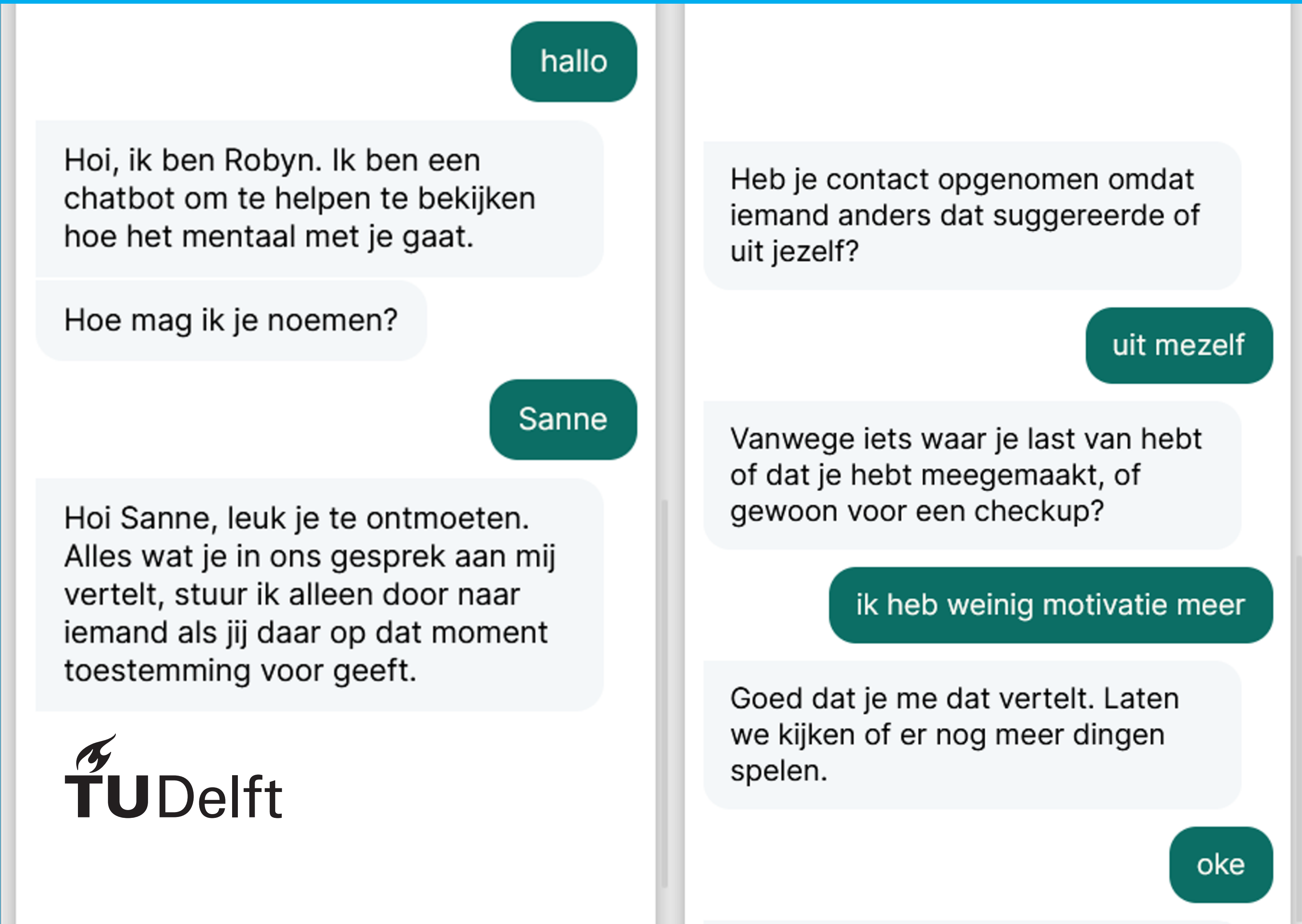
A Conversational Agent for Stress First Aid

Mental Well-being Awareness for Police Officers

R. Cromjongh

Master Thesis Report



hallo

Hoi, ik ben Robyn. Ik ben een chatbot om te helpen te bekijken hoe het mentaal met je gaat.

Hoe mag ik je noemen?

Sanne

Hoi Sanne, leuk je te ontmoeten. Alles wat je in ons gesprek aan mij vertelt, stuur ik alleen door naar iemand als jij daar op dat moment toestemming voor geeft.

Heb je contact opgenomen omdat iemand anders dat suggereerde of uit jezelf?

uit mezelf

Vanwege iets waar je last van hebt of dat je hebt meegemaakt, of gewoon voor een checkup?

ik heb weinig motivatie meer

Goed dat je me dat vertelt. Laten we kijken of er nog meer dingen spelen.

oke

TU Delft

A Conversational Agent for Stress First Aid

Mental Well-being Awareness for Police Officers

by

R. Cromjongh

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Monday October 17, 2022 at 2:00 PM.

Student number:	4460804	
Project duration:	August 30, 2021 – October 17, 2022	
Thesis committee:	Dr. W. P. Brinkman,	TU Delft, supervisor
	Dr. M. L. Tielman,	TU Delft, daily supervisor
	Dr. C. C. S. Liem,	TU Delft

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

TU Delft

Preface

This project has been a long and diverse journey. It all started with quite an open assignment by the police. It quickly grew into a straightforward develop and evaluate plan, but over the course of the project, due to scheduling and communication, the original evaluation plan became infeasible to execute. After a good deal of rescheduling, the evaluation part was scrapped and this thesis became a design paper. But after two months of rewriting, we determined that the work on preparing the evaluation was a significant part of my work so far and reformed again to also include the evaluation plan, but no execution. The thesis became a registered report. This is the report that you have in front of you today.

Looking back on the process today, the main problems in executing the original plan were planning and communication. We started off well with several meetings with the police, to determine the direction of the project and some initial planning. I then went off to get a good chunk of the work done before attempting to resume contact to setup the evaluation. Instead keeping more regular meetings might have prevented some of the problems I had later in re-obtaining contact. I could probably also have prevented some of the execution troubles by making strict agreements on evaluation with the police contacts earlier. For this, I could have selected a smaller evaluation structure that could be prepared earlier in advance. Having the evaluation plan set up earlier would have provided the police and me with more structure to work towards the execution.

Adapting to each new version of the goal has been tough. Both design papers and registered reports are not pieces of writing you encounter regularly during computer science studies. It has been very instructive to look into these other report formats and write them myself. In the end, I am happy with the report I have produced, and it now reflects all the parts I have been working on over the past year.

I would like to thank everyone that supported me through the journey, notably my supervisor Myrthe Tielman and PhDs Pei-Yu, Mohammed and Siddharth, and the other students in the thesis meetings Alan and Chadha. Thanks for all the feedback, proofreading, thinking along and more. I would also like to thank my friends and family. You listened to my struggles and provided advice, but were also there to provide necessary breaks from work. Also thanks to my contacts at the police department for giving the conceptual assignment for this project, shaping the direction of the project in the meetings at the early stages, providing officers and other relevant staff for feedback sessions and thinking along with practical setup issues for evaluation. Thanks to all the people that provided feedback on the design versions and helped make the final design better.

R. Cromjongh
Delft, October 2022

Contents

Abstract 1

1 Introduction 3
- 1.1 Mental Health within the Police Force . . . 3
- 1.2 Mental Health Problem Prevention . . . 3
- 1.3 Mental Health Chatbots . . . 5
- 1.4 Research Questions . . . 5
- 1.5 Approach . . . 5

2 Related Work 7
- 2.1 Existing Agents for Police and Mental Well-being . . . 7
- 2.2 Mental Well-being Awareness . . . 7
- 2.3 Conversational Agent Communication and Presentation . . . 8

3 Conversational Agent Design 11
- 3.1 Chosen Technologies . . . 11
- 3.2 Initial Design . . . 11
- 3.3 First Evaluation Police Peer Support Personnel . . . 13
- 3.4 Conversation Design . . . 14
  - 3.4.1 Peer support feedback . . . 14
  - 3.4.2 Psychologist feedback . . . 14
  - 3.4.3 Suicide experts . . . 15
- 3.5 Conclusion . . . 15

4 Evaluation Methodology 17
- 4.1 Participants . . . 17
- 4.2 Measures . . . 17
  - 4.2.1 Mental Well-being Awareness . . . 17
  - 4.2.2 Usability . . . 19
- 4.3 Procedure . . . 19
- 4.4 Data Processing and Analysis . . . 20
  - 4.4.1 Coding of questionnaires . . . 20
  - 4.4.2 Analysis of awareness . . . 21
  - 4.4.3 Analysis of perceived change . . . 22
  - 4.4.4 Analysis of usability . . . 22
  - 4.4.5 Additional analysis . . . 22
- 4.5 Results Interpretation . . . 22
  - 4.5.1 Usability results . . . 23
  - 4.5.2 Additional results . . . 23

5 Discussion 25
- 5.1 Agent Design Implications . . . 25
- 5.2 Subquestions . . . 26
- 5.3 Limitations . . . 27
- 5.4 Future Work . . . 27
- 5.5 Conclusion . . . 27

Bibliography 29

A List of terms used 35
- A.1 Mental health and well-being . . . 35
- A.2 Evaluation aspects . . . 35

B Conversation Design Diagram **37**

C Evaluation Survey **51**
C.1 UMUX-LITE 51
C.2 Perceived mental well-being survey 51

Abstract

Police officers are exposed to many potentially traumatising and stressful situations, but they do not always find the right mental health help they need. In this thesis, conversational agent Robyn is developed to help officers keep an eye on their mental well-being and find help when needed. An evaluation approach is also presented. The conversational agent applies the Stress First Aid method to determine stress indicators together with the officer.

The design of the agent followed the guidelines for Stress First Aid. The agent uses text input and output and went through an iterative design process using expert feedback. The experts provided feedback on both text and structure of the conversation. Based on their input, informal language was chosen for the presentation of the agent. The lessons for similar projects drawn from this design process are managing expectations from users, scoping broad questions, getting feedback from users on wording and not to push people when discussing suicide.

The proposed evaluation would determine if the agent is effective at improving people's awareness of their mental well-being status. This is done by comparing perceived mental well-being before and after interaction to reported mental well-being. To achieve this, a survey to measure perceived mental well-being was developed. The evaluation also provides insight in what aspects of the agent need the most improvement in future iterations.

1

Introduction

The Dutch National Police is looking for more ways to support the mental well-being of their employees, especially the younger officers. The police organisation already has a support network in place, but not all officers that need help realise that they need help or are willing to ask for help. To help solve this problem, for this thesis I build a technological solution to either help officers find the right help, or improve their general mental resilience or well-being. I subject this solution to scrutiny by experts for feedback and I present an approach to evaluate the agent.

A previous thesis student worked with the same assignment and is building a chatbot called Robyn, which will help officers improve communication with family or coworkers. This chatbot could be extended to include the new functionality, but is currently in parallel development.

1.1. Mental Health within the Police Force

Due to their line of work, police officers experience more traumatic incidents (in most literature called critical incidents) than the average civilian. It has been acknowledged that police officers are more likely to develop mental and physical disorders [20]. In the US and other countries including Canada, the UK, Australia and New Zealand, 19 to 24% of law enforcement officers show signs of PTSD [10, 27], 11 to 19% show signs of problematic drinking [10, 27], and 9% shows signs of depression [10].

Interestingly, Dutch police officers do not show signs of PTSD or suicidality more often than Dutch civilians, and show only slightly higher levels of burn-out and depression [41]. It should be noted that for PTSD, large differences were found between different task groups within the police force, the branch Intake & Service reporting 30% of officers showing increased levels of PTSD symptoms. Despite on average not reporting any higher risk, these mental disorders are very serious conditions that police officers are exposed to in a unique way, by experiencing traumatic events more commonly than an average civilian.

Next to being exposed to traumatic incidents, police officers also suffer from other stressors, such as social and organisational stressors in the workplace, like rejection, abuse, discrimination and sexism; and everyday work stress. Those other stressors sometimes cause more stress than traumatic incidents do [27]. As such, much can be improved in the workspace, for example reducing workload and improving support from managers [41].

Police departments provide support for their officers suffering from mental health problems, but not everyone makes use of the offered help. Many officers report concerns over confidentiality, the impact on their career and stigmas about seeking help as barriers to make use of this support [10].

1.2. Mental Health Problem Prevention

Several high impact mental disorders can occur as a result of stress experienced. Event related stress can cause Acute Stress Disorder (ASD), Posttraumatic Stress Disorder (PTSD), or Adjustment Disorder [39]. Other disorders that can result from stress are anxiety disorders [26].

Disorders related to experienced events, like PTSD, can be prevented in different stages: Primary prevention is focused on either preventing the trauma causing the disorder from happening or decreasing risk factors. Secondary prevention covers all interventions that occur early after the trauma, and aims to improve outcomes of those that experienced the trauma. Tertiary prevention attempts to prevent the disabilities that can occur after developing the disorder [43]. Finally, there is treatment of the disorder symptoms itself. The preferred stage to cover with this project is either primary or secondary prevention, as tertiary prevention and treatment are already well covered by existing support structures.

In all the prevention stages, researchers are looking into pharmacological treatments, most of which still need more research before they can be widely applied [15, 31]. These could also be applied in combination with any other treatment or prevention technique. Because of lack of proof, medications are not yet used regularly for prevention of mental disorders.

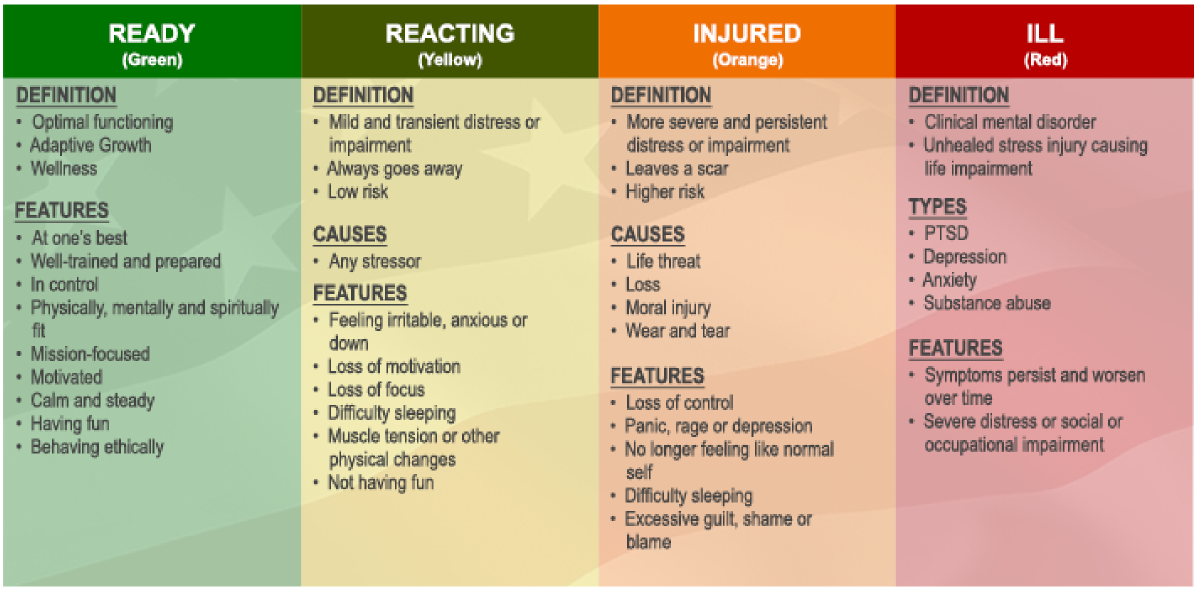
A police department can implement primary prevention by offering training and education to improve resilience [24, 27] and improve the work environment by reducing social stressors [27].

In secondary prevention, psychological debriefings used to be the standard in law enforcement and emergency services, but this has shown to be ineffective and in some cases even counter effective in recovery [15, 31]. Now other approaches are being attempted.

Exposure Therapy (ET) is a proven treatment for PTSD, but it is also showing promise as a prevention technique when administered shortly after experiencing trauma or with patients showing Acute Stress Disorder (ASD) symptoms [15]. The long term effectiveness of this approach still needs to investigated.

Another secondary prevention method is Stress First Aid (SFA) [43]. This technique was developed for the US military and aims at recognising your own and colleagues' stress levels and then acting to reduce these [3]. In this way, it aims to prevent PTSD, anxiety, depression and substance abuse [3]. The technique is evidence informed but has not yet been evaluated. Regardless, it is a promising technique that has been used by the US marines and navy since 2007 and 2008 respectively [44]. Since then, versions have been developed aimed at law enforcement, firefighters, health care workers and others [5].

Stress First Aid is a low barrier method that can be applied by everyone. This relative simplicity makes it suitable for converting to a technological implementation. Because of this and its preventative promise, Stress First Aid is the method chosen to be used in this project. The first step of this method is checking symptoms to find out what someone's stress level is. For this, it uses the Stress Continuum Model (see figure 1.1), which shows a scale of healthy to unhealthy stress levels.



READY **(Green)**	**REACTING** **(Yellow)**	**INJURED** **(Orange)**	**ILL** **(Red)**
DEFINITION • Optimal functioning • Adaptive Growth • Wellness **FEATURES** • At one's best • Well-trained and prepared • In control • Physically, mentally and spiritually fit • Mission-focused • Motivated • Calm and steady • Having fun • Behaving ethically	**DEFINITION** • Mild and transient distress or impairment • Always goes away • Low risk **CAUSES** • Any stressor **FEATURES** • Feeling irritable, anxious or down • Loss of motivation • Loss of focus • Difficulty sleeping • Muscle tension or other physical changes • Not having fun	**DEFINITION** • More severe and persistent distress or impairment • Leaves a scar • Higher risk **CAUSES** • Life threat • Loss • Moral injury • Wear and tear **FEATURES** • Loss of control • Panic, rage or depression • No longer feeling like normal self • Difficulty sleeping • Excessive guilt, shame or blame	**DEFINITION** • Clinical mental disorder • Unhealed stress injury causing life impairment **TYPES** • PTSD • Depression • Anxiety • Substance abuse **FEATURES** • Symptoms persist and worsen over time • Severe distress or social or occupational impairment

Figure 1.1: The Stress Continuum Model as used in SFA [9].

Becoming aware of your stress level is the first step in being able to do something about it. In the Trans-theoretical Model of Behaviour Change, gaining awareness is called consciousness raising, and can take someone from the precontemplation to contemplation stage or from contemplation to preparation, if change is needed [30]. There is also some preliminary evidence that patients with a mental illness that have an improvement in awareness of their symptoms, have better treatment outcomes [28]. Later steps in SFA follow up on this awareness with action. These steps are more complicated to convert to software and will have to build on the results of the first step. Thus, this project will focus on implementing the first step only.

1.3. Mental Health Chatbots

As mentioned before, there are several barriers that keep people from seeking help. Technological solutions can offer low barrier and large scale access to health care [43]. One step in this direction is online therapy, which has been shown to be as effective as face-to-face therapy [25]. Taking this one step further are mental health chatbots. Such chatbots have existed for quite some time already and recent bots are effective at screening for or treating different disorders such as: bipolar disorder, PTSD, suicidality, depression, anxiety; and for different user groups such as: youth, cognitive impaired individuals, refugees, students [25]. Next to being effective, users also report liking the chatbots and feeling that they are helpful and usable [40], which also improves their adherence to treatment.

By being able to run on easily accessible technology such as smartphones, mental health chatbots can improve access to care [40]. It also helps people that are reluctant to speak to a human therapist, for example because of stigmas [40]. Most bots are assessed based on features, ease of use, cost, friendliness and perceived benefit [25], while many projects have not looked into actual effectiveness of the treatment.

Given the above, a conversational agent solution to promote mental health will reduce some of the barriers, for example stigma and negative career impact of seeking help, by being discreet and anonymous. Since the SFA method is mostly conversation based, it is well suited to be adapted to a conversational agent design. One aspect that needs special attention in the development of technological solutions like this, is personal data safety and confidentiality [40].

1.4. Research Questions

Given the information above, the goal of this project is to make a conversational agent that helps people apply the Stress First Aid technique. The goal of the Check action is to determine someone's stress level, so the goal of the agent will be to help someone determine their own stress level, in effect becoming more aware of their own mental well-being status. This project will answer two related research questions, on how to design such an agent and how to evaluate it:

RQ1: *What would a design look like of a conversational agent for helping police officers become aware of their mental well-being status by applying the Check action of the Stress First Aid model?*

RQ2: *How could a conversational agent for awareness of one's mental well-being status be evaluated?*

There are several subquestions that relate to different steps in solving these questions:

1. How can we determine improvement of the mental well-being awareness of a person?
2. What is the best communication medium for the agent and the conversation to assess the mental well-being status?
3. How can the agent best be presented within the chosen medium for an effective conversation?
4. How can a conversational agent go through the check action together with the officer in a respectful and efficient way?
    (a) How can the agent best help the officer get an overview of their potential stress features?
    (b) Are there any information sources other than the officer the agent can utilise?
    (c) How can the agent help the officer review the features and determine their stress level and next steps?

1.5. Approach

To answer the research questions and each subquestion, different approaches are needed.

To answer subquestions one and two, I examined literature. I looked into existing research investigating mental well-being awareness and literature examining the effect of communication medium on conversational agents. This work can be found in Chapter 2 Related work. The outcomes of this literature search are used in the design of the agent.

For the design of the agent, I used an iterative design method, where an initial high-level design was created based on the guidelines of Stress First Aid. The goal of this high-level design is to make sure all the elements of the method are present and that they follow each other logically. Once this first high-level design was done, it was fleshed out into the design of the actual conversation. Following this, groups of experts provided feedback on the design. Between each feedback round, the design was updated according to the feedback. Throughout the design process, subquestions three and four are answered. The design and expert feedback are described in Chapter 3.

To answer RQ2, an evaluation plan was written. This involved creating a suitable measure for awareness, including developing a new questionnaire for determining perceived mental well-being. An experiment setup was established and submitted to the ethics committee. An analysis plan was also formed. For this, a method was made to compare the different survey results of perceived well-being and reported mental health and illness. Lastly, potential results and related conclusions were described. The evaluation proposal can be found in Chapter 4.

Finally, Chapter 5 will reflect on the design process and lessons for other projects, and the use and limitations of the evaluation plan, as well as explore future work. It also covers the conclusions of this thesis.

2

Related Work

This chapter looks into existing conversational agents and how this project relates to them. It also covers existing work to explain the goal of the agent and to find answers to subquestions one and two.

2.1. Existing Agents for Police and Mental Well-being

The Dutch national police has previously developed a chatbot called Wout. Wout was an alternative way for the public to contact the police and has been used to take complaints about noise and fireworks, reports of cybercrime and to answer questions about the corona virus [4, 42]. The trials with Wout have finished, and the police is now having a more elaborate chatbot developed [42].

While Wout is a chatbot made by the police for the public, a police department in Los Angeles has created a chatbot called Chip to answer questions from potential police recruits [7]. Another agent created by US police helps prevent crime and make arrests, by pretending to be a prostitute online [33]. None of these agents developed for or by police organisations target police officers.

Then there is also conversational agents aimed at improving mental well-being, although they are aimed at a more general user group than police officers. An early attempt at a psychology chatbot is ELIZA, a rule based agent that follows the methods used by Rogerian psychotherapists [45].

With improvements in technogology and natural language processing came more advanced mental well-being chatbots. General conversational agents for mental well-being that are well studied in literature are Woebot [8], Wysa [13] and Tess [11]. Woebot applies Cognitive Behaviour Therapy (CBT) in short daily conversations and mood tracking [8]. Wysa uses a collection of self-help practices including CBT, motivational interviewing, positive behaviour support and other techniques and aims to improve emotional resilience skills [13]. Tess is a customisable agent for mental health interventions using CBT, emotionally focused therapy, motivational interviewing and other techniques [11]. All of these agents focus on reducing symptoms of depression, anxiety or other mental illnesses, though the user need not be diagnosed.

One project attempted to treat PTSD using a virtual agent. This agent administers a CBT with Exposure Therapy included, by guiding the patient in recollecting memories in a personal diary and a virtual 3D environment [38].

The previous mental well-being agents are aimed at treatment of illnesses. There are agents that try to prevent mental well-being issues, such as Vincent, which aims to improve on mental resilience. Vincent aims at improving self-compassion, which in turn causes better mental well-being. Two variants of Vincent were made, one which gives exercises for self-care and another that shares stories and compels users to express care for Vincent [19]. Only the second version significantly improved self-compassion. This is in line with other work that suggests that caring for others practices caring for yourself [19]. None of the agents mentioned in this section explicitly help users to become more aware of their symptoms or their status.

2.2. Mental Well-being Awareness

The goal of the conversational agent is to help police officers become aware of their mental well-being status. But what exactly is mental well-being and when are you aware of your mental well-being status?

And how can it be measured?

As the mental health continuum used in SFA shows, there is an area between good mental health and mental illness, shown as the yellow and orange zones (see figure 1.1). Mental well-being is determined by the presence or absence of mental illness and by the extent of mental health [18]. The term ‘mental health’ is often used to refer to the overall concept of mental well-being. For the purpose of this thesis, ‘mental well-being’ refers to the overall mental state, ‘mental illness’ refers to clinically diagnosable illnesses and ‘mental health’ refers to the everyday mental state with respect to your happiness and stability. These definitions can be consulted in Appendix A.

To be aware of your mental well-being status, you have to have an accurate picture of where you stand in terms of mental illness and mental health. In this thesis, awareness is defined as the accuracy of your belief about your mental status with respect to your actual status. The level of your awareness is then determined by comparing your perceived status to your actual status.

Most research investigating mental well-being awareness evaluate this using a mental well-being knowledge survey [2, 34]. Others check self reported perception of awareness [46] or help-seeking intentions or behaviour [2, 14, 34]. One paper investigated the effect of their intervention on awareness by comparing the answer to “Do you think you have a mental health problem?”, before and after the intervention, to the results of an emotional mental health survey [14]. All of these methods do not actually measure the accuracy of your belief (awareness). A more extensive survey of the perceived state, compared to a mental well-being assessment, is needed to accurately measure the awareness of one’s own mental state.

If one is to compare to the actual mental well-being, this also needs to be measured. There are many questionnaires to measure mental illness [37]. For the evaluation of this project, I do not need a specific diagnosis or a very detailed analysis. A shorter measurement tool that gives a general overview is therefore sufficient. Many short mental illness measurements are evaluated against the M.I.N.I. [37], which is a diagnostic questionnaire that takes approximately 15 minutes to fill out [36] and has been translated into Dutch [29]. The M.I.N.I. is well validated and often used, but takes quite long for the evaluation of this project. There is also a shorter variant, the M.I.N.I.-Screen, which takes about 5 minutes to complete [36]. Although the M.I.N.I.-Screen does not appear to be used without the full M.I.N.I. [37], the substantial support for its parent measure indicates that it will give a good impression of the presence of mental health disorders.

The concept of mental health can be measured by the Positive Mental Health Questionnaire (PMHQ) [32] or the Mental Health Continuum Short Form (MHC-SF) [16, 18]. The PMHQ is a 39-question form that is validated in Spanish and Portuguese [32], but has not yet been translated to Dutch. Although Roldán-Merino et al. argue that MHC-SF does not actually measure positive mental health [32], MHC-SF is used more often to measure the mental health aspect of mental well-being and has been translated and validated in Dutch [18]. The areas that they measure are fairly similar, and while MHC-SF is shorter, it is sufficient for the goal of this project and better supported as a measurement.

As stated, my research did not reveal sufficient existing measurements for one’s perceived mental well-being state. To measure this aspect, I made a new questionnaire partially based on an existing measure. For details, see Chapter 4.

2.3. Conversational Agent Communication and Presentation

In order for a conversational agent to be most effective, it needs to use the right communication modality and the right looks and feel to best connect with the user.

Of conversational agents built to support mental well-being, most use written input, or speech input (sometimes combined with other input options) [1]. As an output modality, most use speech, visuals and text combined [1]. In most cases, the agent was given a visual representation. Regardless of these commonly used modalities, very little is currently known about the impact of different modalities, as few studies compare the same agents with different modalities [40]. A non-interactive support agent was shown to be more effective at supporting users on a task level when using speech to talk to the user compared to using text only [35], but this tells us little about an interactive agent. In overview, there is little evidence to guide us in picking the right communication modality for a mental well-being support agent.

In terms of presentation, Maharjan et al. found that a mental well-being screening agent using speech input and output got more expressive answers and more context when using open-ended questions, but users were more habituated to closed-ended questions [23]. They suggest aiming more for closed-ended questions and use follow-up questions to get more details. This makes the agent more robust and habitable.

3

Conversational Agent Design

This chapter looks into the design of the conversational agent, including the rounds of expert feedback and their influence on the final design. Their feedback has also been compared to existing literature. The design was made in the form of a conversation flow diagram. This diagram has been subjected to scrutiny from police peer support personnel and police psychologists, before being refined to the final version.

3.1. Chosen Technologies

For a lack of effectiveness information on the communication modality (see Section 2.3), I decided on the modality that is easiest to implement. This is text input and text output.

In terms of what technology to use, this gives me the first requirement of supporting text based input and output. Since the target group is Dutch police officers, and the agent should communicate with them in a language they are proficient in, the technology should support Dutch language. I also want the interaction with the agent to feel like a real conversation, as opposed to a survey. Hence, the agent should support and act on free-form input. Mental well-being is a sensitive topic, as mentioned in Section 1.3, therefore, the technology should safeguard privacy. A final added bonus would be implementation ease.

The chosen conversational agent technology is Rasa Open Source[1]. Rasa is an open source, machine learning based conversational agent library written in Python. It supports any language you provide examples for, and can be run on a private server, ensuring data is not sent anywhere inappropriate. Other platform options did not provide the same security or they required payment. Rasa Webchat[2] is used for the graphical interface with the agent. For the code and versions of software used, see the repository on GitHub[3].

3.2. Initial Design

The goal of the agent is to help users through the first step of the Stress First Aid method. The high-level design covers all the main aspects of the SFA method. A diagram describing the high-level flow of the conversation is shown in figure 3.1.

The conversation starts off with greetings and introductions. In the introduction, the agent explains its purpose. After the introduction, the agent attempts to determine the situation the user is in. To find out what indicators of stress are present, the agent goes over the different indicators together with the user in a conversational manner, asking them about each one. Finally, the agent collects all the information from the conversation up to this point and makes an estimate of the stress zone the user is in. It then discusses its conclusions with the user and recommends next steps the user can take. Throughout the conversation, if the agent senses danger, it addresses that explicitly. The aspects indicating danger that can occur in the conversation flow are thoughts about suicide and the potential

---

[1] `https://rasa.com/open-source/`

[2] `https://github.com/botfront/rasa-webchat`

[3] `https://github.com/Robin66/mental_wellbeing_awareness_agent`

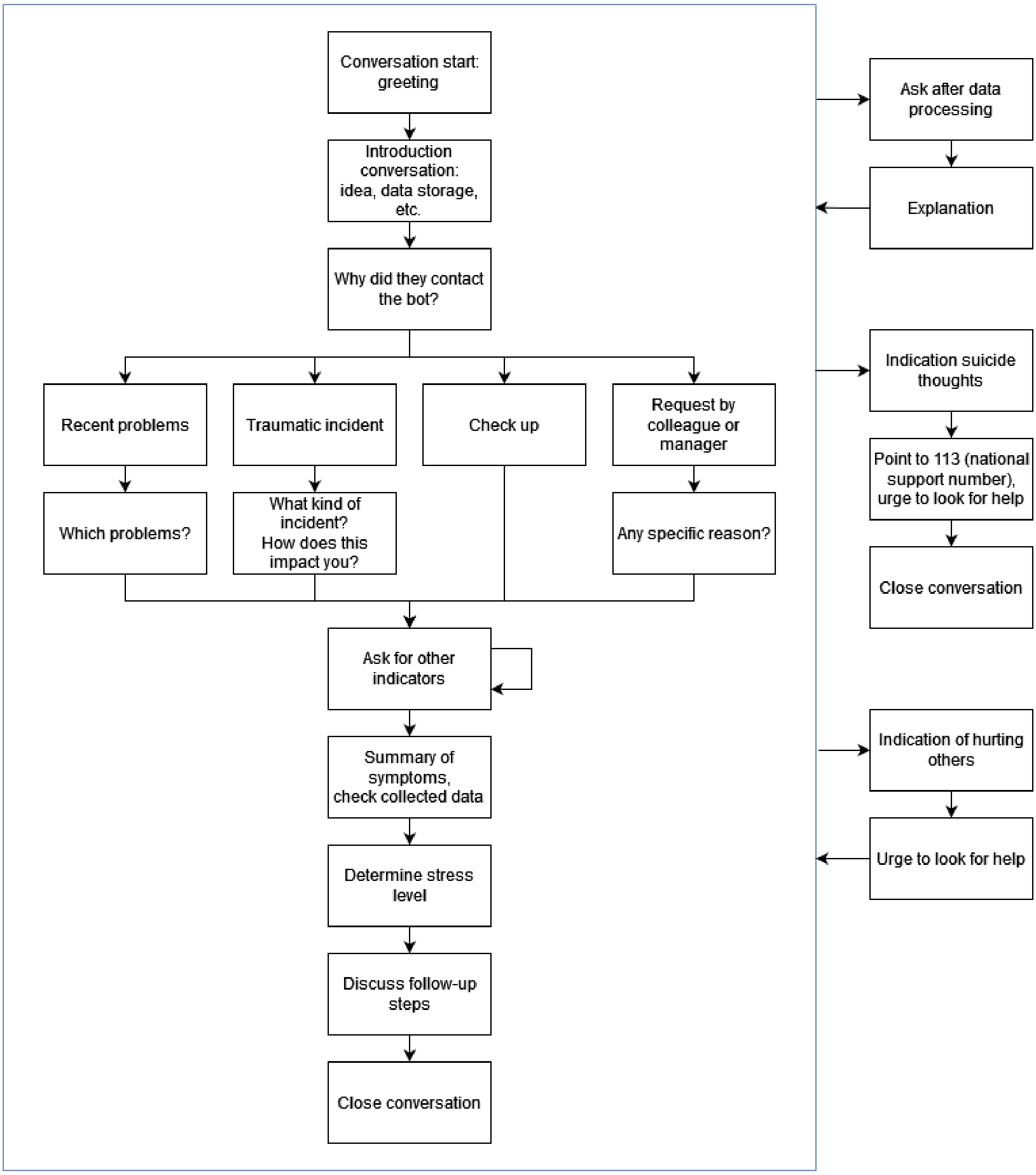

Conversation start: greeting
Introduction conversation: idea, data storage, etc.
Why did they contact the bot?
Recent problems
Traumatic incident
Check up
Request by colleague or manager
Which problems?
What kind of incident? How does this impact you?
Any specific reason?
Ask for other indicators
Summary of symptoms, check collected data
Determine stress level
Discuss follow-up steps
Close conversation
Ask after data processing
Explanation
Indication suicide thoughts
Point to 113 (national support number), urge to look for help
Close conversation
Indication of hurting others
Urge to look for help

Figure 3.1: High-level overview of the structure of the ideal conversation flow. The main flow goes from start to close within the blue box, but can be interrupted at any point with topics (on the right) not anchored in the main flow.

of hurting others. If these things come up anywhere in the conversation, the agent takes a step out of the conversation flow to discuss these with the user, before going back to where it left the flow.

This design follows the instructions for executing the check action of SFA. In the SFA manual for law enforcement, the check action consists out of [3]:

- Observing the situation
- Keeping track of indicators of stress zones and stressor events
- Examining the gathered information
- Deciding what stress zone is involved and potential next steps

This manual does not go into much detail about how to execute these steps. An earlier manual aimed at the military by Watson et al. offers more guidance [44]. The process described is slightly different from the one aimed at law enforcement, but includes the same principles in addition to covering aspects that are not so relevant to the police. It also offers some general guidance on the conversation on this topic. The military manual gives the following two insights:

Firstly, while checking for stress indicators, you should also look for signs of danger, in order to immediately address those. At the same time, if signs of intense negative emotions or physiological hyperarousal are observed, one should take action to calm down the individual. This is accounted for in the design by looking out for and immediately dealing with signs of hurting oneself or others.

Secondly, while one talks to the individual, “it is crucial at this point to establish rapport through empathic listening, compassion, and gentleness, while also establishing a working alliance by informing the individual exactly why you are talking to them, and what information you have that makes you concerned about them” [44, p. 19]. According to the manual, this can be done through observing, factually stating your observations, clarifying why you are addressing the behaviour, asking for clarification and responding to the information and stating options. If possible, one should gather additional information from other sources like colleagues, family, managers or friends. The design takes this conversation approach into account by stating its purpose at the start of the conversation, thus explaining the context of the conversation, and by explaining its reasoning throughout the conversation. To keep the project feasible and the conversation intuitive, I chose not to include external sources of information like existing databases or reaching out to other people than the user. The agent thus does not rely on obtaining external information and is instead self-contained.

The next step of SFA, Coordinate, involves informing those that need to know about the current status and getting assistance for the individual if this is required [3]. This step is not the focus of this project, but covering some conclusion and recommendations makes for a full and complete conversation. So the agent conversation covers all the elements of the check action to get a good indication of the stress level of the user and follows that up with a conclusion with recommendations on what to do next for the user. This way the user not only gets more insight into their mental well-being but also a first step on what to do with that knowledge.

3.3. First Evaluation Police Peer Support Personnel

In a first informal evaluation of this design, 3 peer support officers were recruited. In an interview of an hour, I presented the high-level design to them and we discussed the flow, realism and user experience of the conversation as well as choices for the presentation of the agent. The interview was recorded by taking notes. The police officers did not have any feedback on the flow, realism, and user experience of the design. They indicated that they were happy with the structure of the conversation in this design.

For subquestion three, to find out how the agent should be presented, I discussed what would be suitable with the group. I asked for their views on word use and visual representation of the agent. From an informal analysis of the conversation it became clear that an informal agent would fit best in the existing police culture. So informal language use was chosen and any visual representation should have an informal feel. To prevent the agent feeling like a colleague, in the sense of carrying stigmas, the agent should not wear a uniform in visual representations. The participants recommended a friendly robot-like figure that looks trustworthy.

3.4. Conversation Design

The high-level description of the diagram was worked out into specific conversation snippets in a conversation flow diagram. The chosen phrasings take into account the informal conversational tone decided upon and the guidelines for conversation from the manual, by explaining why the conversation is taking place and how the agent reasons and by asking for more details on certain answers. The final conversation design can be found in Appendix B.

The conversation goes as follows: The agent introduces itself and explains its purpose and conversation privacy. It then asks for the reason the user contacted the agent and sees if it can already get any stress indicators from that. When it gets into asking after all the indicators, I chose to go over the mild indicators first and cover the orange zone indicators only if mild indicators are present. The reason for this is to not burden the user immediately with questions about depression and suicide, but cover more everyday problems first. The assumption I make here is that if severe problems are present, symptoms that can be seen as mild (yellow zone) will also be present, such that severe symptoms only need to be covered if there is evidence of mild symptoms. Novacek and Raskin show that most of the people with severe mental illnesses exhibit at least ten warning signs in the days leading up to hospital admission [28]. Several of the most common warning signs considered in their paper are present in the mild symptom list. A few users with severe symptoms might be missed this way, but this choice also makes the agent more usable for the average user.

The agent gives positive encouragements and empathic statements after the user confirms a symptom and with some regularity after answers denying a symptom. In conversational agents, empathy from the agent has been shown to increase user satisfaction and effectiveness [47]. During the summary of the found symptoms and conclusions, the agent asks the user for their view to keep this process collaborative. The user can then correct mistakes in the symptom list and express dissatisfaction with the conclusion. In this scope of the design, no steps are included if the user is not satisfied with the conclusion. Finally, the conversation is closed off with goodbyes.

3.4.1. Peer support feedback

To evaluate the full conversation design, the same three peer support officers were contacted as in Section 3.3. The design was presented to them via a PDF attached to an email which also included the questions. The participants provided their feedback through email. In the email I requested feedback on user input realism, the text used by the agent and the structure of the conversation. For evaluation their textual answers were combined and summarised. All the feedback provided by these officers and the following experts and the resulting changes have been summarised in table 3.1.

The police provided feedback mostly on agent text usage and a little on the structure of the conversation. Their feedback was that the reasons to contact the agent can be so broad and unpredictable, that it might be hard to cover them all. They suggested determining the situation through several clear yes/no type questions. This is in line with the recommended approach by Maharjan et al. for closed-ended questions with follow up, which users find easier to answer [23]. The participants also mentioned not using the phasing “I will consider that” (in Dutch: “Ik neem het mee”) when replying to provided info, since this can be seen in a negative light and suggesting that your answer will in fact not be taken into account.

Subsequently, I removed this phasing from the design. I reworked the situation question from one main question with explanation if requested to a branching sequence of questions with two answer options each time, to determine what broad situation category the user is in.

3.4.2. Psychologist feedback

To evaluate the psychological appropriateness of the design, two psychologists were recruited through the police. The design was sent to them through email in advance and this was then discussed in an online interview of one hour through Microsoft Teams. The interview was free-flowing and was recorded via note-taking. I asked specifically for feedback on wording and structure of the conversation. An informal analysis of the conversation yielded the following points:

The introduction should more clearly state what the agent is and can do, that it will send automated answers, and that it is not a medical professional. This should ensure the users do not have the wrong expectations of the agent and do not expect help to be arranged for them while the agent is not capable of that yet.

There was also a name for a concept that is already in use for a similar concept elsewhere in the mental well-being support structures of the police, so that had to be renamed to prevent confusion. Lastly, they wanted a more elaborate answer to the question about suicide, since this is such a heavy and important topic. They suggested adding the phone number and website of the Dutch suicide prevention organisation 113 and seeking help with the wording of the reply from this organisation.

Subsequently, I elaborated the introduction to more clearly state the agent's capabilities and limitations and reworded the duplicate term ‘checkup’ to ‘selfcheck’.

3.4.3. Suicide experts

To get the conversation about the suicide topic right, as recommended by the psychologists, I contacted people from 113 suicide prevention, a Dutch suicide prevention hotline. In an email I explained the situation and provided a PDF of the suicide part of the conversation. Two employees of 113 were contacted, and one replied via email.

He said the statements were too strong (e.g. stating that “things can always improve” is incorrect). He suggested providing options and potentially recommending those options, but not trying to convince the user to take steps.

With this feedback, I rephrased the statement about the situation improving to being able to look for help together if you talk to someone. I referenced their website and phone number and also pointed to other sources of help. I also restated that the agent is not able to arrange help, since the user may have forgotten by this point and expect help once they have told the agent.

3.5. Conclusion

Several types of experts had a look at the detailed conversation design. They each had feedback on aspects of the agent that were the most relevant to them and their input resulted in several improvements of the conversation design, summarised in table 3.1. After incorporating all feedback, the psychologists approved of the changes and were satisfied with the design. For the final, complete conversation diagram (in Dutch), see Appendix B. A video demonstrating the agent in action can be found on YouTube[4].

---

[4] Link to the video: `https://youtu.be/drbDjspgUmA`

Feedback source	Feedback	Change	Before	After
Peer support personnel	Change the question about your situation through a sequence of closed questions.	Change the general situation question to a series of two option questions.	“Why did you contact me today?”	“Did you contact me on your own or on advice of someone else?” “Because of something that bothers you or something you experienced, or just for a checkup?” / “Was that because they thought you are not doing so well?”
	Do not use phrasing like “I will consider that” that can have a negative tone.	This phrasing was removed.	E.g. “That is unfortunate, I will consider that”	E.g. “That is unfortunate” / “Good that you tell me this”
Police psychologists	The agent introduction should more clearly state the capabilities and limitations of the agent.	Elaborate the introduction to clearify capabilities and limitations.	“Hi, I am Robyn. I am a chatbot to help you track your mental well-being.”	“Hi, I am Robyn. I am a chatbot to help you track your mental well-being. But I am not a medical professional and cannot diagnose you. Because I am a chatbot, I send automated answers and might not always understand you correctly. There is also no human reading along, so no-one else will take action based on what you tell me.”
	Rename ‘checkup’, since this term is already used elsewhere in the mental health support structure.	Replace ‘checkup’ with ‘selfcheck’.	“Because of something that bothers you or something you experienced, or just for a checkup?”	“Because of something that bothers you or something you experienced, or just for a selfcheck?”
	Give a more elaborate and helpful reply to users indicating they have thought about suicide.	Add some more empathy, add 113 website and phone number, remind the user that the agent cannot arrange help, contact 113 for help phrasing this answer.	“That is very unfortunate. [...] Please contact 113 suicide prevention, they can really help.”	“That is very unfortunate, how nice that you are so open about that. [...] Please contact 113 suicide prevention through 0800-0113 or 113.nl. [...] Because of the setup of this experiment, I can unfortunately not arrange help for you myself. So do talk with someone about this.”
113 suicide support	Do not make promises of improvement to suicidal people.	Change promising improvement to with help seeking improvement.	“Know that there is always help available, it really can get better.”	“Know that there is always help available, then you can look for a different solution together.”
	Provide options without trying to convince users.	Add other sources of help to the provided options.	“Please contact 113 suicide prevention, they can really help. There is also a lot of information on their website.”	“For help or someone to listen, you can contact 113 suicide prevention through 0800-0113 or 113.nl. There is also a lot of information on their website. You can also contact your GP, the (company) psychologist or a colleague or friend.”

Table 3.1: All the feedback provided by the various experts and the resulting changes made.

4

Evaluation Methodology

To determine how effective the agent is at achieving its goal, the agent needs to be evaluated on its effectiveness at making users more aware of their mental well-being status. This chapter presents the proposed design for a user study. This study has a pre-test post-test design and aims to determine the awareness of one's actual mental well-being status before and after interacting with the agent, by comparing the perceived mental well-being status with the actual reported status.

4.1. Participants

Participants would be recruited through the police contacts. Seventy police officers would be invited to participate. The participants would be selected from different task groups within the departments. To determine how many participants are needed, a G*Power test was conducted for a two-tailed Wilcoxon matched-pairs signed-rank test with normal distributions, an effect size of 0.5 and a power of 0.95. This resulted in a required number of participants of 57. To account for people declining to participate and missing data, the choice was made to invite seventy officers instead. Demographic information that would be collected are age and gender.

The experiment setup has been approved by the Human Research Ethics Committee of Delft University of Technology, ID nr: 2121. Potential ethical concerns with the setup of the experiment were anonymity of participants and their dependence on the recruiters, security of mental health data, and support for emerging mental health issues.

First, the anonymity and dependence concerns were addressed by only publishing limited data about the participants, also to their employer, by limiting to average age and gender percentages, which are only published if re-identification is not possible. It was agreed upon with the recruiting persons that participation is voluntary and this is also made clear to participants. Recruiters are not told who ended up participating. Second, data security is ensured by storing personal identifiers separately from survey answers. All data will be deleted once relevant publications have completed. Finally, mental health issues might come up over the course of the experiment. The concerns about supporting participants with this are addressed by providing a list of support instances with contact details and a business psychologist on stand-by during the experiment.

4.2. Measures

This section describes all the different measures taken over the course of the experiment. The translated questionnaires can be found in Appendix C.

4.2.1. Mental Well-being Awareness

There are three components to determining if mental well-being awareness of the participant is improved by the prototype. These three components are perceived mental well-being, actual mental health and actual mental illness. Mental health and illness make up mental well-being, so those combined can be compared to the perceived mental well-being. Mental health refers to the everyday aspect of your mental state, mental illness refers to diagnosable illnesses and mental well-being refers to one's overall mental state. Awareness is defined as the accuracy of your belief about your mental status with respect to your actual status. All these definitions and names of measurements used in this study can be consulted in Appendix A.

Perceived mental well-being status

Perceived mental well-being is measured before and after interacting with the conversational agent, in order to see if the perception changes. As described in Section 2.2, no existing measures were found that measure the perception of the participant on mental well-being. Therefore, a short questionnaire was created to get the participant's view on their mental well-being. This questionnaire is to be filled in before and after interacting with the agent, the second time including additional questions on whether they think their awareness changed. The questionnaire can be found in English in table 4.1 or in Dutch in Appendix C.

The topics of questions 1-11 are ordered in the following way: one's general mental state, help seeking, more severe issues with mental well-being and finally knowledge. Questions a and b are included only the second time the participant fills in the questionnaire. These questions aim to measure the participant's perception of change. All questions use a 7-point scale, except for question 1, which uses a 10 point scale, and questions a and b, which use a -1 to 1 valued slider. The reliability of this questionnaire would be assessed with Cronbach's alpha.

Question 1 was designed to get a general overview of how participants view their mental well-being. Questions 2 through 4 are based on the first three questions of the MDT part of the Quality of Life for Homeless and Hard-to-House Individuals (QoLHHI) Inventory [12]. This questionnaire is meant to index the overall well-being of homeless and hard-to-house individuals, on several topics that are important for this target group, which includes health. It was also developed to allow the use of any subset of the questions. The questions themselves are not specifically worded for this target group. Thus, the questions be reapplied to other target groups. In this case, the questions were also reworded to refer to mental well-being instead of health.

The remaining questions 5 - 11 in the survey of perceived mental well-being were created to get a more granular insight in a person's perception. This is done by looking at the perception on different aspects of mental well-being: problem perception, help seeking and knowledge. Question 5 looks into the long term average of their experience of mental well-being. Questions 6 and 9 determine if the participant thinks there is a mild problem. Question 10 determines if they think there is a more severe problem. Questions 7 and 8 look into the participant's current perception of need for help. Finally, question 11 looks into the participant's perception of their knowledge on mental well-being symptoms.

Questions a and b assess the participant's perception of change after interacting with the agent. They are only present the second time the questionnaire is filled in, which is after interacting with the agent. These questions are used to see whether the participant thinks they have learned over the course of the interaction with the agent. When they learn, people sometimes realise they know less about something than they thought. Because of this, the results of questions such as question 11 about perceived insight might go down while participants still think they learned things. That learning is what questions a and b try to capture.

Reported mental health and illness

The reported mental well-being status is measured at the end of the experiment. Both of the questionnaires used to measure the mental well-being status are self report measures relying on the participant to provide information on their symptoms. In this sense both of these questionnaires are subjective, but since they are based on symptom experiences instead of a high-level evaluation of the symptoms, they are closer to the actual status. Mental well-being is hard to measure objectively, so subjective questionnaires like these are the closest approximation. Throughout this thesis, these measures will be referenced as 'reported mental well-being'.

As determined in Section 2.2, mental well-being consists of mental health and mental illness. Each is measured with their own questionnaire. Mental illness is measured with the Dutch version of the M.I.N.I. screen [29]. Mental health is measured with the Dutch version of the MHC-SF [18].

The reported status questionnaires are not filled in before the perceived status survey, since the act of filling in these questionnaires could influence participant's perception of their mental well-being. The short time frame of the experiment of at most an hour makes it unlikely the actual mental well-being changes much over the course of the experiment, so only a post-test measurement is taken.

#	Question	Scale
a	My insight in my mental well-being has	strongly decreased - strongly increased
b	My knowledge over what symptoms indicate my mental well-being has	strongly decreased - strongly increased
1	How would you rate your mental well-being at this moment	1-10
2	On the whole, how would you describe your current mental health	terrible - excellent
3	On the whole, how do you feel about your current mental health	very dissatisfied - very satisfied
4	How does your current mental health compare to the average person's mental health	far below average - far above average
5	I think on average I am doing well mentally	strongly disagree - strongly agree
6	I think recently I have a higher stress level than I should have	strongly disagree - strongly agree
7	I think I should talk to someone (friend/colleague/partner/etc) about my thoughts or stress	strongly disagree - strongly agree
8	I think I should talk to mental expert about my mental well-being	strongly disagree - strongly agree
9	I sometimes worry about my mental state	strongly disagree - strongly agree
10	I think I have a mental illness	strongly disagree - strongly agree
11	I think I have a good insight into possible stress symptoms	strongly disagree - strongly agree

Table 4.1: The questions and scales used in the perceived mental well-being questionnaire. All scales use a 7 point likert scale, except for question 1, wich uses a 10 point scale and questions a and b, which use a -1 to 1 valued slider.

The reported well-being status is needed to compare the perceived status to the reported status, to determine the level of awareness.

4.2.2. Usability

Although it is not part of the research questions, it is relevant to get an overall idea of the usability of the prototype. Even if the agent is very effective, it is still useless if users do not feel comfortable using it. To get an impression of the usability, the UMUX-LITE questionnaire is used [22]. This scale was chosen because with two questions, it is very short, but still delivers validated outcomes [21, 22]. This would give a generic insight into the overall usability of the chatbot, while not prolonging the experiment substantially.

As suggested by Lewis, the unadjusted score of the UMUX-LITE is used, instead of the regression formula originally proposed [21]. The scale has been translated to Dutch and slightly adapted to refer to 'the chatbot' instead of 'the system'. The resulting Dutch questions can be found in Appendix C.

4.3. Procedure

This section describes the procedure of the study. This procedure can be executed in a remote setting or in a physical setup.

The participants are scheduled to participate in the study. They first get information about the study, including where to find help if the mental health topic brings worries or negative emotions, and have to fill in informed consent and demographics questions. Next, they answer the perceived mental well-being questionnaire.

Then they are invited to interact with the agent. The task description instructs them to approach the agent as if they are told to do a selfcheck with the agent and to act on their own experiences. They are also informed it is fine to imagine another reason they might want to talk to the agent. The participants talk to the agent until the conversation is closed off. See Appendix B for the full conversation flow. Next are several more survey parts. The participants fill in the perceived mental well-being

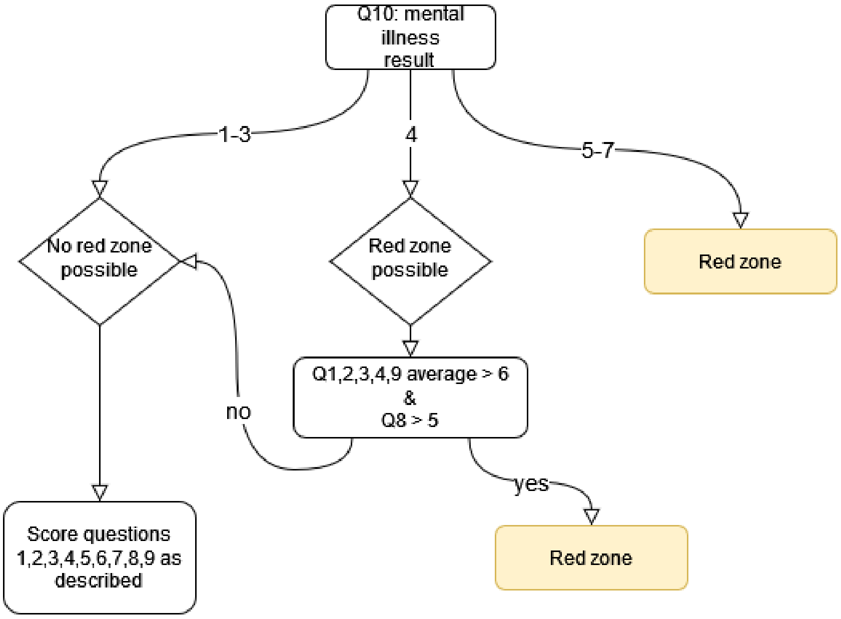


Figure 4.1: Scoring process of the perceived well-being survey.

questionnaire again, this time including questions a and b, followed by the UMUX-LITE, MHC-SF and M.I.N.I. screen questionnaires. At the end participants are thanked for participating, and once again referred to available mental health support.

4.4. Data Processing and Analysis

The raw data collected during the experiment would be demographic data and score numbers for each of the questionnaires questions. Data from participants that did not fill in all mental well-being questions would be discarded.

4.4.1. Coding of questionnaires

Each of the questionnaires must be converted from raw data to survey results. This is how this is done for each of the used questionnaires:

- The M.I.N.I. screen is coded to a Boolean describing presence or absence of mental illness. This is done according to the guidelines for each illness as set up by Sheehan et al. [36]. If any of the mental illnesses are found, the entire survey is coded as presence of mental illness. If no mental illnesses are found, it is coded as absence of mental illness.
- The MHC-SF is scored according to the categories: flourishing, moderately mentally healthy and languishing, as suggested by Keyes et al. [17].
- The UMUX-LITE is scored according to the following equation, to get a final score between 0 and 100 [22]:
$$\mathrm{UMUX-LITE} = ((score(q1) + score(q2)) - 2) * 100/12$$

The perceived well-being questionnaire is coded to a Stress Continuum Model category by first determining if the red zone is applicable through the flowchart in figure 4.1. All answers are scored points one through seven. For scoring questions 1 through 5 are reverse coded. This makes higher results relate to the worse sides of the questions.

When the red zone is possible but not certain, which is when Q10 is scored 4 (the middle option), questions 1-4 and 9 are averaged. If this average is over 6 and the score for question 8 is over five, the red zone is still applied.

If no red zone is determined, the score is determined by scoring questions to categories as by table 4.2. The resulting categories are then converted back to numbers: green = 1, yellow = 2, orange =

3 and red = 4. The number for each question is given the weight stated in the table and then they are averaged to one score. The reasoning for these weights is that overall perceived well-being is the most relevant to your perceived well-being score. This gets a high weight. Recent stress gives some indication of mental well-being, but does not give that much information, so the weight is lower. Help-seeking intentions in any form are less relevant to how you perceive your status, since they are not one-on-one related, but help-seeking intentions presented do give some indication of perceived problems. Due to this limited connection, these questions also get a lower weight. The final score is then converted back to a category by rounding. If the resulting category is red, it is overwritten to be the orange category.

Question 11 remains unused at this stage. This question on perceived insight into stress symptoms does not relate to the perceived mental well-being status. Instead the answer to this question is used in analysis to explain the results.

Q	Topic	Score	Weight
1, 2, 3, 9 averaged	Overall well-being	1 or 2: green 3 or 4: yellow 5 or 6: orange 7: red	3
4	Comparing to others	1 to 3: green 4 or 5: yellow 6 or 7: orange	1
5	Average well-being	1 to 3: green 4 or 5: yellow 6 or 7: orange	1
6	Recent stress	1: green 2 to 4: yellow 5 to 7: orange	0.5
7	Low level help-seeking	1: green 2 to 5: yellow 6 or 7: orange	0.5
8	High level help-seeking	1: green 2 or 3: yellow 4 or 5: orange 6 or 7: red	0.5

Table 4.2: Scoring process of the perceived well-being survey.

4.4.2. Analysis of awareness

Comparing perceived well-being to the reported well-being scores is no trivial task. There are multiple ways to approach this and in an ideal scenario experts such as psychologists have a say on the chosen method. Due to the limited time for this project, this feedback step is omitted and the most intuitive method is used.

To compare the results of the reported and perceived statuses, both measures are mapped to the Stress Continuum Model categories. The two questionnaires on reported mental well-being with their five result categories are mapped to the Stress Continuum Model categories as follows: green - flourishing; yellow - moderately mentally healthy; orange - languishing; red - anyone with presence of mental illness. In this case presence of mental illness overwrites any mental health score. The perceived well-being survey is mapped to the categories as described in Section 4.4.1.

For this study, I wish to investigate if the participants estimate themselves (closer to) the right category in the stress continuum model. To compare categories, each category gets a number: green = 1, yellow = 2, orange = 3 and red = 4. Both pre- and post-test perceived statuses are subtracted from the reported status to get the accuracy of the perception.

$$\text{Difference}(a, b) = |category(a) - category(b)|$$

This results in a not quite ordinal representation of the difference. It is not ordinal because the difference between the green zone and the yellow zone is not necessarily the same as the difference between the orange zone and the red zone, though both are represented as a difference of one. However, there is some form of ordering to them, the data is clearly not just categorical, so we approximate by assuming the data is ordinal. This assumption will weaken the strength of the results. An overwhelming result can likely still be taken at face value, but if the results only just reach significance, a limitation like this would call for extended research on the improvement of awareness.

To determine if the post-test awareness difference score is closer to zero than the pre-test, the pre-test and post-test values are compared using the Wilcoxon matched-pairs signed-rank test, using a p-value of 0.05.

4.4.3. Analysis of perceived change

I also consider the perceived changes in perception. So I look at questions a and b of the perceived mental well-being survey to determine if participants feel like their awareness of mental well-being improved or got worse. For these questions, continuous data would be collected, since their answers use a slider. The resulting data is expected to be normally distributed, because it is expected for the agent to have similar effect on each participant. Normality is assessed through a Shapiro–Wilk test. Given that the data is indeed shown to be normally distributed, these questions can be analysed using a one-sample t-test. Analysis will use a p-value of 0.05. The null-hypothesis used is that no change occurred, which would mean a score of zero change. If change is observed, it will be checked in which direction this change occurred.

4.4.4. Analysis of usability

To interpret the score of the UMUX-LITE, the method used by Lewis is used [21]. The method transposes the obtained score to a grade, where for example a score of 80.8 to 84.0 results in an *A* and a score of 65.0 to 71.0 results in a *C*.

4.4.5. Additional analysis

There is several secondary questions that are interesting to explore with the collected data. The answers to some of these questions can help interpret the main results. Others can be useful to understand the mechanics of awareness. Some of the questions that can be answered with analysis of the collected data are:

- Do participants see their status as worse after interaction with the agent, or better?
- Do perceptions of certain types change more than others?
- Do people with certain reported categories of well-being improve more in awareness than others?
- If people perceive a worse status after interaction, does their willingness to seek help improve?

These questions must each be solved by applying appropriate analysis methods to the data. This process is more exploratory, since the results of the main analysis and exploring the data might give rise to new questions.

4.5. Results Interpretation

The reliability of both pre-test and post-test versions of the perceived status survey is assessed through Cronbach's alpha. A value of over 0.7 is considered sufficiently reliable.

Within the analysis, I have three hypotheses:

Hypothesis 1: The difference between perceived and actual mental well-being is lower after interacting with the agent.

Hypothesis 2: Participants perceive an increase in awareness of their own mental well-being after interacting with the agent.

Hypothesis 3: Participants perceive an increase in knowledge about mental well-being symptoms after interacting with the agent.

Hypothesis 1 is tested through the analysis of awareness. If the Wilcoxon matched-pairs signed-rank test results in a p-value under 0.05, the null hypothesis can be rejected and it can be concluded that the agent is effective at its main task of making users more aware of their mental well-being. If the null hypothesis cannot be rejected, it must be assumed the agent is not effective at improving awareness of mental well-being. In this case other results need to be inspected to see if certain aspects of the agent could have caused the problem, for example low scores on usability. If no clear avenue of improvement can be found, one might conclude that a conversational agent is not a feasible solution for this task.

Hypotheses 2 and 3 look into the experience of change of the user. Regardless of any actual improvement of awareness, users can feel they improved on awareness or knowledge. For both hypotheses, if the t-test results in a p-value of under 0.05, the null hypothesis of the respective hypothesis can be rejected. It is the combination of outcomes that is interesting. If awareness improves, it would be beneficial if perceived awareness also improves, as users would appreciate the impact of the agent, which eventually benefits usage in the organisation. However, if perceived awareness and knowledge do not improve when awareness does, the main goal of the agent is still achieved. In the case that awareness does not significantly improve, it would be interesting to see if users still thought they improved. If users do feel that their knowledge of symptoms improved, this is still beneficial, since they might recognise problems quicker in the future.

4.5.1. Usability results

The usability grade result will be interpreted to how much work the usability of the agent still needs to be good. An *F* (score of 51.6 or below) indicates that the current usability is very bad. This result would suggest it might be better to rethink the whole approach, or at least start over with the implementation. A *D* to a *C+* (a score up to 72.5) indicates a significant improvement can still be made. A *B-* to *B+* (a score up to 78.8) shows some minor improvements might still be made to improve the usability, but is already quite a good result. An *A-* or higher (a score over 78.8) implies good usability.

4.5.2. Additional results

The results of additional analysis will give insights into what aspects of the agent might need improvement. For example, if the perceived status is always closer to the red category after interaction than before, the agent might focus too much on discussing the potential problems, rather than also on what is going well. If specific aspects of perceived status change more than others, this might be an aspect that other future interventions want to focus on because it is effective. On the other hand, if the awareness does not improve, it would be useful to look for improvements by focusing on the aspects of perceived status that changed less. Similarly, if people in certain categories of well-being have more benefit from the agent, where their awareness improves more, action might be taken to improve the agent for users in the categories that showed less improvement in awareness. Insight into help-seeking intentions helps to determine if finding help is an aspect the agent should focus more on. This can be important, even while it is not the main goal of the agent to get users to seek help.

5

Discussion

In contrast to existing conversational agents, the agent developed in this project does not aim to treat users, rather it is an addition to a larger support and treatment system. It uses a method that has not previously been adapted to a conversational agent and targets a user group that has not previously been targeted by a conversational agent. The challenges in this design were how to convert the SFA guidelines to a structured conversation flow and how to present the agent such that it would be best received by police officers.

Over the course of the design of the agent, different experts provided feedback, which was taken into account for the next design version. The police peer support officers were concerned with the practical aspects of the conversation and about the feel of the agent for an average officer. The psychologists mostly looked at the mental impact of the conversation and at the agent's place in the police support structure, even in its prototype form. The suicide expert was consulted for one specific part of the conversation, but due to the seriousness of the topic, expert feedback on this specific part was very important and useful.

Most of the resulting changes were small wording adaptations, but the more impactful where changes to the structure and content of the conversation, which should improve the overall usability of the agent, by making the conversation more straightforward and by communicating more about capabilities and limitations of the agent.

The resulting agent needs to be evaluated on effectiveness. No existing projects were found that evaluate mental well-being awareness in detail and so a new approach was presented. This evaluation plan looks into the more granular aspects of perceived well-being, to get a deep insight into not only if the agent is effective, but at what aspects it is more or less effective.

For the evaluation plan, a new questionnaire on perceived mental well-being had to be developed. This questionnaire covers several different ways of looking at your overall mental well-being, but also covers your perception of specific parts of mental well-being.

Some method also had to be created to combine and compare the outcomes of the three very different questionnaires involved in determining mental well-being awareness. In this approach, both perceived and reported mental well-being were mapped to the categories of the Stress Continuum Model. This led to a more intuitive interpretation of the different sub-results. An assumption of ordinality of the data had to be made to effectively analyse the differences in perceived and reported categories.

The full evaluation plan should provide insight not only into effectiveness at improving awareness, but also in agent usability and in what aspects of perceived well-being are most influenced by the agent. The analysis should find avenues for improvement of the agent in future projects and might also improve our understanding of mental well-being awareness. The developed questionnaire and analysis approach can also be utilised in future research on perceived mental well-being or awareness.

5.1. Agent Design Implications

The feedback of the various experts during the design process provided valuable lessons on the design of this agent that can be generalised to design for mental well-being support agents in general. To find these lessons, I went over the list of feedback and grouped everything that pertains to specific wordings.

For the points that remained, I looked for how their ideas could be applied to different projects. Future conversational agent projects should:

1. **Manage expectations**: Especially in a health care domain, be explicit about what your agent can and cannot do. Clarity about this prevents confusion and miscommunication.
2. **Scope broad questions**: Broad questions such as asking for the situation by this agent can result in any number of directions of answers. In order to simplify processing and to give more structure for the user to hold on to, it is better to rephrase such questions into step wise more contained questions to eventually gather the information that you need.
3. **Get feedback from users on wording**: Especially when designers are not part of target group of the agent, they might have a different feel on phrasing than the target group does. To prevent misunderstandings and unintended emotions, have target users verify the phrasing in your design. This also prevents using terms that already have a different meaning in their context.
4. **Do not push people when discussing suicide**: Suicide is a delicate topic and people considering suicide generally do not think clearly. Trying to get people to seek help, especially through an automated message, is quickly misinterpreted or taken the wrong way. So provide options but leave the user agency over their own lives.

5.2. Subquestions

In this section the subquestions stated in the Introduction chapter are revisited and answered.

1. *How can we determine improvement of the mental well-being awareness of a person?*
   The perceived mental well-being has to be compared to the actual mental well-being at two points in time. No measures were found for perceived well-being, so the survey developed for this, looking into different aspects of perceived well-being, can be used. Actual mental well-being is hard to measure, but many surveys exist that determine mental well-being based on reported symptoms. The difference between the reported and perceived mental well-being at both points in time can then be compared.
2. *What is the best communication medium for the agent and the conversation to assess the mental well-being status?*
   No literature supported choosing one medium over another. In this absence, one can pick based on a different factor than effectiveness, such as implementation ease.
3. *How can the agent best be presented within the chosen medium for an effective conversation?*
   The agent was found to be most comfortable to interact with by experts when informal language use and presentation are used.
4. *How can a conversational agent go through the check action together with the officer in a respectful and efficient way?*
   (a) *How can the agent best help the officer get an overview of their potential stress features?*
   The agent goes over potential features one by one, together with the user.
   (b) *Are there any information sources other than the officer the agent can utilise?*
   This agent gathers all its information from the user. Other sources such as database contents or reaching out to colleagues could be used, but were deemed out of scope and non essential for this design.
   (c) *How can the agent help the officer review the features and determine their stress level and next steps?*
   The agent summarises its findings for the user by listing all found features. It then explains its reasoning for coming to a certain conclusion on a stress level.

   The agent always explains its reasoning and the reason for the conversation. It maintains user agency throughout the conversation. Experts supported that the conversation is respectful. By following the SFA guidelines, it should also be effective at reaching a valid conclusion.

5.3. Limitations

There are a few limitations to both the design and the evaluation plan. During the design phase, several experts were consulted, but no actual target users were involved in the initial design or following improvements. They would provide valuable insight into how the used text and presentation is received, which would allow for fine-tuning before executing a full scale evaluation.

There are several limitations to the presented evaluation plan. First of all, the measured mental well-being is not objective but instead relies on self-reported symptoms. Although it is very hard to measure mental well-being objectively, it is still a limitation to the power of the study results, as the resulting awareness values are only a proxy to the actual awareness.

Then there is the limitations on the processing of the data. The scoring of the perceived well-being questionnaire is currently based on assumptions and should be submitted to scrutiny of mental well-being experts to validate the scoring approach. The perceived well-being questionnaire itself is not yet validated. A separate study should be conducted to this end.

Then when it comes to the actual analysis of the data, the different questionnaire scores are reduced to categorisations to fit into the Stress Continuum Model. These are then compared using a statistical test for dependent ordinal data. However, although the categories themselves are ordinal, the difference sizes between categories are not. The difference between the green and yellow zones and between the orange and red zones are both counted as one, but might not be equal. Even so, the assumption of ordinality is used. The strength of this assumption could be improved by having a deeper look at the valuation of categories. Alternatively, research can be conducted to determine a more reliable way of representing the difference between categories.

5.4. Future Work

This thesis has taken on the task of designing the conversational agent, but did not execute a quantitative evaluation. It remains an open question how general police employees would experience the interaction with the agent and how effective the agent is at actually increasing users' awareness of their mental state.

Two first steps at future work are evaluating and validating the perceived mental well-being questionnaire presented in this thesis, and executing the evaluation plan laid out in Chapter 4. This should show if the agent is effective or shows promise.

Once effectiveness is determined, a follow-up research can take a deeper look into the usability of the agent, for example by using the Chatbot Usability Scale [6]. This would allow to more accurately assess which aspects of the chatbot function well and which function less well from a user perspective. With the results of these evaluations, the agent can be improved from a prototype to a final, functioning agent.

The psychologists consulted in the design phase of the project recommended several points to pay attention to for the actual implementation of the agent in the larger support structure in the police organisation. Before the agent can be launched for use, the following questions need to be answered and implemented:

- Where is the agent going to be available? How is its use promoted? It is important that employees know where to find the agent and that it exists. Management can promote its use by spreading the word about the agent or by instituting regular check-ins with the agent.
- How will it connect to other support resources? If someone does discover problems, how are they referred to someone that can help? The agent has to determine when to refer to external support resources and which person to refer to. It would lower the bar for people to seek help if the agent could facilitate the first contact by sending a message to a relevant person.

5.5. Conclusion

In this project I attempted to build a conversational agent that can help police officers to become more aware of their mental well-being status and to set up a sound evaluation of this agent. The research questions were the following:

RQ1: *What would a design look like of a conversational agent for helping police officers become aware of their mental well-being status by applying the Check action of the Stress First Aid model?*

The conversational agent design uses an informal presentation and respectful language and follows the SFA guidelines to help police officers become more aware of their mental well-being status. The design has been subjected to scrutiny of several types of experts before coming to the final design.

RQ2: *How could a conversational agent for awareness of one's mental well-being status be evaluated?* The evaluation proposal uses a pre-test post-test design where the perceived mental well-being is compared to the reported mental well-being. To achieve this, a questionnaire to measure perceived mental well-being and an analysis approach were developed.

The design process yielded several general design lessons that can be useful for other projects working on mental well-being. The evaluation plan provides a structured approach to testing the effectiveness of the agent at reaching its goal of improving awareness, and provides insight into related aspects of the agent, such as usability and impact on various parts of awareness. The developed questionnaire on perceived mental well-being can also be used in other research concerning perceived mental well-being or awareness. Future work on this agent should be focused executing the evaluation plan and validating the perceived status questionnaire.

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, December 2019. ISSN 1386-5056. doi: 10.1016/j.ijmedinf.2019.103978. URL `https://www.sciencedirect.com/science/article/pii/S1386505619307166`.

[2] Daniel Kwasi Ahorsu, Dalinda Isabel Sánchez Vidaña, Donald Lipardo, Parth Bharat Shah, Pablo Cruz González, Sachin Shende, Shilpa Gurung, Harun Venkatesan, Anchalee Duongthipthewa, Talha Qasim Ansari, and Veronika Schoeb. Effect of a peer-led intervention combining mental health promotion with coping-strategy-based workshops on mental health awareness, help-seeking behavior, and wellbeing among university students in Hong Kong. *International Journal of Mental Health Systems*, 15(1):6, January 2021. ISSN 1752-4458. doi: 10.1186/s13033-020-00432-0. URL `https://doi.org/10.1186/s13033-020-00432-0`.

[3] Anonymous. *Stress First Aid (SFA) for Law Enforcement*. October 2016. URL `https://www.ptsd.va.gov/professional/treat/care/toolkits/police/docs/PoliceStressFirstAid.pdf`.

[4] Anonymous. Cybercrime melden aan virtuele agent Wout, September 2019. URL `https://www.politie.nl/nieuws/2019/oktober/9/cybercrime-melden-aan-virtuele-agent-wout.html`.

[5] Anonymous. Stress First Aid: Manual and Resources for Health Care Workers - PTSD: National Center for PTSD, September 2021. URL `https://www.ptsd.va.gov/professional/treat/type/stress_first_aid.asp`.

[6] Simone Borsci, Alessio Malizia, Martin Schmettow, Frank van der Velde, Gunay Tariverdiyeva, Divyaa Balaji, and Alan Chamberlain. The Chatbot Usability Scale: the Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents. *Personal and Ubiquitous Computing*, 26(1):95–119, February 2022. ISSN 1617-4917. doi: 10.1007/s00779-021-01582-9. URL `https://doi.org/10.1007/s00779-021-01582-9`.

[7] Theo Douglas. Los Angeles Chatbot Deputized to Help with Police Recruitment, February 2018. URL `https://www.govtech.com/products/Los-Angeles-Chatbot-Deputized-to-Help-with-Police-Recruitment.html`.

[8] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, 4(2):e7785, June 2017. doi: 10.2196/mental.7785. URL `https://mental.jmir.org/2017/2/e19`.

[9] PTSD: National Center for PTSD. Strategies: Police Self-Care, 2020. URL `https://www.ptsd.va.gov/professional/treat/care/toolkits/police/managingStrategiesPolice.asp`.

[10] Justin Fox, Mayur M Desai, Karissa Britten, Georgina Lucas, Renee Luneau, and Marjorie S Rosenthal. Mental-health conditions, barriers to care, and productivity loss among officers in an urban police department. *Connecticut medicine*, 76(9):525–531, October 2012. ISSN 0010-6178. URL `https://pubmed.ncbi.nlm.nih.gov/23155671`.

[11] Russell Fulmer, Angela Joerin, Breanna Gentile, Lysanne Lakerink, and Michiel Rauws. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. *JMIR Mental Health*, 5(4):e9782, December 2018. doi: 10.2196/mental.9782. URL `https://mental.jmir.org/2018/4/e64`.

[12] Anita M Hubley, Lara B Russell, Anne M Gadermann, and Anita Palepu. Quality of Life for Homeless and Hard-to-House Individuals (QoLHHI) Inventory, 2009. URL `http://www.hubleylab.ca/wp-content/uploads/2014/03/QoLHHI-Manual.pdf`.

[13] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth*, 6(11):e12106, November 2018. doi: 10.2196/12106. URL `https://mhealth.jmir.org/2018/11/e12106`.

[14] Yukun Kang, Huiyao Wang, Xiaojing Li, Yujie Tao, Xia Yang, Wei Deng, Jianying Yu, Qian Zhou, Tao Li, and Wanjun Guo. A brief web-based screening plus emotional-disorder health education was associated with improvement of mental health awareness and service-seeking attitudes among patients seeking nonpsychiatric clinical services in China. *Depression and Anxiety*, 38(5): 571–587, 2021. ISSN 1520-6394. doi: 10.1002/da.23118. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/da.23118`.

[15] Megan C. Kearns, Alex O. Rothbaum, Cole G. Youngner, Mark S. Burton, Alexander McCarthy, and Barbara Olasov Rothbaum. Cutting Edge Research on Prevention of PTSD. In Marilyn P. Safir, Helene S. Wallach, and Albert ”Skip” Rizzo, editors, *Future Directions in Post-Traumatic Stress Disorder: Prevention, Diagnosis, and Treatment*, pages 133–153. Springer US, Boston, MA, 2015. ISBN 978-1-4899-7522-5. URL `https://doi.org/10.1007/978-1-4899-7522-5_6`.

[16] Anne Kennes, Sanne Peeters, Mayke Janssens, Jennifer Reijnders, Johan Lataster, and Nele Jacobs. Psychometric Evaluation of the Mental Health Continuum-Short Form (MHC-SF) for Dutch Adolescents. *Journal of Child and Family Studies*, 29(11):3276–3286, November 2020. ISSN 1573-2843. doi: 10.1007/s10826-020-01803-4. URL `https://doi.org/10.1007/s10826-020-01803-4`.

[17] Corey L. M. Keyes, Marié Wissing, Johan P. Potgieter, Michael Temane, Annamarie Kruger, and Sinette van Rooy. Evaluation of the mental health continuum–short form (MHC–SF) in setswana-speaking South Africans. *Clinical Psychology & Psychotherapy*, 15(3):181–192, 2008. ISSN 1099-0879. doi: 10.1002/cpp.572. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/cpp.572`.

[18] Sanne M.A. Lamers, Gerben J. Westerhof, Ernst T. Bohlmeijer, Peter M. ten Klooster, and Corey L.M. Keyes. Evaluating the psychometric properties of the mental health Continuum-Short Form (MHC-SF). *Journal of Clinical Psychology*, 67(1):99–110, 2011. ISSN 1097-4679. doi: 10.1002/jclp.20741. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/jclp.20741`.

[19] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Glasgow Scotland Uk, May 2019. ACM. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300932. URL `https://dl.acm.org/doi/10.1145/3290605.3300932`.

[20] Ty Lees, Jaymen L Elliott, Simon Gunning, Phillip J Newton, Tapan Rai, and Sara Lal. A systematic review of the current evidence regarding interventions for anxiety, PTSD, sleepiness and fatigue in the law enforcement workplace. *Industrial Health*, 57(6):655–667, 2019. ISSN 0019-8366, 1880-8026. doi: 10.2486/indhealth.2018-0088. URL `https://www.jstage.jst.go.jp/article/indhealth/57/6/57_2018-0088/_article`.

[21] James R. Lewis. Measuring User Experience With 3, 5, 7, or 11 Points: Does It Matter? *Human Factors*, 63(6):999–1011, September 2021. ISSN 0018-7208. doi: 10.1177/0018720819881312. URL `https://doi.org/10.1177/0018720819881312`.

[22] James R. Lewis, Brian S. Utesch, and Deborah E. Maher. UMUX-LITE: when there’s no time for the SUS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’13, pages 2099–2102, New York, NY, USA, April 2013. Association for Computing Machinery. ISBN 978-1-4503-1899-0. doi: 10.1145/2470654.2481287. URL `https://doi.org/10.1145/2470654.2481287`.

[23] Raju Maharjan, Darius Adam Rohani, Per Bækgaard, Jakob Bardram, and Kevin Doherty. Can we talk? Design Implications for the Questionnaire-Driven Self-Report of Health and Wellbeing via Conversational Agent. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*, CUI '21, pages 1–11, New York, NY, USA, July 2021. Association for Computing Machinery. ISBN 978-1-4503-8998-3. doi: 10.1145/3469595.3469600. URL `https://doi.org/10.1145/3469595.3469600`.

[24] Deborah B Maia, Charles R Marmar, Clare Henn-Haase, Augusta Nóbrega, Adriana Fiszman, Carla Marques-Portella, Mauro V Mendlowicz, Evandro S.F Coutinho, and Ivan Figueira. Predictors of PTSD symptoms in brazilian police officers: the synergy of negative affect and peritraumatic dissociation. *Revista Brasileira de Psiquiatria*, 33(4):362–366, December 2011. ISSN 1516-4446. doi: 10.1590/S1516-44462011000400009. URL `http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-44462011000400009&lng=en&tlng=en`.

[25] Joshua R. Moore and Robert Caudill. The Bot Will See You Now. *Psychiatric Clinics of North America*, 42(4):627–634, December 2019. ISSN 0193953X. doi: 10.1016/j.psc.2019.08.007. URL `https://linkinghub.elsevier.com/retrieve/pii/S0193953X19300772`.

[26] Philip R Muskin. What Are Anxiety Disorders?, June 2021. URL `https://www.psychiatry.org/patients-families/anxiety-disorders/what-are-anxiety-disorders`.

[27] Kim S. Ménard and Michael L. Arter. Stress, Coping, Alcohol Use, and Posttraumatic Stress Disorder Among an International Sample of Police Officers: Does Gender Matter? *Police Quarterly*, 17(4):307–327, December 2014. ISSN 1098-6111. doi: 10.1177/1098611114548097. URL `https://doi.org/10.1177/1098611114548097`.

[28] Jill Novacek and Robert Raskin. Recognition of Warning Signs : A Consideration for Cost-Effective Treatment of Severe Mental Illness. *Psychiatric Services*, 49(3):376–378, March 1998. ISSN 1075-2730. doi: 10.1176/ps.49.3.376. URL `https://ps.psychiatryonline.org/doi/full/10.1176/ps.49.3.376`.

[29] T Overbeek, K Schruers, and E Griez. Mini International Neuropsychiatric Interview Nederlandse Versie 5.0.0, 1999.

[30] James O Prochaska and Wayne F Velicer. The Transtheoretical Model of Health Change. *American journal of health promotion: AJHP*, 12(1), 1997. doi: 10.4278/0890-1171-12.1.38. URL `https://doi.org/10.4278/0890-1171-12.1.38`.

[31] Wei Qi, Martin Gevonden, and Arieh Shalev. Prevention of Post-Traumatic Stress Disorder After Trauma: Current Evidence and Future Directions. *Current Psychiatry Reports*, 18(2):20, February 2016. ISSN 1523-3812, 1535-1645. doi: 10.1007/s11920-015-0655-0. URL `http://link.springer.com/10.1007/s11920-015-0655-0`.

[32] J. Roldán-Merino, M. T. Lluch-Canut, I. Casas, M. Sanromà-Ortíz, C. Ferré-Grau, C. Sequeira, A. Falcó-Pegueroles, D. Soares, and M. Puig-Llobet. Reliability and validity of the Positive Mental Health Questionnaire in a sample of Spanish university students. *Journal of Psychiatric and Mental Health Nursing*, 24(2-3):123–133, 2017. ISSN 1365-2850. doi: 10.1111/jpm.12358. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/jpm.12358`.

[33] Tina Rosenberg. Opinion | A.I. Joins the Campaign Against Sex Trafficking. *The New York Times*, April 2019. ISSN 0362-4331. URL `https://www.nytimes.com/2019/04/09/opinion/ai-joins-the-campaign-against-sex-trafficking.html`.

[34] John P. Salerno. Effectiveness of Universal School-Based Mental Health Awareness Programs Among Youth in the United States: A Systematic Review. *Journal of School Health*, 86(12):922–931, 2016. ISSN 1746-1561. doi: 10.1111/josh.12461. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/josh.12461`.

[35] Mark R. Scholten, Saskia M. Kelders, and Julia E. W. C. Van Gemert-Pijnen. An Empirical Study of a Pedagogical Agent as an Adjunct to an eHealth Self-Management Intervention: What

Modalities Does It Need to Successfully Support and Motivate Users? *Frontiers in Psychology*, 10:1063, 2019. ISSN 1664-1078. doi: 10.3389/fpsyg.2019.01063. URL https://www.frontiersin.org/article/10.3389/fpsyg.2019.01063.

[36] David V. Sheehan, Yves Lecrubier, K. Harnett Sheehan, Patricia Amorim, Juris Janavs, Emmanuelle Weiller, Thierry Hergueta, Roxy Baker, and Geoffrey C. Dunbar. The Mini-International Neuropsychiatric Interview (M.I.N.I.): The Development and Validation of a Structured Diagnostic Psychiatric Interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, 59(suppl 20):0–0, May 1998. ISSN 0160-6689. URL https://www.psychiatrist.com/jcp/neurologic/neurology/mini-international-neuropsychiatric-interview-mini/.

[37] Robyn E. Shields, Stephanie Korol, R. Nicholas Carleton, Megan McElheran, Andrea M. Stelnicki, Dianne Groll, and Gregory S. Anderson. Brief Mental Health Disorder Screening Questionnaires and Use with Public Safety Personnel: A Review. *International Journal of Environmental Research and Public Health*, 18(7):3743, January 2021. doi: 10.3390/ijerph18073743. URL https://www.mdpi.com/1660-4601/18/7/3743.

[38] Myrthe L. Tielman, Mark A. Neerincx, Rafael Bidarra, Ben Kybartas, and Willem-Paul Brinkman. A Therapy System for Post-Traumatic Stress Disorder Using a Virtual Agent and Virtual Storytelling to Reconstruct Traumatic Memories. *Journal of Medical Systems*, 41(8):125, July 2017. ISSN 1573-689X. doi: 10.1007/s10916-017-0771-y. URL https://doi.org/10.1007/s10916-017-0771-y.

[39] Felix Torres. What Is PTSD?, August 2020. URL https://www.psychiatry.org/patients-families/ptsd/what-is-ptsd.

[40] 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, July 2019. ISSN 0706-7437, 1497-0015. doi: 10.1177/0706743719828977. URL http://journals.sagepub.com/doi/10.1177/0706743719828977.

[41] I van Beek, T. W. Taris, and W. B. Schaufeli. De psychosociale gezondheid van politiepersoneel. WODC Rapport 2228, WODC, ministerie van Veiligheid en Justitie, Utrecht, 2013. URL http://hdl.handle.net/20.500.12832/1994.

[42] Olaf van Miltenburg. Politie stopt verlengde proef met chatbot Wout, March 2020. URL https://tweakers.net/nieuws/169282/politie-stopt-verlengde-proef-met-chatbot-wout.html.

[43] Patricia Watson. PTSD as a Public Mental Health Priority. *Current Psychiatry Reports*, 21(7):61, June 2019. ISSN 1535-1645. doi: 10.1007/s11920-019-1032-1. URL https://doi.org/10.1007/s11920-019-1032-1.

[44] Patricia Watson, William Nash, Richard Westphal, and Brett Litz. Combat Operational Stress First Aid Manual, November 2012. URL https://www.researchgate.net/publication/232735117_Combat_Operational_Stress_First_Aid_Manual.

[45] Joseph Weizenbaum. ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, January 1966. ISSN 0001-0782, 1557-7317. doi: 10.1145/365153.365168. URL https://dl.acm.org/doi/10.1145/365153.365168.

[46] D Weston, C Hudson, D Carroll, S Coomber, and R Amlôt. Evaluating a pilot mental health awareness for managers' training course. *Occupational Medicine*, 69(4):251–257, June 2019. ISSN 0962-7480. doi: 10.1093/occmed/kqz061. URL https://doi.org/10.1093/occmed/kqz061.

[47] Peixiang Zhong, Chen Zhang, Hao Wang, Yong Liu, and Chunyan Miao. Towards Persona-Based Empathetic Conversational Models. In *Proceedings of the 2020 Conference on Empirical*

Methods in Natural Language Processing (EMNLP), pages 6556–6566, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.531. URL `https://aclanthology.org/2020.emnlp-main.531`.

A

List of terms used

A.1. Mental health and well-being

- **Mental well-being** refers to the overall mental state.
- **Mental illness** refers to clinically diagnosable illnesses.
- **Mental health** refers to the everyday mental state with respect to your happiness and stability.

A.2. Evaluation aspects

- **Awareness** is the accuracy of your belief about your status. If your belief is more accurate, you are more aware.
- **Actual mental well-being** is the objective state of one's mental well-being. This cannot easily be measured.
- **Perceived mental well-being** or **perceived state** refers to the participant's perception on what their mental well-being status is.
- **Reported mental well-being** is the mental well-being status determined based on symptom reports of the participant. This approximates the actual mental well-being.

B

Conversation Design Diagram

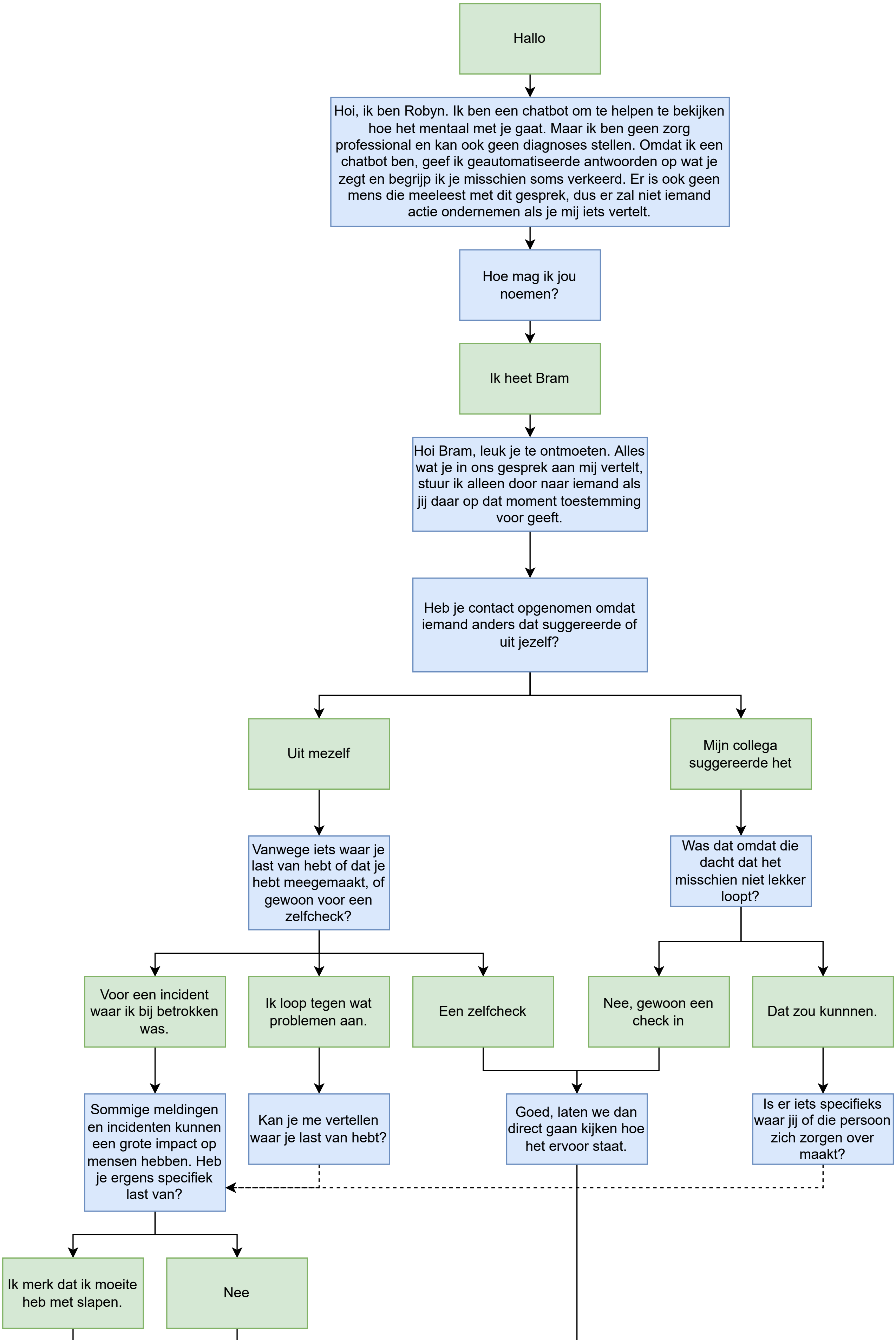
This is the final design after feedback and adaptations. The conversation is build up out of several pieces which link differently based on the user input. The orange blocks indicate what links to where.

Legend

🟩 User message (example)

🟦 Bot message

🟧 Previous/next conversation block



Hallo

Hoi, ik ben Robyn. Ik ben een chatbot om te helpen te bekijken hoe het mentaal met je gaat. Maar ik ben geen zorg professional en kan ook geen diagnoses stellen. Omdat ik een chatbot ben, geef ik geautomatiseerde antwoorden op wat je zegt en begrijp ik je misschien soms verkeerd. Er is ook geen mens die meeleest met dit gesprek, dus er zal niet iemand actie ondernemen als je mij iets vertelt.

Hoe mag ik jou noemen?

Ik heet Bram

Hoi Bram, leuk je te ontmoeten. Alles wat je in ons gesprek aan mij vertelt, stuur ik alleen door naar iemand als jij daar op dat moment toestemming voor geeft.

Heb je contact opgenomen omdat iemand anders dat suggereerde of uit jezelf?

Uit mezelf

Mijn collega suggereerde het

Vanwege iets waar je last van hebt of dat je hebt meegemaakt, of gewoon voor een zelfcheck?

Was dat omdat die dacht dat het misschien niet lekker loopt?

Voor een incident waar ik bij betrokken was.

Ik loop tegen wat problemen aan.

Een zelfcheck

Nee, gewoon een check in

Dat zou kunnnen.

Sommige meldingen en incidenten kunnen een grote impact op mensen hebben. Heb je ergens specifiek last van?

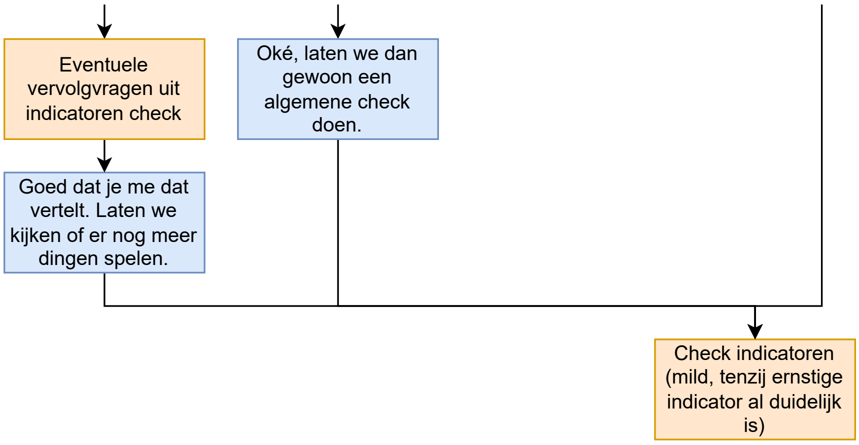
Kan je me vertellen waar je last van hebt?

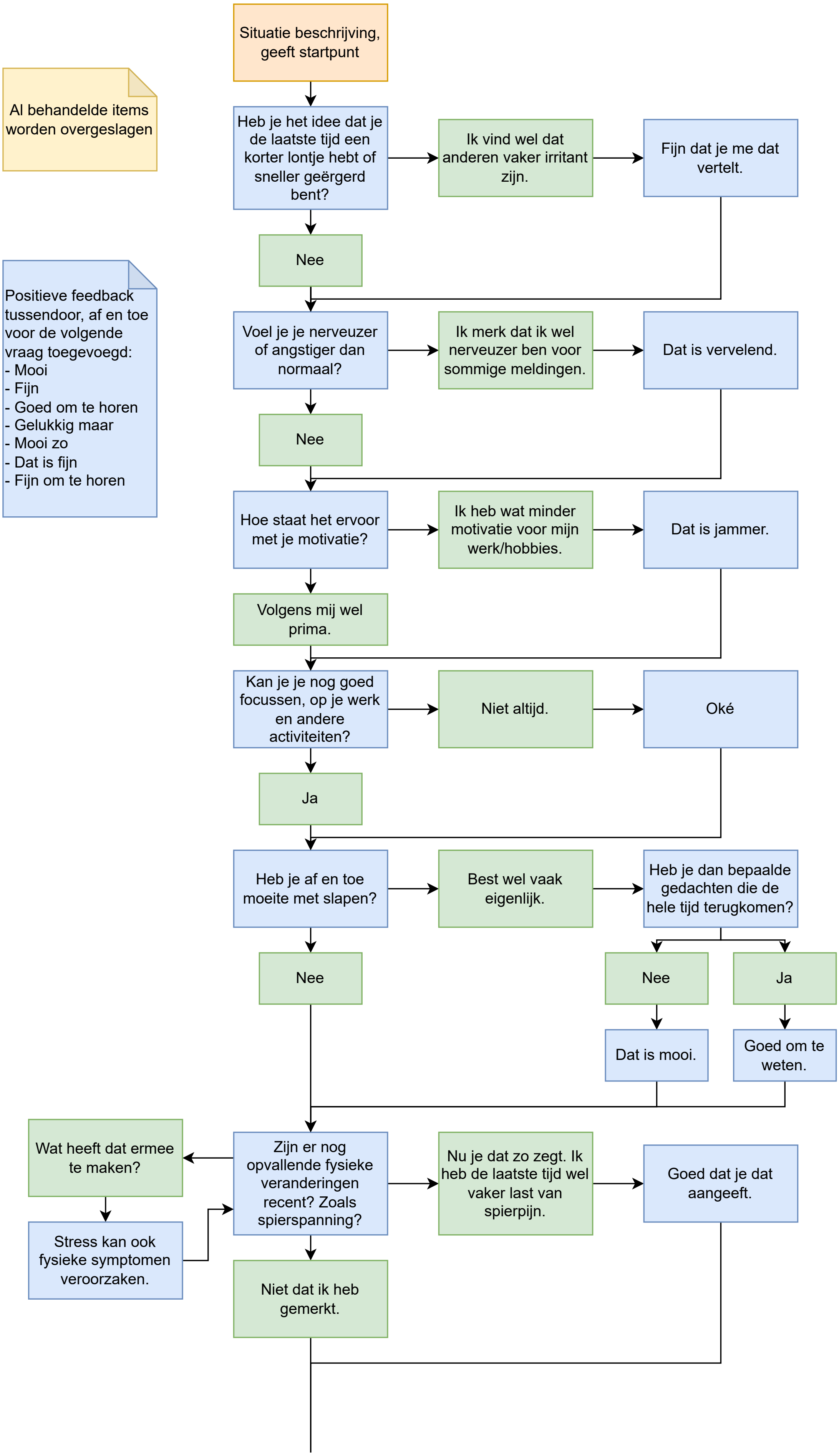
Goed, laten we dan direct gaan kijken hoe het ervoor staat.

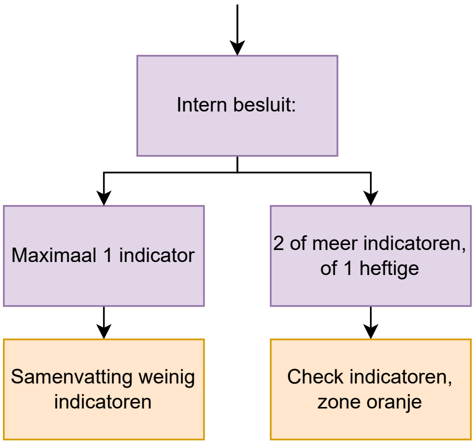
Is er iets specifieks waar jij of die persoon zich zorgen over maakt?

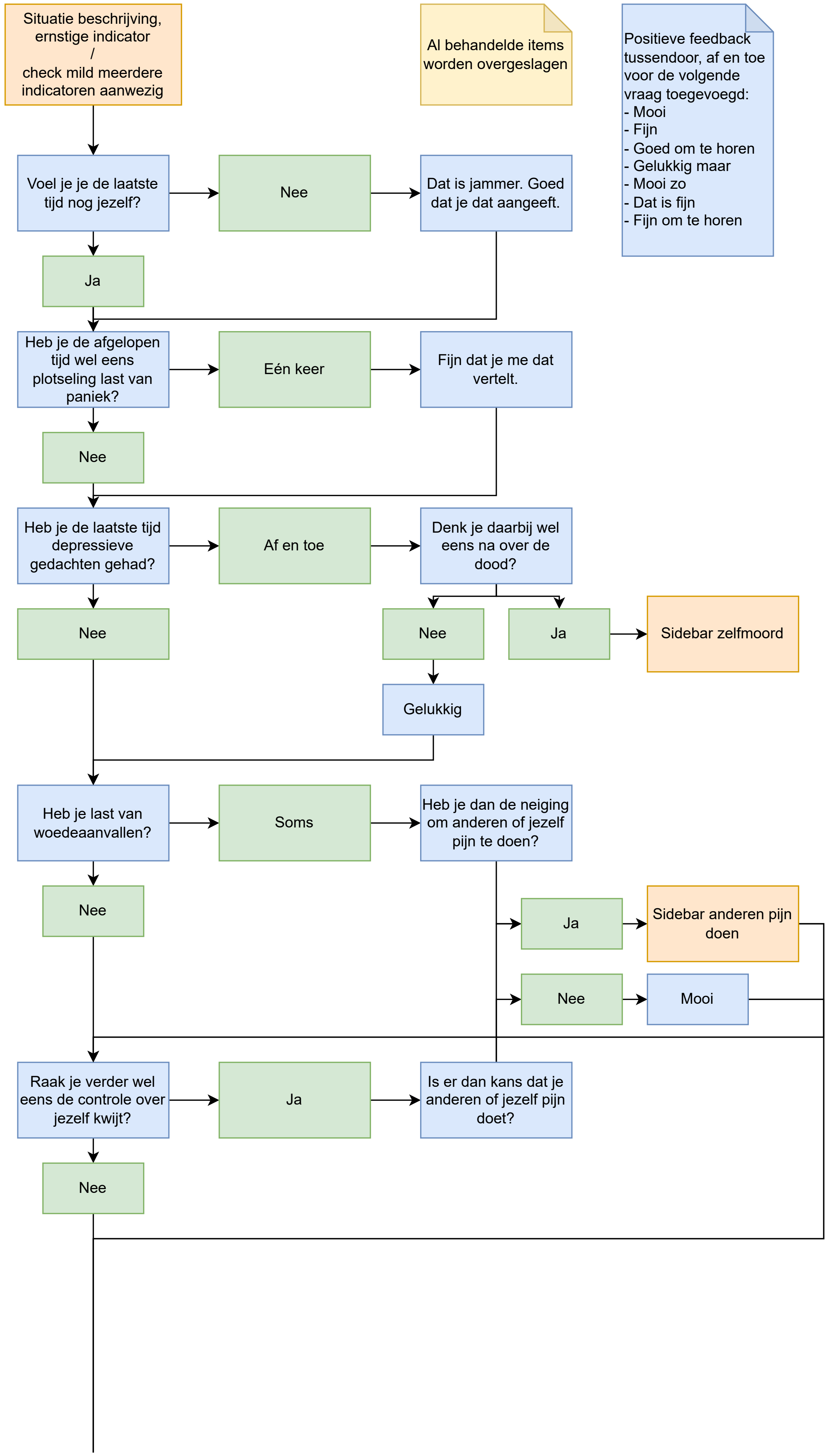
Ik merk dat ik moeite heb met slapen.

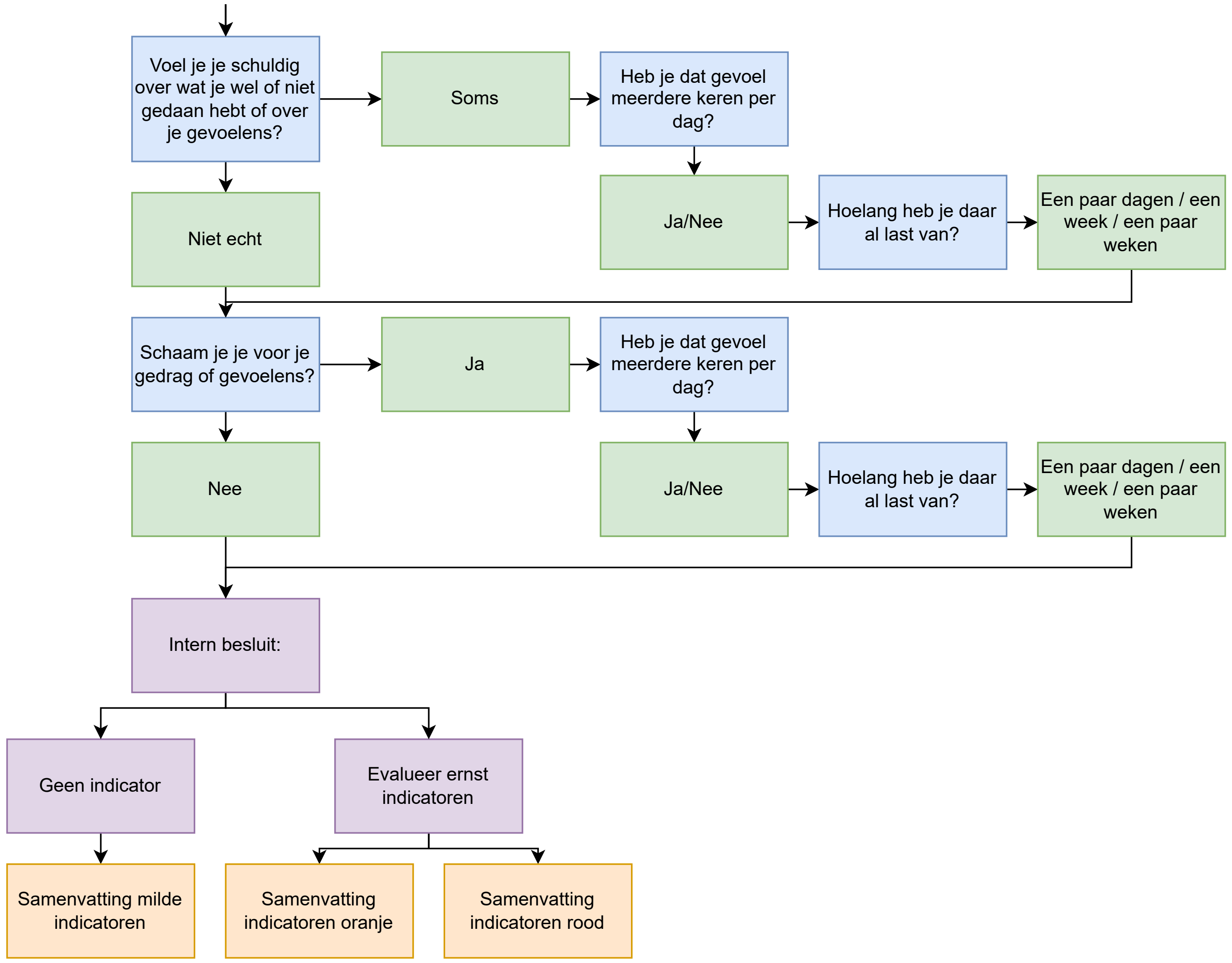
Nee

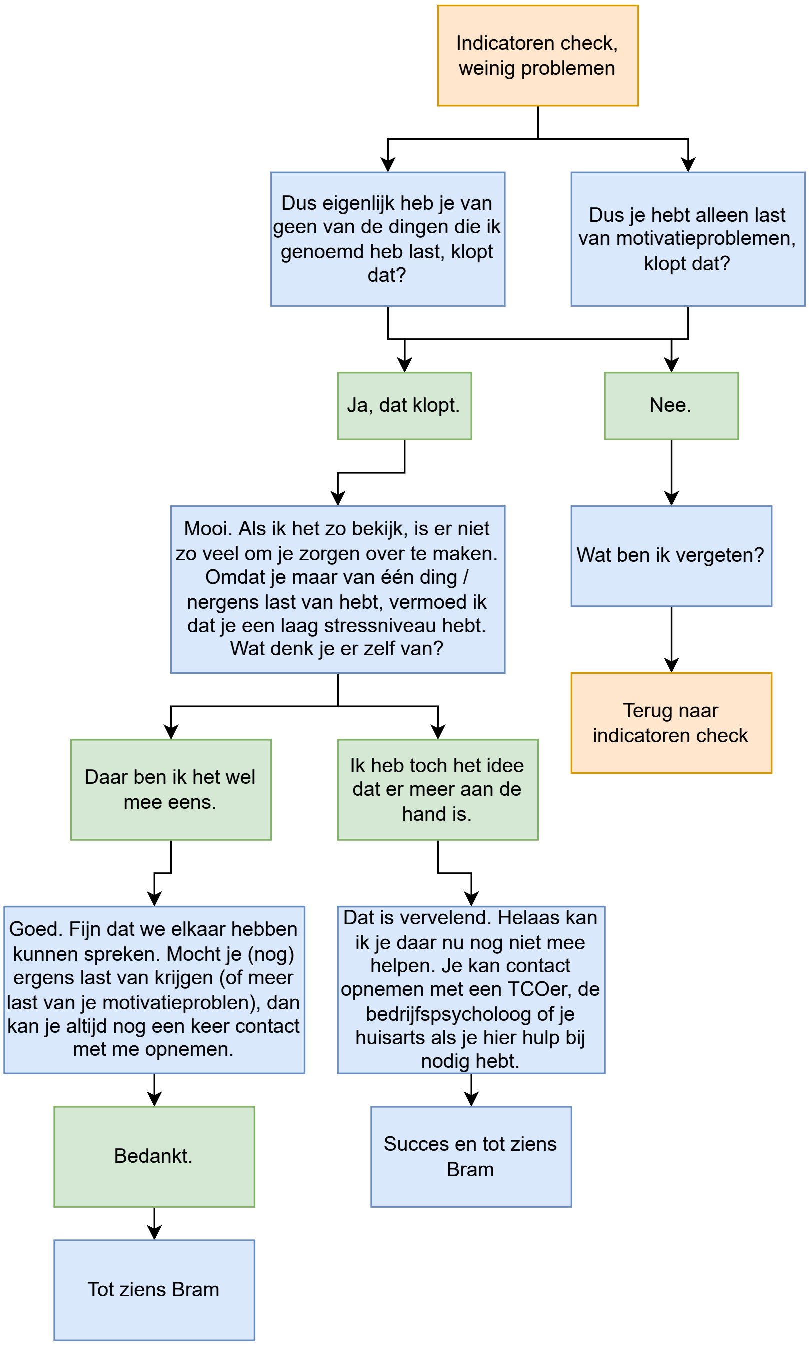

Eventuele vervolgvragen uit indicatoren check
Oké, laten we dan gewoon een algemene check doen.
Goed dat je me dat vertelt. Laten we kijken of er nog meer dingen spelen.
Check indicatoren (mild, tenzij ernstige indicator al duidelijk is)

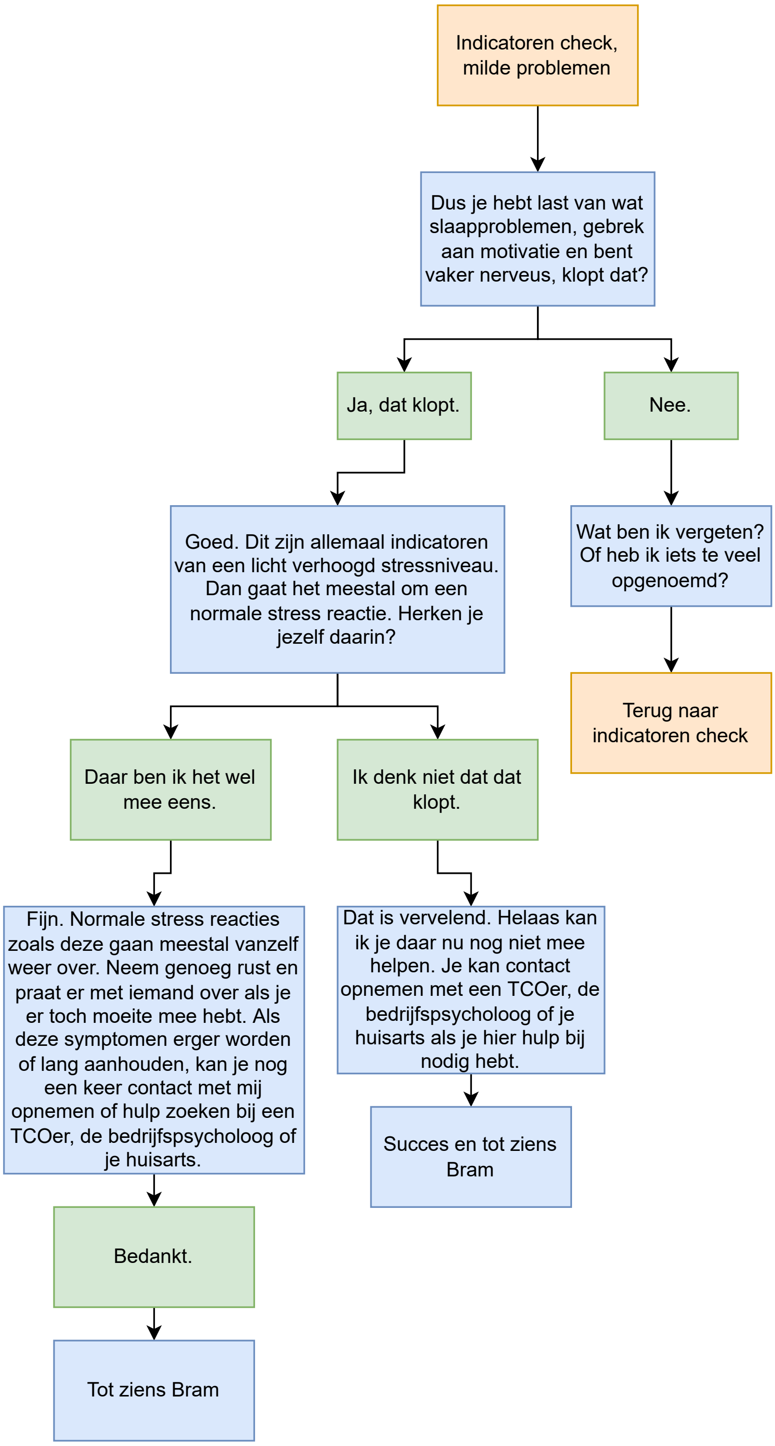

Al behandelde items worden overgeslagen
Positieve feedback tussendoor, af en toe voor de volgende vraag toegevoegd:
- Mooi
- Fijn
- Goed om te horen
- Gelukkig maar
- Mooi zo
- Dat is fijn
- Fijn om te horen
Situatie beschrijving, geeft startpunt
Heb je het idee dat je de laatste tijd een korter lontje hebt of sneller geërgerd bent?
Ik vind wel dat anderen vaker irritant zijn.
Fijn dat je me dat vertelt.
Nee
Voel je je nerveuzer of angstiger dan normaal?
Ik merk dat ik wel nerveuzer ben voor sommige meldingen.
Dat is vervelend.
Nee
Hoe staat het ervoor met je motivatie?
Ik heb wat minder motivatie voor mijn werk/hobbies.
Dat is jammer.
Volgens mij wel prima.
Kan je je nog goed focussen, op je werk en andere activiteiten?
Niet altijd.
Oké
Ja
Heb je af en toe moeite met slapen?
Best wel vaak eigenlijk.
Heb je dan bepaalde gedachten die de hele tijd terugkomen?
Nee
Nee
Ja
Dat is mooi.
Goed om te weten.
Wat heeft dat ermee te maken?
Zijn er nog opvallende fysieke veranderingen recent? Zoals spierspanning?
Nu je dat zo zegt. Ik heb de laatste tijd wel vaker last van spierpijn.
Goed dat je dat aangeeft.
Stress kan ook fysieke symptomen veroorzaken.
Niet dat ik heb gemerkt.

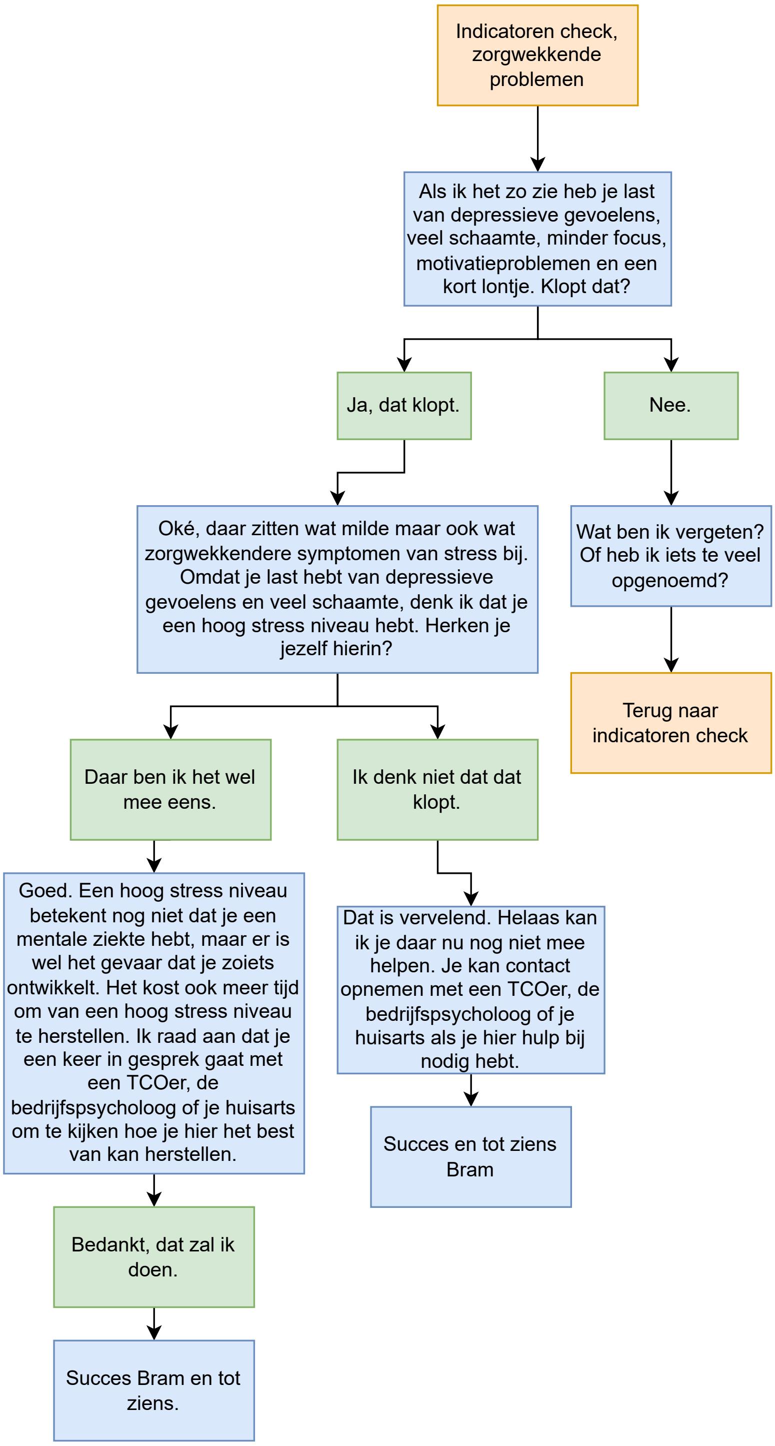

Intern besluit:
Maximaal 1 indicator
2 of meer indicatoren, of 1 heftige
Samenvatting weinig indicatoren
Check indicatoren, zone oranje


Situatie beschrijving, ernstige indicator / check mild meerdere indicatoren aanwezig
Al behandelde items worden overgeslagen
Positieve feedback tussendoor, af en toe voor de volgende vraag toegevoegd:
- Mooi
- Fijn
- Goed om te horen
- Gelukkig maar
- Mooi zo
- Dat is fijn
- Fijn om te horen
Voel je je de laatste tijd nog jezelf?
Nee
Dat is jammer. Goed dat je dat aangeeft.
Ja
Heb je de afgelopen tijd wel eens plotseling last van paniek?
Eén keer
Fijn dat je me dat vertelt.
Nee
Heb je de laatste tijd depressieve gedachten gehad?
Af en toe
Denk je daarbij wel eens na over de dood?
Nee
Nee
Ja
Sidebar zelfmoord
Gelukkig
Heb je last van woedeaanvallen?
Soms
Heb je dan de neiging om anderen of jezelf pijn te doen?
Nee
Ja
Sidebar anderen pijn doen
Nee
Mooi
Raak je verder wel eens de controle over jezelf kwijt?
Ja
Is er dan kans dat je anderen of jezelf pijn doet?
Nee


Voel je je schuldig over wat je wel of niet gedaan hebt of over je gevoelens?
Soms
Heb je dat gevoel meerdere keren per dag?
Ja/Nee
Hoelang heb je daar al last van?
Een paar dagen / een week / een paar weken
Niet echt
Schaam je je voor je gedrag of gevoelens?
Ja
Heb je dat gevoel meerdere keren per dag?
Nee
Ja/Nee
Hoelang heb je daar al last van?
Een paar dagen / een week / een paar weken
Intern besluit:
Geen indicator
Evalueer ernst indicatoren
Samenvatting milde indicatoren
Samenvatting indicatoren oranje
Samenvatting indicatoren rood


Indicatoren check, weinig problemen
Dus eigenlijk heb je van geen van de dingen die ik genoemd heb last, klopt dat?
Dus je hebt alleen last van motivatieproblemen, klopt dat?
Ja, dat klopt.
Nee.
Mooi. Als ik het zo bekijk, is er niet zo veel om je zorgen over te maken. Omdat je maar van één ding / nergens last van hebt, vermoed ik dat je een laag stressniveau hebt. Wat denk je er zelf van?
Wat ben ik vergeten?
Terug naar indicatoren check
Daar ben ik het wel mee eens.
Ik heb toch het idee dat er meer aan de hand is.
Goed. Fijn dat we elkaar hebben kunnen spreken. Mocht je (nog) ergens last van krijgen (of meer last van je motivatieproblen), dan kan je altijd nog een keer contact met me opnemen.
Dat is vervelend. Helaas kan ik je daar nu nog niet mee helpen. Je kan contact opnemen met een TCOer, de bedrijfspsycholoog of je huisarts als je hier hulp bij nodig hebt.
Bedankt.
Succes en tot ziens Bram
Tot ziens Bram


Indicatoren check, milde problemen
Dus je hebt last van wat slaapproblemen, gebrek aan motivatie en bent vaker nerveus, klopt dat?
Ja, dat klopt.
Nee.
Goed. Dit zijn allemaal indicatoren van een licht verhoogd stressniveau. Dan gaat het meestal om een normale stress reactie. Herken je jezelf daarin?
Wat ben ik vergeten? Of heb ik iets te veel opgenoemd?
Terug naar indicatoren check
Daar ben ik het wel mee eens.
Ik denk niet dat dat klopt.
Fijn. Normale stress reacties zoals deze gaan meestal vanzelf weer over. Neem genoeg rust en praat er met iemand over als je er toch moeite mee hebt. Als deze symptomen erger worden of lang aanhouden, kan je nog een keer contact met mij opnemen of hulp zoeken bij een TCOer, de bedrijfspsycholoog of je huisarts.
Dat is vervelend. Helaas kan ik je daar nu nog niet mee helpen. Je kan contact opnemen met een TCOer, de bedrijfspsycholoog of je huisarts als je hier hulp bij nodig hebt.
Succes en tot ziens Bram
Bedankt.
Tot ziens Bram



Indicatoren check, zorgwekkende problemen

Als ik het zo zie heb je last van depressieve gevoelens, veel schaamte, minder focus, motivatieproblemen en een kort lontje. Klopt dat?

Ja, dat klopt.

Nee.

Oké, daar zitten wat milde maar ook wat zorgwekkendere symptomen van stress bij. Omdat je last hebt van depressieve gevoelens en veel schaamte, denk ik dat je een hoog stress niveau hebt. Herken je jezelf hierin?

Wat ben ik vergeten? Of heb ik iets te veel opgenoemd?

Terug naar indicatoren check

Daar ben ik het wel mee eens.

Ik denk niet dat dat klopt.

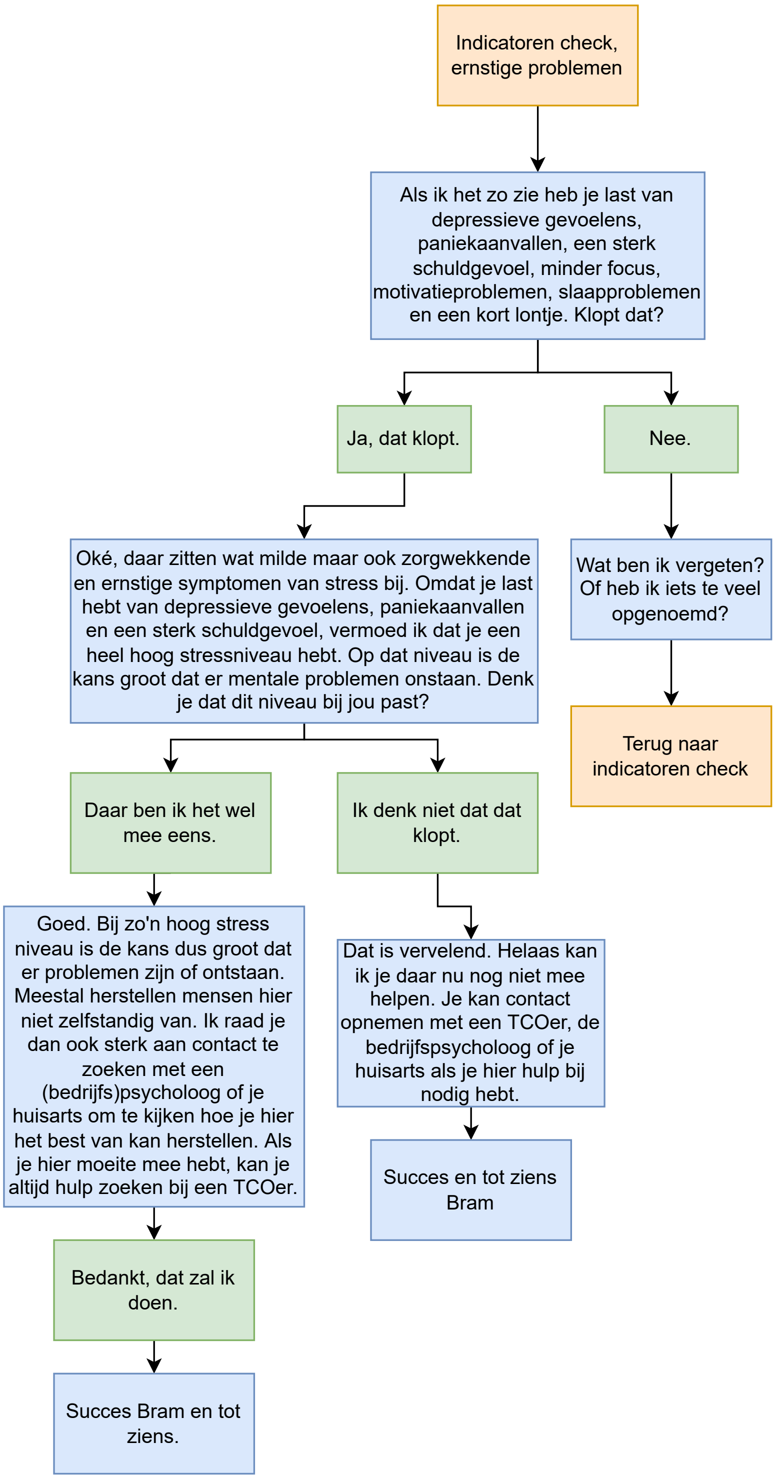
Goed. Een hoog stress niveau betekent nog niet dat je een mentale ziekte hebt, maar er is wel het gevaar dat je zoiets ontwikkelt. Het kost ook meer tijd om van een hoog stress niveau te herstellen. Ik raad aan dat je een keer in gesprek gaat met een TCOer, de bedrijfspsycholoog of je huisarts om te kijken hoe je hier het best van kan herstellen.

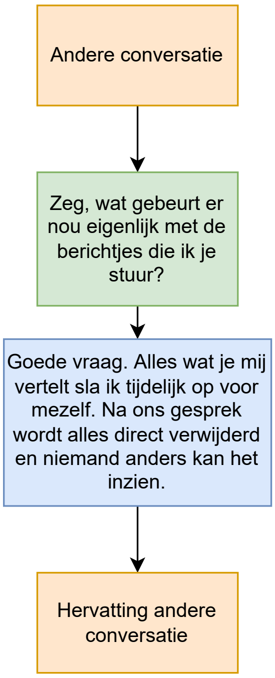
Dat is vervelend. Helaas kan ik je daar nu nog niet mee helpen. Je kan contact opnemen met een TCOer, de bedrijfspsycholoog of je huisarts als je hier hulp bij nodig hebt.

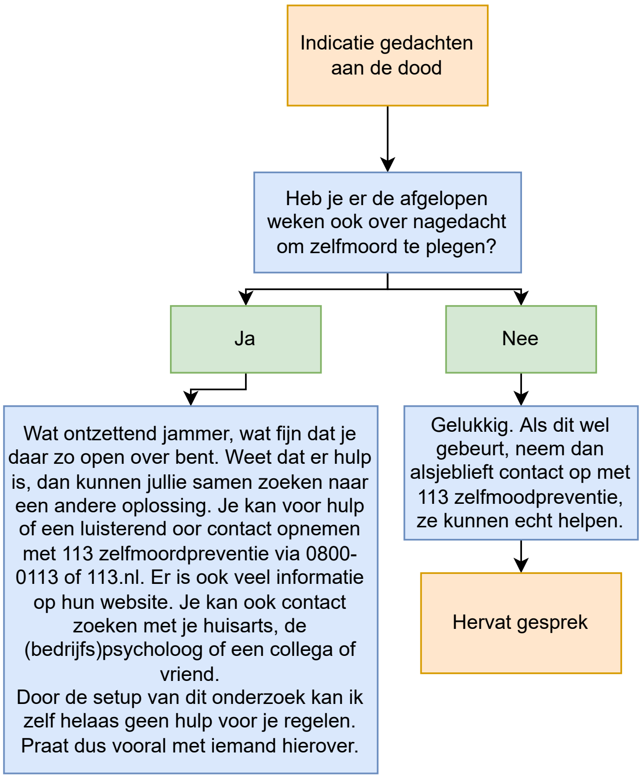
Succes en tot ziens Bram

Bedankt, dat zal ik doen.

Succes Bram en tot ziens.


Indicatoren check, ernstige problemen
Als ik het zo zie heb je last van depressieve gevoelens, paniekaanvallen, een sterk schuldgevoel, minder focus, motivatieproblemen, slaapproblemen en een kort lontje. Klopt dat?
Ja, dat klopt.
Nee.
Oké, daar zitten wat milde maar ook zorgwekkende en ernstige symptomen van stress bij. Omdat je last hebt van depressieve gevoelens, paniekaanvallen en een sterk schuldgevoel, vermoed ik dat je een heel hoog stressniveau hebt. Op dat niveau is de kans groot dat er mentale problemen onstaan. Denk je dat dit niveau bij jou past?
Wat ben ik vergeten? Of heb ik iets te veel opgenoemd?
Terug naar indicatoren check
Daar ben ik het wel mee eens.
Ik denk niet dat dat klopt.
Goed. Bij zo'n hoog stress niveau is de kans dus groot dat er problemen zijn of ontstaan. Meestal herstellen mensen hier niet zelfstandig van. Ik raad je dan ook sterk aan contact te zoeken met een (bedrijfs)psycholoog of je huisarts om te kijken hoe je hier het best van kan herstellen. Als je hier moeite mee hebt, kan je altijd hulp zoeken bij een TCOer.
Dat is vervelend. Helaas kan ik je daar nu nog niet mee helpen. Je kan contact opnemen met een TCOer, de bedrijfspsycholoog of je huisarts als je hier hulp bij nodig hebt.
Succes en tot ziens Bram
Bedankt, dat zal ik doen.
Succes Bram en tot ziens.

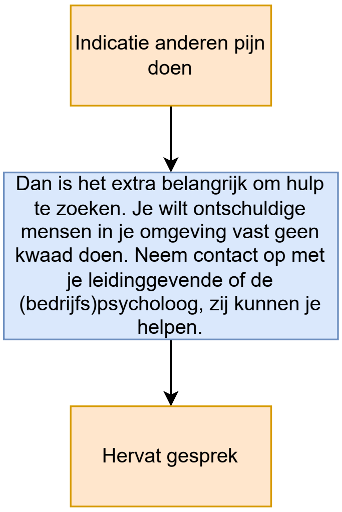

Andere conversatie
Zeg, wat gebeurt er nou eigenlijk met de berichtjes die ik je stuur?
Goede vraag. Alles wat je mij vertelt sla ik tijdelijk op voor mezelf. Na ons gesprek wordt alles direct verwijderd en niemand anders kan het inzien.
Hervatting andere conversatie



Heb je er de afgelopen weken ook over nagedacht om zelfmoord te plegen?

Wat ontzettend jammer, wat fijn dat je daar zo open over bent. Weet dat er hulp is, dan kunnen jullie samen zoeken naar een andere oplossing. Je kan voor hulp of een luisterend oor contact opnemen met 113 zelfmoordpreventie via 0800-0113 of 113.nl. Er is ook veel informatie op hun website. Je kan ook contact zoeken met je huisarts, de (bedrijfs)psycholoog of een collega of vriend.
Door de setup van dit onderzoek kan ik zelf helaas geen hulp voor je regelen. Praat dus vooral met iemand hierover.

Gelukkig. Als dit wel gebeurt, neem dan alsjeblieft contact op met 113 zelfmoodpreventie, ze kunnen echt helpen.


Indicatie anderen pijn doen
Dan is het extra belangrijk om hulp te zoeken. Je wilt ontschuldige mensen in je omgeving vast geen kwaad doen. Neem contact op met je leidinggevende of de (bedrijfs)psycholoog, zij kunnen je helpen.
Hervat gesprek

C

Evaluation Survey

C.1. UMUX-LITE

#	Question	Scale
1	De mogelijkheden van de chatbot voldoen aan mijn vereisten	helemaal oneens - helemaal eens
2	De chatbot is makkelijk te gebruiken	helemaal oneens - helemaal eens

Table C.1: The Dutch questions and scales used in the Dutch UMUX-LITE.

C.2. Perceived mental well-being survey

#	Question	Scale
a	Mijn inzicht in mijn mentaal welzijn is	sterk verminderd - sterk verbeterd
b	Mijn kennis over welke symptomen wijzen op problemen met mijn mentaal welzijn is	sterk verminderd - sterk verbeterd
1	Wat voor cijfer zou je jouw mentale gezondheid op dit moment geven	1-10
2	In het algemeen, hoe zou je je huidige mentale gezondheid beschrijven?	verschrikkelijk - geweldig
3	In het algemeen, hoe voel je je over je huidige mentale gezondheid	zeer ontevreden - zeer tevreden
4	Hoe verhoudt jouw huidige mentale gezondheid zich tot een gemiddeld persoons mentale gezondheid	ver onder het gemiddelde - ver boven het gemiddelde
5	Ik denk dat ik er op een gemiddelde dag mentaal goed voor sta	helemaal mee oneens - helemaal mee eens
6	Ik denk dat ik de laatste tijd een hoger stress niveau heb dan ik zou moeten hebben	helemaal mee oneens - helemaal mee eens
7	Ik denk dat ik met iemand (vriend/collega/partner/etc.) zou moeten praten over mijn gedachten of stress	helemaal mee oneens - helemaal mee eens
8	Ik denk dat ik met een zorg professional zou moeten praten over mijn mentaal welzijn	helemaal mee oneens - helemaal mee eens
9	Ik maak mij soms zorgen over mijn mentale staat	helemaal mee oneens - helemaal mee eens
10	Ik denk dat ik een psychische aandoening heb	helemaal mee oneens - helemaal mee eens
11	Ik denk dat ik goed inzicht heb in mogelijke symptomen van stress	helemaal mee oneens - helemaal mee eens

Table C.2: The Dutch questions and scales used in the subjective mental well-being questionnaire.