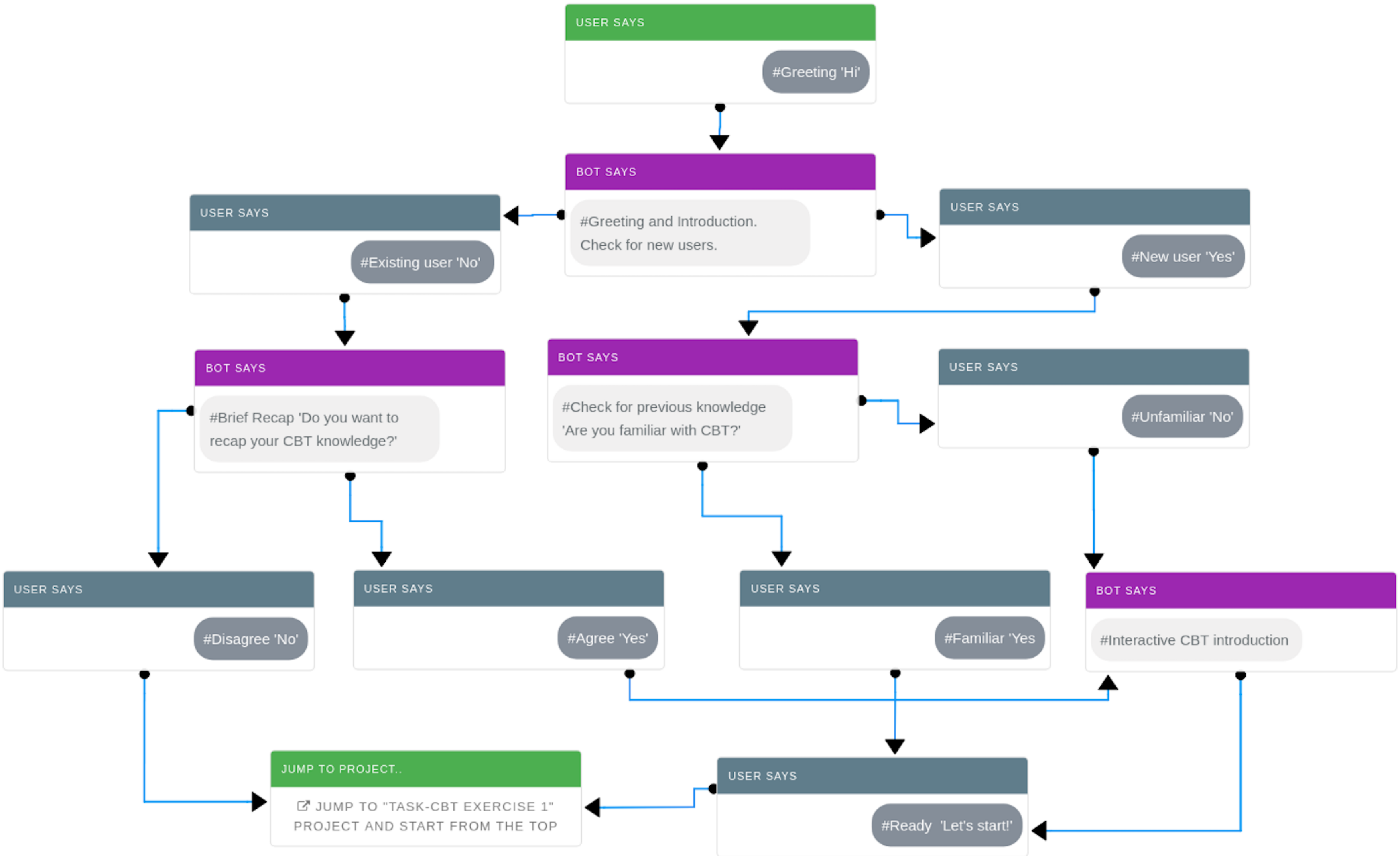

Figure 4.1: Exemplary Dialogue Flow of the Chatbot Onboarding. The figure illustrates an initial sketch of the user onboarding dialogue pathway.

User Journey

The visualisation of the user journey beginning from hospital discharge to the point where stroke survivors seek psychological support highlights the different stages where a TASK-CBT chatbot can intervene to assist the patient. Understanding the circumstances and the history of events that stroke survivors have dealt with is fundamental for designing a good conversational experience. The primary use case of the TASK-CBT chatbot involves patients who have recently suffered a mild stroke or a TIA and have been dealing with mild to moderate anxiety ever since. Hence, the user journey is a summary of what patients thought, felt or did to cope with their mental health problems during this period (figure 4.2). Information was retrieved from user interviews (section 3.2.2) and background research (chapter 2).

4.1.2 Data Enrichment and Corpus Building

Based on the TASK Therapist's Manual and Record, a former anonymised dataset retrieved from the TASK-CBT trial [18] and user research findings (section 3.2.2), possible user scenarios for TASK exercise 1 were outlined. Based on each user scenario, a range of diverse user inputs were created to enrich the dataset for the machine learning model of the chatbot. The training data consisting of user inputs is strongly based on the vocabulary and phrases used by the participants. For each user scenario, a set of example chatbot responses were generated to ensure variety during the chatbot interaction. The responses were closely aligned with the guidelines of the TASK Therapist's Manual which was originally developed for the web-based therapy intervention [18]. In accordance with previously identified user requirements and expectations towards the chatbot's characteristics, responses were then slightly modified.

4.1.3 Chatbot Framework

The chatbot was implemented using RASA open source, a python library [10]. The machine learning framework can be used to build conversational AI assistants based on natural language understanding (NLU), supervised learning and dialogue policies [74, 10]. In particular, the machine learning pipeline utilises the Dual Intent and Entity Transformer (DIET) classifier to recognise and classify user inputs ('intents') [39]. Thereby, the Python library spacy_sklearn is used to preprocess text data [10] and entities are recognised via a Conditional Random Field (CRF) based interpreter [10].

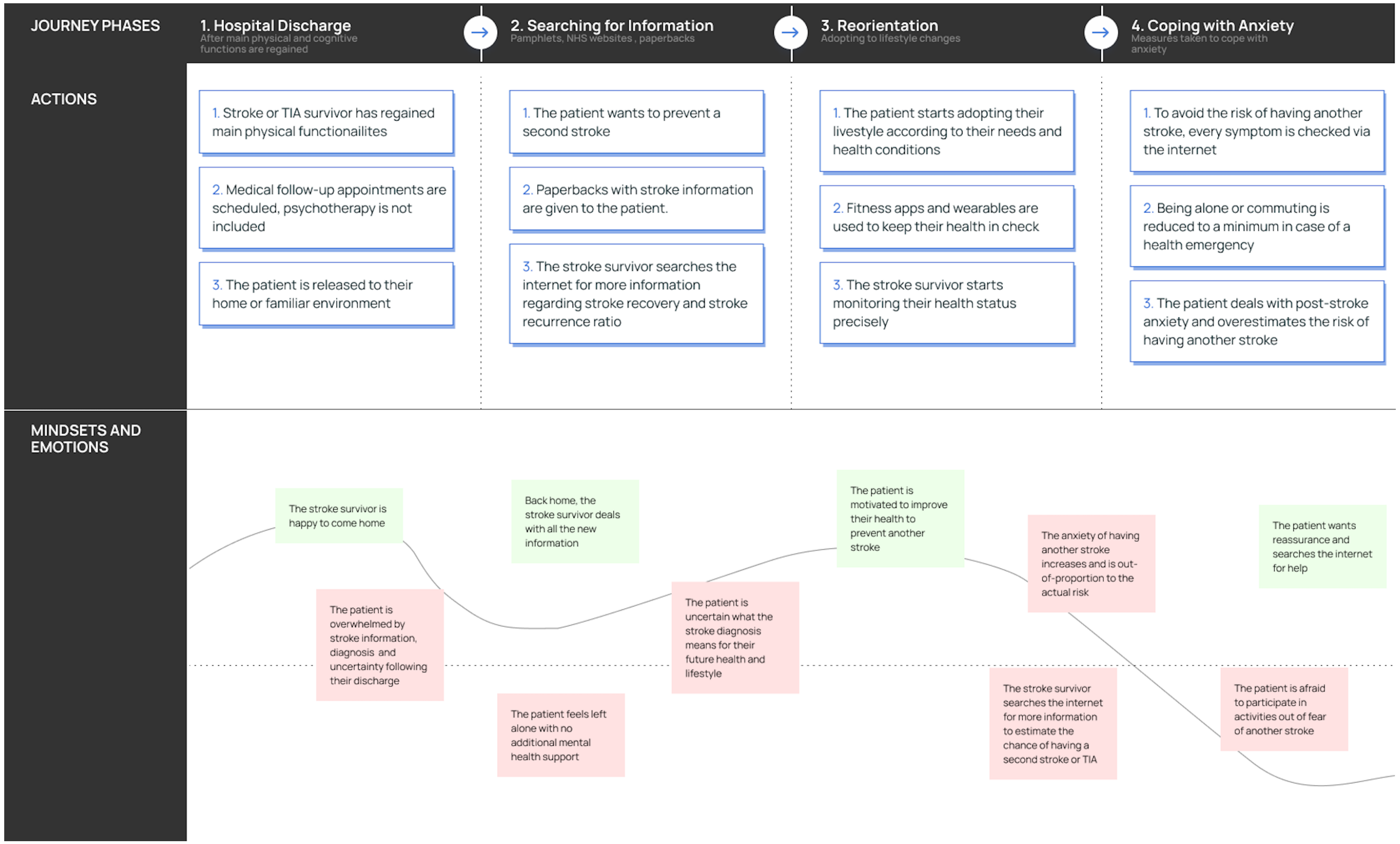

Figure 4.2: User Journey. The figure illustrates stroke survivors' mental journey starting from hospital discharge to coping with PSA.

Because the framework performs both natural language understanding and conversation management by learning from dialogue data, it can also handle conversations that have not been pre-scripted. This allows for some flexibility during the patient-chatbot interaction in cases where users deviate from the intended conversation pathway. Moreover, a machine learning approach generally allows for more precise personalisation and more nuanced responses once training datasets become larger.

4.2 Chatbot Implementation

4.2.1 First Implementation Phase

The implementation of the chatbot started with the definition of the ideal happy path where users are briefly introduced to the TASK exercise 1 before they answer a set of questions related to an anxious situation they have recently encountered. Therefore, the expected user intents, chatbot responses and conversation rules were implemented to define the dialogue flow. The key elements of the first user story are represented by the following action points triggered by the chatbot (figure 4.3):

1. action: utter_welcome_new. The chatbot introduces itself.
2. action: utter_task1_explanation. TASK exercise 1 is explained.
3. action: utter_anxious_moment. The chatbot starts with the first exercise question.

In between, several single-turn conversation units were implemented to make the dialogue flow less robotic and more humane (UR4) through questions such as whether the user is ready or if the user needs a break in between. Based on the findings in section 3.2.2, the different questions of TASK exercise 1 were divided into smaller sections, so that users can sort out their thoughts step by step (UR3). Moreover, button style selection options were implemented throughout the conversation at various points to handover dialogue control to the users (UR2). Most buttons serve as a checkpoint for users to decide whether they want to progress to the next step. This decision was based on the user requirement UR2 that users should be able to control the conversation pace.

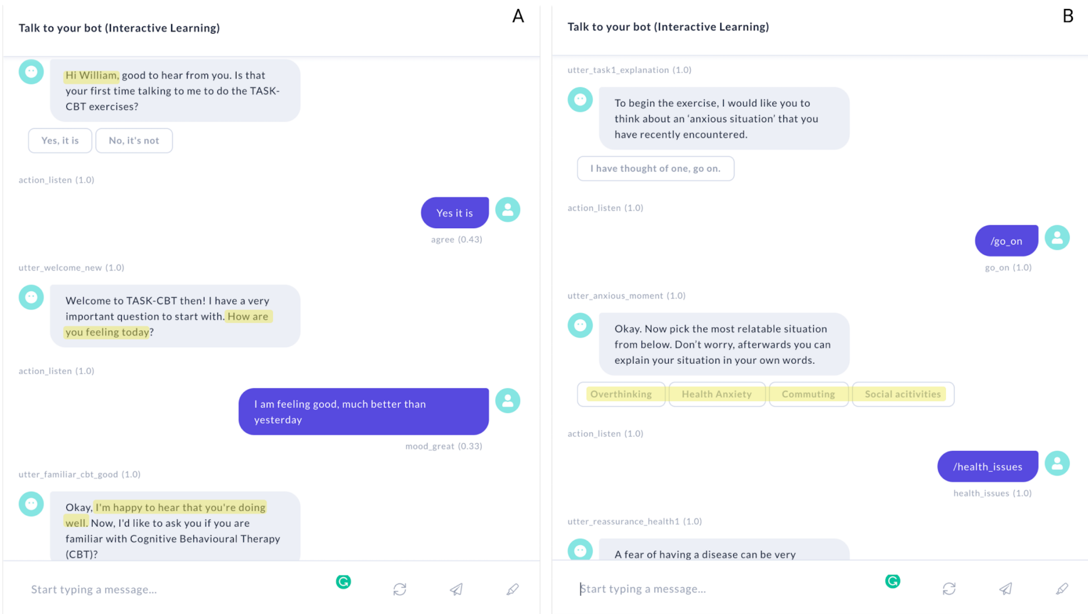

Figure 4.3: Exemplary Chatbot Screenshots. Figure A displays the initial interaction between chatbot and user. Figure B shows different button selection options for choosing an 'anxious situation'. Parts highlighted in yellow represent key user requirements outlined in this section.

During the initial implementation phase, the selection choices for choosing between different anxious situations consisted of the options 'Overthinking', 'Health Anxiety', 'Commuting' and 'Social activities' (figure 4.3). This was grounded in previous findings from user research (section 3.2.2.6) and the web-based TASK-CBT datasets. For each anxious theme, the user was then given the opportunity to describe the anxious situations in their own words (figure 4.4). Slightly nuanced feedback responses were enabled based on the intent training data which was populated with more detailed user scenarios. As displayed in figure 4.4, both conversations deal with the user experiencing anxiety due to overthinking.

During the first scenario, the user experiences high episodes of anxiety before going to sleep. The machine learning model correctly classified this intent as overthinking scenario 2 (overthinking_2) which was trained with different intent examples that specifically dealt with nighttime anxiety. Tailored responses to individual user scenarios help to build trust and are perceived as more authentic so that the human-computer interaction is not experienced as shallow (UR4).

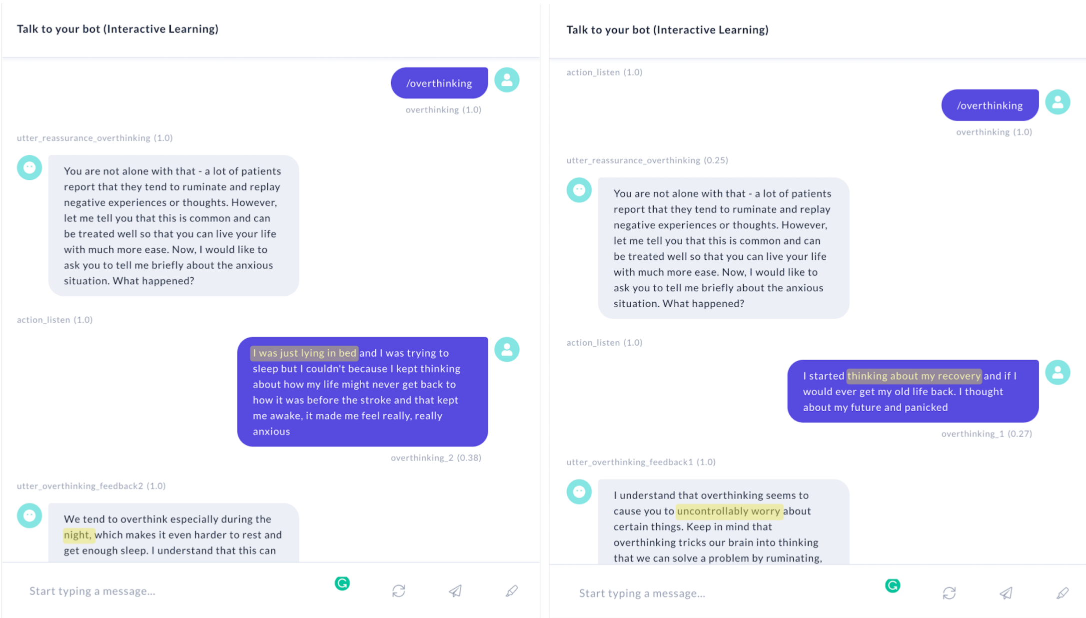

Figure 4.4: Nuanced Chatbot Feedback for Specific User Inputs. The figure highlights how the chatbot responds to different user scenarios addressing the issue 'overthinking'.

The subsequent dialogue flow was implemented closely based on the web-based TASK-CBT protocol (UR8). Users were asked to break their anxious moments down into feelings, thoughts and actions. For each user input scenario, different examples were trained to allow for some degree of personalisation similarly to the example outlined in figure 4.4.

Issues and Troubleshooting

After each training session of the model, the conversation flow was tested using the RASA X user interface as displayed in figure 4.3. RASA X allows interactive learning so that the chatbot model can be tested, corrected and retrained [75]. Additionally, it offers conversation-driven development by saving new intents while testing story pathways and reviewing new conversations. That way, corrections and manual adjustments can be quickly made through the graphical user interface. Other contextual errors identified through RASA X were adjusted and retrained in the terminal.

One major issue identified within the first implementation phase was the low confidence score for chatbot actions whenever user messages were incorrectly classified. Particularly when asked to describe their anxious situations or thoughts using free text forms, user intents displayed low confidence scores, mainly below 0.25. User inputs following the different inquiries 'Describe your anxious situation' and 'Describe your thoughts during that moment' appeared to be very similar to each other in informal evaluations with 2 fellow students, therefore user messages were often incorrectly classified. Because wrongly classified intents disrupted the predicted conversation pathway, these low confidence scores triggered action default fallback options. In instances, where the next chatbot action could not be predicted with a confidence score higher than 0.2, the chatbot would respond with 'Sorry, can you rephrase that?'. To avoid this issue, 3 measurements were taken:

1. The threshold for a default action response was decreased from 0.25 to 0.1.
2. Training data was enriched with further user input examples to increase intent recognition confidence scores.
3. Rules were implemented that allowed the conversation flow to move on even if the user input did not exactly match the expected intent.

4.2.2 Second Implementation Phase

Following the initial implementation, further evaluations and implementation phases were carried out as described in the WCD framework (section 2.5). During the second implementation phase, the overall conversation flow was improved by splitting chatbot responses into smaller parts, providing more visual material and adding empathetic features to increase patients' trust towards the chatbot, especially during the introduction. Moreover, a dialogue flow guiding new patients through the initial onboarding was implemented (section 4.2.2).

As described in UR1, the provision of psychoeducational material can help users to better understand the different steps required to achieve their goals. To facilitate their memorisation process by the end of the virtual therapy session, a visual summary was displayed to help users self-manage their anxiety in situations where they cannot access the chatbot instantly (UR5). Additionally, the recapitulation of the tasks completed might reinforce a sense of achievement by emphasising even small accomplishments (UR5). The following changes were implemented based on informal feedback given by 2 fellow students after testing the chatbot:

1. Less text-heavy chatbot outputs and multiple responses divided through little breaks in between (figure 4.5). The overall balance between user and chatbot input was adjusted to reduce the extend of feedback.

2. Manual intent recognition in addition to button selection options (figure 4.5). Instead of having to select buttons to move on or to choose an option, users were enabled to manually type in their responses if they preferred to do so. However, the button options still remained visible to give users cues about the next step.

3. Visual media to underline exercise questions and examples (figure 4.6). To help users understand certain exercises more easily, visual cues were represented at different conversation steps. This helped to reduce the overall text volume and was intended to encourage users to reflect their thoughts mindfully by giving them visual examples (UR3). Finally, a visual summary of all steps completed during TASK exercise 1 was implemented to stimulate the memorisation process (UR5).

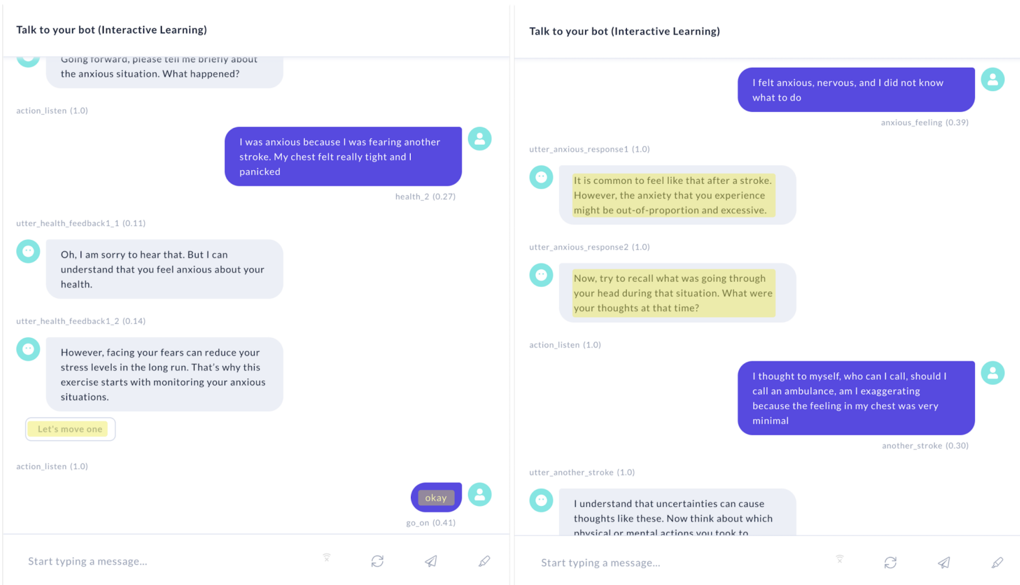

Figure 4.5: Manual user input and multi-response feedback. The figure shows the changes made in the second implementation phase. Users have the option to provide individual input instead of selecting button options. Text-heavy chatbot feedback is resolved into multi-responses.

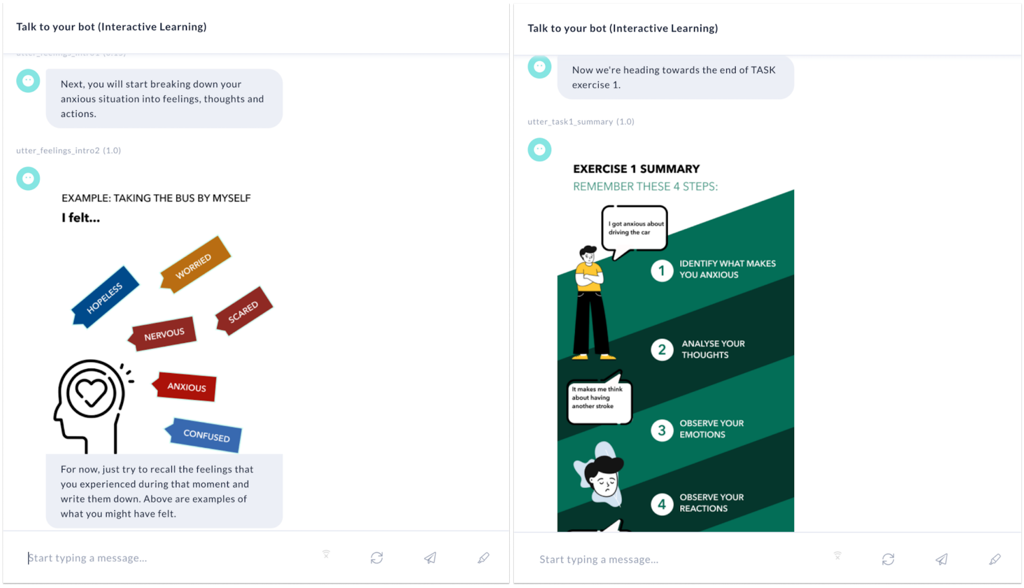

Figure 4.6: Using Visual Media to Reduce Cognitive Load. Visual materials are displayed to communicate psychoeducational and TASK-CBT exercise related information.

Chatbot Onboarding

The onboarding dialogue was implemented for new users who are unfamiliar with TASK-CBT (UR1). During the onboarding and introduction to TASK-CBT, the chatbot would prompt the user to interact with psychoeducational content, for example through open-ended questions such as asking the user how they feel about another patient's thought process (figure 4.7). In accordance with UR4, the questions and statements posed by the chatbot were formulated using empathetic language to convey a caring personality and to act as a 'shoulder to cry on', as seen in the statement 'Well, it can be really hard to live with constant fear, no matter what kind of, right?' (figure 4.7). Finally, the provision of psychoeducational content was implemented to help position the chatbot as a virtual counsellor with a professional attitude (UR1).

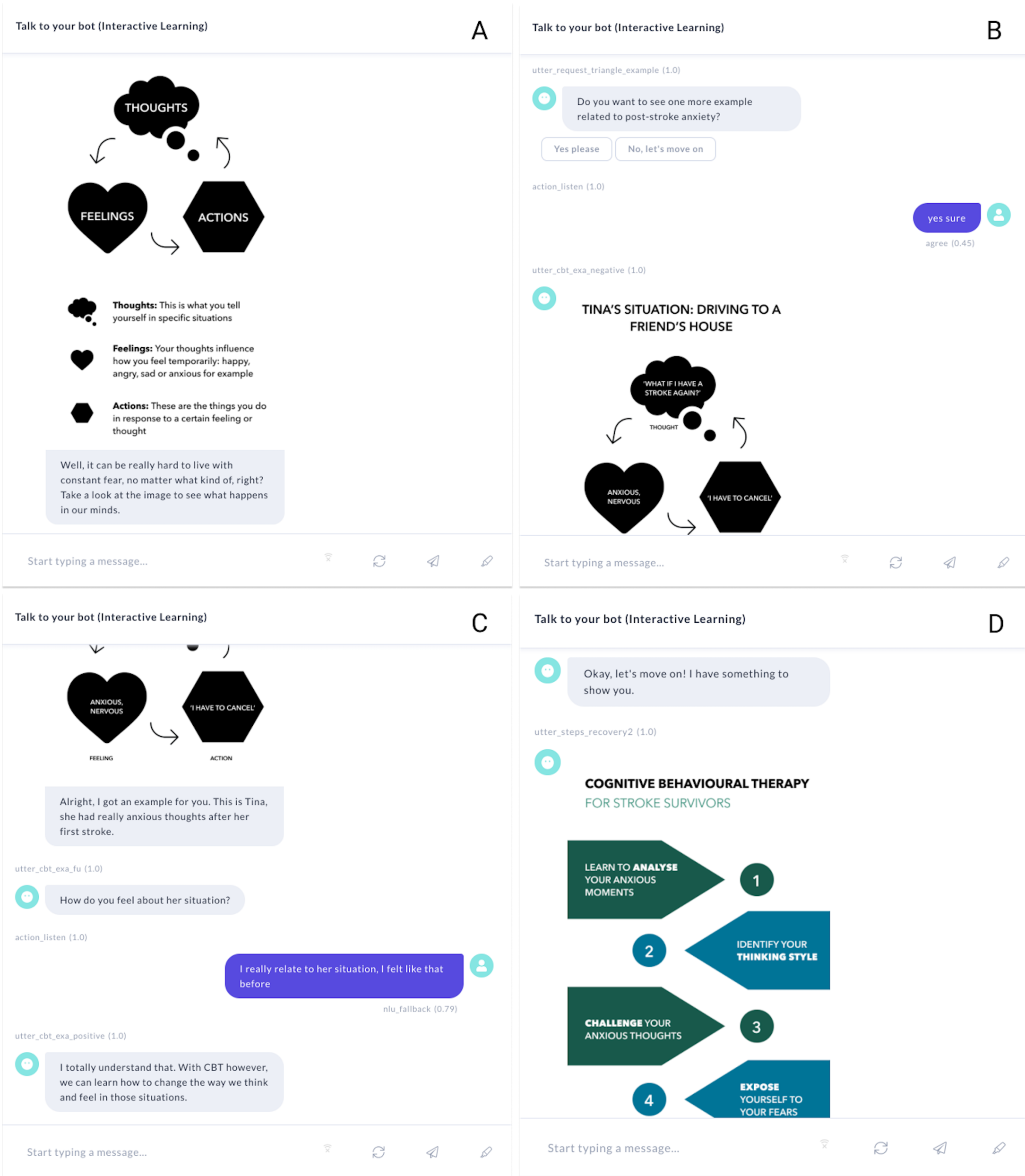

Figure 4.7: Chatbot Onboarding Dialogue. The graphic illustrates how the chatbot guides users through the basics of TASK-CBT (figure A-C) before summarising following exercises (figure D).

4.2.3 Final Adjustments

Based on the second informal user testing session with two fellow design research students, a few more changes were implemented to improve the usability and usefulness of the chatbot. The following issues were identified:

Miscategorisation of Anxious Situations. The designation of selection options for different anxiety scenarios was perceived as misleading because the option 'Overthinking' does not account for an anxious situation but rather an underlying thinking style (figure 4.3). This conceptual issue was identified through informal user testing sessions and discussed with chief investigator and stroke physician Dr. Yvonne Chun [18]. The button option 'Overthinking' was placed as an option next to categories such as 'Commuting', 'Social Activities' and 'Health Anxiety' when users were prompted to select the most comparable anxious situation they have faced. This initial design decision was mainly based on the fear situations mentioned by users during the semi-structured interviews (section 3.2.2.6). However, this categorisation is inaccurate considering that the TASK-CBT protocol requires patients to first explain their situational circumstances before their thoughts and feelings are identified in the next step. Moreover, the identification of thought patterns such as catastrophising and overthinking is done in TASK exercise 3 based on the TASK-CBT protocol (section 2.3.3).

Due to time constraints, the classification of anxiety scenarios was not entirely redesigned but only slightly modified by replacing the option 'Overthinking' with 'Overthinking about over things' (figure 4.8). This was done due to two reasons: First, student researchers noticed that some patients might think of anxious situations that have nothing to do with common triggers in post-stroke anxiety. Secondly, the majority of participants from user interviews have stated that they felt anxious before going to sleep when they were deep in thought (section 3.2.2.6). The decision on which overarching categorisations should be displayed in future developments is further addressed in section 6.2.

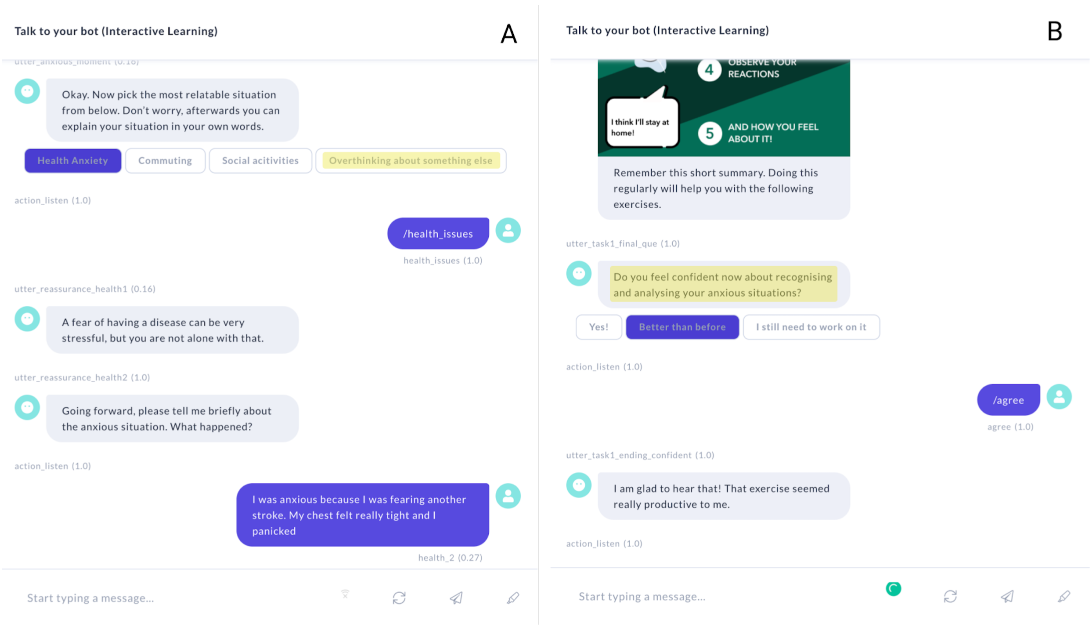

Figure 4.8: Fine-tuning the Dialogue Design. Figure A displays the small change made upon the categorisation of anxious situations in button selection options. Figure B shows the concluding feedback question posed by the chatbot.

Concluding Chatbot Feedback. Moreover, a final conclusion was missing after the end of the virtual therapy session. Both students noticed that a closing feedback was lacking that encouraged users to stick to their exercises. Therefore, questions asking how users perceived the virtual therapy session (UR5) were incorporated. Therefore, concluding words were added to the training example and users were asked to reflect on whether they were able to identify their anxious situations better than before (figure 4.8).

Coping with Errors and Skipping Onboarding. Additionally, students reported that the chatbot did not cope well with errors or unexpected free text user inputs during the introduction and onboarding phase. If certain intents were not classified correctly, the chatbot still got stuck in a loop or discontinued the flow of conversation. This was mainly due to small training datasets. To avoid extensive use of buttons, free text options for personal questions such as 'How are you feeling today?' remained but additional default action responses were implemented (figure 4.9). Another student also reported that existing users should be able to skip the TASK-CBT introduction to save time and avoid unnecessary repetition. Existing users were therefore given the chance to start with the exercises right away as shown in figure 4.9.

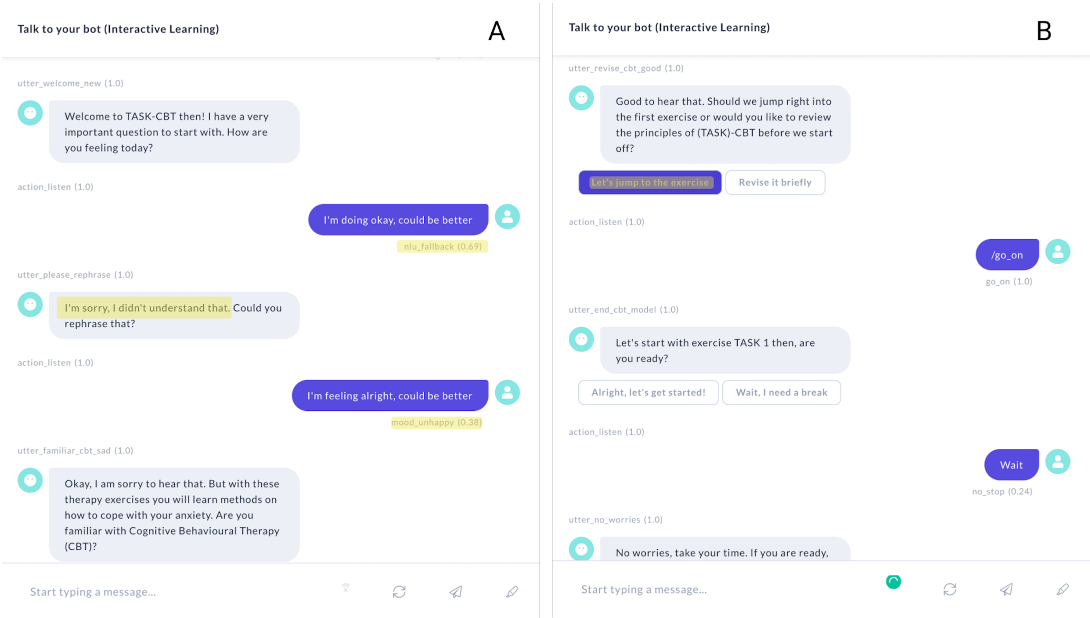

Figure 4.9: Final Adjustments. Figure A displays the chatbot's ability to cope with non-identifiable user intents. Figure B displays an option for existing users to skip the onboarding.

Chapter 5

Evaluation

This section outlines the evaluation methodologies implemented to assess the usability and value provided by the chatbot prototype (section 2.5). In total, two stroke physicians and two stroke survivors participated in the evaluation under patient and public involvement. For the usability testing part, participants interacted with the chatbot and completed the first TASK exercise remotely. The interaction was observed via screen sharing on Microsoft Teams [40]. Minimal moderation during the usability testing took place at the beginning of the testing and at instances where users wanted to clarify or challenge their interactions. Thus, participants were not asked to explain their actions in order not to influence their behaviour and to keep the cognitive load as low as possible [80]. This decision was also made due to time constraints. However, most participants verbalised the majority of their thoughts throughout the process. Live usability testing was subsequently followed up by two structured questionnaires and a debriefing session. This was crucial for the evaluation to ensure that both concurrent and retrospective results were collected to obtain a comprehensive understanding of the user experience [13]. Each evaluation lasted for about an hour. Quantitative analysis was conducted using Microsoft Excel [84].

5.1 User Experience Questionnaire

The User Experience Questionnaire (UEQ) is a structured questionnaire based on six subscales as seen in table 5.1 [48, 73]. The perceived user experience and usability of the system was assessed through 26 questions asking users to rate their experiences on a range from 1 to 7 [73]. The rating stage represents two bipolar adjectives, where 1 indicates full agreement with one adjective and 7 indicates full agreement with the opposite adjective (see Appendix chapter C). The resulting UEQ scores reflect users' perceptions towards the system such as its likeability, usefulness and enjoyability [48]. Notably, resulting scores range from -3 to +3, whereby values in the range of -0.8 and +0.8 are interpreted as neutral (table 5.1). Moreover, the analysis outcomes were benchmarked against a dataset of 468 UEQ evaluations addressing different digital products to investigate the relative usability of the chatbot prototype (figure 5.2).

Scale	Mean	Confidence Interval
Attractiveness	1.71	1.43-1.99
Perspicuity	1.94	1.63-2.25
Efficiency	2.00	1.80-2.20
Dependability	1.31	1.19-1.44
Stimulation	1.31	0.95-1.68
Novelty	1.44	1.20-1.67

Table 5.1: UEQ Mean Values and Confidence Intervals. The confidence intervals are given with two standard deviations so that the real value is captured with 95% certainty.

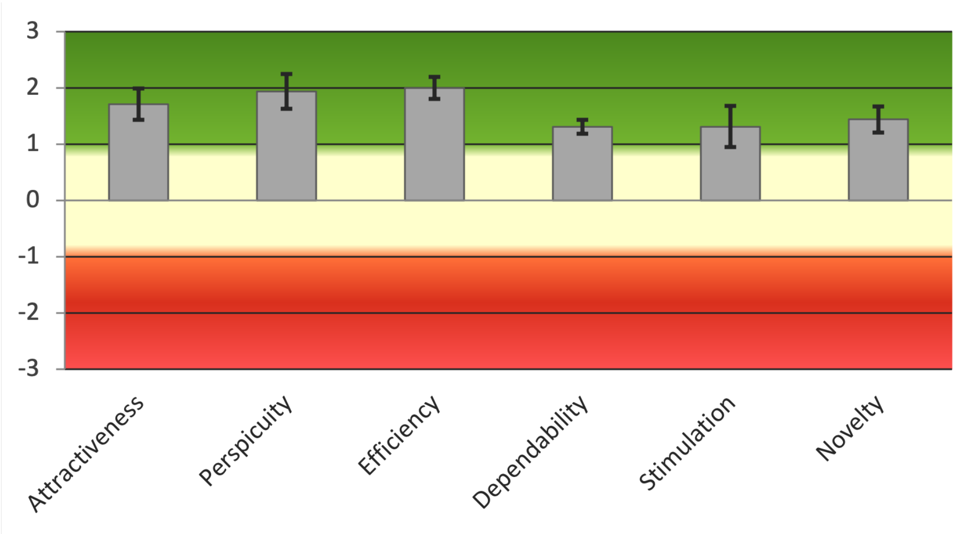

Figure 5.1: Bar chart displaying mean values and error bars of the UEQ subscales. The error bars represent the 95% confidence intervals displayed in table 5.1.

The analysis shows that all mean values including their confidence intervals (p=0.05) are in the range between 0.9 and 2 (table 5.1). This suggests that the system was overall perceived positively rather than neutral ($>$ -0.8 and $<$ 0.8) or even negatively ($<$ -0.8).

The scales that stand out most are ‘Perspicuity’ and ‘Efficiency’ with average values of 1.9 and 2.0, respectively, as visualised in figure 5.1. While ‘Perspicuity’ accounts for items such as understandability, learnability, ease and clarity of use, ‘Efficiency’ accounts for factors such as speed, task efficiency, practicability and system organisation of the chatbot prototype [48]. This indicates that the strongest features of the prototype are its ease of use and its practical application. This might be due to the low complexity of the chatbot interface powered by RASA X, the familiarity with chatbots (section 3.2.2.8) and the implementation of simple conversational structures. The lowest rated scales were ‘Dependability’ and ‘Stimulation’ with a mean value of 1.31 each. While items addressing ‘Dependability’ stand for the perceived reliability and safety of the system, ‘Stimulation’ stands for user engagement and motivation. Notably, values higher than 0.8 are still regarded as ‘good’ rather than neutral. However, these findings suggest that the perceived robustness, security and engaging incentives require more emphasis in future improvements. Notably, the sample size was very low (N=4) and therefore error bars (figure 5.1) indicating the 95% confidence interval turned out large for some scales such as ‘Stimulation’ (Confidence Interval= 0.94 - 1.68). To increase the precision of the estimated mean values and thus, the reliability of the evaluation study, a larger sample size is required.

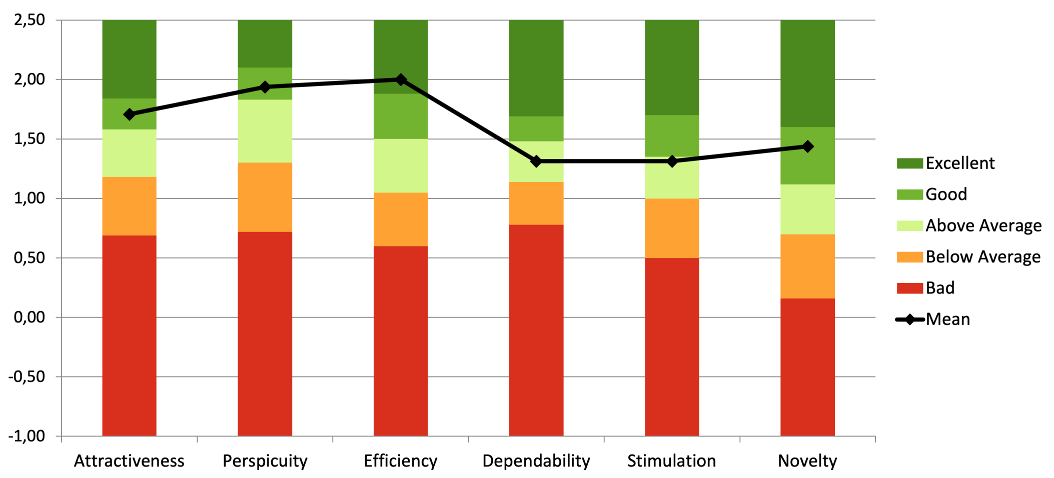

Figure 5.2: Bar Chart Displaying How the Chatbot Was Assessed Against the UEQ Benchmark. The black line represents the average UEQ scores for the chatbot prototype. The coloured bar sections refer to the benchmark intervals shown in the legend on the right.

Figure 5.2 shows how the chatbot prototype compared to 468 other digital products such as software programmes and web applications [72]. It is noteworthy that all scales were rated at least above average (only 25% of products in the UEQ dataset achieved higher ratings), with "Efficiency" being rated excellent, representing 10% of the best results within the UEQ dataset [72].

5.2 Chatbot Usability Questionnaire

On the basis of the System Usability Scale (SUS) for conventional interaction systems, the Chatbot Usability Questionnaire (CUQ) was developed to assess the usability of chatbots on a scale from 0-100 [38]. This allows the comparison with industry standards and benchmarks generated by SUS. Moreover, multiple measurements are helpful to generate a more holistic understanding of a system's usability [44]. In particular, the CUQ evaluates chatbot specific features such as the chatbot's character, onboarding and comprehension of user input [38]. The questionnaire consists of 16 statements that were rated from strongly disagree (1) to strongly agree (5) as illustrated in table 5.2. The calculated CUQ score for the TASK-CBT chatbot prototype is 69.5 ± 8.5 with the lowest score being 62.5 and the highest score being 81.3. The median score (67.2) is slightly lower than the mean value. Compared to the SUS benchmark [50], a score of 69.5 sits just above the median value of 241 previously assessed computer-human systems (68). This means that 50% of prototypes were rated higher than 68. A score of 69.5 falls into the grading scale 'C' which is classified as marginally acceptable [7].

No.	Question	Mean Value
1	The chatbot's personality was realistic and engaging	3.8±0.5
2	The chatbot seemed too robotic	2.5±1.0
3	The chatbot was welcoming during initial setup	4.3±0.5
4	The chatbot seemed very unfriendly	1.5±0.6
5	The chatbot explained its scope and purpose well	3.5±1.3
6	The chatbot gave no indication as to its purpose	2.5±1.3
7	The chatbot was easy to navigate	3.8±1.3
8	It would be easy to get confused when using the chatbot	2.3±0.5
9	The chatbot understood me well	3.3±1.0
10	The chatbot failed to recognise a lot of my inputs	2.0±0.8
11	Chatbot responses were useful, appropriate and informative	4.0±0.8
12	Chatbot responses were irrelevant	2.0±0.0
13	The chatbot coped well with any errors or mistakes	3.0±0.8
14	The chatbot seemed unable to handle any errors	2.8±1.0
15	The chatbot was very easy to use	4.3±0.5
16	The chatbot was very complex	1.8±0.5

Table 5.2: CUQ Mean Values of 16 Items. The scale for mean values and standard deviations range from 1 (strongly disagree) to 5 (strongly agree). Calculations were done accordingly to [38].

However, the calculated CUQ score cannot be blindly compared to SUS benchmark values. Although the evaluation concept behind CUQ is based on SUS calculations, it is rather intended for a direct comparison with a general usability evaluation such as the above shown UEQ (section 5.1). By comparing multiple metrics, usability issues were highlighted that would not be revealed under general assessments. While the UEQ scores indicate that the chatbot prototype performed much better than average systems, the CUQ score indicates that there are still significant usability problems to be solved. A closer look into the subscales reveals that one feature had particularly poor ratings. Participants neither agreed nor disagreed with statement 13 ('The chatbot coped well with any errors or mistakes', mean= 3.0 ± 0.8) and 14 ('The chatbot seemed unable to handle any errors', mean= 2.8 ± 1) as shown in table 5.2. This suggests that the chatbot intelligence was perceived as inferior and did not meet users' expectations regarding the robustness and error management of the system.

5.3 Retrospective Debriefing

After the completion of the questionnaire, participants were asked how they felt during the chatbot conversation and if they would use the chatbot in a post-stroke scenario. The goal of this was to further investigate whether participants felt confident using the chatbot prototype and which suggestions for improvement they have.

Limitations of Use Cases

All participants agreed that the chatbot was easy to use and that they felt confident during the interaction. In particular, they felt gradually more relaxed and reassured as the conversation went on. The participants also agreed that the chatbot was appropriate for recent stroke sufferers that require immediate support, as there is usually no mental support provided following the discharge from the hospital. Two participants concluded that the chatbot was appropriate for someone who required reassurance.

It's a reassurance. I suppose that's because it's a reassurance that there's a communication going on 'cause you can feel very isolated after it. (P23)

Furthermore, two participants stated that the chatbot was only appropriate for someone with minor strokes and few physical or cognitive impairments due to the model's limited understanding of user input. Moreover, it was questioned whether older people were digitally literate enough to use the application. Due to the low complexity of the chatbot, questions were raised about how the chatbot would respond to longer, detailed user inputs. Moreover, ethical concerns were raised such as instances where users would need to be referred to human clinicians for more advanced support.

Chatbot Language

Two participants reported that the chatbot response patterns seemed repetitive and could be perceived as ’talking down’ to stroke survivors, especially for someone with Aphasia. One participant further noted that the degree of emotion required from the chatbot probably varies from person to person and from culture to culture. Another participant mentioned personalisation in a similar context: the chatbot should be able to adjust its features according to different user preferences.

Chapter 6

Discussion

The aim of this project was to design a TASK-CBT chatbot for PSA patients based on user requirements identified through initial user research (section 3.2.3). The evaluations were conducted to assess the user experience and usability of the chatbot against the project objectives and user expectations. While the chatbot achieved above average, satisfying results according to UEQ scores (section 5.1), the CUQ score revealed features that should be improved (section 5.2). Notably, no major errors or misunderstandings were detected during the live usability tests. Interestingly, both the UEQ and CUQ revealed that efficiency and clearity were among the strongest features of the chatbot prototype. The capability to cope with errors on the other hand was rated much lower. Importantly, these findings only reflect the engagement of users with the chatbot prototype during the live usability testing session. Due to time constraints, participants did not have the possibility to interact with the chatbot beforehand to test multiple conversation pathways. Due to the brief interaction between users and the chatbot prototype, crucial shortcomings and error pathways may have remained undetected. Moreover, the live usability testing session was conducted remotely via screen sharing, therefore an unmoderated approach was not feasible. The researcher's presence, observations and briefings have possibly influenced the behaviour or thoughts of users during the interaction which also affected the evaluation questionnaires.

Notably, only four participants evaluated the chatbot. Therefore, the mean values calculated for UEQ and CUQ display broad confidence intervals and standard deviations (table 5.1 and table 5.2), meaning that dispersion among participants was high. Therefore, the generated results are less certain. To generate statistically more robust findings, more participants should be included in quantitative evaluation studies. Moreover, recorded, unmoderated usability evaluations could be implemented to observe how participants engage with the chatbot in their familiar environment.

In this project however, the evaluations were followed up by a retrospective debriefing session where certain usability aspects were addressed by participants. Due to time limitations, the retrospective debriefing questions were kept short. The most important findings dealt with the language of the chatbot and its limited capability to respond appropriately to complex user inputs. Future development of the chatbot requires larger data samples to train the chatbot model and a more precise definition of user scenarios for a personalised chatbot experience. This includes better understanding of different user groups such as elderly people, young stroke survivors, and patients with specific cognitive impairments. Furthermore, it is suggested to run Co-Designing workshops with stroke survivors and CBT therapists to generate and critically assess useful chatbot responses and actions.

6.1 Ethical Considerations

This project also raises ethical questions. During the evaluations, one participant questioned whether the chatbot should provide the functionality to handover the interaction to a human clinician. Because the chatbot deals with PSA patients, sensitive topics are touched upon. Although the chatbot is intended for patients with mild to moderate PSA symptoms, serious mental health crises cannot be ruled out. The precise detection of alarming contents [86] and subsequent prompts urging users to call the emergency number in such cases should be considered in future developments. In addition to clinical symptom reduction, future trials should also assess any possible long-term harms caused by mental health chatbots [46].

6.2 Conceptual Issues

One conceptual issue identified during the final implementation phase in section 4.2.3 was the categorisation of anxious situations 'Overthinking', 'Health Anxiety', 'Commuting' and 'Social Activities' (figure 4.8). While overthinking and worrying were mentioned as anxious situations in TASK-CBT datasets retrieved from the TASK randomised controlled trials [18] and during semi-structured interviews (section 3.2.2.6), it does not represent a triggering situation per se. However, it is plausible that triggering thoughts appear in health anxiety patients without any external influences or conditions such as commuting or social gatherings. In fact, patients with health anxiety can be preoccupied with intrusive thoughts and worries by just recalling an unwanted traumatic event [63]. It is therefore not clear whether the predominant fear of suffering another stroke ('health anxiety') is the underlying cause for PSA patients to also develop Agoraphobia or whether Agoraphobia is the predominant disorder right from the outset. Notably, a study conducted by Rudaz [67] examined the relationship between these two disorders and has found that health anxiety can influence the onset of Agoraphobia. To provide more specific feedback and personalised chatbot responses, it is suggested to investigate co-morbidity of health anxiety and Agoraphobia. This does not only influence the precision of the chatbot but also the degree of reassurance provided. Although study participants have expressed their need for reassurance during treatment (section 3.2.2.9), in health anxiety patients, this 'safety-seeking behaviour' [32, p.619] can lead to exactly the opposite of the desired effect. A patient feeling relieved for a short-time after seeking reassurance [32, p.619] that the experienced bodily symptoms are not characteristic for a stroke might reinforce the anxiety in the long run. Therefore, the personalisation of the chatbot, including the degree of reassurance, depends on the exact subtype of the anxiety disorder.

Chapter 7

Conclusions

This study examined the user requirements of PSA patients as well as their attitudes towards technology. Building on identified user needs, previous work from Lohse [51] and the TASK-CBT protocol developed by Chun et al. [18], a chatbot protoype was developed to guide PSA patients through the basics of CBT and TASK-CBT exercise 1. Furthermore, this project builds on the premise that conventional therapy for post-stroke patients is difficult to access and that a TASK-CBT chatbot can support patients instantly after their discharge accompanied by weekly telemedical therapy sessions with clinicians. Previous work has investigated the attitude towards technology and mental health chatbots in younger non-stroke populations [51]. This study provides comprehensive results in elderly [+55 years] stroke patients with PSA. Moreover, the implemented chatbot was assessed by a group of stroke physicians and stroke survivors. It was found that the participants possessed high levels of digital literacy and they were more likely to use smartphones and tablets than computers. Moreover, they had an overall positive attitude towards chatbots for mental health, whilst having a good understanding of the technical limitations. In particular, relevant chatbot features identified included empathetic language, reassurance, some degree of personalisation and a professional attitude. The implemented chatbot was perceived as friendly, useful and efficient. Evaluations have further revealed that the chatbot is lacking complexity and has difficulty handling errors. Furthermore, conceptual work (section 6.2) and a more precise understanding of different user personalities and impairments is required to personalise the chatbot experience. Additionally, the chatbot can be used to accumulate more training data to increase chatbot robustness and reliability.

Bibliography

[1] Alaa A Abd-Alrazaq, Mohannad Alajlani, Ali Abdallah Alalwan, Bridgette M Bewick, Peter Gardner, and Mowafa Househ. An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics*, 132:103978, 2019.

[2] Alaa A Abd-Alrazaq, Mohannad Alajlani, Nashva Ali, Kerstin Denecke, Bridgette M Bewick, and Mowafa Househ. Perceptions and opinions of patients about mental health chatbots: Scoping review. *Journal of medical Internet research*, 23(1):e17828, 2021.

[3] Eleni Adamopoulou and Lefteris Moussiades. An overview of chatbot technology. In Ilias Maglogiannis, Lazaros Iliadis, and Elias Pimenidis, editors, *Artificial Intelligence Applications and Innovations*, pages 373–383, Cham, 2020. Springer International Publishing.

[4] Berit Ahlsiö, Mona Britton, Veronica Murray, and Tores Theorell. Disablement and quality of life after stroke. *Stroke*, 15(5):886–890, 1984.

[5] Marie Almkvist Muren, Matthias Hütler, and Julie Hooper. Functional capacity and health-related quality of life in individuals post stroke. *Topics in stroke rehabilitation*, 15(1):51–58, 2008.

[6] Gerhard Andersson, Pim Cuijpers, Per Carlbring, Heleen Riper, and Erik Hedman. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. *World Psychiatry*, 13(3):288–295, 2014.

[7] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.

[8] Aaron T Beck. Cognitive therapy: Nature and relation to behavior therapy. *Behavior therapy*, 1(2):184–200, 1970.

[9] Ekaba Bisong. *Matplotlib and Seaborn*, pages 151–165. Apress, Berkeley, CA, 2019.

[10] Tom Bocklisch, Joey Faulkner, Nick Pawlowski, and Alan Nichol. Rasa: Open source language understanding and dialogue management. *arXiv preprint arXiv:1712.05181*, 2017.

[11] Niall M Broomfield, Terence J Quinn, Azmil H Abdul-Rahim, Matthew R Walters, and Jonathan J Evans. Depression and anxiety symptoms post-stroke/tia: prevalence and associations in cross-sectional data from a regional stroke registry. *BMC neurology*, 14(1):1–9, 2014.

[12] Niall M Broomfield, Anne Scoular, Peter Welsh, Matthew Walters, and Jonathan J Evans. Poststroke anxiety is prevalent at the population level, especially among socially deprived and younger age community stroke survivors. *International Journal of Stroke*, 10(6):897–902, 2015.

[13] Anders Bruun and Simon Ahm. Mind the gap! comparing retrospective and concurrent ratings of emotion in user experience evaluation. In *IFIP Conference on Human-Computer Interaction*, pages 237–254. Springer, 2015.

[14] C Alexia Campbell Burton, Jenni Murray, John Holmes, Felicity Astin, Darren Greenwood, and Peter Knapp. Frequency of anxiety after stroke: a systematic review and meta-analysis of observational studies. *International Journal of Stroke*, 8(7):545–559, 2013.

[15] Fatoumata Camara and Gaëlle Calvary. Worth-centered design in practice: Lessons from experience and research agenda. In Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2015*, pages 123–139, Cham, 2015. Springer International Publishing.

[16] Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. Best practices for designing chatbots in mental healthcare–a case study on ihelpr. In *Proceedings of*

the 32nd International BCS Human Computer Interaction Conference 32, pages 1–5, 2018.

[17] Ho-Yan Yvonne Chun, Alan J Carson, Martin S Dennis, Gillian E Mead, and William N Whiteley. Treating anxiety after stroke (task): the feasibility phase of a novel web-enabled randomised controlled trial. *Pilot and feasibility studies*, 4(1):1–11, 2018.

[18] Ho-Yan Yvonne Chun, Alan J Carson, Athanasios Tsanas, Martin S Dennis, Gillian E Mead, Clementina Calabria, and William N Whiteley. Telemedicine cognitive behavioral therapy for anxiety after stroke: proof-of-concept randomized controlled trial. *Stroke*, 51(8):2297–2306, 2020.

[19] Ho-Yan Yvonne Chun, William N Whiteley, Martin S Dennis, Gillian E Mead, and Alan J Carson. Anxiety after stroke: the importance of subtyping. *Stroke*, 49(3):556–564, 2018.

[20] Gilbert Cockton. Designing worth is worth designing. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles*, NordiCHI '06, page 165–174, New York, NY, USA, 2006. Association for Computing Machinery.

[21] Cheryl A Cott, Rose Wiles, and Rachel Devitt. Continuity, transition and participation: preparing clients for life in the community post-stroke. *Disability and rehabilitation*, 29(20-21):1566–1574, 2007.

[22] Simon D'Alfonso, Reeva Lederman, Sandra Bucci, and Katherine Berry. The digital therapeutic alliance and human-computer interaction. *JMIR Mental Health*, 7(12):e21895, 2020.

[23] E De Beurs, AJLM Van Balkom, R Van Dyck, and A Lange. Long-term outcome of pharmacological and psychological treatment for panic disorder with agoraphobia: a 2-year naturalistic follow-up. *Acta Psychiatrica Scandinavica*, 99(1):59–67, 1999.

[24] Mauro De Gennaro, Eva G Krumhuber, and Gale Lucas. Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. *Frontiers in psychology*, 10:3061, 2020.

[25] Liesbet De Wit, Peter Theuns, Eddy Dejaeger, Stefanie Devos, Andreas R Gantenbein, Eric Kerckhofs, Birgit Schuback, Wilfried Schupp, and Koen Putman. Long-term impact of stroke on patients' health-related quality of life. *Disability and rehabilitation*, 39(14):1435–1440, 2017.

[26] Diana Degan, Raffaele Ornello, Cindy Tiseo, Federica De Santis, Francesca Pistoia, Antonio Carolei, and Simona Sacco. Epidemiology of transient ischemic attacks using time-or tissue-based definitions: a population-based study. *Stroke*, 48(3):530–536, 2017.

[27] Nancy E. Mayo, Sharon Wood-Dauphinee, Sara Ahmed, Gordon Carron, Johanne Higgins, Sara Mcewen, and Nancy Salbach. Disablement following stroke. *Disability and rehabilitation*, 21(5-6):258–268, 1999.

[28] Fifth Edition et al. Diagnostic and statistical manual of mental disorders. *Am Psychiatric Assoc*, 21, 2013.

[29] Danielle Elmasri and Anthony Maeder. A conversational agent for an online mental health intervention. In *International Conference on Brain Informatics*, pages 243–251. Springer, 2016.

[30] Uwe Flick. *The SAGE handbook of qualitative data analysis*. Sage, 2013.

[31] Andrew T Gloster, Christina Hauke, Michael Höfler, Franziska Einsle, Thomas Fydrich, Alfons Hamm, Andreas Sthröhle, and Hans-Ulrich Wittchen. Long-term stability of cognitive behavioral therapy effects for panic disorder with agoraphobia: a two-year follow-up study. *Behaviour research and therapy*, 51(12):830–839, 2013.

[32] Brynjar Halldorsson and Paul M Salkovskis. Why do people with ocd and health anxiety seek reassurance excessively? an investigation of differences and similarities in function. *Cognitive Therapy and Research*, 41(4):619–631, 2017.

[33] Madeleine Harrison, Tony Ryan, Clare Gardiner, and Amanda Jones. Psychological and emotional needs, assessment, and support post-stroke: a multi-perspective qualitative study. *Topics in stroke rehabilitation*, 24(2):119–125, 2017.

[34] Ken Hillis, Michael Petit, and Kylie Jarrett. *Google and the Culture of Search*. Routledge, 2012.

[35] Jacqueline Hochstenbach, George Prigatano, and Theo Mulder. Patients' and relatives' reports of disturbances 9 months after stroke: subjective changes in physical functioning, cognition, emotion, and behavior. *Archives of physical medicine and rehabilitation*, 86(8):1587–1593, 2005.

[36] Stefan G Hofmann, Gordon JG Asmundson, and Aaron T Beck. The science of cognitive therapy. *Behavior therapy*, 44(2):199–212, 2013.

[37] Stefan G Hofmann, Anu Asnaani, Imke JJ Vonk, Alice T Sawyer, and Angela Fang. The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive therapy and research*, 36(5):427–440, 2012.

[38] Samuel Holmes, Anne Moorhead, Raymond Bond, Huiru Zheng, Vivien Coates, and Michael McTear. Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? In *Proceedings of the 31st European Conference on Cognitive Ergonomics*, pages 207–214, 2019.

[39] Jonas Hülsmann, Lennart J Sieben, Mohsen Mesgar, and Florian Steinke. A natural language interface for an energy system model. 2021.

[40] Balu N Ilag. Microsoft teams overview. In *Understanding Microsoft Teams Administration*, pages 1–36. Springer, 2020.

[41] HN Io and CB Lee. Chatbots and conversational agents: A bibliometric analysis. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 215–219. IEEE, 2017.

[42] Derek King, Raphael Wittenberg, Anita Patel, Zahid Quayyum, Vladislav Berdunov, and Martin Knapp. The future incidence, prevalence and costs of stroke in the uk. *Age and ageing*, 49(2):277–282, 2020.

[43] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian E Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica B Hamrick, Jason Grout, Sylvain Corlay, et al. *Jupyter Notebooks-a publishing format for reproducible computational workflows.*, volume 2016. 2016.

[44] Ahmet Baki Kocaballi, Liliana Laranjo, and Enrico Coiera. Understanding and measuring user experience in conversational interfaces. *Interacting with Computers*, 31(2):192–207, 2019.

[45] Matthias Kraus, Philip Seldschopf, and Wolfgang Minker. Towards the development of a trustworthy chatbot for mental health applications. In *International Conference on Multimedia Modeling*, pages 354–366. Springer, 2021.

[46] Kira Kretzschmar, Holly Tyroll, Gabriela Pavarini, Arianna Manzini, Ilina Singh, and NeurOx Young People's Advisory Group. Can your phone be your therapist? young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical informatics insights*, 11:1178222619829083, 2019.

[47] Rita V Krishnamurthi, Valery L Feigin, Mohammad H Forouzanfar, George A Mensah, Myles Connor, Derrick A Bennett, Andrew E Moran, Ralph L Sacco, Laurie M Anderson, Thomas Truelsen, et al. Global and regional burden of first-ever ischaemic and haemorrhagic stroke during 1990–2010: findings from the global burden of disease study 2010. *The Lancet Global Health*, 1(5):e259–e281, 2013.

[48] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*, pages 63–76. Springer, 2008.

[49] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. " i hear you, i feel you": Encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–12, 2020.

[50] James R Lewis and Jeff Sauro. Item benchmarks for the system usability scale. *Journal of Usability Studies*, 13(3), 2018.

[51] Kim-Morgaine Lohse. Designing a prototype chatbot for people with minor strokes.

[52] Gale M Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis-Philippe Morency. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI*, 4:51, 2017.

[53] Kien Hoa Ly, Ann-Marie Ly, and Gerhard Andersson. A fully automated conversational agent for promoting mental well-being: a pilot rct using mixed methods. *Internet interventions*, 10:39–46, 2017.

[54] Audrey Mariamo, Caroline Elizabeth Temcheff, Pierre-Majorique Léger, Sylvain Senecal, and Marianne Alexandra Lau. Emotional reactions and likelihood of response to questions designed for a mental health chatbot among adolescents: Experimental study. *JMIR Human Factors*, 8(1):e24343, 2021.

[55] Wes McKinney. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc.", 2012.

[56] Robert R Morris, Kareem Kouddous, Rohan Kshirsagar, and Stephen M Schueller. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. *Journal of medical Internet research*, 20(6):e10148, 2018.

[57] Rodrigo A Munoz. American psychiatric association practice guidelines for the treatment of psychiatric disorders: compendium 2000. *American Journal of Psychiatry*, 159(6):1066–1066, 2002.

[58] Jooyoung Oh, Sooah Jang, Hyunji Kim, and Jae-Jin Kim. Efficacy of mobile app-based interactive cognitive behavioral therapy using a chatbot for panic disorder. *International journal of medical informatics*, 140:104171, 2020.

[59] Kyo-Joong Oh, Dongkun Lee, Byungsoo Ko, and Ho-Jin Choi. A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation. In *2017 18th IEEE International Conference on Mobile Data Management (MDM)*, pages 371–375. IEEE, 2017.

[60] Józef A Opara and Krystyna Jaracz. Quality of life of post–stroke patients and their caregivers. *Journal of medicine and life*, 3(3):216, 2010.

[61] Nancy A Pachana. Ageing and psychological disorders. 2014.

[62] Adam Palanica, Peter Flaschner, Anirudh Thommandram, Michael Li, and Yan Fossat. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. *Journal of medical Internet research*, 21(4):e12887, 2019.

[63] Stanley Rachman. Health anxiety disorders: A cognitive construal. *Behaviour Research and Therapy*, 50(7):502–512, 2012.

[64] Lyn Richards. *Using NVivo in qualitative research*. Sage, 1999.

[65] Larry D Rosen, Kelly Whaling, L Mark Carrier, Nancy A Cheever, and Jeffrey Rokkum. The media and technology usage and attitudes scale: An empirical investigation. *Computers in human behavior*, 29(6):2501–2511, 2013.

[66] Peter M Rothwell, AJ Coull, MF Giles, SC Howard, LE Silver, LM Bull, SA Gutnikov, P Edwards, D Mant, CM Sackley, et al. Change in stroke incidence, mortality, case-fatality, severity, and risk factors in oxfordshire, uk from 1981 to 2004 (oxford vascular study). *The Lancet*, 363(9425):1925–1933, 2004.

[67] Myriam Rudaz, Michelle G Craske, Eni S Becker, Thomas Ledermann, and Jürgen Margraf. Health anxiety and fear of fear in panic disorder and agoraphobia vs. social phobia: a prospective longitudinal study. *Depression and Anxiety*, 27(4):404–411, 2010.

[68] Ralph L Sacco, Scott E Kasner, Joseph P Broderick, Louis R Caplan, JJ Connors, Antonio Culebras, Mitchell SV Elkind, Mary G George, Allen D Hamdan, Randall T Higashida, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the american heart association/american stroke association. *Stroke*, 44(7):2064–2089, 2013.

[69] Mark R Scholten, Saskia M Kelders, and Julia EWC Van Gemert-Pijnen. Self-guided web-based interventions: scoping review on user needs and the potential of embodied conversational agents to address them. *Journal of medical Internet research*, 19(11):e383, 2017.

[70] Henning Schöttke and Claire-Marie Giabbiconi. Post-stroke depression and post-stroke anxiety: prevalence and predictors. *International Psychogeriatrics*, 27(11):1805–1812, 2015.

[71] Margrit Schreier. Qualitative content analysis. In *The SAGE Handbook of Qualitative Data Analysis*, pages 170–183, London, Sep 2014. SAGE Publications Ltd.

[72] Martin Schrepp. User experience questionnaire. `https://www.ueq-online.org/`, 2018.

[73] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. Construction of a benchmark for the user experience questionnaire (ueq). *Int. J. Interact. Multim. Artif. Intell.*, 4(4):40–44, 2017.

[74] Rakesh Kumar Sharma and Manoj Joshi. An analytical study and review of open source chatbot framework, rasa. *International Journal of Engineering Research and*, 9(06), 2020.

[75] Abhishek Singh, Karthik Ramasubramanian, and Shrey Shivam. *Introduction to Microsoft Bot, RASA, and Google Dialogflow*, pages 281–302. Apress, Berkeley, CA, 2019.

[76] Colm Sweeney, Courtney Potts, Edel Ennis, RR Bond, Maurice Mulvenna, Siobhan O'Neill, Martin Malcolm, Lauri Kuosmanen, Catrine Kostenius, Alex Vakaloudis, et al. Can chatbots help support a person's mental health?: Perceptions and views from mental healthcare professionals and experts. *ACM Transactions on Computing for Healthcare*, 2(3):1–16, 2021.

[77] Robert W Teasell, Marc P McRae, and Hillel M Finestone. Social issues in the rehabilitation of younger stroke patients. *Archives of physical medicine and rehabilitation*, 81(2):205–209, 2000.

[78] Robert Thornberg, Kathy Charmaz, et al. Grounded theory and theoretical coding. *The SAGE handbook of qualitative data analysis*, 5:153–69, 2014.

[79] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S. Kashavan, and John Blake Torous. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, 64(7):456–464, 2019. PMID: 30897957.

[80] Maaike Van Den Haak, Menno De Jong, and Peter Jan Schellens. Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & information technology*, 22(5):339–351, 2003.

[81] Guido Van Rossum et al. Python programming language. In *USENIX annual technical conference*, volume 41, page 36, 2007.

[82] CP Warlow. Epidemiology of stroke. *The Lancet*, 352:S1–S4, 1998.

[83] Carla Willig. Interpretation and analysis 1. In *The SAGE Handbook of Qualitative Data Analysis*, pages 136–150, London, Sep 2014. SAGE Publications Ltd.

[84] Kevin Wilson. Microsoft office 2013. In *Using Microsoft Office 2013*, pages 1–5. Springer, 2014.

[85] Jacqui Wise. Stroke care audit shows major staff shortages. 2014.

[86] Bei Xu and Ziyuan Zhuang. Survey on psychotherapy chatbots. *Concurrency and Computation: Practice and Experience*, page 6170, 2020.

Appendix A

Ethics

A.1 Participant Information Sheet

accord
Academic and Clinical Central Office for Research and Development
NHS Lothian

Chatbot for Anxiety after St4oke
PIS v3.0_12_07_20
IRAS ID:264524

Participant Information Sheet

Co-designing a Chatbot for Anxiety after Stroke

A chatbot is a computer program that you can talk to – it asks you a question, and you type the response.

We hope to create a chatbot that can help people after stroke with their self-help exercises.

Have you experienced a stroke or 'mini-stroke'?

Would you be interested in helping us design a chatbot for stroke patients in a telephone research project?

- Please take time to read the following information carefully.
- We hope this information sheet will help you understand why the research is being done and what it will involve.
- Talk to others about the study if you wish.
- Contact us if there is anything that is not clear, or if you would like more information.
- Take time to decide whether or not you wish to take part.

What is the purpose of the study?

- We invite you to help us design a chatbot to improve the way we deliver support for people experiencing anxiety after stroke.
- We would like to design this chatbot together with people who have experienced a stroke or a mini-stroke (also called a TIA).
- We are inviting volunteers to have one telephone interview with our researchers.
- The interview will help us learn about:
  i. the ways you use technology
  ii. the features you would find helpful in a chatbot

Why have I been invited to take part?

- You have been asked to take part as you have received a diagnosis of stroke or 'mini-stroke' at least one month ago.

OR

- You previously took part in our stroke research and agreed to be contacted again for further research which might be relevant to you.
- We are aiming to recruit 10 participants with a spread of age groups into this study:

  At least 2 adults in age group <50
  At least 2 adults in age group 50-70
  At least 2 adults in age group >70

  The remaining 4 adults can fall into any of the above age groups

Do I have to take part?

- No. Taking part in this study is entirely voluntary.
- Deciding not to take part or withdrawing from the study will not affect the healthcare that you receive, or your legal rights.

How long do I have to decide to take part?

- You can take as long as you need to decide whether to take part.
- We aim to recruit participants from May to August 2020.
- Please contact us if you require more information to help you decide.

What will happen if I take part?

- We will ask you to complete an informed consent form online or by post.
- Our research team will contact you by telephone within 5 working days to confirm your eligibility and schedule your telephone interview.
- There is one telephone interview with our researcher in this study.
- The interviewer will guide you through a series of questions on

  i. the ways you use technology
  ii. how to make a chatbot useful for people with anxiety after stroke

- The telephone interview will last approximately 1.5 hours.
- You will be able to take breaks anytime you need.
- The interview will be audio-recorded using a digital recorder for transcribing into text for analysis in this study.

What if I have questions about my health?

- The interviewers are not medically trained. They will not be able to answer questions about your health.
- If you have questions regarding your health, we would advise you to contact your GP.

What are the possible benefits of taking part?

- There are no direct benefits to you taking part in this study, but the results from this study might help to improve the healthcare of patients in the future.
- Many people enjoy taking part in research.
- The results from this study could potentially contribute to the future development of a new treatment. Your participation in this study will not entitle you to benefit financially from the commercial development of the product or treatment.

What are the possible disadvantages of taking part?

- The study will take up an hour and a half of your day.

What happens when the study is finished?

- We will summarise and analyse the interview data.
- Audio-recording of the interview will be passed onto 1st Class Secretarial Services for transcription into text.
- We do not pass on your personal details to the transcription service.
- Transcripts do not contain personal details and will be erased after a year.

Will my taking part be kept confidential?

- All the information we collect during the course of the research will be kept confidential and there are strict laws which safeguard your privacy at every stage. For details on what data will be held about you and who will hold and store this information please refer to the Data Protection Information Sheet. This can be found on www.task4stroke.org.

What will happen if I want to withdraw from the study?

- Deciding not to take part or withdrawing from the study will not affect the healthcare that you receive, or your legal rights.
- We will use any data already collected with consent. No further data would be collected.

What will happen to the results of the study?

- This study will be written up as a publication for scientific journals and academic conferences.
- You will not be identifiable from any published results.
- Once published in a journal, we will also publish a public summary on our website: www.task4stroke-co-design.org

Who is organising and funding the research?

- This study has been organised jointly by the Centre for Clinical Brain Sciences and Informatics Forum at the University of Edinburgh.
- This study is sponsored by University of Edinburgh and NHS Lothian.
- The study is part of a Masters project in design informatics at the University of Edinburgh

Who has reviewed the study?

- All research in the NHS is looked at by an independent group of people called a Research Ethics Committee. A favourable ethical opinion has been obtained from <insert REC name>. NHS Management Approval has also been given.

Researcher Contact Details

- Dr Yvonne Chun, honorary research fellow and a medical doctor in stroke and medicine of the elderly, NHS Lothian.
  - Centre for Clinical Brain Sciences, 49 Little France Crescent, University of Edinburgh, EH14 4SB
  - ho-yan.chun@nhslothian.scot.nhs.uk

- Dr Maria Wolters, Reader in Design Informatics, School of Informatics,

Academic Associate of the School for Philosophy, Psychology, and Langauge Science, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, maria.wolters@ed.ac.uk, Tel 07914 600 458

- [Name to be inserted], MSc student in Design Informatics, School of Informatics, University of Edinburgh

What if there are any problems?

- In the unlikely event that something goes wrong and you are harmed during the research and this is due to someone's negligence then you may have grounds for a legal action for compensation against NHS Lothian but you may have to pay your legal costs. The normal National Health Service complaints mechanisms will still be available to you.

- If you have a concern about any aspect of this study please contact our research team who will do their best to answer your questions.

Complaints

- If you would like to speak to someone who is independent and not part of the research team, you can contact: Dr Sarah Keir, Consultant stroke physician, NHS Lothian. Sarah.keir@nhslothian.scot.nhs.uk

- Or contact Patient Experience Team, NHS Lothian, 2nd Floor, Waverly Gate, 2-4 Waterloo Place, Edinburgh, EH1 3EG (Tel: 0131 536 3370). feedback@nhslothian.scot.nhs.uk

A.2 Participant Consent Form

Academic and Clinical Central Office for Research and Development

TASK Chatbot
CF v3.0_12_07_2020
IRAS Project ID 264524

Participant ID: **Centre ID (if applicable)**

CONSENT FORM

Co-designing a Chatbot for Anxiety after Stroke

	Please **initial** box
1. I confirm that I have read and understand the information sheet (12/07/2020 Version 3.0 or the electronic information 12/07/2020 Version 3.0) and the Data Protection Information Sheet (30/04/2019 Version 1.0) for the above study. I have had the opportunity to consider the information, ask questions and have had these questions answered satisfactorily. (Link to documents: www.task4stroke-co-design.org)	☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without my medical care and/or legal rights being affected.	☐
3. I understand that relevant sections of data collected during the study may be looked at by individuals from the Sponsor (University of Edinburgh and/or NHS Lothian), from regulatory authorities or from the NHS organisation where it is relevant to my taking part in this research. I give permission for these individuals to have access to my data.	☐
4. I give permission for my personal information (including name, email, telephone number and consent form) to be passed to the University of Edinburgh for administration of the study.	☐
5. I agree to my telephone interview being audio recorded.	☐
6. I understand that data collected about me during the study may be converted to anonymised data.	☐

Participant ID: **Centre ID (if applicable)**

Please **initial** box

7. I agree to my audio-recorded interview being transcribed by a third party contractor- 1st Class Secretarial service. (No personal data are passed onto the third party) ☐

8. I understand that the data generated during this study may be used for future development of products/ treatments and I will not benefit financially from this. ☐

9. I agree to my anonymised data being used for future research. Yes ☐ No ☐

I agree to take part in the above study. ☐

Name of Person Giving Consent	Date	Signature
Name of Person Receiving Consent	Date	Signature

1x original – into Site File; 1x copy – to Participant

Appendix B

User Research

B.1 Structured Questionnaire

Confidential

TASK Chatbot
Page 1 of 5

Face to face interview structured questionnaire

Study ID	________________
Face-to-face Interview questionnaire **v1.0 30_4_2019**	
sex	◯ male ◯ female
Age in years	________________
Time since last stroke	(Time since last stroke (days/months/years))
Highest level of education	________________ (e.g. high school, university)
Occupation	________________ (If retired, your last occupation)

Use of technology:

Which of the following devices do you have? If yes, how often do you use it? (daily, weekly, rarely, never):

	daily	weekly	rarely	never / don't have one
Smartphone	◯	◯	◯	◯
Mobile phone	◯	◯	◯	◯
iPad or other tablet	◯	◯	◯	◯
Kindle or other eReader	◯	◯	◯	◯
Computer	◯	◯	◯	◯
Smart Watch	◯	◯	◯	◯
Digital TV	◯	◯	◯	◯
Activity Tracker (e.g. FitBit)	◯	◯	◯	◯
Video / DVD recorder	◯	◯	◯	◯
Games Console (e.g. Nintendo Wii)	◯	◯	◯	◯

How often do you do each of the following activities?

	all the time	several times a day	several times a week	several times a month	rarely	never
Watch TV shows, movies, etc. on a TV set						
Watch video clips etc. on a TV set						

How often do you do each of the following activities on your phone?

	all the time	several times a day	several times a week	several times a month	rarely	never
Send and receive text messages						
Make and receive mobile phone calls						
Check for text messages on a mobile phone						
Check for voice calls on a mobile phone						
Read email on a mobile phone (never if no smartphone)						
Get directions or use GPS on a mobile phone (never if no smartphone)						
Browse the web on a mobile phone (never if no smartphone)						
Listen to music on a mobile phone. (never if no smartphone)						
Take pictures using a mobile phone. (ask always)						
Record video on a mobile phone. (ask always)						
Use apps (for any purpose) on a mobile phone.						
Search for information with a mobile phone.						
Use your mobile phone while you do other activities.						

How often do you do each of the following activities?

(Mark as never if the person does not have a computer or tablet)

	all the time	several times a day	several times a week	several times a month	rarely	never
Watch TV shows, movies, etc. on a computer or tablet	◯	◯	◯	◯	◯	◯
Watch video clips on a computer or tablet	◯	◯	◯	◯	◯	◯
Download media files from other people on a computer or tablet	◯	◯	◯	◯	◯	◯
Share your own media files on a computer or tablet	◯	◯	◯	◯	◯	◯
Search the Internet for news (can be on any device, including library computers)	◯	◯	◯	◯	◯	◯
Search the Internet for information (can be on any device, including library computers)	◯	◯	◯	◯	◯	◯
Search the Internet for videos (can be on any device, including library computers)	◯	◯	◯	◯	◯	◯
Search the Internet for images or photos (can be on any device, including library computers)	◯	◯	◯	◯	◯	◯

Do you use social media (Twitter, Facebook, Instagram, Snapchat, etc.?)

◯ Yes
◯ No

If yes, how often do you use social media?

◯ never
◯ monthly
◯ weekly
◯ daily
◯ hourly

What is the main purpose?
(e.g. follow interesting accounts, talk to friends, keep up with family)

____

How many contacts do you have on social media?

____

How many people have you met online that you have never met in person?

____

How many people do you regularly interact with online that you have never met in person? ________________________________

Please state to what extent you agree with the following statements.

	strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
I feel it is important to be able to find any information whenever I want online.	○	○	○	○	○
I feel it is important to be able to access the Internet any time I want.	○	○	○	○	○
I think it is important to keep up with the latest trends in technology.	○	○	○	○	○
I get anxious when I don't have my mobile phone.	○	○	○	○	○
I get anxious when I don't have the Internet available to me.	○	○	○	○	○
I am dependent on my technology.	○	○	○	○	○
Technology will provide solutions to many of our problems.	○	○	○	○	○
With technology anything is possible.	○	○	○	○	○
I feel that I get more done because of technology.	○	○	○	○	○
New technology makes people waste too much time.	○	○	○	○	○
New technology makes life more complicated.	○	○	○	○	○
New technology makes people more isolated.	○	○	○	○	○
I prefer to work on several projects in a day, rather than completing one project and then switching to another.	○	○	○	○	○
When doing a number of tasks, I like to switch back and forth between them rather than do one at a time.	○	○	○	○	○

I like to finish one task completely before focusing on anything else.	○	○	○	○	○
When I have a task to complete, I like to break it up by switching to other tasks intermittently.	○	○	○	○	○

B.2 Semi-structured Interview Questions

TASK Chatbot

Semi-structured interview schedule

Coping with Stroke

How do you get information about stroke? (clinicians, other stroke survivors, carers, GP, online) Probe whether they are part of any online forums or support groups.

How do you get in touch with people who help you? Phone, letter, email …

Coping with Anxiety

What are their best strategies for coping with anxiety? What works for them? Listen for any strategies that are covered in the treatment manual

Follow up on observations from Media Usage scale.

If people have a smartphone/tablet

What are their favourite apps? When did they get the smartphone? What make and model is it? What do they like about it, what do they dislike about it?

If they had it before the stroke, did anything change after the stroke? (e.g. having to relearn how to use the device, using the device differently now, using different apps)

Do they use any health related apps? If yes, which? Mention as examples mindfulness, food intake, exercise tracker … If patient has an iPhone, check whether they use Apple Health.

If participant mentions gaming: What games do you play? Are any of those multiplayer games?

If people have a computer

What do they mainly use it for? When did they get it? What is the make and model? How long have they been using a computer for?

If they had it before the stroke, did anything change after the stroke? (e.g. having to relearn how to use the device, using the device differently now, using different apps)

Do they access any health related web sites? If yes, which? Mention as examples mindfulness, food intake, exercise tracker.

If people have a Smartwatch or an activity tracker

What do they mainly use it for, or did they stop using it? When did they get it? What is the make and model?

Follow up questions from Attitudes subscales

Ask people to expand on the key positive attitudes that they subscribe to
When was the last time they felt that technology was helpful? Was there a situation where technology made them feel calmer and less stressed / anxious?

Ask people to expand on the key negative attitudes that they subscribe to
When was the last time they felt frustrated with technology? Was there a situation where technology made them feel stressed or anxious?

How do they feel about talking to computers? Have they spoken to a computer before, in what circumstances?

What personality should a talking chatbot have?

B.3 Codebook

Main Categories

Code	Description	Files	References
Activity tracking	How and why stroke survivors started to monitor their fitness and health	6	24
Anxiety themes	This category applies to all anxious moments or situations mentioned by stroke survivors	5	13
Attitude towards exploring technologies	This category refers to participants' behavior when using web technologies or when exposed to technical novelties.	6	34
Curious and confident	Digital literacy, confident usage of different technologies or applications	4	19
Less risk tolerant	Keen to follow instructions instead of exploring features	3	4
Safety and anonymity	Perceive technology as a safe space due to anonymity	2	4
Technology as company	Relies on technology to feel less isolated	3	7
Attitudes towards operating systems	Attitude towards iOS, Android, and Windows systems	4	26
Chatbot experience	This category focuses on participants' experiences and existing knowledge about chatbots in general.	5	22
Computer model	Different computer models in use	4	8
Coping with post-stroke anxiety	Helpful technologies, applications and websites, physical and mental coping strategies	6	60
Expectations towards a TASK chatbot	Statements concerning the chatbot personality, usability features, and general attitude towards a TASK chatbot	6	74
Favourite apps	Participants' favorite mobile applications	6	15
Information retrieval	This category applies if participant explain their underlying reasons for why they gather relevant information about stroke, anxiety, or health from specific sources	6	37
Post-stroke changes in technology usage	Differences before and after stroke regarding technology usage	6	38
Preference among computers, laptops and tablets	This category deals with statements regarding the advantages and disadvantages of computers, laptops and tablets.	6	26
Preferred representational systems	Auditory and visual preferences	6	19
Smartphone	Statements dealing with smartphone models, positive and negative features	6	26
Technology frustration	Frustrating incidents related to technology usage	6	30
Willingness to share personal (stroke) information	This category applies whenever participants explain how they share personal information such as health status, health anxieties, general thoughts about their feelings in relation to the stroke.	6	21

Appendix C

Evaluation

C.1 Usability Evaluation Questionnaire

[1]

Please make your evaluation now.

For the assessment of the product, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the product. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

<u>Example:</u>

attractive	○	⊗	○	○	○	○	○	unattractive

This response would mean that you rate the application as more attractive than unattractive.

Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression.

Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular product. Nevertheless, please tick a circle in every line.

It is your personal opinion that counts. Please remember: there is no wrong or right answer!

[1]The UEQ was downloaded from www.ueq-online.org, a cooperation project developed by Martin Schrepp, Jörg Thomaschewski and Andreas Hinderks.

Please assess the product now by ticking one circle per line.

	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

C.2 Chatbot Usability Questionnaire

[2]

Chatbot Usability Questionnaire

Usage Guide

Thank you for downloading the Chatbot Usability Questionnaire, a new bespoke tool for measuring chatbot usability.

Please read these instructions carefully before using the questionnaire.

The CUQ consists of sixteen balanced questions related to different aspects of chatbot usability. Eight of these relate to positive aspects of chatbot usability, and eight relate to negative aspects. Scores are calculated out of 100. See the **Calculation** section of this guide for information on how to calculate CUQ scores.

This questionnaire may be administered in printed form, however for convenience you are free to digitise the questionnaire if you wish to use a web-based survey platform such as Qualtrics. You should consult the user guide for your chosen platform to determine the best way to do this.

How to use the CUQ

All sixteen questions are scored using a five-point Likert-type scale. Odd-numbered questions relate to positive aspects of the chatbot, and even-numbered questions relate to negative aspects. Respondents should read each item carefully and decide the extent to which they agree with the statement. Respondents should then indicate their level of agreement by placing a tick (✓) or a cross (×) in the circle that best matches how they feel about the statement.

Example

	Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5
The chatbot's personality was realistic and engaging	○	○	○	○	⊗

In the example above the respondent strongly agreed with the statement *The chatbot's personality was realistic and engaging*, therefore they marked circle number 5 (Strongly Agree).

[2]The CUQ was developed by: Samuel Holmes, Anne Moorhead, Raymond Bond, Huiru Zheng, Vivien Coates, and Michael Mctear. 2019. Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces?. In Proceedings of the 31st European Conference on Cognitive Ergonomics (ECCE 2019), Maurice Mulvenna and Raymond Bond (Eds.). ACM, New York, NY, USA, 207-214.

CHATBOT USABILITY QUESTIONNAIRE

Please complete this questionnaire by reading each statement carefully and placing a tick (✓) or a cross (×) in the circle that best matches how you feel about the statement. Remember that there are no right or wrong answers!

	Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5
The chatbot's personality was realistic and engaging	○	○	○	○	○
The chatbot seemed too robotic	○	○	○	○	○
The chatbot was welcoming during initial setup	○	○	○	○	○
The chatbot seemed very unfriendly	○	○	○	○	○
The chatbot explained its scope and purpose well	○	○	○	○	○
The chatbot gave no indication as to its purpose	○	○	○	○	○
The chatbot was easy to navigate	○	○	○	○	○
It would be easy to get confused when using the chatbot	○	○	○	○	○
The chatbot understood me well	○	○	○	○	○
The chatbot failed to recognise a lot of my inputs	○	○	○	○	○
Chatbot responses were useful, appropriate and informative	○	○	○	○	○
Chatbot responses were irrelevant	○	○	○	○	○
The chatbot coped well with any errors or mistakes	○	○	○	○	○
The chatbot seemed unable to handle any errors	○	○	○	○	○
The chatbot was very easy to use	○	○	○	○	○
The chatbot was very complex	○	○	○	○	○

Score Calculation

NOTE: For best results it is recommended that the CUQ Calculation Tool (available from the CUQ website) be used to calculate CUQ scores. However, should you wish to do this manually, please follow the instructions below (see page 4 for a worked example).

1. For each question, assign a score from 1 to 5 based on the level of agreement with the statement in the question (i.e. “Strongly agree” is worth 5 points, “Neutral” is worth 3 points, “Strongly disagree” is worth 1 point)
2. Calculate the sum of all the **odd-numbered** (positive) questions.
3. Calculate the sum of all the **even-numbered** (negative) questions.
4. Subtract 8 from the score you got at *step 2.*
5. Subtract the score you got at *step 3* from 40.
6. Add the scores you got at *steps 4 and 5*. You should now have a score out of 64.
7. Divide the score you got at *step 6* by 64 and multiply the answer by 100. This will give you a CUQ score out of 100.

Assuming you will be testing with multiple participants, you may wish to calculate the mean CUQ score, find the median score, and so on. This may be accomplished using the CUQ Calculation Tool or may be done manually using a spreadsheet package such as Microsoft Excel.

Example

1. Assume the CUQ questions have been scored as follows:

Question	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Score	*4*	*4*	*5*	*1*	*4*	*4*	*5*	*1*	*4*	*1*	*4*	*1*	*3*	*2*	*5*	*1*

2. Calculate the sum of all the **odd-numbered** questions:

   4 + 5 + 4 + 5 + 4 + 4 + 3 + 5 = 34

3. Calculate the sum of all the **even-numbered** questions:

   4 + 1 + 4 + 1 + 1 + 1 + 2 + 1 = 15

4. Subtract 8 from (2):

   34 – 8 = 26

5. Subtract (3) from 40:

   40 – 15 = 25

6. Add (4) and (5):

   26 + 25 = 51 (out of 64)

7. Convert (6) to a score out of 100:

   (51 / 64) * 100 = 79.7