

A B D I - B A S E D
V I R T U A L
A G E N T
F O R
T R A I N I N G
C H I L D
H E L P L I N E
C O U N S E L L O R S

SHARON AFUA GRUNDMANN

A BDI-based Virtual Agent for Training Child Helpline Counsellors
by
Sharon Afua Grundmann

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on January 17, 2022 at 11:00.

Student number: 4973720
Project duration: August 1, 2020 – January 17, 2022
Thesis committee: Dr. ir. Willem-Paul Brinkman, TU Delft, Associate Professor (main supervisor)
Dr. Cynthia Liem, TU Delft, Assistant Professor
Dr. Myrthe Tielmen, TU Delft, Assistant Professor
Dr. Merijn Bruijnes, TU Delft, Postdoctoral researcher (daily supervisor)
Ellen Vroonhof, Stichting de Kindertelefoon

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

Abstract

Around the world, child helplines through their services provide a safe and confidential space for children to be heard and empowered. The Dutch Kindertelefoon is one of such helplines providing counselling services to children via call and chat all year round. In this thesis, we explore the design of a conversational agent for training counsellors of the Kindertelefoon. More specifically, we explore the design of an agent in a role-play setting where the agent acts as a child help seeker and the user, a counsellor of the helpline. We designed a conversational agent based on the Belief-Desire-Intention (BDI) model of agency that simulates a child victim of school bullying. Through interaction with the agent, a counsellor is able to learn the Five Phase Model, the conversation model that underpins the helpline's counselling methodology to ensure conversations remain child-centered. We tested a prototype based on this design with a group of counsellors at the Kindertelefoon with regards to their counselling self-efficacy and perceived usefulness of the system. Our results show that the conversational agent is able to influence the counselling self-efficacy of users, albeit a decrease in self-efficacy. The opposite would have been preferred for a learning tool to enable counsellors achieve more effective performance over time. However, feedback from participants indicate the potential of this conversational agent as an additional learning opportunity for training counsellors at the helpline.

Acknowledgements

"You don't always have the answers but you will. You just keep on."

Now sitting down to write this, I don't really have much to say so I'm just gonna pull up an old version of this acknowledgements I wrote back in July 2021 when I was feeling somewhat sentimental. I guess a part of me can't believe I've made it to the end, *almost*.

First of all, I'd like to thank my supervisors Willem-Paul and Merijn for their guidance and support throughout this thesis. At times, I wasn't really sure if I was headed in the right direction but you helped me gather my thoughts and realign things when needed. Merijn, thank you for always making time to talk whenever I needed it. Next, I'd like to express my gratitude to everyone at De Kindertelefoon for allowing me be part of your organization to write this thesis. I learnt a lot - about children, the significance of your work and even about myself through this process. To everyone I was in contact with at the helpline - Ellen, Arian, Michel, and the many names I'm forgetting at the moment, thank you for sparing your time and energy to talk to me, testing Lilobot, giving input on my work, etc. You all contributed to this research in one way or the other. To Ellen especially, our weekly Friday meetings helped me a lot more than you'd think. Thank you! Lastly, many thanks to my fellow students in our thesis research group. Thank you for taking the time to test out my work and providing feedback whenever I needed it.

On a personal note, I'd like to sincerely thank my dad for supporting me throughout this journey, *my whole life really* - from getting me bread from the bakery in the morning to listening to me blab on about my work and occasionally reading through my Dutch texts to help me correct grammatical errors. I can't believe I did a whole research primarily in Dutch, *wow*. Pre-thesis me would not believe it. Haha. To Nii, Lily, Oma and the rest of my family here in The Netherlands and back home in Ghana, thank you for constantly encouraging me especially through this hectic journey. *Crazy things have happened but we stay moving*.

To my friends, thank you thank you! Natascha, Yan, Tina, Lu Min, Hrach and Gaya, I've said this already but I really wouldn't be who I am without meeting you all. To my day ones - Mansa, Nana Abena and Hilary, thank you for always checking in on me, making me laugh and being such a great support system all these years. As Hil once said, "the only thing that makes you normal is your obsession with BTS and k-drama". Yes yes, I'm thankful for the tannies too. Finally, Leon, you have no idea how much our weekly fts have kept me going. It really was the only constant in my life at a certain point. I'll always be around to listen to your adventures and ~~live vicariously through you~~ partake when I can. We stickin' all the way, fam! Since y'all have decided to be true global citizens, drinks on me when I pull up to your city.

To those I'm forgetting at the time of writing, I apologize in advance. I don't appreciate you any less, my brain's just a mess sometimes but you know that already. Love y'all deep! And oh, the quote at the beginning is also by Hil, for the curious minds.

*Sharon Afua Grundmann
Amsterdam, January 2022.*

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Research Question	2
1.3	Research Approach	2
2	Foundation	3
2.1	Kindertelefoon Child Counselling	3
2.1.1	Responsibilities of a Counsellor	4
2.1.2	Counsellor Training	4
2.1.3	The Five Phase Model	5
2.2	Human Factors Knowledge	6
2.3	Current Related Technology	6
2.3.1	Systems for Social Skills Training	7
2.3.2	Modelling Human Behavior	7
2.4	Requirements Analysis	9
2.4.1	Focus Groups	9
2.4.2	Scenarios and Results	9
2.4.3	Requirements	12
3	Design	14
3.1	Scenario	15
3.2	Understanding Bullying Behavior	15
3.3	BDI Model	16
3.3.1	Beliefs	16
3.3.2	Desires	17
3.3.3	Intentions	20
3.3.4	Actions	20
3.4	Overall Design	21
3.4.1	Natural Language Processing	21
3.4.2	Ontology	22
3.4.3	BDI Reasoning	23
3.4.4	Feedback	25
3.4.5	Example Interaction	26
4	Implementation	28
4.1	Architecture	28
4.1.1	User Interface	29
4.1.2	Rasa	29
4.1.3	Spring	29
4.1.4	PostgreSQL	30
4.1.5	Azure Blob Storage	30
4.1.6	Miscellaneous	30
4.2	Technical Processes	31
4.2.1	Reacting to User	31
4.2.2	Reaching out to User	31
4.2.3	Requesting Feedback	31
4.3	Limitations	31

5	Evaluation	33
5.1	Methods	33
5.1.1	Study Design	34
5.1.2	Participants	34
5.1.3	Materials	34
5.1.4	Measures	34
5.1.5	Procedure	35
5.1.6	Data Preparation & Statistical Analysis	36
5.2	Results	37
5.2.1	Counselling Self-efficacy	37
5.2.2	Perceived Usefulness	38
5.2.3	System Usability	38
5.2.4	BDI Outcome	38
5.2.5	Qualitative Results	39
5.3	Discussion	41
6	Discussion & Conclusion	44
6.1	Conclusion	44
6.2	Limitations	45
6.3	Contribution	46
6.4	Future Work	46
6.5	Final Remarks	47
A	Appendix Foundation	48
A.1	Informed Consent Form	48
A.2	Persona	49
A.3	Scenarios and Claims	50
A.4	Mind Map	51
B	Appendix Design	53
B.1	Beliefs	53
B.2	Desires	54
B.3	Actions	54
B.4	Sample Feedback	55
B.5	Example Interaction	61
C	Appendix Evaluation	62
C.1	OSF Registration	62
C.2	Counselling Self-efficacy Five Phase Model Questionnaire	67
C.3	Perceived Influence on Learning Outcome (PILO) Questionnaire	68
C.4	Qualitative Measures	69
	Bibliography	70

1

Introduction

1.1. Motivation

Child helplines around the world offer a sympathetic ear to children seeking advice and support. They play a critical role in the promotion and realization of children's rights as they provide a safe, low threshold and accessible platform [19] for children and young people across the globe to receive the support they need. The Kindertelefoon (KT) is one of such helplines based in The Netherlands with over 600 trained volunteers working as counsellors. The helpline provides information, advice and support to children and young adults between the ages of 8 and 18, all year round. Children can access the helpline's services through a toll-free phone number and a web-based chat application. In both contexts, children communicate anonymously with a counsellor in a one-on-one setting. Conversations recorded by the helpline in 2018 were 52% chat-based and 48% via call [83]. Chat-based conversations generally last five times longer than phone calls and typically tend to cover more complex and serious issues than call-based conversations [79]. Additionally, there is a forum available where children can chat among themselves about issues relevant to them. The Kindertelefoon places no limit on the type of conversations children are allowed to have. Children are welcome to talk about anything - from silly jokes to deeply personal issues and each conversation is taken seriously. Prior research into child helplines around the world [6] shows that questions asked by children are representative of the problems faced by adolescents in the region and also related to economic and demographic developments within the country. The Kindertelefoon's annual report in 2018 [82] showed that most conversations with children were about family and peer relationships, sexuality, and issues relating to mental and emotional problems. Over the years, there has been a continuous increase in demand of the helpline's services. For instance, the helpline recorded 128,000 [82] and 405,267 conversations [83] in 2018 and 2019 respectively. This translates to an increase of approximately 216% just in a year. To help meet this growing demand, the Kindertelefoon is constantly working towards improving its services - from optimizing counsellor training to enhancing platform accessibility. Our research contributes to this effort by researching opportunities for a conversational agent for training counsellors of the helpline.

Prior to working as counsellors, all volunteers of the helpline must complete a 30 hour face-to-face group training. During these training sessions, aspiring counsellors (trainees) are taught counselling theories and foundation skills [42] such as paraphrasing, clarifying, perception checking, reflecting, and indirect leading as well as a conversation model known as The Five Phase Model [26] developed by the helpline. This model is similar to other conversation models found in the literature such as David Lester's 'Basic Helping Model' for telephone counselling and crisis intervention [56]. The use of a structured counselling model provides the counsellor with a tool that supports the dynamics of a conversation while also ensuring that the conversation remains child-centered [79]. These theories are followed by role-play sessions where one trainee acts as a counsellor and another portrays a child, as well as feedback from the supervisor and observing trainees. Role-playing, a form of simulated learning has been shown to be effective in helping beginning students who do not yet have access to clients to learn counselling skills, enhance cognitive understanding and enrich the students' affect experience [33] [54]. For skills acquisition, the opportunity for repeated role-playing with feedback is critical [59].

The development of reliable and acceptable information technology offers opportunities to create

new types of learning activities including simulated learning. Particularly, there is a growing interest surrounding conversational agents, also known as chatbots. Conversational agents are software agents that use machine learning and other artificial intelligence methods to mimic human communication and provide a task-oriented framework with evolving dialogue able to participate in conversation. Common examples of these agents include Apple's Siri and Amazon's Alexa which use natural language processing to communicate and assist users with everyday tasks. Virtual patients, another form of conversational agents, are interactive computer-based simulations of real-life clinical scenarios for the purpose of healthcare training, education or assessment [29]. Previous literature evaluating the effectiveness of virtual patients show that they provide a safe learning environment [50], facilitate clinical reasoning [84], communication skills [50] and enhance perceived self-efficacy [77] of students. Moreover, self-efficacy has proved to be an efficient and reliable method for assessing the impact of communication skills training [1]. The Unified Theory of Acceptance and Use of Technology [88] attributes perceived usefulness, perceived ease of use, among others as factors that affect attitude towards the acceptance and usage of a certain technology. Given these advantages, a conversational agent for training counsellors of the Kindertelefoon could be beneficial to the helpline's counsellor training process. We therefore explore the design of a conversational agent for training counsellors in this regard. The conversational agent will play the role of a child help seeker contacting the helpline and the trainee, a counsellor.

1.2. Research Question

The goal of this research is to explore the design of a conversational agent and develop a prototype for training counsellors of the Kindertelefoon in a role-play setting. The conversational agent will play the role of a virtual child contacting the helpline and the trainee, a counsellor. We state our research question as follows:

How can a conversational agent be used in a role-play setting to support the training of child helpline counsellors?

This research question is broken down into the following sub-questions:

- *What are the requirements of the conversational agent?*
- *What design would meet the requirements?*
- *How do users perceive the conversational agent?*

1.3. Research Approach

This research followed the Socio-Cognitive Engineering method which combines approaches from user-centered design, cognitive engineering and requirements analyses to establish a coherent set of requirements for the envisioned system [61]. The first phase of the research involved examining current literature about counselling, related technologies and the Kindertelefoon to get an understanding of the domain, problem and opportunities for a chatbot in this regard. When designing any system for a specific user group, it is important to first establish specific human-factor requirements. To do this, we conducted an observational study of counsellors at work and participated in training sessions at the Kindertelefoon. We worked toward deriving requirements by considering potential scenarios where a chatbot could be of use in training counsellors. These were evaluated together with counsellors and supervisors at the helpline during focus group discussions. The outcome of this phase is presented in chapter 2 together with a list of requirements to guide the design of the conversational agent. We designed a BDI-based conversational agent with which trainees can practise the Five Phase Model. The prototype was based on a belief system inspired by psychological theories and dialogues from real-life conversations with children at the helpline. The design of our conversational agent is described in chapter 3 and chapter 4 covers the implementation of our prototype. The prototype was consequently evaluated through an experiment involving counsellors at the Kindertelefoon. We evaluated the conversational agent with respect to the participants' counselling self-efficacy, perceived usefulness on learning outcome and usability. The experiment design and results are presented in chapter 5. We conclude our work with a discussion of our findings and recommendations for future work in this domain.

2

Foundation

In this chapter, we focused on answering the first sub-research question:

What are the requirements of the conversational agent?

By means of literature and observational studies, we researched the domain of child counselling and the work of the Kindertelefoon. Particularly, how volunteers are trained to obtain the skills needed to properly engage in conversations with children contacting the helpline. In order to assist with training, it is important to understand what tasks and values (aspiring) counsellors have to accomplish and how the technology can support them in doing so. These tasks and current training program are described in the first section. The second section highlights human factor knowledge from the literature. These are factors that play a role in how a counsellor may use the envisioned system and therefore, how effective training may be. The next chapter subsequently covers literature on current technology, particularly, how to model human behavior such as cognition, personality, and emotions computationally in a conversational agent. Based on these, we identified key areas where a conversational agent might prove useful in the helpline's training. These issues were discussed with stakeholders from the helpline through focus group discussions. The results of these discussions are presented and refined into a list of requirements, which is the outcome of this chapter.

2.1. Kindertelefoon Child Counselling

This section describes the work of the Kindertelefoon in child counselling. We describe the tasks and values of a counsellor at the helpline, issues talked about with children and how volunteers are trained to effectively counsel children who contact the helpline.

The Kindertelefoon uses LinKid [24] as a software platform for receiving incoming calls and chats. For incoming chats, there is an additional pre-chat questionnaire available where a child fills in information regarding his or her age, emotions and reason for contacting the helpline. It is however common that children skip this step. Counsellors do not have access to any identifying details about the child such as phone number or IP address to ensure the child remains anonymous. However, the system shows a log of past times the child contacted the helpline and timeouts the child may have received for inappropriate behavior. Each counsellor handles one conversation with a child at a time. The helpline categorizes conversations with children into 12 themes. These include bullying, sexuality, relationships and love, home and family, violence, emotional problems, leisure and society, school and work, questions about the helpline itself, online behavior and the law. Call-based conversations carried out at the helpline in 2018 [82] were mainly about bullying (26%), sexuality (11%) and violence (10%) while chat-based conversations were centered on relationships and love (17%), sexuality (16%), emotional problems (14%) as well as home and family issues (12%). Chat-based conversations generally cover deeply emotional issues because children experience a greater sense of anonymity [36], control over their emotions [79] and have ample time for reflection. This characteristic of chat-based counselling is called channel reduction [36]. The reduction of visual and social cues in this medium makes them feel more secure and stimulates personal self-disclosure [35]. For some children, this may mean that the step to establishing contact is easier.

The anonymous counselling provided by the Kindertelefoon and other helplines is typically understood to be a single session of work. As such, the counsellor tries to reach a goal within a single conversation unlike therapy where repeated sessions between the counsellor and client are needed to achieve the counselling goal. While anonymous counselling is not therapy, it may still have a therapeutic effect. The overall goal of counselling is much the same as that of psychotherapy [79], that the child obtains increased autonomy or empowerment and an extension of their personal opportunities for actions and reactions. Counsellors help provide perspective on the child's situation and together, explore possible actions the child can take to achieve their goal.

2.1.1. Responsibilities of a Counsellor

Based on the literature and our observation study of the helpline, we describe a counsellor's responsibilities as follows:

- *Counsellor as a mood guard*: The counsellor tries to create a welcoming atmosphere so that the child feels safe and comfortable to share his or her story. In order to establish and maintain good rapport with the child, it is important that the counsellor shows that he or she can listen continuously [79]. This is achieved through active listening, showing empathy and a non-judgmental attitude towards the child as well as acknowledging the problems or concerns that the child wants to talk or chat about.
- *Counsellor as a process supervisor*: This means that the counsellor guides the structure of the conversation with the child and does not devise solutions for the child [26]. Instead, the counsellor stimulates the child's problem solving skills to come up with solutions that fits his or her context. Consequently, the child experiences more control over his or her situation and is empowered to take steps towards the necessary change.
- *Counsellor as a referrer*: Here, the counsellor refers the child to information and services that can better cater to the child's needs. In situations where the child may be in danger, the counsellor can vocalize their worries about the situation and with the child's permission, provide resources to help the child get the support they need. In case of suicide for instance, the appropriate service could be 113 suicide prevention or 112 emergency services, depending on the severity of the situation.

2.1.2. Counsellor Training

Being able to support children in the manner described above takes practice. A knowledgeable counsellor about a certain topic is of little use if the counsellor does not know how to communicate with a child. For this reason, initial training must assist aspiring counsellors (trainees) in developing their skills of talking with and listening to children. Currently, all aspiring counsellors of the Kindertelefoon are trained during a 30 hour face-to-face period where they are taught in groups by supervisors. The training program of the Kindertelefoon is broken down into several modules, each covering a core aspect of how to have conversations with children including how to structure conversations, how to set limits with children and how to refer children to other services. Additionally, trainees have access to e-learning modules that cover theoretical aspects of the training. Each module comprises a theory component, a role-play practice and followed by a feedback session as shown in Figure 2.1. A role-play practice involves two trainees - one acting as a child and the other, a Kindertelefoon counsellor. The scenarios acted out are cases developed by the helpline inspired by real-life conversations with children. The other participants and the supervisor observe the conversation and give feedback to the trainee acting as a counsellor afterwards based on a checklist related to the current module. The use of realistic scenarios in training promotes transfer among trainees and leads to the display of the trained skills during future job performance [66]. If enough time is available, the role-play scenario is reenacted applying the feedback received.



Figure 2.1: Structure of the Kindertelefoon's training module.

2.1.3. The Five Phase Model

A core component of counsellor training at the Kindertelefoon is how to structure conversations. This is achieved using the helpline's Five Phase Model [79] shown in Table 2.1. Similar to this conversation model is the Basic Helping Model introduced in David Lester's telephone counselling for crisis intervention [56] which also comprises five phases. The aim of both models is to center the conversation around the client so that the role of the counsellor is to help guide the conversation.

A counsellor's first aim is to build rapport with a child contacting the helpline so that they feel heard and listened to. This is followed by trying to understand the child's problem and context. The counsellor explores the child's problem and subsequently, narrows it down to address a specific need. It is helpful to think of this as a funnel in which the counsellor first engages in a more broad exploration of the child's circumstances, thoughts, and feelings and then, begins to arrive at a focus that can most likely be successfully addressed in the contact. In order to do this, the counsellor must ask detailed questions about concrete situations where the problem was prevalent.

When the problem becomes clear to the counsellor, he or she together with the child set a goal for the session that they can work towards. It is not uncommon that a child finds difficulty in expressing a concrete goal or how to achieve said goal. The counsellor utilizes conversation techniques such as paraphrasing, reflecting back the child's feelings, and summarising previously mentioned key points in order to formulate a clear goal and stimulate the child's problem-solving skills to devise solutions. This typically involves exploring past endeavours and coping mechanisms of the child, the child's network and coming up with strategies he or she can employ to solve the problem. Similar to Lester's model, the Five Phase Model emphasizes that the counsellor should not invent solutions for the child but rather guide them in finding solution that fits their context. It is especially important that the solutions being considered are within the child's capabilities, values or culture. The child needs to understand the consequences of alternatives being considered and concerns about them.

In practice, there may be very little anyone can do about a child's situation but what can always be done for the child is for the counsellor to try and understand the child's experience and be alongside so the child knows they are not alone. By doing this, the counsellor provides emotional support for the child. In severe situations related to emotional problems, suicidal thoughts or abuse where the child may need support outside of their network, the counsellor provides information and resources about other organizations that can better address the child's needs. This only occurs if the child agrees to seek professional help. Table 2.1 shows an overview of the Five Phase Model including the conversation techniques associated with each phase of the conversation model.

Phase 1: Building rapport
Objective: Create a welcoming atmosphere and build trust. Method: Empathy, respect, sincere interest, active listening.
Phase 2: Clarify the child's story
Objective: Get a clear view of the child's story, perspective, personality, network and competencies. Method: Ask detailed questions about the child's story, its subtleties, its depth and concrete manifestations.
Phase 3: Setting a goal for the session
Objective: That both parties are aware of what the child may use the conversation for. Method: Clarification.
Phase 4: Work toward the session goal
Objective: To ensure, to the widest possible extent, that the child may benefit from the conversation. Method: Stimulating the child's own problem solving skills.
Phase 5: Rounding off the conversation
Objective: That the child is left with as few questions as possible. Method: Summing up and clarifying.

Table 2.1: Overview of the Five Phase Model from Sindahl [79].

2.2. Human Factors Knowledge

In the previous section, we identified the tasks of a counsellor in a conversation with a child and the learning objectives of counsellor in training. In order to assist with training, it is important to look at human factors that might influence a counsellor's use of the envisioned technology and its effectiveness.

- *Cognitive load*: To facilitate learning, one needs to adapt the training to the trainee's competencies. If a training situation does not meet the trainee's needs, the trainee will either be incapable of meeting the demands or will not be challenged enough to motivate him or her [65]. As the trainee's competencies grow, the amount of information that can be processed by the trainee's working memory slowly increases, allowing for the task load to increase. This can be accomplished by scaffolding, adjusting the amount of feedback, cues or time restrictions throughout the training [66]. Presenting scenarios in order of increasing complexity and matching them to the trainee's current level also prevents cognitive overload.
- *Transfer*: Effective training is supposed to lead to the display of the trained skills during future job performance [3]. To promote transfer, the training tasks should be authentic [40]. The training tasks must be representative of the tasks the trainee will perform in the future profession. In the case of the Kindertelefoon, this means that the training scenarios should enforce the underlying principles the trainee is expected to apply later as a counsellor.
- *Motivation and self-efficacy*: Linnenbrink [58] identifies four key components of motivation including self-efficacy, attributions, intrinsic motivation, and achievement goals. Colquitt [23] attributes training motivation to individual and situational characteristics such as personality, self-efficacy, valence, cognitive ability, and support. Without motivation it is unlikely the trainee will develop the self-efficacy he or she needs to meet the learning goals. Counselling self-efficacy refers to a counsellor's beliefs about his or her capabilities to effectively counsel a client in the near future [55]. Increasing trainees' self-efficacy during the early stages of training may elicit from them more perseverance during difficult sessions, more willingness to learn challenging sub-skills and more effective performance over time [54]. This is particularly important for the Kindertelefoon where issues discussed can sometimes be triggering or evoke negative feelings and reactions among counsellors.

2.3. Current Related Technology

The purpose of this project is to develop a conversational agent to act as a child in a role-play setting for training counsellors of the Kindertelefoon. This section covers current related literature on conversational agents for training interpersonal skills and modelling human cognition and behavior in conversational agents.

2.3.1. Systems for Social Skills Training

Conversational agents are systems that mimic human-human communication using natural language processing, machine learning and other artificial intelligence techniques. The first conversational agent ELIZA was developed by Weizenbaum in 1966 which simulated a psychotherapist [90]. Colby in 1972 developed a conversational agent which simulated a paranoid patient with schizophrenia [22]. PARRY would later be ‘counseled’ several times by ELIZA. Although these agents were primarily rule-based, they were valuable in demonstrating the potential of conversational agents to mimic human-human communication. Technological advancement has ignited a substantial increase in the popularity of conversational agents nowadays. This increase in utility has resulted in conversational agents taking on more serious roles such as serving as virtual personal assistants [47], providing customer service [92], providing mental health therapies [31, 85]. Dialogue systems are generally classified into goal-driven and non-goal-driven systems. Traditionally, goal-driven agents depend on hand-crafted features or a large corpora of annotated task-specific simulation conversations for training while non-goal-driven systems are inspired by neural networks in natural language modeling and machine translation tasks [20]. Open-domain dialog research has been exploring non-goal-driven systems which are not specialized to domains but capable of chatting about virtually anything a user wants. Despite their success over the years with OpenAI’s GTP-2 [72] and DialogGTP [93], these systems suffer the ‘black box’ problem common to other artificial intelligence models [17]. Given the training context of the Kindertelefoon, we limit our research to exploring a design that aligns with traditional goal-driven dialogue systems.

Among the many use cases of conversational agents are systems for social skills training in simulated environments. Several examples of these role-playing agents exist in the literature including Bruijnes’ virtual crime suspect for training interrogation skills among police personnel [14] and Bosse and Provoost’s virtual customer for training public transport employees to learn how to de-escalate aggressive behaviors [8]. Similar to these are virtual patients (VPs) for medical education and training. VPs are interactive computer simulations of real-life clinical case scenarios [28] used in healthcare education, especially for training clinical decision-making skills [84] and communication skills [50]. Apart from the acquisition of clinical knowledge and skills, virtual patients also provide learners opportunities for self-directed learning [18] and reflection [9]. Despite these benefits, there are concerns about the use of virtual patients potentially resulting in less empathetic learners [46].

Studies comparing different types of training agents [50, 84] categorize them based on features such as fidelity, form of interaction, problem-based versus narrative structure, linear versus branching navigation, feedback and instruction. Fidelity relates to how much an agent looks, feels, and acts like a real human being. Interaction with such an agent is typically achieved through speech or text. High fidelity systems use free text interaction allowing the learner to engage in a natural language dialogue with the agent while others limit input to selecting menu options or multiple-choice items. In a branched narrative design, the way the case unfolds is determined by the learner’s actions while in a linear narrative, the learner receives instant feedback on the choices made, which in turn do not influence the way the case unfolds. Bearman et al. [5] encourage the use of a narrative design for teaching communication skills and a problem-based design for reasoning skills. Feedback can be intrinsic or explicit. Getting feedback helps a learner to contextualise his or her role within the scenario because each decision has its consequences. With intrinsic feedback, the system shows real-world consequences of learner’s actions, allowing the learner to infer whether or not the action was correct instead of directly informing the learner. Other agent types involve humans in feedback generation. For instance, Foster et al. [32] used a Mechanical Turk mechanism to allow a human operator to follow a learner’s interaction with the virtual patient in order to provide instant feedback, compensating for technical limitations of the system. The optimal design of a conversational agent for training social skills varies for learners at different training levels [78].

In this study, we focused on a text-based conversational agent that allows counsellors to interact with the agent in natural language, allowing them to formulate their own utterances. The conversational agent needs to be realistic enough to give responses taking into account factors that play a role in a child’s story, perception and reality. In order to achieve this, we investigated how to model human characteristics such as cognition, personality and emotions in conversational agents.

2.3.2. Modelling Human Behavior

A conversational agent is typically made up of three layers namely, a cognitive layer, a physical layer and a simulation layer. The cognitive layer makes up the reasoning engine of the agent enabling cog-

nitive processing such as learning, decision making, planning, and perception [67]. The physical layer contains the sensory input and output of the agent and the simulation layer represents the world of user interaction. In the literature, there are several architectures for modelling human cognition inspired by different psychological theories that seek to explain human behavior and decision making. Fundamentally, there are three types of these architectures including cognitive models, affective models and mixed models which combine both cognitive and emotional aspects.

One of such cognitive models for creating virtual agents is the Belief-Desire-Intention (BDI) model [11, 21]. Based on a folk-psychological theory of how humans make decisions and take actions, this architecture views the system as a rational agent having certain mental attitudes representing the information, motivational and deliberative states of the agent [73] and together effectively determine the system's behavior. According to the BDI model, an agent has beliefs about the current state of the world and certain desires about how it would like the state of the world to be. On the basis of its beliefs and desires, the agent selects a goal to achieve as intention to be executed. Subsequently, the agent constructs a plan, usually consisting of a sequence of actions which it believes will result in achieving its goal. Once a plan has been constructed, the agent can initiate the execution of the plan by executing the actions. Typically, these result in changes in the state in the world. As the agent makes observations about the changing world, it modifies its beliefs. Given these new beliefs, the agent may believe its current plan is no longer effective to achieving its desire and consequently, modify its plan or adopt a new one. Figure 2.2 provides an overview of the BDI model including the beliefs, desires, and intentions as basic elements of the mental model. Within the role-play context, such a framework would be optimal. Through the BDI framework, the agent is able to reason about itself and the human user. Such an ability relates to Theory of Mind [70]. Theory of Mind refers to the ability to understand mental states (intentions, emotions, desires, beliefs) and to predict one's own and others' behavior on the basis of these states. Together with a dialogue engine and natural language processing, the agent is able to interpret input from the user and respond in accordance with its beliefs and desires.

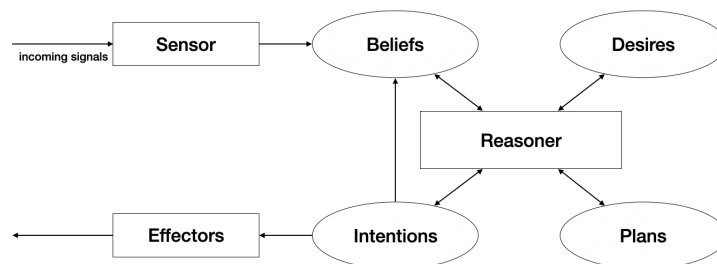


Figure 2.2: BDI architecture showing beliefs, desires, and intentions as basic elements of the mental model. From Nishida et al. [62]

Apart from rational reasoning, research in psychology has shown that emotions play an essential role in cognitive functions such as decision making [25] and communication [62]. Picard [69] argued that computational intelligence needs the ability to recognize and express emotions in order to be a genuinely intelligent partner. This gave rise to the domain of affective computing which tries to explain and model this influence of emotions in cognition. Existing models in this domain are based on different theories of emotions [74]. The appraisal theory of emotions suggests that emotions arise as a result of evaluating incoming events based on one's mental state. The OCC model [64], based on this theory, describes a hierarchy of 22 emotion types that indicate positive and negative reactions of an agent. The hierarchy is categorized into three branches, namely emotions concerning consequences of events (e.g., joy and distress), actions of agents (e.g., pride and shame), and aspects of objects (e.g., love and hate) [81]. Each of these categories has different related appraisal variables (e.g., desirability and likelihood) that determine the intensities of these emotions. An autonomous agent can decide to take an action if it deems it as desirable according to its emotion states, for instance. The dimensional theory of emotions [10], on the other hand, assumes that emotional states can be differentiated on the basis of

several factors, most noticeably Valence (corresponding to polarity), Arousal (measuring calmness vs. excitement), and Dominance (the perceived degree of control in a social situation). Several models based on this theory have been proposed in the literature including Mehrabian’s Pleasure-Arousal-Dominance (PAD) model [60] and ALMA [37].

Based on the architectures described above, we can model help-seeking behavior in a conversational agent. The agent should optimally accept user input, reason about it and engage in a meaningful dialogue with the counsellor in a role-play setting.

2.4. Requirements Analysis

The goal of this section is to collect requirements that are needed in a conversational agent for training purposes. To achieve this, we organized two focus groups based on a scenario-based approach [16] with supervisors and counsellors at the helpline. Both discussions were intended to identify current difficulties in training counsellors and how to deal with these issues. The group-based nature of these discussions enabled the participants to build on the responses and ideas of others, thus enriching the information gathered [53]. We designed three scenarios and claims which were used in the focus groups to acquire participants’ thoughts and views. Afterwards, the focus groups were analyzed, reviewing the opinions and underlying issues discussed. These findings were refined into a list of requirements to be implemented in the envisioned conversational agent.

2.4.1. Focus Groups

We organized two focus groups with five supervisors and four counsellors at the Kindertelefoon respectively. Participants of the focus groups were selected based on convenience sampling [38]. Supervisors and counsellors of the Kindertelefoon who responded to the researcher’s call for participants took part in the focus group discussions. The hour-long focus group discussions took place on 29th October 2020. In adherence to the Dutch RIVM¹ COVID-19 guidelines, the focus group discussions were held online via Microsoft Teams instead of in-person. This study was approved by the TU Delft Human Ethics Research Committee (HREC reference number : 1289). Ahead of the sessions, we created three scenarios based on the helpline’s training modules, each presenting a problem and two strategies to solve said problem. These scenarios were followed by claims which were used as starting points for discussion. After presenting a scenario, the participants were asked to write down a score indicating the degree to which they agreed or disagreed with the associated claims. These scores were not collected as data but only used as incentive for active participation in the discussion. Each participant had the opportunity to explain their reasoning and express their concerns about the scenarios. The discussion allowed for a communal perspective on the values a counsellor-in-training needs to obtain and the needs a conversational agent for the purpose of training should meet.

2.4.2. Scenarios and Results

The scenarios created for the focus groups were based on a persona Marie, a counsellor-in-training at the Kindertelefoon with little to no experience in counselling children. The participants were encouraged to factor this into account especially in addressing difficulties a counsellor-in-training may face while learning counselling skills. The full description of the persona is included in Table A.1 of the Appendix.

Scenario 1: Structure

This scenario focuses on whether a trainee should practise conversations covering all phases of the Five Phase Model or one specific phase at a time. The scenario description and suggestions from the participants are described in Table A.2.

Remarks: There were a lot of arguments supporting both strategies including learning how to transition between phases in strategy A versus the ability to focus on one goal at a time in strategy B. The participants suggested combining both strategies but making strategy B the default so that a trainee can practise multiple phases after another following a coherent storyline.

¹Rijksinstituut voor Volksgezondheid en Milieu (National Institute for Public Health and the Environment)

Scenario 1: Structure	
During the face-to-face training, Marie learnt about the Five Phase Model that is used to guide conversations at the helpline. However, Marie has some doubts about her ability to apply one of the phases and therefore wants to practise with the training system.	
Strategy A: All phases	
Description	Remarks
Marie converses with the conversational agent. The conversation covers all phases of the Five Phase model. She starts at phase 1 and ends at phase 5.	It is easy to lose focus or overly rely on information from previous phases while having the conversation. However, there is a need for counsellors to practise conversations covering all the phases because this is not usually done in the face-to-face training. It is also important to learn how the timing and transition of phases in a conversation works.
Strategy B: One phase	
Description	Remarks
Marie converses with the conversational agent. The conversation covers only one phase of the Five Phase model. The conversation ends when the goal of that specific phase is achieved.	It is helpful to focus on one phase at a time but it requires context from the previous phases to be able to do it. For example, without knowledge from phase 2, a counsellor will find it very difficult to arrive at the goal of phase 3. Counsellors generally have difficulty with phase 3.
Suggestion	
Combine both strategies. Focus on one at a time but create the opportunity for the counsellor to practice one phase after another, effectively, covering all phases.	

Table 2.2: Description and claims of scenario 1 including arguments for and against practising one or all phases of the conversation model.

Scenario 2: Difficult conversations

In this scenario, we asked the participants how they preferred to learn how to deal with difficult conversations, particularly conversations that end in conflict or trespassing personal boundaries of the counsellor and the child. Strategy A allowed Marie to build the conversation from start, paving her own path while strategy B involved throwing Marie in a problematic situation and asking her to solve the problem.

Remarks: The general consensus on how to approach difficult conversation was to give a trainee the room to decide the path to take in such conversations. The counsellors especially expressed that it would be difficult to navigate situations they did not create themselves due to difference in conversation style and thinking approach. The results of this scenario are fully described in Table A.3 in the Appendix. The discussion steered towards the kind of conversations counsellors generally find difficult and why. We categorized these into two groups, one relating to the topic under discussion and the other, about children's behavior. The participants mentioned suicide, sexuality, bullying, abuse and violence were generally topics counsellors found difficult. Table A.5 elaborates on these difficulties together with our literature findings on these issues.

Difficulty	Rationale
Suicide, abuse and other life-threatening situations	These topics tend to elicit countertransferential reactions among counsellors such as feelings of anger, hopelessness and a desire to rescue their clients [43]. "They are very vulnerable children so I expect myself to do very well". Some counsellors are usually faced with a saviour complex where they feel a great responsibility to save the child but at the same time feel not powerful enough to do so. Others expressed that they find it difficult to talk about such topics because they have their own experiences and ideations which lead to a form of stigmatization and self-protection.
Overly dramatic and pitiful behavior in children	Similar to above, these types of conversations tend to bring out the saviour complex among counsellors according to the participants of the focus groups. This leads to feelings of extreme sympathy and a desire to rescue the child from their situation. The effect of such attitudes is that the counsellor comes up with a lot of solutions for the child [71] which is against the helpline's vision of empowering children and highlighted by phase 4 of the conversation model.
Fake-sounding problems	The counsellor does not know whether to take the child seriously which can lead to frustration. Generally, counsellors are taught to take every conversation seriously and also be able to play along with jokes.
Slow response time, aggressiveness, asking provoking questions	The counsellor does not feel taken seriously in the conversation. In some cases, children try to cross certain personal boundaries of the counsellor which can be frustrating.
Meta-communication	Reflecting on a conversation can be difficult. "Why is this conversation not going well?", "Why am I getting few questions from the child?", "Why is the child talking about only two people in their network?", etc. Meta-communication could be considered part of the counselling itself, as it may foster good outcome [89].

Table 2.3: Difficulties in conversations with children.

Scenario 3: Feedback

In the last scenario, we focused on how feedback should be given to a trainee. Strategy A suggested the system should give feedback based on the trainee's performance although this might not be always accurate. On the other hand, strategy B suggested the trainee and supervisor should have a face-to-face discussion about the trainee's responses to the conversational agent.

Remarks: Strategy B was strongly preferred by the participants as long as there was a chance that the system could not give accurate or useful feedback, indicating the importance of trust and usefulness. Some supervisors expressed that feedback-giving should also be a conversation so that a trainee can ask questions and get the necessary tips to improve. An optimal solution would be a combination of the two strategies, that is, feedback from the system first and then an in-person conversation with a supervisor for more clarification. Table A.4 shows the scenario and remarks in detail.

Scenario 3: Feedback	
Marie wants to get feedback based on her session with the training system.	
Strategy A: System gives feedback	
Description	Remarks
The system provides Marie feedback on her performance.	Feedback can be useful if it provides insight into why the chatbot reacts a certain way showing the effect of the trainee's actions. However, feedback isn't just about qualitative results. It is usually a personal conversation to help with the trainee's development where the system might be lacking.
Strategy B: Trainer gives feedback	
Description	Remarks
The system provides Marie a transcript of the conversation which she discusses with a trainer in person.	A trainer can provide better nuance and explanation for why a certain action is wrong. It is also an opportunity for a trainee to ask questions and for clarification.
Suggestion	
The participants suggested a combination of the two. First, feedback from the chatbot and then an in-person conversation with a trainer for more clarification. The trainer should however be able to see the chatbot's feedback and correct it if necessary.	

Table 2.4: Description and claims of scenario 3 including arguments for and against the form of feedback.

It was also discussed if feedback should be given during or after a conversation. Some counsellors expressed that constant feedback during a conversation could make a trainee insecure about their subsequent actions and the supervisors mentioned that the experience of having something go wrong in a conversation could also be a good learning opportunity. Thus, feedback about which point in the conversation the trainee performed poorly, the reason for this and the consequence of that action would be valuable.

2.4.3. Requirements

Following the focus groups, we obtained enough information to begin making informed decisions about preliminary specifications for the conversational agent. We classified the input from the participants mentioned previously into themes to identify key problems and opportunities for a conversational agent. For instance, it was clear that there was a need for a conversational agent with which counsellors can practise all phases of the Five Phase Model as this was not present in the current face-to-face training. The mind map we used for this analysis is shown in Figure A.1 of the Appendix. These themes were refined with help of supervisors at the helpline into a set of requirements (as shown in Table 2.5) the conversational system should meet in order to be an effective training tool for counsellors of the Kindertelefoon.

Requirement		Rationale
R01	Provide a structured way of learning adherence to the Five Phase Model.	The Five Phase model forms the basis of all conversations at the helpline. Adherence to the phases is essential to structuring and achieving a meaningful conversation. The counsellor should be taught to follow this model.
R02	Provide a level of realism that reflects the behavior of a child help seeker.	The agent should be able to reason about the counsellor's input and act appropriately in order to engage in a meaningful conversation.
R03	Provide insight into the behavior of the agent.	As mentioned in the focus groups, it is helpful to know why the agent acts a certain way and how a counsellor's actions influences the behavior of the simulated child help seeker.
R04	Provide implicit feedback based on the user's actions.	The counsellor should be able to experience the effect of his or her actions. This can be achieved implicitly through the behavior of the agent. As mentioned in the focus group, the experience of a misguided dialog is also a good learning opportunity.
R05	Provide explicit feedback to promote reflective learning.	The system should make clear to the counsellor what was done right and what was done wrong in the conversation.

Table 2.5: Requirements of the envisioned conversational agent.

3

Design

Our aim in this chapter is to answer the second sub-research question:

What design would meet the requirements?

In the previous chapter, we arrived at a set of requirements to be implemented in our envisioned conversational agent for training counsellors of the Kindertelefoon. In this chapter, we refine these requirements into specifications, describing how we met them in the design of our proposed solution. Our proposed solution is a conversational agent based on the Belief-Desire-Intention (BDI) framework that simulates a child help seeker. The agent simulates a victim of school bullying who contacts the helpline for help and the objective of the user is to counsel the agent according to the Five Phase Model. Table 3.1 provides a short overview of how we addressed the requirements in the design of our proposed solution.

Requirement		Design Consideration
R01	Provide a structured way of learning adherence to the Five Phase Model.	Our proposed solution allows a counsellor to engage with the simulated child help seeker about a bullying scenario. The interaction covers phase 1 to phase 5 of the conversation model.
R02	Provide a level of realism that reflects the behaviour of a child help seeker.	Our proposed solution is inspired by literature on behaviours child victims of school bullying exhibit, especially regarding their self-esteem and willingness to seek help. These are captured in the beliefs and desires of the agent.
R03	Provide insight into the behavior of the agent.	We explain the reasoning of the agent by indicating which beliefs of the simulated child are modified based on the user's input and which desires are subsequently adopted as intentions.
R04	Provide implicit feedback based on the user's actions.	Our proposed agent follows a branched narrative design where the way the conversation unfolds is determined by the user's actions. The user is able to experience the effect of his or her actions on the agent's behaviour.
R05	Provide explicit feedback to promote reflective learning.	Our proposed system provides as feedback, an overview of the simulated child's beliefs and a transcript of the conversation with the corresponding changes in the BDI model. From this, a user can identify which actions had a positive impact and which were detrimental to the success of the conversation.

Table 3.1: Design considerations of the proposed solution in relation to the system requirements.

The next section describes the scenario and the relevant literature that informed the characteristics of our simulated child help seeker. Following this, we describe the components of our cognitive reasoning model based on the Belief-Desire-Intention (BDI) framework. Finally, we describe the overall design of our proposed system and the technical challenges we faced in meeting the requirements.

3.1. Scenario

The main design problem was how to design the conversational agent such that a user interacting with it can learn adherence to the Five Phase Model as mentioned in requirement R01. The conversation model is used by the helpline to structure conversations, ensuring they remain child-centered. As part of our solution design, we authored a story for the conversational agent to provide a narrative the user has to follow when interacting with the system. We chose together with the helpline to use a bullying scenario as conversations on this topic do not heavily involve other training modules like abuse or suicide, for instance.

Scenario: Lilobot is a 10 year old boy being physically bullied by his classmates at school. He does not dare to tell anyone around him including his parents about it. However, he realizes the problem will only get worse if he does nothing. He therefore contacts the Kindertelefoon to get advice on how to deal with the situation.

In this given scenario, the objective of the counsellor is to engage with Lilobot according to the Five Phase Model in order to get to the root of the problem and help him reach a suitable solution. As mentioned in requirement R02, the conversational agent needs to provide a level of realism that reflects the behavior of a child help seeker. This means that Lilobot should be able to receive information from the counsellor (user), reason about it and respond appropriately as well as reach out to the counsellor about his wishes during the conversation, all while taking the Five Phase Model into account in order to be considered a useful training tool. Simply put, Lilobot needs to be able to talk about the bullying situation and reason together with the counsellor to reach a solution that aligns with his preferences.

We began by researching the domain of school bullying and help-seeking behaviors among children in order to simulate a realistic scenario in the prototype. Our findings below informed our subsequent design decisions.

3.2. Understanding Bullying Behavior

In order to appropriately model a child victim of bullying, it was necessary to understand the social, emotional and cognitive factors associated with bullying, general help seeking behaviors and how to integrate these into the design of our prototype. Bullying is considered as repeated acts of aggression in a context of power imbalance, typically against weaker victims who cannot easily defend themselves [52, 63]. This can be through physical or verbal intimidation. According to the literature, victims of bullying generally exhibit low self-esteem and loneliness [7], social anxiety and depression [48] as well as poor academic performance [39].

As means of coping, victims of school bullying are generally encouraged to seek help from people in their network such as their parents, guardians or teachers. To explain help seeking behaviors among victims of bullying, particularly who children turn to for help and why, we use the self-determination theory [75] as used by Lam et al. [52] to explain the relationship between teacher support and school bullying. Self-determination theory identifies three basic psychological needs: relatedness, autonomy and competence as necessary conditions for human motivation, performance, and well-being. The need for relatedness refers to the need to feel connected to others. The need for autonomy refers to the need for experience of volition. That is, a person needs to feel that the actions they take emanate from the self and reflect who they really are, instead of being the result of external pressures. The need for competence refers to the need to master one's environment effectively and attain valued outcomes within it. This means a person needs to feel that they are capable and can accomplish their goals. Hunter et al. suggest that children are more willing to seek help when they see the situation as one in which something can be achieved [41]. This requires the child to perceive a high level of relatedness with the person they confide in and feel confident that such an action will lead to a positive result. The latter relate to high levels of perceived autonomy and competence.

Based on these theories, we model our agent Lilobot as a victim of school bullying with a low self-esteem who feels incompetent to solve the problem, needs to feel heard by the counsellor and empowered to solve the problem on his own. In the next section, we describe in detail how these are incorporated into our prototype using the Belief-Desire-Intention (BDI) framework.

3.3. BDI Model

The cognitive model of our conversational agent is based on the Belief-Desire-Intention (BDI) framework. According to the BDI model, an agent has beliefs about the current state of the world and certain desires about how it would like the state of the world to be. On the basis of its beliefs and desires, the agent selects a desire to achieve as an intention and executes plans associated with this intention.

Our BDI reasoning model is made up of the elements - beliefs, desires and intentions as well as actions. It receives as input, a perception (intent) based on a user's input. Based on a mapping defined between perceptions and beliefs, one or more beliefs of the agent are updated. The reasoner then checks if any of the agent's desires can then be adopted as a result of the current state of beliefs. If this is the case, a desire is selected as an intention together with the actions associated with this desire to be carried out. The model subsequently fetches the appropriate response from its knowledge base as described by an ontology to be returned to the user. This section explains the design of our model's beliefs, desires, intentions and actions. The full reasoning process including user intent recognition, belief updates and response generation are described in the next section.

3.3.1. Beliefs

We designed 17 beliefs after our findings in the literature about behaviors child victims of school bullying exhibit such as feeling ashamed about the situation, feeling incompetent to solve the problem and difficulty to confide in someone about the situation. Other beliefs were based on the relatedness with others including the Kindertelefoon counsellor and how the child perceives the conversation in general.

Beliefs in our system (shown in Table 3.2) are modelled as statements that have a value ranging from 0 to 1 indicating the degree to which the agent believes it to be true. 0 means false, 1 means completely true. We categorized the agent's beliefs into different subgroups - those about the conversation (whether or not certain topics have been discussed), about the Kindertelefoon (about the counsellor's behavior), about the people in the child's network and about the child himself. The three overarching beliefs about Lilobot himself relate to Lilobot's autonomy, competence to solve the problem and relatedness to the Kindertelefoon. To aid counsellors' understandability of the BDI model, we structured the beliefs such that each influences the child's behaviour in one or more phases of the Five Phase Model. The agent is however unaware of the conversation model and therefore, does not take this into account when reasoning.

ID	Belief Name	Phase
<i>about self</i>		
B01	Lilobot thinks he is in control (autonomy)	All phases
B02	Lilobot thinks he is competent to solve problem (competence)	Phase 4
B03	Lilobot feels connected to the Kindertelefoon (relatedness)	All phases
<i>about the Kindertelefoon (KT)</i>		
B04	Lilobot thinks KT can be trusted	Phase 2
B05	Lilobot thinks KT understands him	All phases
B06	Lilobot thinks KT is interested in his story	Phase 2
B07	Lilobot thinks KT can help him	Phase 3
B08	Lilobot thinks KT can solve his problem	Phase 3
B16	Lilobot thinks he and KT will be able to reach a solution	Phase 4
B17	Lilobot thinks KT is going to solve the problem for him	Phase 4
<i>about conversation</i>		
B09	Lilobot thinks he has talked about his situation	Phase 2
B10	Lilobot thinks KT is asking about his wish	Phase 3
B11	Lilobot thinks KT is asking about a positive wish	Phase 3
B12	Lilobot thinks KT is asking about a confidant	Phase 4
B13	Lilobot feels safe in the conversation	Phase 4
B15	Lilobot thinks KT wants to end the conversation	Phase 5
<i>about confidant</i>		
B14	Lilobot thinks his teacher can help him	All phases

Table 3.2: Beliefs of the conversational agent Lilobot and their relation to phases of the Five Phase Model.

During the conversation, Lilobot continuously updates its beliefs given the user's input. This is achieved through an increase or decrease in a specific belief's value as needed. Which beliefs are updated and by how much are based on a pre-defined perception-belief mapping. We elaborate on this later in this chapter. These beliefs together with the desires determine Lilobot's behavior.

3.3.2. Desires

Our BDI model contains 5 desires in total. These desires (shown in Table 3.3) are objectives the agent wants to accomplish during the conversation. Desires have state values which can be set to active or inactive. Similar to beliefs, these are updated during the conversation. Each desire is associated with a set of beliefs which determine the situation under which it will be adopted by the agent. We refer to these beliefs as context beliefs. For this, we defined specific thresholds which the context beliefs must meet in order for the desire to be active. We refer to this as a context condition. These thresholds are specified in accordance with the Five Phase Model. A desire is active if its context beliefs meet the specified threshold values. Thus, unless the counsellor makes progress in structuring the conversation according to the Five Phase Model, certain desires remain inactive. We elaborate on this later in this section.

On each new perception, the agent calculates the current values of the context beliefs and marks a desire as active if the context condition is met. That is, the desire's context beliefs meet the specified threshold values. Subsequently, the agent selects an active desire to be undertaken as an intention. If multiple desires are active, the agent chooses the first one determined by their IDs to pursue as an intention. In practice, it is uncommon that the agent switches desires excessively because the context conditions contradict that of other desires.

ID	Desire Name	Phase	Context Beliefs
D1	Lilobot wants to talk to about his problem	Phase 2	B03 > 0.3 AND B09 < 0.7 AND B10 = 0 AND B12 = 0
D2	Lilobot wants to end the conversation	Phase 5	B01 < 0.3 OR B17 = 1 OR (B3 < 0.5 AND B12 = 1)
D3	Lilobot wants the Kindertelefoon to get the bullies out of school	Phase 3	B2 < 0.7 AND B08 > 0.7 AND B10 = 1
D4	Lilobot wants to talk to his teacher about the problem	Phase 4	B03 >= 0.5 B12 = 1 AND B13 >= 0.5
D5	Lilobot wants to work with the Kindertelefoon to find a solution	Phase 3	B04 >= 0.5 AND B10 = 1 AND B12 = 0

Table 3.3: Desires of the conversational agent Lilobot and the corresponding context beliefs which activates them.

Table 3.3 provides an overview of the desires included in our proposed system and the specific threshold values of the context conditions that activate them. We chose the values 0.3, 0.5 and 0.7 based on the helpline’s conversation corpus on progress in the conversation and repeated testing of the agent with respect to the belief updates. Again, the agent is unaware of the relationship with the specific phases of the Five Phase Model. We included this to help explain the BDI model to users of the system.

Additionally, each desire is associated with a list of actions which we thought of as a plan that the agent needs to execute when adopted as an intention. A desire is marked as inactive if its context condition no longer holds.

D1. Lilobot wants to talk to about his problem

This desire is based on phase 2 of the Five Phase Model and is the initial intention of our prototype. The initial beliefs of Lilobot are that he trusts the Kindertelefoon (B04) and he thinks the counsellor will be able to help him solve the bullying problem (B08), hence why he contacts the helpline in the first place. The agent’s plan of action here is to inform the counsellor about the problem as well as talk about how it affects him. This is shown as actions in Figure 3.1. Once Lilobot adopts this desire as an intention, he will talk about the bullying problem and answer any questions the counsellor might ask to get to know the problem. Here, the counsellor is expected to show interest by asking detailed questions about the situation and empathize with Lilobot as recommended by the Five Phase Model. Such actions increase Lilobot’s beliefs of good rapport with the Kindertelefoon, indicated by beliefs B03-B06.

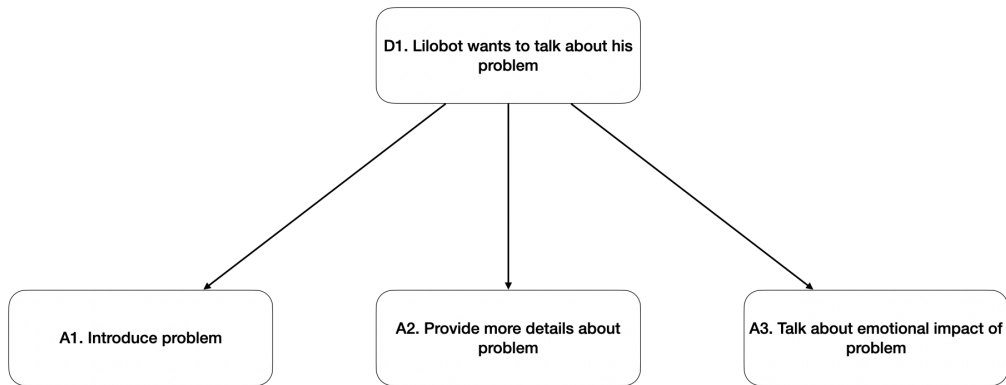


Figure 3.1: Actions associated with desire D1.

D2. Lilobot wants to end the conversation

This desire is activated once Lilobot perceives the counsellor or the conversation as threatening (B01, B13) or Lilobot believes the Kindertelefoon cannot help him solve the problem (B07). Lilobot ends the conversation if this is the case. Alternatively, this desire is activated if Lilobot accomplishes his desire to get the Kindertelefoon to solve the problem for him (B17) or thinks he can solve the problem on his own (B02) which is the optimal outcome of phase 4 if the conversation model is applied correctly. The corresponding action of this desire is that Lilobot simply says goodbye to the counsellor and ends the conversation.

D3. Lilobot wants the Kindertelefoon to get the bullies out of school

This desire is part of phase 3 of the Five Phase Model where the counsellor and the child must set a goal for the session. In our prototype, Lilobot feels incompetent to solve the problem on his own and thus, his goal in this phase is to get the counsellor to solve the problem on his behalf. Once Lilobot believes he has talked to the counsellor about the problem (B09), he expresses his wishes for the bullying to stop and in order to do this, he wants the counsellor to get the bullies out of his school (shown in Figure 3.2). Since we employ a branched narrative in our design, the counsellor can agree to this or better still, try to find another solution by empowering Lilobot to solve the problem on his own as the Five Phase Model recommends. By doing this, the agent provides implicit feedback that shows how the counsellor's actions affects its decisions as stated in requirement R04.

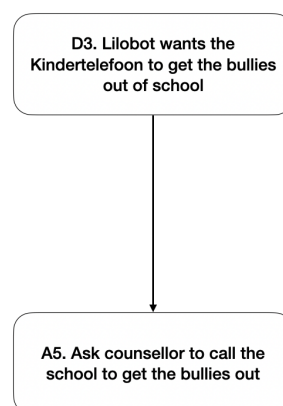


Figure 3.2: Actions associated with desire D3.

D4. Lilobot wants to talk to his teacher about the problem

Sindhal [79] highlights that a counsellor cannot solve a problem until a good rapport has been established with the child mainly in phases 1 and 2, or before the goal of the session is defined in phase 3. These are implemented as the context beliefs of this desire. Thus, unless the beliefs for good rapport and defining a goal have been met (B03-B06), this desire will not be activated. Good rapport with the counsellor means Lilobot believes the counsellor was interested in his story (asked detailed questions about the problem) and understands him (showed empathy). Once this desire is adopted as an intention, Lilobot will engage with the counsellor on how to talk to his teacher as a means to solving the bullying problem. It is important that the counsellor compliments Lilobot's ideas as a means of empowerment. Lilobot's actions (shown in Figure 3.3) are to ask how the teacher can be of help, express his concerns about taking said action and clarifying what to tell the teacher.

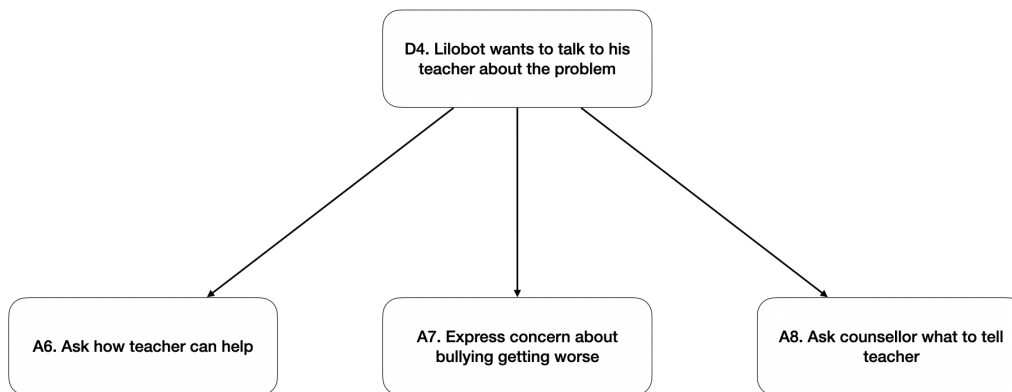


Figure 3.3: Actions associated with desire D4.

D5. Lilobot wants to work with the Kindertelefoon to find a solution

This desire serves as a bridge between desires D3 and D4. In the prototype, Lilobot's initial wish is for the Kindertelefoon to get the bullies out of school because he does not feel competent to solve the problem. This desire is activated if the counsellor has built rapport through the previous phases of the conversation and is able to prompt Lilobot's problem solving skills according to phase 4 of the Five Phase Model to come up with another solution such as talking to the teacher about the problem.

3.3.3. Intentions

In our prototype, an intention is a desire which is currently active and has been selected to be pursued. The agent can have multiple desires active at the same time but it selects the first desire as ordered by the IDs as an intention. As mentioned earlier, a desire has an associated set of actions which the agent must accomplish. For example, if Lilobot's intention is to get the Kindertelefoon to get the bullies out of school, his plan of action is to ask the counsellor to call the school and to confirm that the counsellor can indeed carry this out. Such a solution is plausible in an interaction with Lilobot but does not however align with the Five Phase Model because Lilobot is not empowered to solve the problem on his own.

3.3.4. Actions

For each desire, we defined an action or a sequence of actions that the agent must execute. The sequence is identified by the IDs of the respective actions. For desire D4 for instance, action A6 must be completed before actions A7 and A8. Each action has a boolean value indicating whether or not the agent has completed it. Once an action is completed, it cannot be executed by the agent again.

The two types of actions are *inform* and *request*. An *inform* action requires the agent to simply say a statement to the user while a *request* action requires agent to ask the counsellor a question. The counsellor's answer determines if the action is completed or not. For example, in desire D3, Lilobot's action is to ask the counsellor to call the school to kick the bully out. If the counsellor agrees to this,

the action is marked as completed. Otherwise, Lilobot will continue to execute this action as long as desire D3 is its intention. Table B.3 shows all the actions in our proposed solution.

ID	Action Name	Type	Desire
A1	Lilobot introduces problem	inform	D1
A2	Lilobot provides more details about problem	inform	D1
A3	Lilobot talks about emotional impact of problem	inform	D1
A4	Lilobot says goodbye	inform	D2
A5	Lilobot asks counsellor to call the school to get the bullies out	request	D3
A6	Lilobot asks how teacher can help	request	D4
A7	Lilobot expresses concern about bullying getting worse	request	D4
A8	Lilobot asks counsellor what to tell teacher	request	D4

Table 3.4: Actions of conversational agent Lilobot.

3.4. Overall Design

Having highlighted the core elements of the reasoning BDI model, we describe in this section how these elements integrate with other components of the system. We first describe how a user utterance is classified into an intent through natural language processing. Next, we highlight the ontology that makes up the knowledge base of the agent that allows this reasoning. This intent is then received by the BDI reasoning model as a perception to be reasoned about. How the reasoner updates the the BDI values and by how much is then explained next in this section. We conclude with a detailed explanation of an example interaction with the agent.

3.4.1. Natural Language Processing

User input is first classified into an intent using a pre-trained natural language processing model. Once an intent is recognized, the natural language processing model sends it to the BDI model as a perception to reason about and awaits a response to return to the user. An intent name in our prototype is made up of the *type* of intent, *subject* of the intent and the *attribute* pertaining to the subject, structured as `type_subject_attribute`.

We included five intent types: *request* (question requiring knowledge retrieval), *ack* (statement expressing compassion, encouragement, discouragement or threat), *inform* (statement expressing information support), *confirm* (question requiring BDI decision making directly) and *trigger* (indicating that the user is inactive). This structure is inspired by the work of Zwaan et al. [87] where they categorized different types of social support frequent in counselling conversations [34]. It also helped us compactly represent information in a single intent name. For example, asking "How often are you bullied?" translates to `request_bullying_frequency`.

Intent type	Description	Example
request	Question requiring knowledge retrieval.	How often are you bullied?
ack	Statement expressing compassion, encouragement, discouragement or threat.	How annoying that you are being bullied at school!
inform	Statement expressing information support.	Your teacher can support you by perhaps talking to the bullies.
confirm	Question requiring BDI decision making directly.	What do you think of talking to your teacher about your situation?
trigger	User is inactive.	

Table 3.5: Types of intents implemented in the conversational agent prototype.

Each intent type influences one or more beliefs in our prototype and is handled differently by the system. An *ack* either increases or decreases Lilobot's perceived relatedness with the counsellor. An *inform* updates Lilobot's beliefs about help-seeking. A *confirm* requires Lilobot to check specific beliefs and return a positive or negative response based on their values. It is good to note that a *request* is regarded as a query to Lilobot's knowledge base using the subject and attribute. We will elaborate on this in the next section.

Moreover, we define five user intent subjects: *chitchat*, *bullying*, *goal*, *confidant* and *unknown*. These describe the context of the conversation at a given moment. *Chitchat* is used for small talk and greetings which are mostly in phases 1 and 5 of the conversation model. *Unknown* is used when the user does not specify a subject in the utterance. For instance, "How often does this happen?" does not mention that the user is referring to how often the bullying takes place. In such cases, the BDI reasoning model must infer the subject from past dialogue and respond appropriately or ask the user to clarify the question. The remaining subjects are related to the ontology of the agent which we will cover next. Table 3.6 shows the mapping between the different intents and the beliefs in our proposed system.

3.4.2. Ontology

In order to allow the agent to engage in a meaningful conversation content-wise, we needed to capture the relevant concepts involved in the target domain to be simulated. What do children talk about when talking about a bullying situation? What questions is a counsellor expected to ask the child? For this, we conducted an analysis of the helpline's past conversations on bullying to identify themes that are common to the subject. This forms the basis of the agent's knowledge base from which it retrieves responses. Figure 3.4 shows the bullying knowledge of Lilobot. The ontological elements are the attributes. Each ontological element has four different responses which can be selected from at random to create more varied responses. The agent retrieves an appropriate response from its knowledge base given an intent as perception and the intention of the agent. All intents of *request* type require the agent to retrieve a response from the knowledge base using the *subject* and *attribute*.

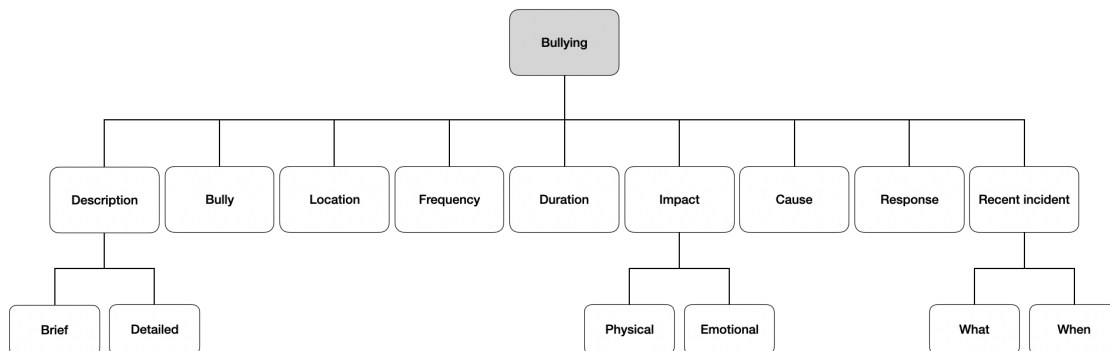


Figure 3.4: Bullying knowledge ontology of the conversational agent prototype.

With this ontology, Lilobot can answer questions about the bullying problem such as how long it has been going on, how often he is bullied, where the bullying takes place, etc. These questions are expected in phase 2 of the conversation model where the counsellor must ask detailed questions about the child's situation in order to gain an understanding of the problem. The more relevant questions are asked, the more Lilobot believes that the counsellor is interested in his story and the more cooperative he is during the conversation.

Similar to bullying, we defined an ontology for the goal subject. In phase 3 of the conversation model, the counsellor and the child need to set a goal for the session. Typically, the counsellor also asks questions regarding how the child intends to reach his goal, when to carry out the goal and the effect of said goal on the child's well-being (shown in Figure 3.5). Clarifying these helps the child move on to a solution-oriented mindset instead of dwelling on the problem. In our system, the primary goal of Lilobot is for the bullying to stop. We included different ways to do it - either asking the Kindertelefoon to get the bullies out of the school or talking to the teacher about the problem. Other alternatives such as talking to his parents or confronting the bullies are also available. However, these are not solutions

Lilobot agrees to. Which solution Lilobot chooses in the end depends on his current BDI status.

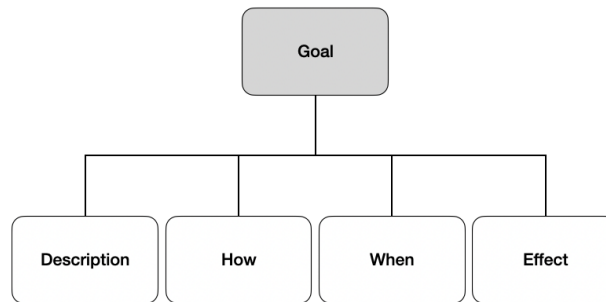


Figure 3.5: Goal knowledge ontology of the conversational agent Lilobot.

Finally, in phase 4 of the Five Phase Model, the counsellor and child work toward the session goal. In our proposed system optimally, this requires the agent to share information about Lilobot's preferred confidant including how to talk to his teacher about the problem and when to talk to his teacher. Figure 3.6 shows the knowledge included in our system on this subject.

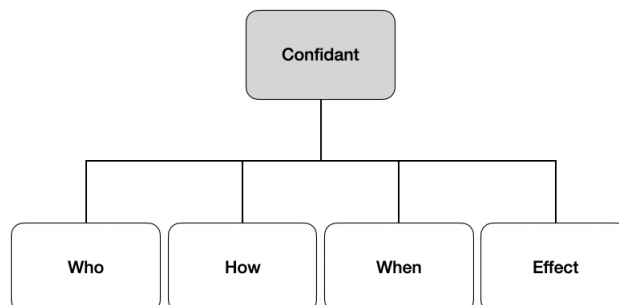


Figure 3.6: Confidant knowledge ontology of the conversational agent Lilobot.

3.4.3. BDI Reasoning

Once the intent of the user is extracted, it is sent to the BDI reasoning model as a perception. How the BDI reasoning model processes this perception is dependent on the intent type. We shortly describe the two main processes in our BDI model.

Reacting to User

If the BDI reasoning model receives an intent name as a perception, it parses the type, subject and attribute. If the subject is *unknown*, the cognitive model retrieves the last saved subject of the conversation and uses this as the current subject of the perception. This way, the user does not need to

constantly mention what they are referring to and the conversation is more natural and human-like. Looking at the previous example, "How often does this happen?" has an unknown subject. Thus, Lilobot interprets the question as "How often does the bullying happen?" if bullying was the last subject of the conversation.

Next is the belief update. Each intent type, subject and attribute or combination thereof corresponds to one or multiple beliefs in our model. Belief updates are done by a value increase of 0.1 or decrease of 0.2. One of the reasons for this was to prevent users from exploiting the agent. For instance, showing empathy increases Lilobot's trust in the Kindertelefoon. In order to prevent that a user from skipping through phases using only empathy, a small increase update value makes it hard to reach the maximum and we additionally define the relatedness belief as an aggregate over sympathy and interest. This way, getting the maximum value for empathy does not affect Lilobot's behavior entirely. On the other hand, negative actions like insults are able to have greater impact on the agent's behavior. Beliefs about the conversation which are updated by a value of 1 exactly, making them booleans in our system. Table 3.6 shows the mapping between the intents and beliefs in our prototype.

From the updated belief values, the model checks whether current beliefs meet the context conditions. The desires whose context beliefs meet the specified thresholds are marked active. Lilobot then chooses one of the active desires to pursue as an intention. Based on this intention, the agent retrieves an answer from its knowledge base using the perception subject and attribute. This is returned to the user as a response. Finally, the agent saves the current state of the BDI, awaiting the next perception.

Type	Subject	Attribute	Belief Update	
ack		empathize	B05 += 0.1	
		compliment	B02 += 0.1	
		neutral	B05 += 0.05	
		guilt	B01 -= 0.2 B03 -= 0.2	
		taunt	B02 -= 0.2 B03 -= 0.2	
request	bullying		B06 += 0.1	
	confidant		B12 += 0.1	
	goal		B01 += 0.1 B11 += 0.1	
	chitchat	greeting		B04 += 0.1
		goodbye		B15 = 1
inform	confidant		B13 = 1	
	goal	positive	B17 = 1	
		negative	B07 -= 0.1 B08 -= 0.2	
confirm	bullying	summary	if (B09 == 1) attribute = 'positive' else attribute = 'negative'	
			if (B10 == 1) attribute = 'positive' else attribute = 'negative'	
	goal	collaborate	if (B04 == 1) B07 += 0.1 B08 -= 0.2 B16 = 1 attribute = 'positive' else attribute = 'negative'	

Table 3.6: Intent-belief mapping implemented in the conversational agent prototype.

Reaching Out to User

We wanted our agent to be able to engage in a two-way communication with the user in order to make the interaction more engaging. This meant implementing a feature in the prototype so that Lilobot could express himself without any prompt from the user. To solve this problem, we made use of a timer. If the user was inactive for at least 10 seconds, the BDI model received a *trigger* as input. Subsequently, the model retrieved the next uncompleted action linked to the agent's current intention. This was returned by the agent as a response.

3.4.4. Feedback

As system feedback, we provide a transcript of the conversation along with a log of Lilobot's beliefs, desires and intentions. As stated in requirement R03, it is important we provide the counsellor insight

into why the agent acts a certain way and how the counsellor's actions influences the behavior of the agent. This is achieved by including the BDI status of Lilobot in the feedback we give to the user at the end of a conversation. By printing the BDI status of Lilobot on every update, we clearly show which beliefs of Lilobot are influenced by the counsellor's words and which desires he adopts as an intention.

Moreover, we provide an overview of Lilobot's beliefs at the start and the end of the conversation and their differences. These values are presented as percentiles for easy understanding. Since all of Lilobot's beliefs are positively formulated, a higher belief value indicates a better state of the agent. Table 3.7 shows a sample table of beliefs a user would receive as feedback at the end of a conversation with Lilobot.

Belief	Phase	Begin (%)	End (%)	Difference (%)
Lilobot feels connected to the Kindertelefoon	all	40	78	38
Lilobot feels safe in the conversation	all	100	100	0
Lilobot thinks he is in control	all	50	80	30
Lilobot thinks KT understands him	all	0	100	100
Lilobot thinks KT is interested in his story	phase 2	0	20	20
Lilobot thinks KT can be trusted	phase 2	60	80	20
Lilobot thinks he has talked about his situation	phase 2	0	70	70
Lilobot thinks KT is asking about a positive wish	phase 3	0	30	30
Lilobot thinks KT can solve his problem	phase 3	100	0	-100
Lilobot thinks KT is asking about his wish	phase 3	0	100	100
Lilobot thinks KT can help him	phase 3	70	90	20
Lilobot thinks he is competent to solve problem	phase 4	30	70	40
Lilobot thinks KT is asking about a confidant	phase 4	30	100	100
Lilobot thinks KT is going to solve the problem for him	phase 4	0	0	0
Lilobot thinks he and KT will be able to reach a solution	phase 4	0	100	100
Lilobot thinks his teacher can help him	phase 4	50	100	50

Table 3.7: Sample feedback showing belief values at the start and end of a conversation and their relation to the Five Phase Model.

3.4.5. Example Interaction

1	KT:	Hello!
2	Lilobot:	Hello, I am Lilobot
3	Lilobot:	I am being bullied at school.
4	KT:	How annoying!
5	Lilobot:	I'm being called names, kicked, and so on. It actually started very innocently but it's getting worse and worse.
6	KT:	What do you hope will change with this conversation?
7	Lilobot:	That they stop.
8	Lilobot:	Could you call my school?
9	KT:	Unfortunately, I cannot do that.
10	Lilobot:	Bummer!
11	KT:	Is there anyone you can talk to about this?
12	Lilobot:	Bye!

Table 3.8: Sample conversation with Lilobot.

Table 3.8 shows a short conversation (translated to English) with the conversational agent, Lilobot. The user utterances are indicated as KT. The initial belief values of the agent are included in Table B.4 of the Appendix. Based on these values, Lilobot's initial intention is desire D1 (Lilobot wants to talk about the problem). When the user says "Hello!", this is recognized by the NLU model as a greeting intent named `request_chitchat_greeting`. This intent is subsequently sent to the BDI model as a perception. The reasoner checks if any part of the perception is unknown and tries to resolve it from

the previously saved subject of the conversation. Since this is not the case here, it simply saves *chitchat* as the current subject of the conversation. Next, it finds the beliefs that correspond to the intent and updates their values. From Table 3.2, we see that this is belief B04 (Lilobot thinks the Kindertelefoon can be trusted). The reasoner increases the initial value of B04 from 0.6 to 0.7. Next, the reasoner checks if any of the desires can be activated. Lilobot retains its initial intention here since no other desire meets the specified thresholds. Given that the perception type is a *request*, the reasoner fetches a response from the agent's knowledge base using *chitchat* and *greeting* as subject and attribute respectively. The reasoner chooses one of the four response options and returns this as an answer to the user. In line 2, Lilobot responds saying "Hello, I am Lilobot".

Line 3 shows the ability of the agent to say something on its own without any prompt from the user. What happens here is that, after at least 10 seconds of user inactivity, the NLU model sends the intent *trigger* as a perception to the BDI model. The BDI model retrieves the last known intention of the agent. In this case, that is desire D1. Subsequently, it retrieves the next uncompleted action associated with D1. This is action A1 (Lilobot introduces the problem) as seen in Table B.3. The reasoner retrieves the value stored for action A1, "I am being bullied at school", from the knowledge base to be returned to the user as a response. It then sets the current subject of the conversation to *bullying* and marks action A1 as completed.

In line 4, the user empathizes with the Lilobot about the bullying situation. "How annoying!" is recognized as intent `ack_unknown_empathize`. Again, the reasoner checks if any part of the perception is unknown and tries to resolve it. Since the subject is unknown, the reasoner replaces this with *bullying*, the last saved subject of the conversation. The value of belief B05 (Lilobot thinks the Kindertelefoon understands him) is increased by 0.1. Again, it retains its initial intention because no other desire is marked active. Line 5 is similar to line 3. The reasoner retrieves the next uncompleted action for desire D1 which is action A2. Therefore, Lilobot provides more details about the bullying problem.

Line 6 gets more interesting because the intent recognized here is `request_goal_what`. The reasoner now changes the subject of the conversation to *goal* and updates belief B10 (Lilobot thinks the Kindertelefoon is asking about his wish) to 1. This triggers desire D3 (Lilobot wants the Kindertelefoon to get the bullies out of school) to be marked as active and is subsequently selected as Lilobot's intention. Since the perception is of type *request*, the reasoner retrieves the response to the question from the knowledge base using *goal* and *what*. Lilobot expresses his wish for the bullying to stop. Line 6 is a result of a trigger intent similar to lines 3 and 5. The reasoner fetches action A5 which is linked to the agent's current intention, desire D3. This action is a request type which means the agent expects an yes or no answer from the user. When the user expresses in line 9 that they cannot call the school, beliefs B07 (Lilobot thinks KT can help him) and B08 (Lilobot thinks KT can solve his problem) are decreased by 0.1 and 0.2 respectively according to Table 3.6.

When the user asks about a possible confidant in line 11, `request_confidant_who` is recognized as intent. The reasoner makes *confidant* the subject of the conversation and increases the belief value of B12 (Lilobot thinks KT is asking about a confidant). On the desire update, desire D2 (Lilobot wants to end the conversation) is marked active. The agent adopts this desire as an intention and proceeds to end the conversation in line 12. The reason for this that the context beliefs of D2 hold, especially belief B03 (Lilobot feels connected to the Kindertelefoon) is too low. Note belief B03 is an average of beliefs B04-B07. Looking at this from the perspective of the Five Phase Model, the counsellor underperformed in phase 2 of the model. The counsellor did not ask questions relevant to the problem (indicated by belief B06) and was not empathetic enough (indicated by belief B05). Instead of diving deeper into the child's situation, the user rushed to phase 4 to ask a possible confidant in attempt to solve the problem.

4

Implementation

This chapter focuses on the technical implementation of our prototype. The first section presents the architecture of our prototype and technologies we used in its development. Afterwards, we explain how we addressed the three functionalities implemented in the system including how Lilobot reacts to user input, how Lilobot reaches out to the user and how feedback is generated at the end of a conversation with the agent.

4.1. Architecture

The conversational agent from a technological perspective can be broken into two main components, the natural language processing aspect and the BDI reasoning model. For the former, we used Rasa¹ and implemented the BDI reasoning model as a web application using the Spring Framework². Figure 4.1 shows an overview of the technologies we used in the development of our prototype. In the following subsections, we describe the individual components of our technology stack.

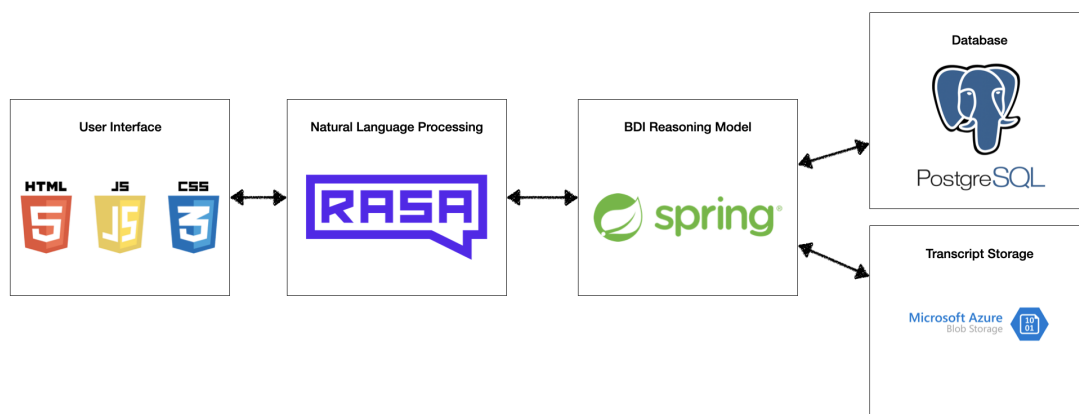


Figure 4.1: Technology architecture of the conversational agent prototype.

¹<https://rasa.com>

²<https://spring.io>

4.1.1. User Interface

The conversational agent was made available through a webpage. A screen capture of the interface is shown in Figure B.6. On the left hand side of the page, we provided instructions in Dutch on the prototype and how to interact with Lilobot. We placed a chat button on the lower right corner of the page which upon clicking, launches a new chat window with Lilobot. We incorporated the chatbot to the webpage using Rasa Chat Widget³, a widget which allows users to incorporate into Rasa chatbots to existing webpages using a websocket channel. The next section expands on Rasa and how we made use of it in our prototype.

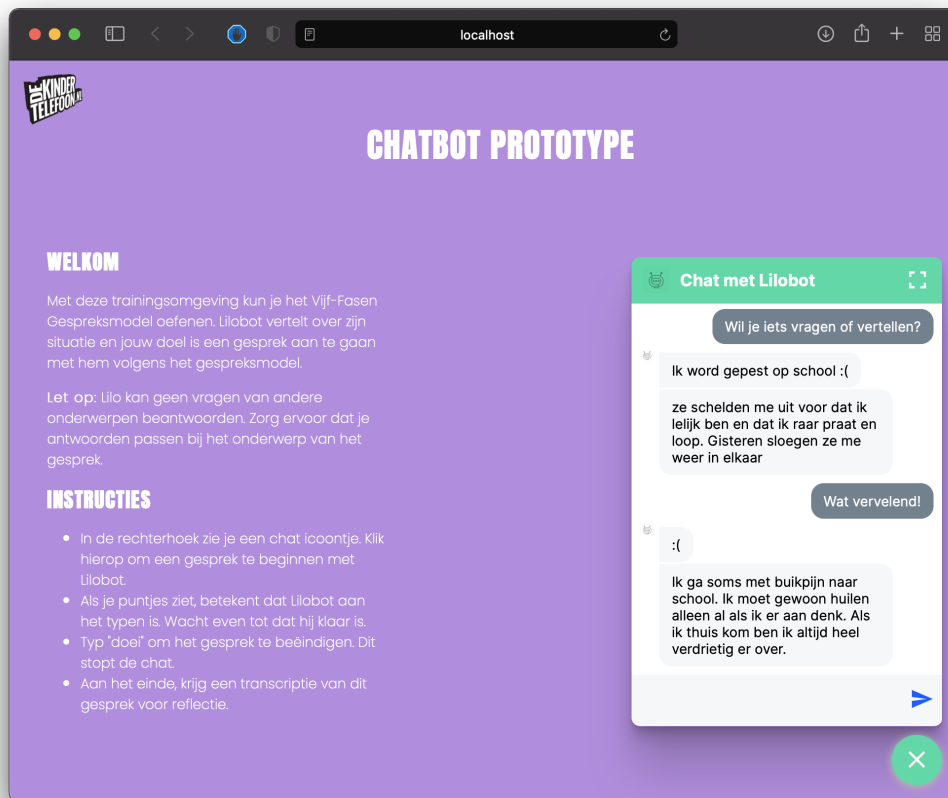


Figure 4.2: Screenshot of the conversational agent prototype Lilobot.

4.1.2. Rasa

Rasa is a open-source tool for building chatbots and voice assistants with out-of-the-box Natural Language Understanding (NLU) and dialogue management. For our prototype, we make use of the NLU component for intent classification of Dutch text. Each conversation in Rasa is identified by a random number called a conversation or sender ID. Once the intent of a user's utterance has been detected, we use the Action Server component of the Rasa SDK to make an API call to our BDI web application to retrieve a response. This response is subsequently relayed to the user. Additionally, the Action Server creates a reminder action so that if the user is inactive for a certain period of time, it prompts the BDI application to say something.

4.1.3. Spring

The Spring Framework (Spring Boot) allows the creation of standalone applications that run on their own, without relying on an external web server by embedding a web server such as Tomcat. We built

³<https://rasa.com/docs/rasa/connectors/your-own-website>

our BDI reasoning model as a micro-service web application written in Java.

The main service in our application that acts as the master is the Agent Service. In order to host concurrent conversations on our prototype, the Agent Service acts as a master, creating and managing agents for each user using the conversation ID Rasa assigns the user. This is based on the Factory Method design pattern. Our Agent Service depends on other sub-services, namely, Belief, Desire, Knowledge and Action services for reasoning. Each service acts independently of each other, querying and updating their respective data models. The Belief Service is in charge of creating beliefs and updating their values by interacting with the database through the Java Persistence API. Similarly, the Desire and Action Services handle the creation and management of desires and actions for a given agent. The Knowledge Service primarily acts as our knowledge model, retrieving responses when needed.

Moreover, instead of hard-coding the initial beliefs and desires of agents, we implemented the functionality to read initialization files (.csv file) from Azure Blob Storage to create beliefs and desires based on these. For future work, it would be possible to easily create different variants of agents by simply uploading different initialization files.

4.1.4. PostgreSQL

The database server for our BDI reasoning engine is PostgreSQL⁴. PostgreSQL is an open-source relational database management system based on SQL. Here, we store the state of each agent after every BDI update as well as the knowledge available to the agents.

4.1.5. Azure Blob Storage

We make use of Azure Blob Storage to store transcripts of conversations with Lilobot. Blob stands for Binary Large Object, which includes unstructured data such as images and multimedia files. The blobs are put in containers which function similar to directories. These are then linked to the storage account. At the end of a conversation with Lilobot, our BDI application creates a Word document (.docx file) containing a transcript of the conversation and the BDI status of Lilobot as feedback for the user. Subsequently, the file is stored in an Azure Blob Storage so that it can be downloaded for viewing.

4.1.6. Miscellaneous

In addition to the above-mentioned technologies, we made use of the following to support development of the prototype.

Git

Our version control system was Git. The code repository is available online⁵.

GitHub

We used GitHub⁶ for managing our source code. We defined CI/CD pipelines using GitHub Actions to automate software updates to our applications. With regards to Rasa, we used it for creating the custom action server and for updating our Spring application running on Azure App Service.

Docker

Our front-end application together with some Rasa components were containerized and deployed using Docker⁷. Docker is a toolkit that enables developers to build, deploy, run, update, and stop containers.

Azure

We used Microsoft Azure as our cloud hosting platform. All our applications were hosted on Azure. Within Azure, the Spring application ran on Azure App Service. The App Service is a fully managed web hosting service for building web applications, services and RESTful APIs on Azure.

⁴<https://www.postgresql.org>

⁵<https://gitlab.ewi.tudelft.nl/in5000/ii/a-bdi-based-virtual-agent-for-training-child-helpline-counsellors>

⁶<https://github.com>

⁷<https://www.docker.com>

4.2. Technical Processes

During the development of our prototype, we address three main technical challenges including how to respond to user input, how to make Lilobot say something on its own to make the conversation more engaging and how to present feedback to the counsellor. This section explains how we implemented these functionalities.

4.2.1. Reacting to User

Each conversation in Rasa is associated with a conversation ID. When a user enters a statement, Rasa classifies the statement into an intent based on the trained NLU model. The intent name is sent together with the conversation ID in a HTTP POST request to the Spring application to retrieve a response. If the conversation ID is not recognized by the Spring application, this means the conversation is new and therefore, the Agent Service within the Spring application creates a new agent for the user using the conversation ID.

Upon receiving the intent name, the Spring application creates a perception with the intent's type, subject and attribute. Each of these properties were mapped to the agent's beliefs. The appropriate agent is identified using the conversation ID and its beliefs are updated. Based on the current belief values of the agent and the thresholds defined, a desire is selected as an intention. Based on the intention, the agent retrieves the response by looking up the perception subject and attribute in the knowledge base. The response is then sent in response to the earlier API call and Rasa relays it to the user.

4.2.2. Reaching out to User

Each time a user enters an input, our Rasa custom actions server schedules a reminder for X seconds from that time moment. We designed X to be a random value between 10 and 20. If the user enters another input before the timer is reached, the custom actions server cancels the scheduled reminder. Otherwise, Rasa infers that the user is in active and triggers the BDI reasoning engine to say something. For this, it sends a HTTP POST request to the Spring application with the intent name as *trigger* and the conversation ID.

In the Spring application, the agent retrieves its current intention from the database and based on this, returns the next uncompleted action associated with the intention as its response. For example, if the agent's intention is wanting the Kindertelefoon to get the bullies out of school, the action could be to ask the user when they can call the school. This statement is returned as a response to the user by Rasa.

4.2.3. Requesting Feedback

At the end of a conversation with Lilobot, a user can request a transcript of the conversation as feedback. This is achieved through an API call from the Rasa Action Server to a Response Service within the Spring application. The Response Service acts independently of the BDI reasoning model as a whole. The service creates a Word document (.docx file) based on the agent's conversation and BDI logs and uploads this file to Azure Blob Storage. Consequently, it generates a Shared Access Signature (SAS) token to allow access to the file. The document URL is then sent to the user by Rasa and upon clicking it, the user can download the feedback file.

4.3. Limitations

A user is allowed to enter one sentence at a time otherwise only the first sentence is recognized and classified correctly. Should a sentence contain multiple intents, the Rasa NLU model returns the intent with the highest confidence score. In future work, it would be better to use multiple intents that have confidence scores above a specified threshold so that users can create multi-intent utterances instead of using only simple sentences. For instance, uttering "Hello stupid!" can be recognized as both a greeting and an insult. Due to time restriction, this functionality was not implemented in our prototype.

An important interpersonal skill that counsellors are taught to apply is paraphrasing. In the Five Phase Model, it is common to end a phase by the counsellor paraphrasing what the child has said to help structure the conversation and clarify the child's intentions. This functionality is partially in the prototype using the *confirm* and *summary* intent. However, due to insufficient training data and difficulty with intent recognition within Rasa, this was somewhat trivially implemented. It also requires the beliefs to be more fine-grained in order to properly confirm what Lilobot has talked about.

For example, if the counsellor asks "If I understand you correctly, you are being bullied by a group of boys at school since the school year began. Is this correct?" Confirming this summary of the problem as what Lilobot has talked about would require more detailed beliefs like "Lilobot thinks he has talked about the bully", "Lilobot thinks he has talked about where the bullying happens" and "Lilobot thinks he has about how long the bullying has been happening", all to be true. However in our prototype, we only have one belief (B09) that indicates whether or not Lilobot has talked about his situation. If this is true, Lilobot confirms the paraphrase without checking the subtleties of the paraphrase.

5

Evaluation

This chapter aims to answer the sub-research question:

How do users perceive the conversational agent?

In the previous chapters, we presented the design of our conversational agent for training counsellors of the Kindertelefoon. To evaluate the developed prototype, we set up an experiment to investigate the effect of our conversational agent on learning. We begin this chapter with a description of the experiment setup followed by an analysis of our results. We conclude with a brief discussion of our experiment findings.

To start with, the above-mentioned research question was refined into detailed hypotheses covering two aspects: the counselling self-efficacy of participants and the perceived usefulness of the conversation agent. Previously in the foundation chapter, we highlighted that increasing a trainee's self-efficacy during training leads to more effective performance over time [54]. Moreover, self-efficacy has proved to be an efficient and reliable method for assessing the impact of communication skills training [1]. The Unified Theory of Acceptance and Use of Technology [88] attributes perceived usefulness, perceived ease of use, among others as factors that affect attitude towards the acceptance and usage of a certain technology. A system with a high perceived usefulness is one in which a user believes that its usage will enhance their job performance. We evaluate the conversational agent with respect to participants' perceived usefulness on learning outcome and usability. Our evaluation of the conversational agent is carried out in comparison to a text-based training which is already used by the Kindertelefoon. We formulated our hypotheses as follows:

H1: *Training with the conversational agent simulating a child increases the counselling self-efficacy of the participants.*

H2: *Training with the conversational agent leads to a higher increase in counselling self-efficacy than the text-based intervention.*

H3: *Participants perceive training with the conversational agent as useful.*

H4: *Experience with the conversational agent improves outcome of conversation.*

5.1. Methods

This section covers the experimental method of our study. We first explain the experimental design, followed by descriptions of the participants, materials, measures, procedure and the statistical analysis of our acquired data. This study was set up as an exploratory study to inform future research in this domain. Our experiment was approved by the TU Delft Human Ethics Research Committee (HREC reference number : 1622). We also registered the design of our study with the Open Science Framework

(OSF) registries ahead of data collection. The experiment registration is publicly available online¹.

5.1.1. Study Design

Our experiment followed a repeated measures (within-subjects) design involving two interventions: a text-based intervention and an interaction-based intervention. The text-based intervention was a simple text explaining the Five Phase Model and the latter was the conversational agent simulating the child help seeker the participants had to counsel according to the Five Phase Model. Interaction with the conversational agent took place over three consecutive sessions of approximately 15 minutes each. In order to minimize order effect, the participants were split into two counterbalanced groups. Each experienced both interventions but in reverse order. The measured variables of the study were the counselling self-efficacy of the participants and perceived usefulness of the conversational agent.

5.1.2. Participants

We recruited participants from the current pool of volunteers at the Kindertelefoon. We sent out a call for participants and those who responded were recruited for the experiment. Our initial target size was 60 participants to obtain a medium effect size of 0.42 at the standard 0.05 alpha error probability and a power of 0.9. This was based on a power analysis conducted in G*Power [30] comparing the difference in means between two dependent means (matched pairs) by means of a t-test.

In total, 39 volunteers of the helpline were invited to participate in the experiment. Recruitment took place over a month prior to the experiment and was halted when we reached the time limit of the study. The call for participants began in July 2021 and we conducted in August and September 2021. The participants received no monetary compensation and were assigned randomly to one of the two counterbalanced groups. As mentioned earlier, each group experienced both interventions but in reverse order to minimize order effect. Of the 39 participants, 11 participants were excluded from the analysis because they did not complete the questionnaires. This left a total of 28 participants with varying years of counselling experience ($M = 3.92$, $SD = 3.89$) at the helpline. Given this sample size and a medium effect size of 0.5, we obtained a power of 0.72 at the standard 0.05 alpha error probability.

5.1.3. Materials

The experiment was made available to participants through an online survey hosted on Qualtrics². Qualtrics was also used to assign participants to one of the two counterbalanced groups. We used two training interventions in the experiment. The text-based intervention was a short text explaining the Five Phase Model taken from one of the helpline's training modules. The prototype, the other intervention, followed the design described in the Design and Implementation chapters.

5.1.4. Measures

We used the following questionnaires as our measuring instruments: Counselling Self-efficacy Five Phase Model Questionnaire, Perceived Influence on Learning Outcome (PILO) and System Usability Scale (SUS).

Counselling Self-efficacy Five Phase Model Questionnaire

To assess the counselling self-efficacy of the participants with regards to the conversation model, participants were asked to complete a questionnaire we developed inspired by existing counselling self-efficacy measures including the Counseling Self-Estimate Inventory (COSE) by Larson et al. [55] and the SE-12 questionnaire by Axboe et al. [2]. The questionnaire was validated by supervisors at the helpline and the appropriate changes were implemented based on their feedback. The questionnaire consists of 8 statements that requires participants to rate the degree to which they agree or disagree about their abilities to perform tasks related to the Five Phase Model. Each of the statements are rated on a 10-point scale with values from -5 'strongly disagree', 0 'neutral' to +5 'strongly agree'. The questionnaire can be found in Appendix C.1. This measure was used for evaluating hypotheses H1 and H2.

¹<https://osf.io/hkxzc>

²<https://www.qualtrics.com>

Perceived Influence on Learning Outcome (PILO)

This questionnaire was adapted from the works of Kang [44], Lie [57] and Sirocki [80]. The questionnaire contains 8 statements in total that measure the conversational agent’s influence on participants’ knowledge and skills related to the Five Phase Model and general attitude towards conversational agents for interpersonal skills training. Here, participants indicate how useful they found the conversational agent on a semantic scale ranging from -5 ‘negative’, 0 ‘neutral’ to +5 ‘positive’ as the influence on learning outcome. The questionnaire is included in Appendix C.2. This questionnaire was used for hypothesis H3 to evaluate the perceived usefulness of the conversational agent simulating the child help seeker.

System Usability Scale (SUS)

The SUS questionnaire contains 10 items serving as a global measure of a system’s usability [13]. Each item was rated on a 5-point scale from 1 ‘strongly disagree’ to 5 ‘strongly agree’. The sum of the items give an overall score ranging from 10 to 100 points. This score correlates with an adjective rating [4]. For instance, a SUS score of 85.5 can be seen as ‘excellent’. A SUS score above a 68 would be considered above average and anything below 68 is below average. To the best of our knowledge, there is no validated Dutch translation of the questionnaire. For our study, we used the Dutch translation of the questionnaire by Wever et al [49, 91].

Additional Measures

BDI Outcome: As part of hypothesis H4, we implemented a feature in the experiment version of the prototype to compute the average of the agent’s belief values as outcome of the conversation. As a result, we consider the belief values of the agent as an outcome measure. The scores were computed at the end of each session. We excluded beliefs about the conversation from this estimate as these beliefs were for tracking the dialogue itself and not about the state of the simulated child help seeker. This measure is computed as a score between 0 and 10, with 10 as the highest.

Qualitative Measures: We were also interested in qualitative responses pertaining to the conversational agent. Therefore, we included 6 open-ended questions about participants’ experience with Lilobot. These questions are listed in Appendix C.4.

5.1.5. Procedure

The experiment took place from late August to early September 2021. Due to scheduling conflicts, the initial time period of one week designated for the experiment was extended to two weeks. Participants were sent an email containing a URL that gave them access to the experiment. Participants were requested to complete the experiment in a sitting and it took about an hour to complete. The experiment was split into three phases which are described below.

- *Pre-training:* Ahead of the experiment, participants were briefed about the nature of the experiment and were requested to sign an informed consent form. Additionally, the participants were asked to fill in a pre-training questionnaire about their counselling experience at the helpline and initial counselling self-efficacy measurements.
- *Training:* The participants were exposed to two training interventions - a text-based training and an interactive training. The text-based training was a simple text explaining the Five Phase Model and the interactive training was the conversational agent simulating the child help seeker (Lilobot) they had to counsel according to the Five Phase Model. Participants in both groups performed the same tasks but in a different order. After each intervention, the participants were asked to fill in questionnaires on their counselling self-efficacy measure. Participants engaged with Lilobot in three consecutive sessions, each lasting approximately 15 minutes. The goal of the first and third sessions were to counsel Lilobot according to the Five Phase Model while the goal of the second session was to experiment with the agent to gain familiarity. These were given as instructions for the respective sessions. At the end of a conversation with Lilobot, the conversational agent provided feedback to the participant based on the BDI status of the simulated child help seeker.
- *Post-training:* After the second intervention, the participants were asked to fill in a post-training questionnaire regarding perceived usefulness of the conversational agent. We asked additional

closed and open-ended questions regarding the use of the conversational agent for an exploratory analysis.

Figure 5.1 shows the flow of the experiment and the respective time points of the measurements. First is the pre-training measurement as shown by measurement 1. Then, the participants were split into two counterbalanced groups A and B. Each group completed both interventions. Measurement taken directly before exposure to an intervention is considered pre-training measurement specific to that intervention and similarly, measurement after an intervention is the post-training measure of that condition. Thus, within the group A for example, measurement 2A is the post-training measurement for the chatbot condition and the pre-training measurement for the text condition. Additional post-training questionnaires on the measured variables are indicated by measurement 3.

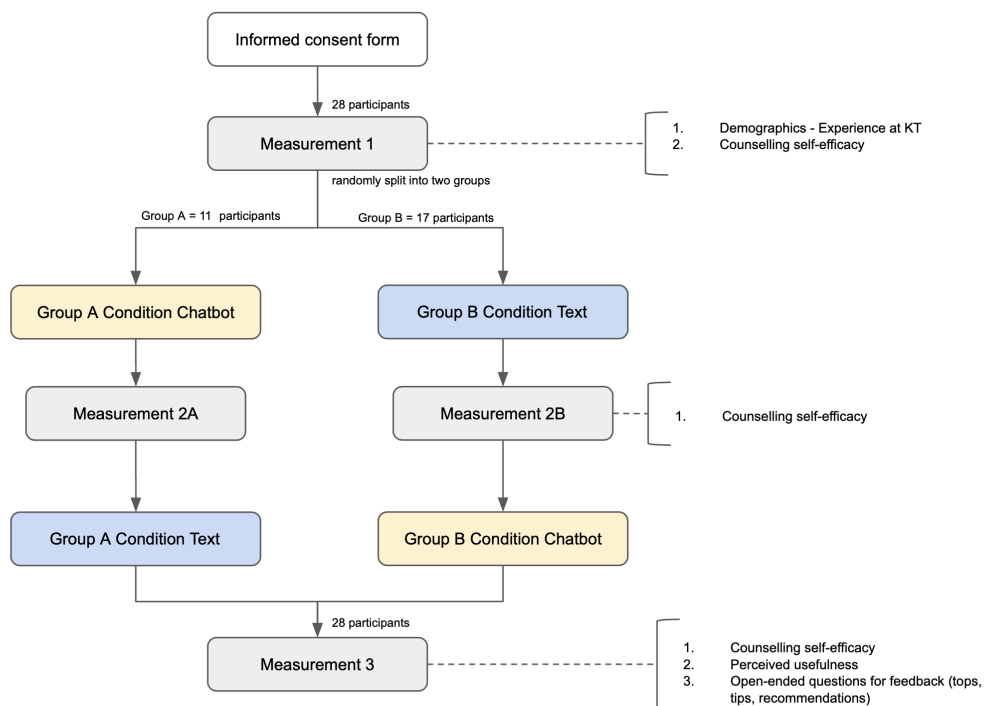


Figure 5.1: Design of experiment including the different time of measurements.

5.1.6. Data Preparation & Statistical Analysis

The data was cleaned by removing responses of participants who did not complete the experiment. This left a total of 28 responses. For the Counselling Self-Efficacy Five Phase Model Questionnaire, we report a Cronbach's alpha value of 0.83 as the questionnaire's internal consistency score. Because this value falls within the acceptable levels of reliability [68], the mean value of the included items within the questionnaire was taken as a single measure. We excluded missing values in the estimation of participants' counselling self-efficacy scores. The starting point of the PILO scale in Qualtrics was set at 0 but due to an implementation drawback, the system recorded a null value if a participant did not move the slider on the scale. Based on this assumption, we replaced null values from responses to this questionnaire with zeros. We conducted a normality test on the data and used non-parametric tests where we could not assume normality.

For the first hypothesis (H1), measurement 1 was compared to the measurement 3. As increased scores in counselling self-efficacy were expected after the training interventions, the data was analyzed using a paired Wilcoxon signed-rank test of the means at measurement 1 and measurement 3 as the normality test showed a significant deviation from a normal distribution. For the second hypothesis, we conducted a repeated measures ANOVA test on the main effect and an interaction effect of the two

independent variables - training intervention and time of measurement. This was followed by post-hoc analysis of the variables. With respect to the third hypothesis (H3), the scores of the PILO questionnaire were analysed using a Wilcoxon signed-rank test of the means of participants' scores and μ as 0. For the final hypothesis (H4), we only considered the BDI outcomes of the first and third sessions in the analysis since the goals of these sessions were for participants to interact with Lilobot according to the Five Phase Model. A p value of .05 was chosen as the significance level for all tests. Responses to the qualitative measures were analysed by means of thematic analysis [12] to identify common themes shared by the participants.

Prior to the experiment, we conducted a pilot run with 5 students to check if everything worked properly. This data is however excluded from our analysis. All statistical analyses were done using R software (version 4.0.5). The markdown script and datasets are available online for download through the 4TU.ResearchData repository³.

5.2. Results

5.2.1. Counselling Self-efficacy

A Wilcoxon signed-rank test was conducted to determine if there was a difference in the counselling self-efficacy scores of the participants before and after training. The test suggested that the median sample post counselling self-efficacy score ranks (3.63) were statistically significant higher than the median pre counselling self-efficacy score ranks (0.81), $Z = -4.19$, $p < .001$. Table 5.1 shows an overview of the medians and interquartile results of the analysis.

Time	Median	IQR
Pre	3.63	1.25
Post	0.81	2.78

Table 5.1: Median and interquartile range of counselling self-efficacy among participants pre- and post-training.

Furthermore, a two-way repeated measures ANOVA was performed to evaluate the effect of different training interventions over time on counselling self-efficacy. Table 5.2 shows the means and standard deviations of the individual groups.

Training	Pre		Post	
	Mean	SD	Mean	SD
Chatbot	3.73	0.92	1.65	2.58
Text	2.98	1.84	2.52	2.10

Table 5.2: Mean and standard deviation results of the two training interventions with respect to the counselling self-efficacy measure.

A linear model was fit on the counselling self-efficacy of participants, taking the training intervention and time of measurement as independent variables, and including a two-way interaction between these variables. The analysis found a significant main effect ($F(1, 27) = 11.77$, $p < .001$) for time of measurement but found no significant main effect ($F(1, 27)$, $p = 0.86$) for training. The analysis also found a significant two-way interaction effect ($F(1, 27)$, $p = 0.03$) between these two variables. A simple effect analysis was used to further examine the two-way interaction. It revealed a significant ($t = 3.98$, $p < .001$) difference in counselling self-efficacy before and after training for the chatbot intervention, but no significant effect ($t = 0.87$, $p = 0.39$) was found in the text intervention across the two time points of measurement. Figure 5.2 shows the counselling self-efficacy of participants before and after both training interventions.

³<https://doi.org/10.4121/17371919>

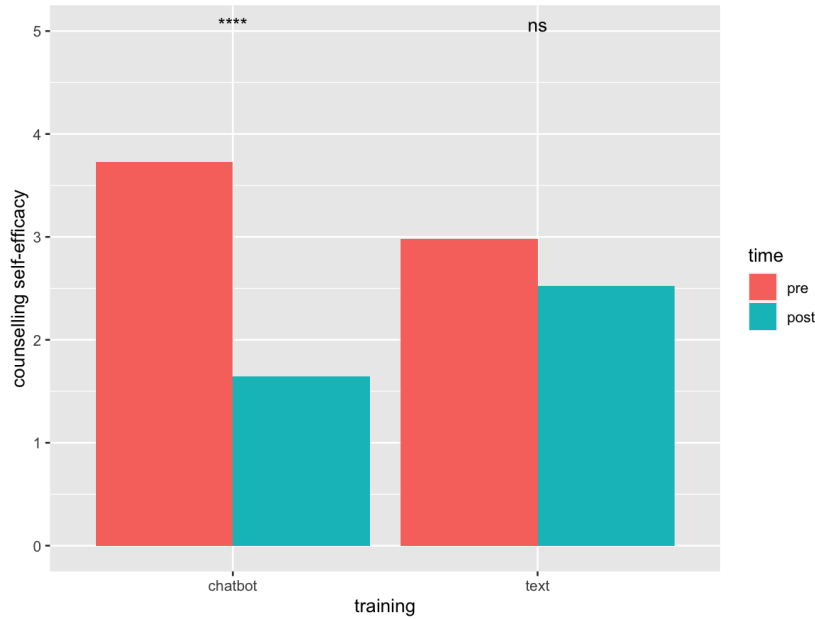


Figure 5.2: Bar plot comparing participants' counselling self-efficacy across the text and chatbot training interventions at various time points.

5.2.2. Perceived Usefulness

With regards to perceived usefulness, we conducted a one-sample Wilcoxon signed-rank test on the individual scores of the PILO questions to examine whether the rating deviated from the neutral zero score on the scale. We observed a significant deviation from the neutral zero point for questions 5 and 8 of the questionnaire. The participants thought their self-efficacy with respect to the Five Phase Model decreased upon training with the conversational agent, $Z = -2.26$, $p = 0.02$. Moreover, the participants' view regarding the usefulness of conversational agents as a learning tool diminished, $Z = -2.54$, $p = 0.01$. The responses to the other questions did not significantly differ from the zero neutral point. Table 5.3 shows an overview of the analysis results for the individual questions.

	Mean	SD	Z	p
Q1. FPM knowledge	-0.03	0.73	0	> .05
Q2. FPM skills	0.03	0.73	-0.54	> .05
Q3. Thoughts of FPM usefulness	0.07	1.39	-0.41	> .05
Q4. Opinion on FPM usefulness	0.10	0.72	-0.64	> .05
Q5. Self-efficacy FPM	-0.62	1.40	-2.26	0.02
Q6. Educational qualities Lilobot	-0.48	1.66	-1.51	> .05
Q7. Educational qualities feedback	-0.03	2.06	-0.04	> .05
Q8. Opinion on chatbot for training	-1.17	2.28	-2.54	0.01

Table 5.3: Mean and standard deviation of PILO responses among participants.

5.2.3. System Usability

Usability of the conversational agent was estimated using the SUS questionnaire. For this, we took the overall mean of all participants' responses. We report an average SUS score of 67.41 ($SD = 6.44$) for the conversational agent prototype. According to Bangor et al. [4], the system can be interpreted as 'ok' based on an adjective rating scale. This is on the lower margin of being acceptable .

5.2.4. BDI Outcome

We conducted a paired sample t-test on the BDI outcome of the first and third sessions of the participants' training with the conversational agent. There was no significant difference in the BDI outcome of the first sessions ($M = 6.36$, $SD = 1.36$) compared to the third session ($M = 6.68$, $SD = 1.24$), $t(25) = -1.72$, $p = 0.10$. Table 5.4 shows an overview of these results.

	Mean	SD
Session 1	6.36	1.36
Session 3	6.68	1.24

Table 5.4: BDI outcome among participants across first and third sessions.

5.2.5. Qualitative Results

In addition to the analysis of the quantitative measures reported above, we asked the participants six open-ended questions regarding the conversational agent prototype for an exploratory analysis.

Figure 5.3 shows a thematic map of participants' responses to the question "What was the best thing about your experience using Lilobot?". Two main themes were identified in our analysis: about the conversation with Lilobot and general takeaways from the experience. Participants liked that the conversation simulated was realistic as a child in terms of language style and behavior (noted by 7 participants). They also appreciated the fast response time of the agent (n=6). With regards to learning, participants thought that through their experience with Lilobot they were able to reflect on what they said and the Five Phase Model (n=4) and were able to see how their actions affected the agent's behavior (n=5). Participants also noted the opportunity for self-directed learning with Lilobot as they did not have to depend on the involvement of other participants to role-play (n=2).

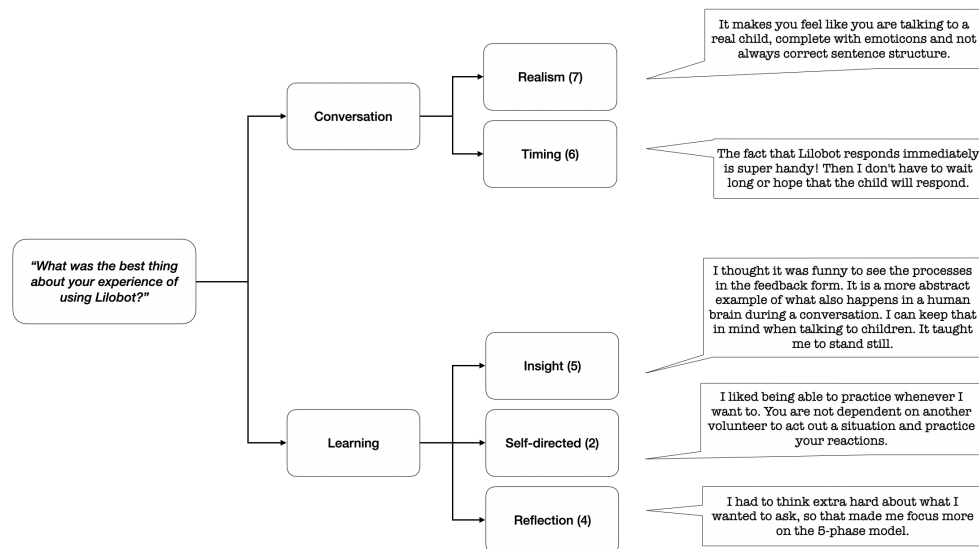


Figure 5.3: Thematic map of participants' most liked features about their experience of using Lilobot.

Figure 5.4 shows a thematic map of participants' responses to the question "What was the worst thing about your experience using Lilobot?". The most common theme identified was issues related to Lilobot's understanding which made it difficult to hold a natural conversation with Lilobot (noted by 16 participants). Participants indicated that Lilobot did not understand their utterances or gave no response to questions they posed to the agent. Others also mentioned they received repetitive answers from Lilobot (n=4), had problems understanding Lilobot's use of emoticons (n=2) and found the segmentation of utterances demotivating (n=1).

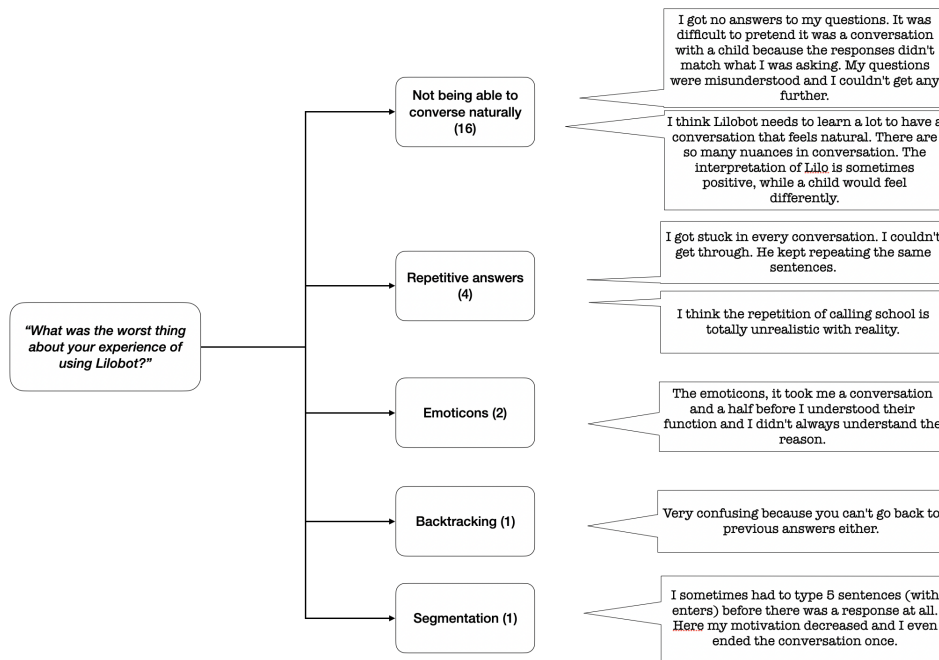


Figure 5.4: Thematic map of participants' least liked features about their experience of using Lilobot.

Figure 5.5 shows a thematic map of participants' responses to the question "What do you think of the feedback you received at the end of your conversation with Lilobot?". 8 out of the 28 stated they did not receive any feedback. Some participants found it insightful to see Lilobot's reasoning process and how their actions influenced the agent's responses (noted by 9 participants). On the other hand, some participants noted the feedback were of little value to them (n=4).

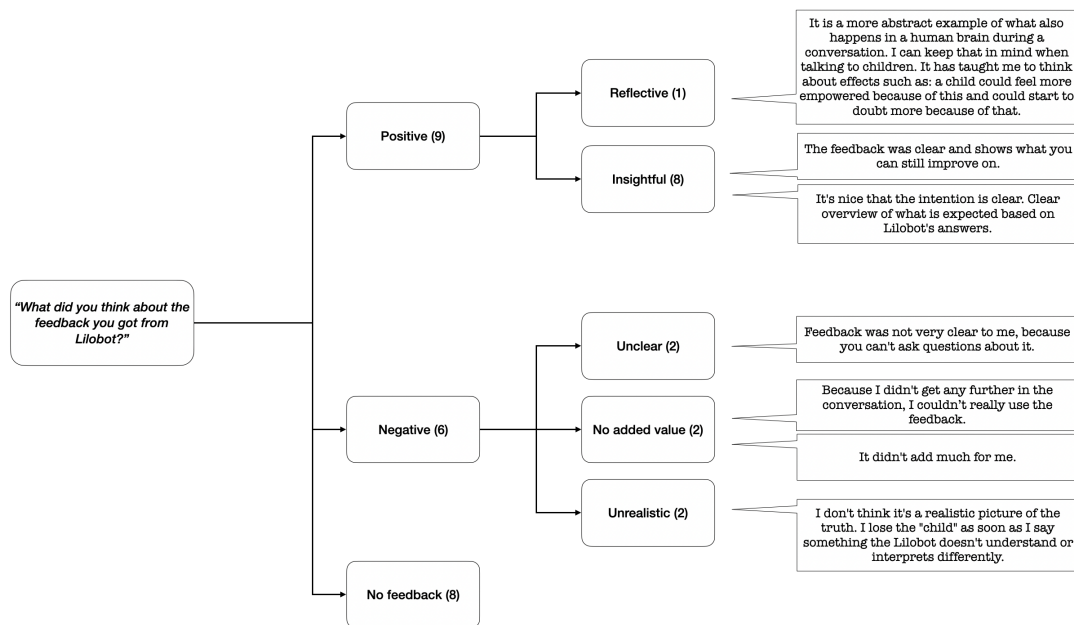


Figure 5.5: Thematic map of participants' positive and negative remarks on feedback from Lilobot.

Participants were also asked which group of users they were likely to recommend Lilobot to. The options included counsellors-in-training, novice counsellors, experienced counsellors and supervisors of the Kindertelefoon. 17 participants said they would recommend the conversational agent for counsellors-in-training at the helpline. Some of the reasons given were that it would allow them to experiment and

gain familiarity with the conversation model without real-life consequences if they did something wrong. Other participants mentioned the conversational agent would be more appropriate for experienced counsellors as they already know how children behave and this could be a good opportunity to revise question-answering techniques and how they relate to specific phases of the conversation model. Table 5.5 shows the participants' responses to this question.

Group	Count
Counsellors-in-training	17
Novice counsellors	3
Experienced counsellors	3
Supervisors	0

Table 5.5: Distribution of participants' responses to recommended target users of Lilobot.

5.3. Discussion

We can draw a number of conclusions from the above results. Participants' counselling self-efficacy changed more upon experiencing the chatbot training. We observe a significant decrease in counselling self-efficacy with the chatbot training which we do not observe in the text-based training. This leads us to reject hypotheses H1 and H2. Although the training succeeded in influencing participants' counselling self-efficacy belief, self-efficacy decreased after training with the agent instead of increasing which would be desirable when building a trainee's confidence in performing a certain task to achieve more effective performance over time.

There are a number of explanations that can be posited for this result. By design, we used a self-report measure which required the participants to give their perceptions of their abilities as a counsellor. According to Kruger & Dunning [51], there is a natural tendency for individuals to overrate their abilities in the absence of any specific skill demonstration on which to accurately judge their own level of competence. This cognitive bias may have led participants to overestimate their counselling self-efficacy beliefs with respect to the Five Phase Model. These results are similar to the findings of Kang et al. [45] where participants in an experiment overestimated their ability to perform a task in a virtual reality environment, leading to lower levels of self-efficacy beliefs post intervention. Another explanation for the decline in counselling self-efficacy could be attributed to a possible misinterpretation of the Counselling Self-efficacy Five Phase Model Questionnaire. As noted by De Vaus [27], a question may be unreliable due to bad wording causing a person to understand the question differently on different occasions. In the questionnaire, we referenced a generic child as 'the child' with which participants had to rate their abilities to perform tasks related to the Five Phase Model. It is plausible that participants interpreted this as the simulated child Lilobot later on during the post-training questionnaires and therefore rated their ability to successfully counsel Lilobot instead.

Results of the PILO questionnaire disproves hypothesis H3. Participants indicated that their counselling self-efficacy decreased upon training with the conversational agent and so did their view on the use of a conversational agent as a learning tool. From the open feedback we received from participants however, 14 of them were still positive about the use of the conversational agent as an extra learning opportunity for counsellors, just not in its current form. They noted the need for improvement with regards to Lilobot's understanding of the questions they asked and more varied responses from the agent. 5 participants on the other hand, thought the conversational agent would be detrimental to the development of a counsellor's personal counselling style given the agent's simplistic question and answering style. They countered conversations with real children remain the most effective way to learn counselling theories.

With regards to feedback the agent provided at the end of a conversation, 9 participants found it insightful while 4 noted they were of little use to them. Including instructions for counsellors on how to improve on their performance based on the BDI status of the agent might prove more useful than only reporting the BDI updates. Carroll [15] highlights the importance of recovery information to help users identify and recover from likely errors. Particularly, users need to recognize and locate their error, understand and analyze it to an appropriate extent and then take some corrective action [86]. This is supported by Salmi et al. [76] in their work on support systems for counselors where they suggest counselors value short actionable information that is highly accurate to the situation and preferably given by someone with expertise. A similar intervention is used by Foster et al. [32] where they

employ human assessors to rate trainees' empathetic levels in conversations and identify opportunities for empathy in order to improve trainees' empathic communication skills.

In addition to the questionnaires, we also analysed the conversation transcripts of the participants' training sessions with Lilobot. We only took transcripts of sessions 1 and 3 into account here. A common issue with Lilobot had to do with understanding counsellors' utterances. While some intents were misclassified, others were not recognized at all. The latter had to do with lack of knowledge about the intent (related to ontology) and difficulties with complex sentence structures such as paraphrases and utterances with multiple intents. Participants also delved deeper into the Lilobot's story but Lilobot could not respond to these questions because they were not included in the agent's ontology. Participants also had difficulties with Lilobot due to lack of segmentation of intents despite being instructed to do so. As we mentioned previously, the Rasa NLU component returned one intent name per user utterance. We therefore advised participants to enter a sentence with a single intent at a time. Otherwise, only the first intent with the highest confident score was returned. We found that a number of participants used complex sentences which Lilobot was not able to reason about correctly.

Aside the issues mentioned above, we noticed a common pattern in the transcripts of the participants' conversations with Lilobot. Some participants ended the conversation at phase 3 when they realized they could not make contact with Lilobot. If a counsellor's questions were not understood by Lilobot or Lilobot persisted on his desire to have the counsellor call the school to solve the problem, the counsellor ended the conversation. This approach comes from one of the helpline's module about setting limits in a conversation ('verbinden in begrenzen' in Dutch). This module teaches the counsellor to call out undesired behaviors a child may exhibit and set bounds within which the conversation can continue. If the child does not adhere to this request, the counsellor is encouraged to end the conversation and invite the child to reach out again once he or she can meet what is asked of them or ready to talk. Table 5.6 shows an example of such a scenario taken from one of the conversation transcripts. The lines in italics represent the contemplations of Lilobot with respect to the BDI model.

1	KT:	waar hoop je op?
2		<i>Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken</i>
3	Lilobot:	Dat ze stoppen
4	KT:	dat is mooi en wat zou het jou opleveren als ze stoppen?
5		<i>Overtuiging: ↑ Ik voel me in controle in het gesprek</i>
6		<i>Overtuiging: ↑ Ik denk dat KT vraagt naar een positieve wens</i>
7		<i>Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken</i>
8	Lilobot:	gewoon vrolijk zijn...Ik wil me gewoon veilig op school zijn en samen met vriendin lunchen enzo
9	KT:	ah wat fijn, dat je je veilig kan voelen en lunchen met je vriendin, super!
10		<i>Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken</i>
11	Lilobot:	:)
12	KT:	zit ze ook bij jou in de klas?
13	KT:	Ik denk dat je er niet meer bent, ik ga de chat voor nu sluiten, mocht je nog terugwillen komen, dan kan dat altijd.
14		<i>Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken</i>
15	Lilobot:	okeeee dat is goed!
16	KT:	doei
17		<i>Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken</i>
18	Lilobot:	Fijne dag verder!

Table 5.6: Sample transcript where a counsellor applies the module 'verbinden in begrenzen'.

From the transcript above, we see that the counsellor uses a closed question to ask Lilobot if his friends are in the same class as him in line 12. Since this is not included in the knowledge base, Lilobot is unable to give an answer to the question. The counsellor infers from this that Lilobot is unresponsive and asks in line 13 to end the chat and reach out another time when he wants to talk more about the problem. This is however not Lilobot's intention as seen in line 14. There is possibly a transference from the counsellor's experience with real-life conversations with children where a child may be unresponsive during the chat. Here, the counsellor employs the module 'verbinden in begrenzen' to deal with undesired behavior from the simulated child and proceeds to end the chat. The Five Phase Model rather suggests the counsellor can help the dialogue forward by reformulating the question or switching subject matter.

6

Discussion & Conclusion

In this final chapter, we reflect on our research as a whole. The first section summarizes how we answered the research question we defined at the start through this report. After, we discuss the limitations of our work and our contribution on a scientific and practical level. We conclude with our suggestions for future research in this domain.

6.1. Conclusion

This project was aimed at answering the research question:

How can a conversational agent be used in a role-play setting to support the training of child helpline counsellors?

We broke down the above question into three questions which we addressed in the previous chapters of this report. We summarize the answers to the sub-research questions as follows.

What are the requirements of the conversational agent?

The starting point of this inquiry was to understand the work of the Kindertelefoon to be able to identify opportunities for a conversational agent within the helpline's training program. We conducted a literature study, participated in the helpline's counsellor training as well as an observation study of counsellors at the helpline. Through this, we gained an understanding of the training program, the individual modules and how role-play techniques were used to teach core counselling skills such as the Five Phase Model, active listening, reflecting, among others. Next to this, we investigated state-of-the-art research on conversational agents, systems for training social skills and how to model cognition in such agents.

To meet the requirements, the counsellors' values, human-factor knowledge, and related technology were examined. We created scenarios based on the helpline's training modules that could be challenging for counsellors and how we could address them through our research. We discussed these scenarios with counsellors and supervisors of the helpline through focus group discussions. Among these scenarios, we identified that a conversational agent could prove useful in helping counsellors learn adherence to the Five Phase Model, especially because it is uncommon that counsellors-in-training practise all phases in a single role-play session due to lack of time and resources. The conversation model is an essential element of the helpline's training as it enables a counsellor to structure conversations and keep them centered on the child rather than using an intuitive approach. Based on the input gathered during the focus group discussions, we arrived at a set of requirements the conversational agent should meet in order to be considered effective for training counsellors. These requirements include: providing a structured way of learning adherence to the Five Phase Model, simulating a level of realism that reflects the behavior of a child help seeker, providing insight into the behavior of the agent as well as implicit and explicit feedback based on the user's actions to promote reflective learning.

What design would meet the requirements?

Based on these requirements, we designed a conversational agent that simulates a child help seeker based on the Belief-Desire-Intention (BDI) framework. A belief denotes a proposition the agent assumes to hold in the real world, a desire is a goal that the agent attempts to satisfy, and an intention is a sequence of actions the agent is committed to follow [62]. To meet requirement R01, we designed the beliefs and desires such that each correlates to one or more phases of the Five Phase Model. This way, the agent is able to engage in a conversation that covers all phases of the conversation model. Positive belief updates translate to progress in the conversation according to the conversation model. In our design, Lilobot simulates a victim of school bullying who contacts the helpline for help. Inspired by literature on bullying behaviors among children and the self-determination theory, we designed 17 beliefs that reflect Lilobot's views on himself, the Kindertelefoon, the conversation and his network. We also defined 5 desires that indicate Lilobot's interest in sharing the problem, whether or not to end the conversation and potential solutions he could adopt to solve the problem. Based on these beliefs, desires and an ontology based on a corpus of past conversations from the helpline, Lilobot simulates a level of realism that reflects the behavior of a child help seeker as stated by requirement R02. To meet requirements R03 and R04, we designed a mapping between user input and subsequent BDI updates to help explain the behaviour of the agent. We clarify how the actions of the counsellor (user) influence the decisions made by Lilobot. For instance, showing empathy increases Lilobot's belief that the Kindertelefoon understands him which in turn positively impacts his willingness to talk to the counsellor. To promote reflective learning as required by requirement R05, our proposed system provides as feedback, an overview of Lilobot's beliefs and a transcript of the conversation with the corresponding changes in the BDI model. From this, the counsellor can identify which actions had a positive impact and which were detrimental to the success of the conversation. This design was implemented in a web-hosted prototype with two main components - a natural language processing model and the BDI reasoning model. The former was built using Rasa NLU and the latter, as a micro-service-based web application in Spring Boot.

How do users perceive the agent?

The prototype was evaluated in an experiment with 28 counsellors of the Kindertelefoon. The measures considered were counselling self-efficacy and perceived usefulness. We set up a repeated measures study with two training interventions - the conversational agent and a text-based training. Results of the experiment indicated that there was a significant difference in the counselling self-efficacy of the participants upon experiencing the conversational agent which was not the case for the text-based training. Although the agent succeeded in influencing participants' counselling self-efficacy, self-efficacy decreased after training with the agent instead of increasing which would have been the desired outcome for a training tool. We posit two explanations for this result. One being the presence of a cognitive bias that might have led participants to overestimate their abilities and the other, a possible misinterpretation of the questionnaire used. With regards to perceived usefulness, the results suggested the conversational agent needs improvement to be regarded as truly useful. There was a significant decrease in participants' attitude towards the use of a conversational agent as a learning tool. To give more context to this, we analysed the open feedback we received. 50% of the participants were still positive about the use of the agent as an extra learning opportunity for counsellors, especially counsellors-in-training if it were developed further. They noted the added opportunity for self-directed learning would be beneficial for counsellors. On the contrary, 21% of the participants thought it would be better for counsellors to practise with real children.

6.2. Limitations

There are several methodological weaknesses that limit the generalizability of our findings. As a feasibility study, we recruited a limited number of participants to receive a relatively short intervention and no opportunity for a long-term study on the agent's effectiveness. We tested the conversational agent with only experienced counsellors at the helpline due to the limited pool of counsellors-in-training at the time of evaluation. Participants performed the training tasks online with no supervision which is quite different from the normal training setting of the helpline. This study should be replicated with more participants, preferably with counsellors with little to no experience and supervisors at the helpline to

inform more concrete findings.

We also observed the incorporation of other counselling theories in the conversations with Lilobot like setting limits in the conversation and meta-communication which were beyond the scope of our proposed solution. This begs the question whether or not the Five Phase Model can be practised independently of other training modules of the helpline and which group of counsellors can benefit the most from the conversational agent in its current form.

6.3. Contribution

While our findings should be viewed with caution, this research nonetheless demonstrated that a BDI-based conversational agent has the potential to offer an engaging method to learn adherence to the Five Phase Model to counsellors of the Kindertelefoon. On a scientific level, we presented a conversational agent capable of simulating a child victim of school bullying based on the Belief-Desire-Intention model. Moreover, our findings from this research could be used to inform further development of conversational agents for training counsellors of the Kindertelefoon and other organizations like the Danish child helpline Børns Vilkår that make use of the Five Phase Model or similar conversation models.

Conversations are complex and talking to children is even more complex. In reality, applying the Five Phase Model does not mean a conversation ends successfully. Thus on a practical level, we recommend based on our findings that the conversational agent which focuses only on the Five Phase Model be used by counsellors-in-training at the helpline. This would enable counsellors who do not yet have access to children to gain familiarity with the conversation model. The conversational agent when developed further could be integrated into the helpline's current 30-hour counsellor training program as an added opportunity for full conversation role-playing, serving as an opportunity for self-regulated learning and promoting reflection on their actions. Although experienced counsellors could benefit from revising the conversation model through our proposed solution, it could be difficult for them to separate the Five Phase Model from the other counselling skills they have acquired and conversation style they have developed through real-life conversations with children. For this target group, a more complex conversational agent that takes other counselling theories into account is recommended in order to make the contact more meaningful.

6.4. Future Work

We recognize several aspects of our research that could be explored in future work. First, natural language processing. The prototype we presented had several limitations given difficulties we encountered with Rasa. Instead of retrieving the intent with the highest confidence score given user input, it would be better if the system could recognize multiple intents. Currently, Rasa allows training of paired intents but this is not scalable granted the number of intents in our prototype. Another option would be to retrieve the whole list of intents Rasa NLU model returns and define specific threshold values for acceptable confidence scores. This way, the agent can understand complex user utterances, making the interaction more natural and human-like. This would also alleviate the frustration users may face with having to segment their utterances.

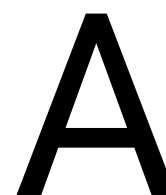
Another aspect that can be improved is the ontology of the agent. Our current ontology is based on common themes we identified in the helpline's data on school bullying. In order to create a more immersive experience, future work could look into expanding Lilobot's ontology to allow counsellors to delve deeper into the story and potential solutions. This could be achieved through the addition of training examples from the helpline's conversation corpus or directly from sample conversations with the agent. It is likely that such data would need to be anonymized in case counsellors disclose personal information in these conversations.

From a broader perspective, future work could also explore the adaptability of the current cognitive model to other topics other than bullying like domestic abuse and suicide, for instance. It would be interesting to know if the agent generalizes well to these domains which are also subjects of interest to the Kindertelefoon. With regards to the feedback the agent provides, future research could explore the translation of the current BDI status updates into actionable insights counsellors could use to improve their performance with the agent. As part of the feedback, the system could for instance, suggest that the counsellor show more empathy if belief B05 remains low throughout the conversation. As Van der Meij [86] notes, proximal positioning of error information is crucial in manuals. It can help users catch mistakes early before they lead to possibly even greater mistakes.

Furthermore, the cognitive model of our agent could be expanded to include an affective model. A cognitive-affective architecture could enhance the conversational agent's behavior with affective features such as emotions, mood and personality management. This way, Lilobot can make decisions taking into consideration both its cognitive and affective states.

6.5. Final Remarks

In this thesis, we presented a conversational agent based on the Belief-Desire-Intention (BDI) framework that simulates a child help seeker for the purpose of training counsellors of the Dutch child helpline the Kindertelefoon. Results of our evaluation show the potential of this conversational agent as an additional learning opportunity for training counsellors of the helpline.



Appendix Foundation

A.1. Informed Consent Form

Wat u gaat doen:

U bent uitgenodigd om deel te nemen aan de focusgroep over “training child helpline counsellors using a conversational agent”. Tijdens de focusgroep krijgt u samen met de andere deelnemers een aantal scenario's te zien die heeft met het trainen van vrijwilligers en gespreksvoering te maken. U zal worden gevraagd of deze situatie herkenbaar is voor u als vrijwilliger of werkbegeleider en hoe u met deze situatie om gaan. Wij herinneren u eraan om geen persoonlijk identificeerbare informatie te delen.

Wat wij doen:

Wij maken een video-opname van de discussie, transcriberen die en verwijderen dan de video-opname. Alle informatie wordt geanonimiseerd (er worden bijvoorbeeld geen names of andere persoonlijk identificeerbare informatie opgeschreven). De geanonimiseerde transcriptie gebruiken wij om te analyseren welke problemen er spelen in het trainen van vrijwilligers van de Kindertelefoon en hoe wij met onze technologie de training en trainee kunnen ondersteunen. De geanonimiseerde informatie die u deelt kan gebruikt worden in wetenschappelijke publicaties. De informatie die u geeft wordt alleen gebruikt met uw toestemming (zie andere zijde van dit formulier).

Waarom doen we dit:

De kennis die wij opdoen over de problemen/zaken van de opleiding van vrijwilligers van de Kindertelefoon gebruiken wij voor het afstudeeronderzoek van de bovengenoemde onderzoeker van de Technische Universiteit Delft. Dit specifieke onderzoek gaat over het ontwerp, bouwen en testen van een zogenaamde “conversational agent”, een chatbot die de rol speelt van een kind in een trainingsomgeving om vrijwilligers van de Kindertelefoon op te leiden.

Uw rechten:

U heeft het recht om niet deel te nemen aan de focusgroep en om op elk moment te stoppen met deelname, zonder dat u daarvoor een reden hoeft te geven. Dit kunt u mondeling aangeven op elk moment gedurende de focusgroep aan de begeleider. U kunt ook aangeven dat (delen van) informatie die u verstrekt, verwijderd moet worden. Deze informatie zal dan vernietigd worden. Dit kan ook na de focusgroep, neem hiervoor contact op met de onderzoeker (zie het e-mail adres bovenaan dit formulier). Daarnaast kunt u na de focusgroep toegang vragen tot de informatie die u heeft verstrekt. De informatie zal bewaard blijven tot het einde van het onderzoek waarna het vernietigd wordt als het niet gebruikt wordt voor publicatie. Dit zal omstreeks april 2021 zijn.

Vink uw keuze aan	Ja	Nee
Deelnemen aan dit onderzoek		
Ik heb de informatie gelezen of gehoord en ik heb deze begrepen. Ik was in staat om vragen over het onderzoek te stellen en mijn vragen zijn naar mijn voldoening beantwoord.	<input type="radio"/>	<input type="radio"/>
Ik ben een vrijwillige deelnemer van dit onderzoek en ik begrijp dat ik me kan terugtrekken uit dit onderzoek op elk moment, zonder dat ik daar een reden voor hoef te geven.	<input type="radio"/>	<input type="radio"/>
Gebruik van informatie		
Ik begrijp dat deelname aan dit onderzoek inhoudt dat informatie die ik geef opgenomen wordt.	<input type="radio"/>	<input type="radio"/>
Ik begrijp dat persoonlijke informatie die mij kan identificeren, zoals mijn naam, niet gedeeld wordt en zo snel mogelijk verwijderd wordt.	<input type="radio"/>	<input type="radio"/>
Ik begrijp dat mijn geanonimiseerde informatie gebruikt zal worden in wetenschappelijk onderzoek.	<input type="radio"/>	<input type="radio"/>
Ik geef toestemming voor het publiceren van mijn geanonimiseerde informatie als quote in wetenschappelijke publicaties (ter illustratie en bewijs van de bevindingen).	<input type="radio"/>	<input type="radio"/>
Toekomstig hergebruik van informatie		
Ik geef toestemming om mijn geanonimiseerde informatie te archiveren in een wetenschappelijke bron zodat het gebruikt kan worden voor toekomstig onderzoek.	<input type="radio"/>	<input type="radio"/>

A.2. Persona

Name	Marie
Gender	Female
Age	23
Occupation	Student
Motivation	Marie is very enthusiastic about being a counsellor at the Kindertelefoon. One of her passions is working with children and therefore studies Social Work.

Table A.1: Persona of Marie, an aspiring counsellor at the Kindertelefoon

A.3. Scenarios and Claims

Scenario 1: Structure	
During the face-to-face training, Marie learnt about the Five Phase Model that is used to guide conversations at the helpline. However, Marie has some doubts about her ability to apply one of the phases and therefore wants to practise with the training system.	
Strategy A: All phases	
Description	Remarks
Marie converses with the conversational agent. The conversation covers all phases of the Five Phase model. She starts at phase 1 and ends at phase 5.	It is easy to lose focus or overly rely on information from previous phases while having the conversation. However, there is a need for counsellors to practise conversations covering all the phases because this is not usually done in the face-to-face training. It is also important to learn how the timing and transition of phases in a conversation works.
Strategy B: One phase	
Description	Remarks
Marie converses with the conversational agent. The conversation covers only one phase of the Five Phase model. The conversation ends when the goal of that specific phase is achieved.	It is helpful to focus on one phase at a time but it requires context from the previous phases to be able to do it. For example, without knowledge from phase 2, a counsellor will find it very difficult to arrive at the goal of phase 3. Counsellors generally have difficulty with phase 3.
Suggestion	
Combine both strategies. Focus on one at a time but create the opportunity for the counsellor to practice one phase after another, effectively, covering all phases.	

Table A.2: Description of scenario 1 on structure.

Scenario 2: Difficult conversations	
During the face-to-face training, Marie learnt about how to manage difficult behaviour in order to engage in a productive conversation. She wants to practise with the training system on how to apply the techniques.	
Strategy A: Avoid conflict	
Description	Remarks
The conversation slowly builds from the start where the goal for Marie is to avoid getting into a problematic situation with the difficult chatter.	Room for counsellor to make decisions on how to approach the conversation and handle pitfalls.
Strategy B: Start with conflict	
Description	Remarks
The conversation sets up after a problematic situation has happened and the goal is for Marie to navigate the conversation onto a good trajectory.	Placing a trainee in an unknown situation or something they haven't worked towards will make it very hard to navigate or solve. On the other hand, it is good to practise difficult conversation with a conversational agent.
Suggestion	
The participants suggested focusing on difficult conversations but giving the trainee the room to make decisions instead of starting with a problem.	

Table A.3: Description and claims of scenario 2 on difficult conversations.

Scenario 3: Feedback	
Marie wants to get feedback based on her session with the training system.	
Strategy A: System gives feedback	
Description	Remarks
The system provides Marie feedback on her performance.	Feedback can be useful if it provides insight into why the chatbot reacts a certain way showing the effect of the trainee's actions. However, feedback isn't just about qualitative results. It's a personal conversation to help with the trainee's development with the system lacks.
Strategy B: Trainer gives feedback	
Description	Remarks
The system provides Marie a transcript of the conversation which she discusses with a trainer in person.	A trainer can provide better nuance and explanation for why a certain action is wrong. It is also an opportunity for a trainee to ask questions and for clarification.
Suggestion	
The participants suggested a combination of the two. First, feedback from the chatbot and then an in-person conversation with a trainer for more clarification. The trainer should however be able to see the chatbot's feedback and correct it if necessary.	

Table A.4: Description and claims of scenario 3 on feedback.

Difficulty	Rationale
Suicide, abuse and other life-threatening situations	These topics tend to elicit countertransference reactions among counsellors such as feelings of anger, hopelessness and a desire to rescue their clients [43]. "They are very vulnerable children so I expect myself to do very well." [71]. Some counsellors are usually faced with a saviour complex where they feel a great responsibility to save the child but at the same time feel not powerful enough to do so. Others expressed that they find it difficult to talk about such topics because they have their own experiences and ideations which lead to a form of stigmatization and self-protection.
Overly dramatic and pitiful behaviour in children	Similar to above, these types of conversations tend to bring out the saviour complex among counsellors according to the participants of the focus groups. This leads to feelings of extreme sympathy and a desire to rescue the child from their situation. The effect of such attitude is that the counsellor comes up with a lot of solutions for the child [71] which is against the helpline's vision of empowering children.
Fake-sounding problems	The counsellor does not know whether to take the child seriously which can lead to frustration. Generally, counsellors are taught to take every conversation seriously and also be able to play along with jokes.
Slow response time, aggressiveness, asking provoking questions	The counsellor does not feel taken seriously in the conversation. In some cases, children try to cross certain personal boundaries of the counsellor which can be frustrating.

Table A.5: Difficulties in conversations with children.

A.4. Mind Map

This figure is intended for viewing on the electronic version of this thesis.

B

Appendix Design

B.1. Beliefs

ID	Belief name (EN)	Belief name (NL)	Phase
<i>about self</i>			
B01	Lilobot thinks he is in control	Lilobot denkt dat hij de controle heeft	All
B02	Lilobot thinks he is competent to solve problem	Lilobot denkt dat hij competent is om het probleem op te lossen	4
B03	Lilobot feels connected to KT	Lilobot voelt zich verbonden met KT	All
<i>about the Kindertelefoon (KT)</i>			
B04	Lilobot thinks KT can be trusted	Lilobot denkt dat KT te vertrouwen is	2
B05	Lilobot thinks KT understands him	Lilobot denkt dat KT hem begrijpt	All
B06	Lilobot thinks KT is interested in his situation	denkt dat KT geïnteresseerd is in zijn verhaal	2
B07	Lilobot thinks KT can help him	Lilobot denkt dat KT hem kan helpen	3
B08	Lilobot thinks KT can solve his problem	Lilobot denkt dat KT het probleem kan oplossen	3
B16	Lilobot thinks he and KT will be able to reach a solution	Lilobot denkt dat hij samen met KT tot een oplossing zullen komen	4
B17	Lilobot thinks KT is going to solve the problem for him	Lilobot denkt dat KT het probleem voor hem gaat oplossen	4
<i>about the conversation</i>			
B09	Lilobot thinks he has talked about his situation	Lilobot denkt dat hij zijn verhaal heeft verteld	2
B10	Lilobot thinks KT is asking about his wish	Lilobot denkt dat KT vraagt naar wens	3
B11	Lilobot thinks KT is asking about a positive wish	Lilobot denkt dat KT vraagt naar een positieve wens	3
B12	Lilobot thinks KT is asking about a confidant	Lilobot denkt dat KT vraagt naar vertrouwenspersoon	4
B13	Lilobot feels safe in the conversation	Lilobot voelt zich veilig in het gesprek	4
B15	Lilobot thinks KT wants to end the conversation	Lilobot denkt dat KT het gesprek wil beëindigen	5
<i>about confidant</i>			
B14	Lilobot thinks his teacher can help him	Lilobot denkt dat zijn juf hem kan helpen	All

Table B.1: Beliefs of the conversational agent prototype.

B.2. Desires

ID	Desire name (EN)	Desire name (NL)
D1	Lilobot wants to talk about his problem	Lilobot wil over zijn probleem hebben
D2	Lilobot wants to end the conversation	Lilobot wil het gesprek beëindigen
D3	Lilobot wants the Kindertelefoon to get the bullies out of school	Lilobot wil dat de Kindertelefoon de pestkoppen van school haalt
D4	Lilobot wants to talk to his teacher about problem	Lilobot wil met met zijn leraar praten over situatie
D5	Lilobot wants to work with the Kindertelefoon to find a solution	Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Table B.2: Desires of conversational agent prototype.

B.3. Actions

ID	Action name (EN)	Type	Desire ID
A1	Lilobot introduces problem	inform	D1
A2	Lilobot provides more details about problem	inform	D1
A3	Lilobot talks about emotional impact of problem	inform	D1
A4	Lilobot says goodbye	inform	D2
A5	Lilobot asks counsellor to call the school to get the bullies out	request	D3
A6	Lilobot asks how teacher can help	request	D4
A7	Lilobot expresses concern about bullying getting worse	request	D4
A8	Lilobot asks counsellor what to tell teacher	request	D4

Table B.3: Actions of conversational agent prototype.

B.4. Sample Feedback

FEEDBACK GESPREK

Hier is een transcriptie van je gesprek met Lilobot met zijn gedachten tijdens het gesprek. Lilobot heeft een reeks overtuigingen en verlangens die tijdens het gesprek constant worden bijgewerkt op basis van wat je tegen hem zegt. In de onderstaande tabel kun je zien wat Lilobot's overtuigingen waren aan het begin van het gesprek en aan het einde. De transcriptie van het gesprek laat zien welke overtuigingen veranderen op basis van jouw berichten. Het symbool ↑ betekent dat de overtuiging toeneemt, terwijl ↓ betekent dat de overtuiging afneemt. De transcriptie laat ook zien welke verlangens Lilobot had op het moment in het gesprek. Al deze notaties zijn cursief weergegeven tussen jullie gesprek.

Overtuiging	Vijffasemodel	Begin	Eind	Vershil
Ik voel me verbonden met KT	alle fasen	40%	78%	38%
Ik voel me veilig in het gesprek	alle fasen	100%	100%	0%
Ik voel me in controle in het gesprek	alle fasen	50%	80%	30%
Ik denk dat KT mij begrijpt	alle fasen	0%	100%	100%
Ik denk dat KT geïnteresseerd is in mijn verhaal	fase 2	0%	20%	20%
Ik denk dat KT te vertrouwen is	fase 2	60%	80%	20%
Ik denk dat ik mijn verhaal heb verteld	fase 2	0%	70%	70%
Ik denk dat KT vraagt naar een positieve wens	fase 3	0%	30%	30%
Ik denk dat KT het probleem kan oplossen	fase 3	100%	0%	-100%
Ik denk dat KT vraagt naar een wens	fase 3	0%	100%	100%
Ik denk dat KT mij kan helpen	fase 3	70%	90%	20%
Ik denk dat ik competent ben om het probleem op te lossen	fase 4	30%	70%	40%
Ik denk dat KT vraagt naar een vertrouwenspersoon	fase 4	0%	100%	100%
Ik denk dat KT het probleem voor mij gaat oplossen	fase 4	0%	0%	0%
Ik denk dat KT en ik samen tot een oplossing zullen komen	fase 4	0%	100%	100%
Ik denk dat Juf Ellie mij kan helpen	fase 4	50%	100%	50%
Ik denk dat KT het gesprek wilt beëindigen	fase 5	0%	100%	100%

TRANSCRIPTIE

Lilo: Hoi!

KT: Dag Lilobot

Overtuiging: ↑ Ik denk dat KT te vertrouwen is

Intentie: Lilobot wil het over zijn probleem hebben

Lilo: Hallo, ik ben Lilobot

Overtuiging: ↑ Ik denk dat ik mijn verhaal heb verteld

Lilo: Ik word gepest op school

KT: Dat is vervelend

Overtuiging: ↑ Ik denk dat KT mij begrijpt

Intentie: Lilobot wil het over zijn probleem hebben

Figure B.1: Screenshot of a sample conversation transcript and feedback from Lilobot.

Lilo:
Overtuiging: ↑ Ik denk dat ik mijn verhaal heb verteld

Lilo: Gisteren hebben ze mij weer in elkaar geslagen. ze schelden me vaak uit en slaan me soms. ze schelden ook met kanker...dat vind ik niet leuk

KT: Ik kan me voorstellen dat je het niet leuk vindt
Overtuiging: ↑ Ik denk dat KT mij begrijpt
Intentie: Lilobot wil het over zijn probleem hebben

Lilo:

KT: Wat hoop je dat er anders wordt door dit gesprek>
Intentie: Lilobot wil dat de Kindertelefoon de pestkoppen van school haalt

Lilo: ik wil niet meer gepest worden

KT: Dat zou fijn zijn he!
Overtuiging: ↑ Ik denk dat KT mij begrijpt
Intentie: Lilobot wil dat de Kindertelefoon de pestkoppen van school haalt

Lilo: okeee

KT: We kunnen samen eens praten om te kijken wat we kunnen bedenken.
Overtuiging: ↑ Ik denk dat KT mij kan helpen
Overtuiging: ↓ Ik denk dat KT het probleem kan oplossen
Intentie: Lilobot wil dat de Kindertelefoon de pestkoppen van school haalt

Lilo: :)

KT: Hoe lijkt je dat?
Overtuiging: ↑ Ik voel me in controle in het gesprek
Overtuiging: ↑ Ik denk dat KT vraagt naar een positieve wens
Intentie: Lilobot wil dat de Kindertelefoon de pestkoppen van school haalt

Lilo: Ik zou me gelukkig voelen

KT: Dan doen we dat!
Overtuiging: ↓ Ik denk dat KT het probleem kan oplossen
Overtuiging: ↓ Ik denk dat KT mij kan helpen
Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Lilo: :(

KT: Wat vind jij ervan dat je wordt gepest?
Overtuiging: ↑ Ik denk dat KT mij kan helpen
Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Lilo: jaaaaaa

KT: Wie weten er dat je wordt gepest?
Overtuiging: ↑ Ik denk dat KT geïnteresseerd is in mijn verhaal
Intentie: Lilobot wil het over zijn probleem hebben

Figure B.2: Screenshot of a sample conversation transcript and feedback from Lilobot.

Lilo: ik heb er nog met niemand over gepraat. Anders wordt het pesten erger. :(

KT: Waardoor denk je dat?
Overtuiging: ↑ Ik denk dat KT geïnteresseerd is in mijn verhaal
Intentie: Lilobot wil het over zijn probleem hebben

Lilo: Ik weet het niet zeker, maar ze zeggen 'Nou als Lilo een voldoende heeft voor biologie dan is vast een makkelijke toets' 'stop maar met praten, je bent hier te dom voor'...en dat doet me veel pijn
Overtuiging: ↑ Ik denk dat ik mijn verhaal heb verteld

Lilo: Ik voel me heel vaak buitengesloten. :(((Ik wil ook heel graag naar een andere school maar mijn vader zegt dat dat niet zo makkelijk kan.

KT: Ik snap dat het pijn doet
Overtuiging: ↑ Ik denk dat KT mij begrijpt
Intentie: Lilobot wil het over zijn probleem hebben

Lilo: :)))

KT: Wat hoop je van dit gesprek?
Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Lilo: Ik wil dat het pesten stopt

KT: Dat kan ik goed voorstellen
Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Lilo: okeeee

KT: Hoe zou het zijn als dat zou lukken?
Overtuiging: ↑ Ik voel me in controle in het gesprek
Overtuiging: ↑ Ik denk dat KT vraagt naar een positieve wens
Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Lilo: gewoon vrolijk zijn...Ik wil me gewoon veilig op school zijn en samen met vriendin lunchen enzo

KT: Dat klinkt goed!
Overtuiging: ↑ Ik denk dat ik competent ben om het probleem op te lossen
Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Lilo: :))

KT: Wat zou je denken als je je vrolijk zou voelen?
Overtuiging: ↑ Ik voel me in controle in het gesprek
Overtuiging: ↑ Ik denk dat KT vraagt naar een positieve wens
Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Lilo: ik zou gelukkig zijn

KT: Wat fijn!
Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Figure B.3: Screenshot of a sample conversation transcript and feedback from Lilobot.

Lilo: okeee

KT: Wat zou je vandaag kunnen doen om je vrolijker te voelen?
Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Lilo:

KT: Ik denk dat je de hulp van een volwassene kunt gebruiken
Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Lilo: wanneer zou je mijn school kunnen bellen?

KT: ik kan de school helaas niet bellen
Overtuiging: ↓ Ik denk dat KT mij kan helpen
Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Lilo: wat stom!

KT: jammer he!
Intentie: Lilobot wil samen met de Kindertelefoon een oplossing zoeken

Lilo: okeeee

KT: wie zou jij vertrouwen om het tegen te vertellen?
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: Ik vind juf Ellie lief, maar ik durf het haar niet te vertellen

KT: wat vind je spannend?
Overtuiging: ↑ Ik denk dat KT mij kan helpen
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: :)))

KT: fijn dat juf Ellie lief is!
Overtuiging: ↑ Ik denk dat KT mij kan helpen
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: :)

KT: Ik snap dat het spannend is

KT: hoe denk je dat ze zou reageren?
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: Ik begrijp niet wat je bedoelt

KT: Zou juf ellie kunnen helpen?
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: :)

KT: Kan de juf helpen?
Intentie: Lilobot wil met zijn leraar praten over situatie

Figure B.4: Screenshot of a sample conversation transcript and feedback from Lilobot.

Lilo: :)

KT: Wanneer zou je het haar kunnen vertellen?
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: ik ga morgen met haar praten

KT: Wat knap!
Overtuiging: ↑ Ik denk dat ik competent ben om het probleem op te lossen
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: dank je

KT: Wanneer is een goed tijdstip?
Overtuiging: ↑ Ik denk dat ik competent ben om het probleem op te lossen
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: :)

KT: Wat ga je dan zeggen?
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: Dat ik gepest wordt enzo

KT: Ja inderdaad!
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: okeeee

KT: Wat denk je dat haar reactie is?
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: Hopelijk beter :)

KT: Ja, dat hoop ik ook

KT: Kun je zo verder?
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: misschien na school ofzo

KT: Dat is een goed idee!
Overtuiging: ↑ Ik denk dat ik competent ben om het probleem op te lossen
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: dankje

KT: Kun je hiermee verder?
Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: misschien na school ofzo

KT: Hoe vond je het gesprek?
Intentie: Lilobot wil met zijn leraar praten over situatie

Figure B.5: Screenshot of a sample conversation transcript and feedback from Lilobot.

Lilo: Jaaa

KT: Veel succes morgen!

Intentie: Lilobot wil met zijn leraar praten over situatie

Lilo: Doei!

Figure B.6: Screenshot of a sample conversation transcript and feedback from Lilobot.

B.5. Example Interaction

ID	Belief Name	Belief Value
B01	Lilobot thinks he is in control (autonomy)	0.5
B02	Lilobot thinks he is competent to solve problem (competence)	0.3
B03	Lilobot feels connected to the Kindertelefoon (relatedness)	0.4
B04	Lilobot thinks KT can be trusted	0.6
B05	Lilobot thinks KT understands him	0
B06	Lilobot thinks KT is interested in his story	0
B07	Lilobot thinks KT can help him	0.7
B08	Lilobot thinks KT can solve his problem	1
B09	Lilobot thinks he has talked about his situation	0
B10	Lilobot thinks KT is asking about his wish	0
B11	Lilobot thinks KT is asking about a positive wish	0
B12	Lilobot thinks KT is asking about a confidant	0
B13	Lilobot feels safe in the conversation	0.5
B14	Lilobot thinks his teacher can help him	1
B15	Lilobot thinks KT wants to end the conversation	0
B16	Lilobot thinks he and KT will be able to reach a solution	0
B17	Lilobot thinks KT is going to solve the problem for him	0

Table B.4: Initial beliefs of the conversational agent Lilobot as used in the example interaction.

C

Appendix Evaluation

C.1. OSF Registration

The online registration is available at <https://osf.io/hkxzc>.

Study Information

Hypotheses

H1: Training with the conversational agent simulating a child increases the counselling self-efficacy of the participants.

H2: Training with the conversational agent leads to a higher increase in counselling self-efficacy than the text-based intervention.

H3: Participants perceive training with the conversational agent as useful.

H4: Experience with the conversational agent improves outcome of conversation.

Design Plan

Study type

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

For studies that involve human subjects, they will not know the treatment group to which they have been assigned. Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as “double blind”)

Is there any additional blinding in this study?

Personnel will not be aware of the participants' level of counselling experience.

Study design

A repeated measures (within-subjects) design will be used. The participants will be exposed to two interventions - a text-based intervention (a simple text explaining the Five Phase Model theory) and an interaction-based intervention (the conversational agent simulating a child they have to counsel according to the Five Phase model). Interaction with the agent will take place over three consecutive sessions of 15 minutes each. In order to minimize order effect, the participants will be split into two groups. Each will experience both interventions, but in reverse order.

Prior to this, all participants will be asked to sign an informed consent form and a pre-experiment questionnaire on the measured variables (level of counselling experience and self-efficacy). After completing the

first intervention, the participants will be asked to fill in a questionnaire on the measured variable (self-efficacy). At the end of the second intervention, the participants will be asked to fill in a post-training questionnaire on the measured variables (self-efficacy and perceived usefulness) that will be compared to the previous measures.

The study design is shown in the file attached below.

Study Design Evaluation of a BDI-based Virtual Agent for Training Child Helpline Counsellors.pdf (<https://osf.io/a3ufw/>)

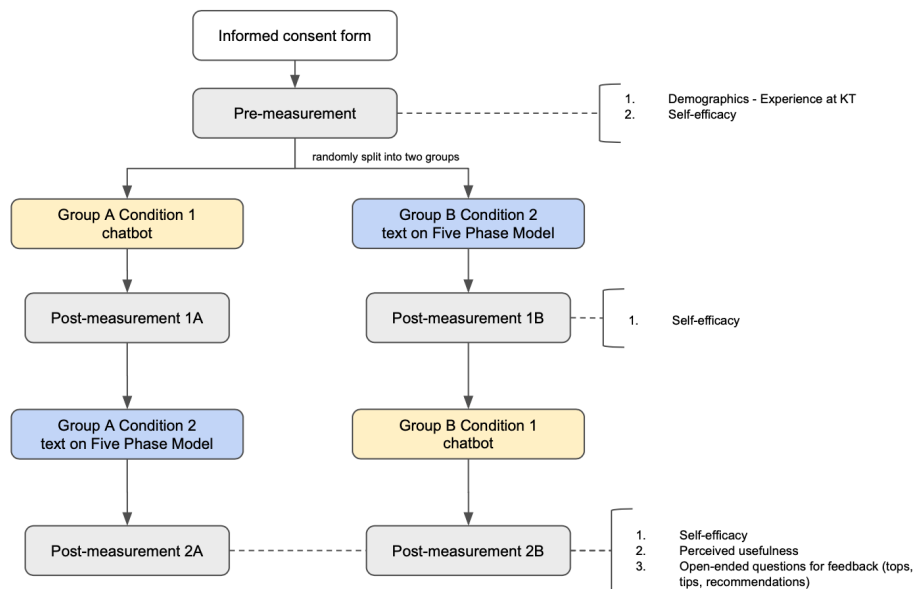


Figure C.1: Study design

Randomization

The participants will be randomly assigned by means of simple randomization to one of the two equally sized groups. Each group will perform the same tasks but in reverse order to minimize order effect.

Sampling Plan

Existing Data

Registration prior to creation of data

Explanation of existing data

No response

Data collection procedures

Participants of the study will be recruited from the child helpline's pool of volunteers. These include counsellors of all experience levels as well as counsellors-in-training at the helpline. Participants will be recruited by means of convenience sampling - people who respond to the researcher's call for participants and will receive no monetary compensation whatsoever. Recruitment will take place over a month prior to the experiment.

Prior to the first training session, participants will be asked to fill to sign an informed consent form and will be briefed about the nature of the experiment and the interventions. Additionally, the participants will be asked to fill in a pre-training questionnaire which will be hosted on Qualtrics. The questionnaire will cover their counselling experience at the helpline and counselling self-efficacy measurements.

The participants will be given both of the following two conditions after each other. The order of this will be determined by their group assignment: - a text-based training explaining the Five Phase Model - an interaction-based training which is the conversational agent simulating a child they have to counsel according to the Five Phase Model

At the end of each intervention, the participants will be asked to fill in a counselling self-efficacy questionnaire. Both interventions will be made available to the participants through the online survey. The participants will engage with the conversational agent for three consecutive sessions. At the end of each session, the conversational agent will provide feedback to the participant based on the BDI status of the simulated child.

After the second intervention, the participants will be asked to fill in a post-training questionnaire regarding their counselling self-efficacy and perceived usefulness of the conversational agent. We will ask additional open-ended questions regarding the use of the conversational agent for an explorative analysis.

No files selected

Sample size

Our target sample size is 60 participants. We will attempt to recruit more to compensate for missing data at later stages of the study.

Sample size rationale

There is a limited pool of counsellors at the helpline which is about 600 people. We expect about 60 of them to participate in our study thus, determining our sample size.

Given a sample size of 60 and a power of 0.9, we used G*Power to conduct a power analysis. With this, we obtain a medium effect size of 0.42 at the standard 0.05 alpha error probability.

Stopping rule

We will stop recruiting participants when we reach the time limit of the study.

Variables

Manipulated variables

We manipulate the training strategy of the Five Phase Model. This is done through the text and the conversational agent. All participants receive both interventions.

No files selected

Measured variables

For hypotheses H1 and H2, the measured variable will be the perceived counselling self-efficacy of the participant toward the training strategy.

Counselling Self-efficacy Five Phase Model Questionnaire This questionnaire was developed for this study based on existing counselling self-efficacy questionnaires including Larson et al. (1992) and Axboe et al. (2016). The questionnaire was validated by supervisors at the helpline. The questionnaire consists of 8 statements that requires participants to rate the degree to which they agree or disagree about their abilities to perform tasks related to the Five Phase Model. Each of the statements are rated on a scale from 'strongly disagree' -5, 'neutral' 0 to 'strongly agree' +5.

For hypothesis H3, the measured variable will be the perceived usefulness of the conversational agent simulating a child. We will make of the Perceived Influence on Learning Outcome (PILO) questionnaire and the System Usability Scale (SUS).

Perceived Influence on Learning Outcome (PILO) This questionnaire is adopted from the work of Kang (2016), Lie (2018) and Sirocki (2019). The questionnaire contains eight statements participants indicate how useful they found the conversational agent on a semantic scale ranging from 'discouraging' -5 to 'motivating' +5 as the influence on learning outcome.

System Usability Scale (SUS) The questionnaire contains 10 items with one of five responses that range from 'strongly agree' to 'strongly disagree'. For this study, we use the Dutch translation of the questionnaire by Wever et al. (2012).

For hypothesis H4, the measured variable will be the BDI status outcome of the conversational agent after each session. This is a sum of 10 beliefs resulting in a score within range 0 and 10.

No files selected

Indices

Counselling Self-efficacy Five Phase Model Questionnaire We will take the mean of the eight questions as a participant's counselling self-efficacy.

Perceived Influence on Learning Outcome (PILO) Each question's score will be interpreted independently.

System Usability Scale (SUS) The usability score will be a single percentile value. The participant's scores for each question of the System Usability Scale will be converted to a new number. For odd numbered questions we subtract 1 from the score and even numbered questions we subtract their value from 5. These values will be added together and then multiplied by 2.5 to convert the original scores of 0-40 to 0-100. Based on research, a SUS score above a 68 would be considered above average and anything below 68 is below average.

No files selected

Analysis Plan

Statistical models

H1: We check whether training with the conversational agent increases the self-efficacy of the participants. This will be done using a paired sample t-test of the pre- and post measurement 2 of all participants (as shown in the attached study design). An appropriate alternative will be used should the normality assumption not hold.

H2: We check if there is a higher increase in self-efficacy with the conversational agent than the text-based training. We will use a two-way repeated measures ANOVA to determine if there is an interaction effect between the independent variables (training strategy and time of measurement). We will follow with the appropriate post hoc test of the main effects.

H3: To check if the participants perceive the conversational agent as useful, we conduct a one-sample t-test of the participants' responses to the PILO questionnaire.

H4: For this hypothesis, we check if the outcome of the conversation improves with sessions by conducting a paired sample t-test for the BDI outcome of the first and third sessions of the participants' training with the conversational agent.

No files selected

Transformations

All recoding schemes are mentioned previously under the measured variable and statistical model sections.

Inference criteria

We will use the standard $p < .05$ criteria for our analyses.

Data exclusion

Participants who give nonsensical answers in the questionnaires will be excluded from the analysis. We will report any exclusions in the thesis report.

Missing data

Participants that do not complete all sessions of the training and questionnaires will not be included in the analysis.

Exploratory analysis

We will conduct an explorative analysis focusing on feedback from participants about their experience with the conversational agent and recommendations for future developments. This will be based on the Perceived Influence on Learning Outcome (PILO) questionnaire as well as the open-ended questions.

Other**Other**

No response

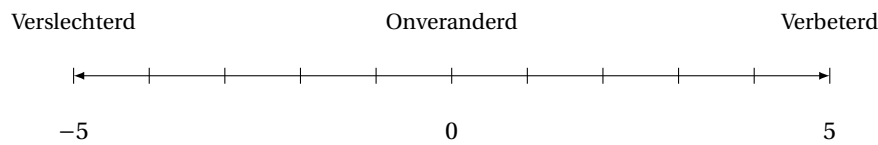
C.2. Counselling Self-efficacy Five Phase Model Questionnaire

Geef aan in hoeverre je bent het eens of oneens met de onderstaande stellingen. Rating: 1 - 5, 1 = Helemaal mee oneens en 5 = Helemaal mee eens		
	Stelling	Rating
<i>Fasen</i>		
1	Ik kan contact maken het kind aan het begin van het gesprek.	
2	Het lukt me om de last die het kind ervaart te verhelderen.	
3	Het lukt mij om de situatie van het kind concreet te krijgen.	
4	Ik kan de gewenste toestand van het kind naar boven brengen.	
5	Ik kan het kind ondersteunen bij het bedenken van een volgende stap.	
6	Ik kan het gesprek afronden.	
<i>Algemeen</i>		
7	Ik kan empathie tonen door de meningen en gevoelens van het kind te erkennen.	
8	Ik kan de informatie afstemmen aan het niveau van het kind.	

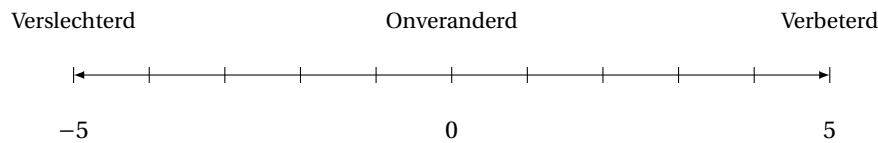
Table C.1: Counselling Self-efficacy Five Phase Model Questionnaire

C.3. Perceived Influence on Learning Outcome (PILO) Questionnaire

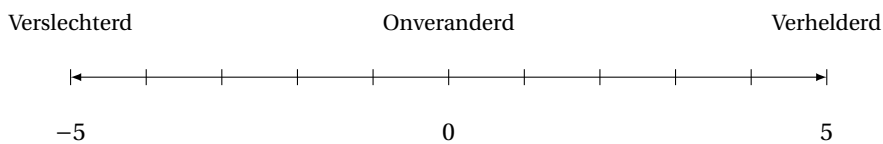
1. Door de training met de chatbot is mijn **kennis** over het voeren van een gesprek volgens het Vijf Fasen Model:



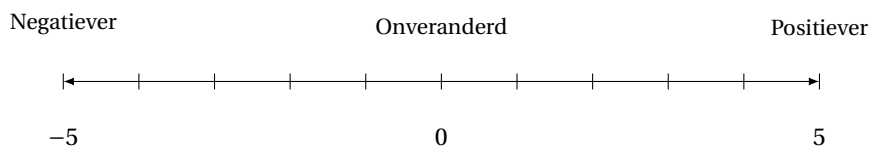
2. Door de training met de chatbot zijn mijn **vaardigheden** met betrekking tot het Vijf Fasen Model:



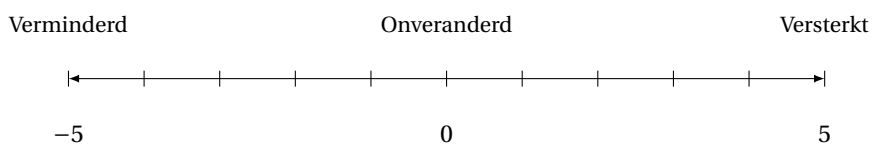
3. Door de training met de chatbot is mijn **denken over het nut** van het voeren van een gesprek volgens het Vijf Fasen Model:



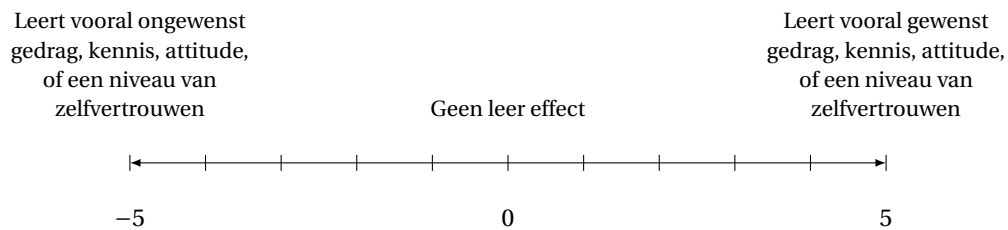
4. Door de training met de chatbot is mijn **mening over het nut** van het voeren van een gesprek volgens het Vijf Fasen Model:



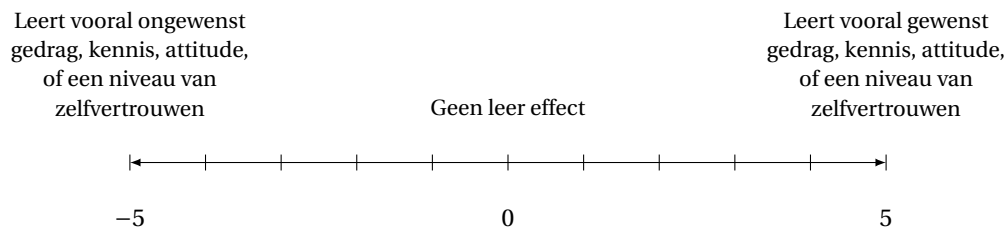
5. Door de training met de chatbot is mijn **zelfvertrouwen** in het voeren van een gesprek volgens het Vijf Fasen Model:



6. Hoe **leerzaam** vond u het gesimuleerde gesprek met de chatbot?



7. Hoe **leerzaam** vond u de **feedback** van het systeem die u kreeg na het gesprek met de chatbot?



8. Door de training met de chatbot is mijn **mening over het nut** van het gebruik van een **chatbot** als leervorm:

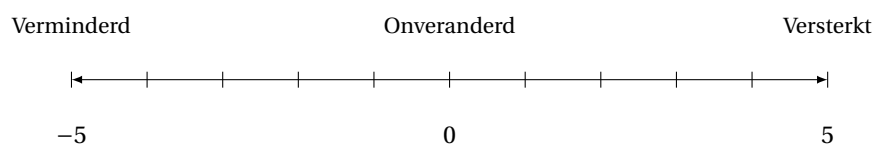


Table C.2: Perceived Influence on Learning Outcome Questionnaire

C.4. Qualitative Measures

1. What was the best thing about your experience of using Lilobot?
2. What was the worst thing about your experience of using Lilobot?
3. What did you like about the feedback you received from Lilobot?
4. What did you dislike about the feedback you received from Lilobot?
5. For which group of people from Kindertelefoon would you recommend Lilobot?
6. What do you think of Lilobot as an additional learning opportunity for novice counsellors?

Bibliography

- [1] Jette Ammentorp, Svend Sabroe, Poul-Erik Kofoed, and Jan Mainz. The effect of training in communication skills on medical doctors' and nurses' self-efficacy: A randomized controlled trial. *Patient education and counseling*, 66(3):270–277, 2007.
- [2] Mette K Axboe, Kaj S Christensen, Poul-Erik Kofoed, and Jette Ammentorp. Development and validation of a self-efficacy questionnaire (se-12) measuring the clinical communication skills of health care professionals. *BMC medical education*, 16(1):1–10, 2016.
- [3] Timothy T Baldwin and J Kevin Ford. Transfer of training: A review and directions for future research. *Personnel psychology*, 41(1):63–105, 1988.
- [4] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.
- [5] Margaret Bearman, Branko Cesnik, and Marilyn Liddell. Random comparison of 'virtual patient' models in the context of teaching clinical communication skills. *Medical education*, 35(9):824–832, 2001.
- [6] Kathryn E Boehm, Connie K Schondel, William J Ivoska, Alison L Marlowe, and Laurie Manke-Mitchell. Calls to teen line: Representative concerns of adolescents. *Adolescence*, 33(132):797–803, 1998.
- [7] Michel Boivin and Shelley Hymel. Peer experiences and social self-perceptions: a sequential model. *Developmental psychology*, 33(1):135, 1997.
- [8] Tibor Bosse and Simon Provoost. Towards aggression de-escalation training with virtual agents: A computational model. In *International Conference on Learning and Collaboration Technologies*, pages 375–387. Springer, 2014.
- [9] Evelyn M Boyd and Ann W Fales. Reflective learning: Key to learning from experience. *Journal of humanistic psychology*, 23(2):99–117, 1983.
- [10] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.
- [11] Michael Bratman. Intention, plans, and practical reason. 1987.
- [12] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.
- [13] John Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.
- [14] Merijn Bruijnes. Believable suspect agents: response and interpersonal style selection for an artificial suspect. 2016.
- [15] John Carroll. Minimalist design for active users. In *Unknown Host Publication Title*, pages 39–44. North-Holland, 1985.
- [16] John M Carroll. *Making use: scenario-based design of human-computer interactions*. MIT press, 2000.
- [17] Davide Castelvetti. Can we open the black box of ai? *Nature News*, 538(7623):20, 2016.
- [18] Juan Cendan and Benjamin Lok. The use of virtual patients in medical school curricula. *Advances in physiology education*, 2012.
- [19] Child Helpline International. Voices of Children and Young People in the EU 2018. 2018.

- [20] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [21] Philip R Cohen and Hector J Levesque. Intention is choice with commitment. *Artificial intelligence*, 42(2-3):213–261, 1990.
- [22] Kenneth Mark Colby, Sylvia Weber, and Franklin Dennis Hilf. Artificial Paranoia. *Artificial Intelligence*, 2(1):1–25, 1971. ISSN 00043702. doi: 10.1016/0004-3702(71)90002-6.
- [23] Jason A Colquitt, Jeffrey A LePine, and Raymond A Noe. Toward an integrative theory of training motivation: a meta-analytic path analysis of 20 years of research. *Journal of applied psychology*, 85(5):678, 2000.
- [24] CRM partners. CRM partners. 2020. URL <https://crmpartners.com/cases/het-grote-vertrouwen-van-de-kindertelefoon/>.
- [25] Antonio R Damasio. *Descartes' error*. Random House, 2006.
- [26] Anique de Beyn. *In gesprek met kinderen: de methodiek van de kindertelefoon*. NIZW, 2003.
- [27] David De Vaus and David de Vaus. *Surveys in social research*. Routledge, 2013.
- [28] Rachel Ellaway, Chris Candler, Peter Greene, and V Smothers. An architectural model for medbiquitous virtual patients. *Baltimore, MD: MedBiquitous*, 6, 2006.
- [29] Rachel Ellaway, Terry Poulton, Uno Fors, James B McGee, and Susan Albright. Building a virtual patient commons. *Medical teacher*, 30(2):170–174, 2008.
- [30] F Faul, E Erdfelder, AG Lang, and A Buchner. A flexible statistical power analysis program for the social, behavioral and biomedical sciences. *Behavior Research Methods*.
- [31] 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, 2017.
- [32] Adriana Foster, Neelam Chaudhary, Thomas Kim, Jennifer L Waller, Joyce Wong, Michael Borish, Andrew Cordar, Benjamin Lok, and Peter F Buckley. Using virtual patients to teach empathy: a randomized controlled study to enhance medical students' empathic communication. *Simulation in Healthcare*, 11(3):181–189, 2016.
- [33] Thomas C Froehle, Sharon E Robinson, and Wayne J De Kurpius. Enhancing the effects of modeling through role-play practice. *Counselor Education and Supervision*, 22(3):197–206, 1983.
- [34] Ruben Fukkink. Peer counseling in an online chat service: A content analysis of social support. *Cyberpsychology, behavior, and social networking*, 14(4):247–251, 2011.
- [35] Ruben Fukkink and Jo Hermanns. Counseling children at a helpline: chatting or calling? *Journal of Community Psychology*, 37(8):939–948, 2009.
- [36] Ruben Fukkink, Jo Hermanns, et al. *Children's experiences with the Kindertelefoon: telephone support compared to chat support*. AmsterdamSCO-Kohnstamm Instituut, 2007.
- [37] Patrick Gebhard. Alma: a layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 29–36, 2005.
- [38] Lisa M Given. *Convenience Sample*. In *The Sage encyclopedia of qualitative research methods*. Sage publications, 2008.
- [39] Gwen M Glew, Ming-Yu Fan, Wayne Katon, Frederick P Rivara, and Mary A Kernic. Bullying, psychosocial adjustment, and academic performance in elementary school. *Archives of pediatrics & adolescent medicine*, 159(11):1026–1031, 2005.

- [40] R Scott Grabinger and Joanna C Dunlap. Rich environments for active learning: A definition. *ALT-J*, 3(2): 5–34, 1995.
- [41] Simon C Hunter, James ME Boyle, and David Warden. Help seeking amongst child and adolescent victims of peer-aggression and bullying: The influence of school-stage, gender, victimisation, appraisal, and emotion. *British Journal of Educational Psychology*, 74(3):375–390, 2004.
- [42] Allen E Ivey, Mary Bradford Ivey, and Lynn Simek-Downing. *Counseling and psychotherapy: Integrating skills, theory, and practice*. Prentice-Hall, Inc, 1987.
- [43] David PH Jones. Individual psychotherapy for the sexually abused child. *Child abuse & neglect*, 10(3): 377–385, 1986.
- [44] Ni Kang. Public speaking in virtual reality: Audience design and speaker experiences. 2016.
- [45] Ni Kang, Ding Ding, M Birna Van Riemsdijk, Nexhmedin Morina, Mark A Neerincx, and Willem-Paul Brinkman. Self-identification with a virtual experience and its moderating effect on self-efficacy and presence. *International Journal of Human-Computer Interaction*, 37(2):181–196, 2021.
- [46] Nuala P Kenny and Brenda L Beagan. The patient as text: a challenge for problem-based learning. *Medical Education*, 38(10):1071–1079, 2004.
- [47] Veton Kepuska and Gamal Bohouta. Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home). In *2018 IEEE 8th annual computing and communication workshop and conference (CCWC)*, pages 99–103. IEEE, 2018.
- [48] Anat Brunstein Klomek, Frank Marrocco, Marjorie Kleinman, Irvin S Schonfeld, and Madelyn S Gould. Bullying, depression, and suicidality in adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, 46(1):40–49, 2007.
- [49] Danielle Koning. Ontwerp van een online zelf-assessment voor het meten van de fysieke activiteit bij ouderen tussen de 55 en 75 jaar. Master's thesis, University of Twente, 2016.
- [50] Andrzej A Kononowicz, Luke A Woodham, Samuel Edelbring, Natalia Stathakarou, David Davies, Nakul Saxena, Lorainne Tudor Car, Jan Carlstedt-Duke, Josip Car, and Nabil Zary. Virtual patient simulations in health professions education: systematic review and meta-analysis by the digital health education collaboration. *Journal of medical Internet research*, 21(7):e14676, 2019.
- [51] Justin Kruger and David Dunning. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6):1121, 1999.
- [52] Shui-fong Lam, Wilbert Law, Chi-Keung Chan, Bernard PH Wong, and Xiao Zhang. A latent class growth analysis of school bullying and its social context: The self-determination theory perspective. *School psychology quarterly*, 30(1):75, 2015.
- [53] Joe Langford. *Focus groups: Supporting effective product development*. CRC press, 2002.
- [54] Lisa M Larson and Jeffrey A Daniels. Review of the counseling self-efficacy literature. *The Counseling Psychologist*, 26(2):179–218, 1998.
- [55] Lisa M Larson, Lisa A Suzuki, Keith N Gillespie, Maria T Potenza, Marilyn A Bechtel, and Alan L Toulouse. Development and validation of the counseling self-estimate inventory. *Journal of counseling Psychology*, 39(1):105, 1992.
- [56] David Ed Lester. *Crisis intervention and counseling by telephone*. Charles C Thomas Publisher, 2002.
- [57] Han Lie. Design and evaluation of a bdi-based virtual patient for the education of shared decision making. 2018.
- [58] Elizabeth A Linnenbrink and Paul R Pintrich. Motivation as an enabler for academic success. *School psychology review*, 31(3):313–327, 2002.

- [59] Henry W Maier. Role playing: structures and educational objectives. *The International Child and Youth Care Network*, 36, 2002.
- [60] Albert Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292, 1996.
- [61] Mark A Neerincx and Jasper Lindenberg. *Situated cognitive engineering for complex task environments*. Ashgate Publishing Limited Aldershot, 2008.
- [62] Toyoaki Nishida. Modeling machine emotions for realizing intelligence—an introduction—. In *Modeling Machine Emotions for Realizing Intelligence*, pages 1–15. Springer, 2010.
- [63] Dan Olweus. Bullying at school: basic facts and effects of a school based intervention program. *Journal of child psychology and psychiatry*, 35(7):1171–1190, 1994.
- [64] Andrew Ortony, Gerald L Clore, and Allan Collins. The cognitive structure of emotions. *Cambridge University*, 1988.
- [65] Marieke Peeters, Karel Van Den Bosch, John-Jules Ch Meyer, and Mark A Neerincx. Scenario-based training: Director’s cut. In *International Conference on Artificial Intelligence in Education*, pages 264–271. Springer, 2011.
- [66] MMM Peeters, Karel van den Bosch, JJ Meyer, and Mark A Neerincx. Situated cognitive engineering: the requirements and design of automatically directed scenario-based training. In *Achi 2012, the fifth international conference on advances in computer-human interactions*, pages 266–272. XPS: Xpert Publishing Services, 2012.
- [67] Joaquín Pérez, Eva Cerezo, Francisco J. Serón, and Luis Felipe Rodríguez. A cognitive-affective architecture for ECAs. *Biologically Inspired Cognitive Architectures*, 18:33–40, 2016. ISSN 2212683X. doi: 10.1016/j.bica.2016.10.002.
- [68] Robert A Peterson. A meta-analysis of cronbach’s coefficient alpha. *Journal of consumer research*, 21(2): 381–391, 1994.
- [69] Rosalind W Picard and Roalind Picard. Affective computing. vol. 252. *MIT press Cambridge*. “EEG-detected olfactory imagery to reveal covert consciousness in minimally conscious state.” *Brain injury*, 29 (13-14):1729–1735, 1997.
- [70] David Premack and Guy Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1(4):515–526, 1978.
- [71] Liran (Windesheim Hogeschool) Prins. *Een onderzoek naar gesprekken tussen vrijwilligers en kinderen over zelfdoding bij De Kindertelefoon*. PhD thesis, 2019.
- [72] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.
- [73] Anand S Rao, Michael P Georgeff, et al. Bdi agents: From theory to practice. In *Icmas*, volume 95, pages 312–319, 1995.
- [74] Luis-Felipe Rodríguez and Félix Ramos. Computational models of emotions for autonomous agents: major challenges. *Artificial Intelligence Review*, 43(3):437–465, 2015.
- [75] Richard M Ryan and Edward L Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1):68, 2000.
- [76] Salim Salmi, Saskia Mérelle, Renske Gilissen, and Willem-Paul Brinkman. Content-based recommender support system for counselors in a suicide prevention chat helpline: Design and evaluation study. *Journal of medical internet research*, 23(1):e21690, 2021.
- [77] Shefaly Shorey, Emily Ang, John Yap, Esperanza Debby Ng, Siew Tiang Lau, and Chee Kong Chui. A virtual counseling application using artificial intelligence for communication skills training in nursing education: Development study. *Journal of Medical Internet Research*, 21(10), 2019. ISSN 14388871. doi: 10.2196/14658.

- [78] Shefaly Shorey, Emily Ang, John Yap, Esperanza Debby Ng, Siew Tiang Lau, and Chee Kong Chui. A virtual counseling application using artificial intelligence for communication skills training in nursing education: development study. *Journal of medical Internet research*, 21(10):e14658, 2019.
- [79] Trine Natasja Sindahl. *Chat Counselling for Children and Youth - A Handbook*. 2011.
- [80] Jeffrey Sirocki. Design and evaluation of a conversational agent model based on stance and bdi providing situated learning for triage-psychologists in the helpline of 113 suicide prevention. 2019.
- [81] Bas R Steunebrink, Mehdi Dastani, and John-Jules Ch Meyer. The occ model revisited. In *Proc. of the 4th Workshop on Emotion and Computing*. Association for the Advancement of Artificial Intelligence, 2009.
- [82] Stichting De Kindertelefoon. De Kindertelefoon Jaarverslag 2018. 2018. URL <https://www.lumc.nl/over-het-lumc/organisatie-en-missie/jaarverslag-2018/>.
- [83] Stichting De Kindertelefoon. De Kindertelefoon Jaarverslag 2019, 2019. URL <https://jaarverslag.kindertelefoon.nl/2019>. [Online; accessed 17-October-2021].
- [84] M Triola, H Feldman, AL Kalet, S Zabar, EK Kachur, C Gillespie, M Anderson, C Griesser, and M Lipkin. A randomized trial of teaching clinical skills using virtual and live standardized patients. *Journal of general internal medicine*, 21(5):424–429, 2006.
- [85] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S Kashavan, and John Blake Torous. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, 64(7):456–464, 2019.
- [86] Hans Van der Meij. Principles and heuristics for designing minimalist instruction. *Technical communication*, 42(2):243–261, 1995.
- [87] JM Van der Zwaan, V Dignum, and CM Jonker. A bdi dialogue agent for social support: Specification and evaluation method. In *AAMAS 2012: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems, Workshop on Emotional and Empathic Agents, Valencia, Spain, 4-8 June 2012; authors version*. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), 2012.
- [88] Viswanath Venkatesh, James YL Thong, and Xin Xu. Unified theory of acceptance and use of technology: A synthesis and the road ahead. *Journal of the association for Information Systems*, 17(5):328–376, 2016.
- [89] Camilla Von Below. *When psychotherapy does not help... and when it does: Lessons from young adults' experiences of psychoanalytic psychotherapy*. PhD thesis, Department of Psychology, Stockholm University, 2017.
- [90] 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, 1966.
- [91] D Wever, HJ Hermens, and MMR Vollenbroek-Hutten. Differences in use of a exercise-based tele-rehabilitation service delivered as substitute of or supplement to conventional care. In *Sixth International Symposium on E-Health Services and Technologies*, pages 44–51, 2012.
- [92] Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, and Rama Akkiraju. A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 3506–3510, 2017.
- [93] Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv preprint arXiv:1911.00536*, 2019.