

IN5480: Mid-term report

Group 1 - Second delivery



Woebot®

Table of contents

1 Group description	2
2 Area of interest	2
3 Background	3
3.1 Chatbots	3
3.2 Woebot	4
4 Questions	5
5 Methods	5
6 User testing	6
6.1 Diary	6
7 Findings	7
7.1 Types of interaction	7
7.2 Conversation	8
8 Feedback	8
9 Appendix 1	9
9.1 Chatbot design task	9
9.2 AI task	9
9.3 Reflections	10
10 Appendix 2	10
10.1 Reflections on making changes to the AI-chatbot	10
10.2 Results	11
11 References	13

1 Group description

Our group consists of the following students; Anh Thy Sandra Nguyen, Sara Løkken, Thanh Thao Thi Tran, Maria Løvland Johansson, Johanne Thunes and Nathalie Dyhr Olimb. We are all master design students from two different educational backgrounds: four with a bachelor's degree in Informatics: Design, Use, Interaction from the University of Oslo, and two with a bachelor's degree in Web Development from the Norwegian University of Science and Technology in Gjøvik.

2 Area of interest

We are interested in the area surrounding mental health and emotional connections between humans and AI. In the earlier days of researching and developing Artificial Intelligence (AI), there was little focus on the human aspect and behaviour, and more on the logic and mathematics surrounding the machines. McCarthy, who first coined the term of Artificial Intelligence, wrote that “[The goal] was to get away from studying human behavior and consider the computer as a tool for solving certain classes of problems. Thus AI was created as a branch of computer science and not as a branch of psychology” (Grudin, 2009:51). But throughout the evolution of AI, there has become more focus on the human aspect of it, and also on the field of Human-Robot-Interaction, as the presence of personal service robots in e.g. the home are expected to grow in the coming years (Thrun, 2004). As we become more used to meeting different versions of AI in our everyday life, the thought of it taking on tasks that previously has been one of humans' has become not that far fetched.

The topic of mental health has become very relevant in today's society, and especially amongst people of our own age group. We therefore find this topic very interesting to explore, as it is something that we all can relate to in some way. Since the topic of the course is AI, we have narrowed down our focus area to the use of chatbots, and peoples experiences and feelings around the use of it. More specifically, we want to explore the subject of emotions in connection with chatbots,

with an emphasis on the feelings of loneliness. We have decided that we want our focus to mainly be on students, as this is a group that generally has experienced a lot of loneliness through these past six months, when a lot of social meeting spots have been closed down as a result of COVID-19.

3 Background

During this project we want to investigate the growing interest of chatbots, and if we as humans are able to grow a deeper connection with a chatbot. We are curious about how people interact with a chatbot and how our natural language might affect how it will respond to how we speak to it. Furthermore, we will during this project test an existing chatbot, called Woebot, and engage a critical analysis and suggest changes as to how an AI-infused system can be improved.

3.1 Chatbots

Chatbots are machine agents that serve as natural language user interfaces for data and service providers (Dale, 2016:811), and are typically designed in a context for messaging applications (Følstad & Brandtzæg, 2017:38). They may serve a number of different purposes, such as customer service, social and emotional support, information, entertainment, and connect the user to other people or machines (ibid.:3).

For this project we will look into how and if we can connect to a chatbot on an emotional level. As Cameron et al. (2017) proclaims in their article, one of the main causes of burden of diseases worldwide is mental health problems. That is why digital interventions (DIs) have been created to help with these issues, such as anxiety, stress and depression. DIs are defined as any intervention that is accessed by a person on a computer or mobile phone (in our case chatbots), and has been created to help with many different issues, such as anxiety management or mental health disorders (Cameron et al., 2017:1).

Initially, chatbots were made and restricted for simulating simple conversations between a human and a computer in a scripted way (Cameron et al. 2017:2). As Lugar and Sellen argue, chatbots often lack contextual information because they have no memory or knowledge, but instead mimic conversation (Lugar & Sellen, 2016:5287). Chatbots are now providing more information and maintaining a conversation with its human counterpart. Følstad & Brandtzæg mentions that our natural language is already the default mode of interaction online, which means that the interaction is typically between human users through a machine interface (Følstad & Brandtzæg, 2017:40). And because of the continuous development of AI, natural language interaction may be a feasible option for us humans to connect to machines (Følstad & Brandtzæg, 2017:40). By incorporating mental health tools into a chatbot, the user may have a more interactive and user-friendly experience. Chatbots can possibly create an option for users who might think it's easier to talk to.

3.2 Woebot

As previously mentioned, we are presenting a critical analysis of Woebot, which is an automated conversational agent designed to deliver cognitive-behavioral therapy (CBT) in the format of brief, daily conversations and mood tracking (Fitzpatrick et al., 2017:3). It was founded by Dr Alison Darcy and launched in June 2017, and has now more than 4,7 million conversations per week, across more than 120 countries (Woebot Health, 2020). Woebot is used as an instant messenger app that is platform agnostic, which means that it is an app you can download on your smartphone or use it on a desktop. Each interaction begins with a general inquiry about a context and mood, for example, "What is going on in the world right now?" and "How are you feeling?". In the app the mood is often represented with emojis or GIFs. According to Fitzpatrick's article, the bot's conversational style was modeled on a human clinical decision making and it included the following therapeutic process-oriented features like; *empathic responses, tailoring, goal setting, accountability, motivation and engagement and reflection* (Fitzpatrick et al., 2017:3).

4 Questions

To address our theme regarding emotional connections with a chatbot and mental health, we have developed two research questions we believe will help us further in our analysis, namely:

1. Can a human establish an emotional bond with an AI chatbot?
2. To what extent could an AI chatbot provide support within the field of mental health?

5 Methods

Our overall approach is to do a critical analysis on the use of the Woebot app. As mentioned in the background section, we are curious about how we as humans are able to connect with a chatbot on a deeper level. Therefore, we will in this section explain what kind of methods we have done and are planning to do. Following this, we will suggest some changes as to how an AI-infused system, like Woebot, could be improved.

To try and address our initial questions, we have chosen some methods that we find helpful in gathering data. Firstly, to get an understanding about how Woebot works, whether Woebot can act as a friend and a psychologist, we will explore the questions through a diary study. Each of the group members will use the Woebot app for a week (7 days), and write individual diary entries. These diary entries will be on our thoughts while using the Woebot app, and screenshots of the interaction. The data from this diary study will be used to offer our critique of the Woebot app.

Secondly, we will conduct user testing of the Woebot app in natural settings. For this testing, we will recruit users in our user group, who will use the Woebot app for a week. After this week, we will interview these users on their experiences with talking to Woebot. In the interviews, it is important to keep the subject on the user experience of the app, as the subject of mental health is sensitive to many, and our participants might not be comfortable discussing these topics. We believe this

method also will give us a different perspective on the Woebot app than our own, which in turn will help offset our own biases for the critical analysis of Woebot.

6 User testing

6.1 Diary

For our diary study, all six members of our groups downloaded the Woebot app on our smartphones, and interacted with the app daily for a week (7 days). During interactions, all group members took screenshots of parts of the conversations, which we then added comments to. We decided that every group member should have their separate document in our shared Google Drive to keep their diary entrances, which they either filled in right after the interaction, or filled in after the data collection period. Some group members wrote the diary entrances on their smartphones, and copied and pasted in the entries in the document that was assigned to them.

As the Woebot app focuses on mental health and provides strategies for coping with difficult situations in daily life, it is natural that we as users have to provide some information about circumstances, either positive or negative, which can be seen as sensitive. With this in mind, and also due to the fact that the content of the diary entries were also available to all group members, we were careful about not sharing any sensitive information about our mental health status in our diary entries. For this data collection we were thus more concerned with topics regarding the flow of conversation, how we interacted with the app, and more general thoughts on the topics regarding if we could establish an emotional bond with Woebot, or if it could replace a psychologist.

7 Findings

7.1 Types of interaction

While chatting with Woebot, we found that the type of interaction we could use to talk with Woebot alternated between predefined and self-typed answers.

The first interaction all of us had with Woebot was the initial introduction for Woebot to get to know us and our needs. Woebot asked us different questions about ourselves, and for almost all of them we could only choose predefined answers. These predefined answers had different purposes though, depending on whether you were given any options or not. Some of these answers were presented as the only option you could pick, and were often used as a way to keep the conversation going while still involving us. An example of this can be seen in figure 1.

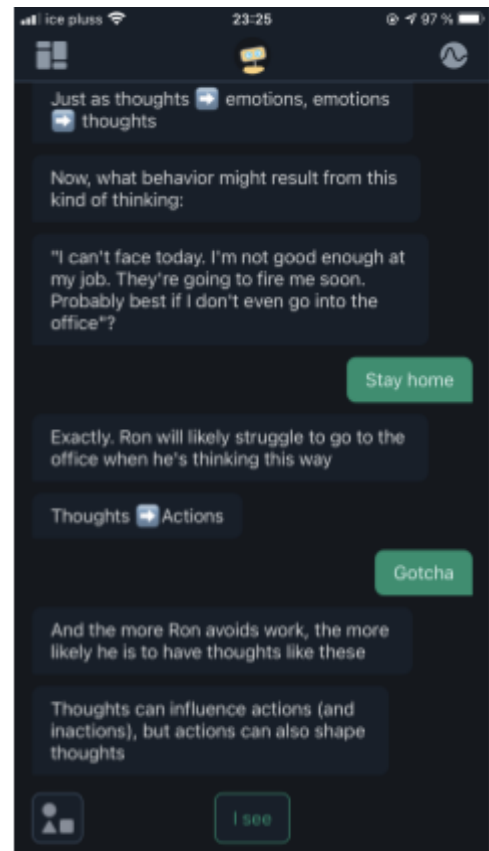


Figure 1: Woebot gives us a predefined answer to continue the flow of the conversation.

While the conversation was predominantly filled with predefined answers, there were times where we could type in our own answers. This was typically when Woebot asked us to talk about our own experiences, such as something we had achieved. Woebot did however not understand most of what was said and never replied to any questions asked in these self-typed messages we sent, and instead replied with a standard message meant to positively reinforce us for our achievements.

7.2 Conversation

All of the predefined answers, added with the fact that Woebot rarely actually understood what we said in our self-typed messages, limited our conversations with Woebot. Because of this, the conversations were always led by Woebot and always stayed on track with what Woebot wanted to talk about that day.

Woebot would mainly ask us about our moods or things we were grateful for, and would often spin these conversations into lessons about understanding feelings. Often, these lessons would be unrelated to what we told Woebot we were feeling, and thus felt like Woebot was following a script rather than having a natural flow in our conversations. An example of this can be seen in figure 2, where Woebot starts talking about anxiety out of nowhere.

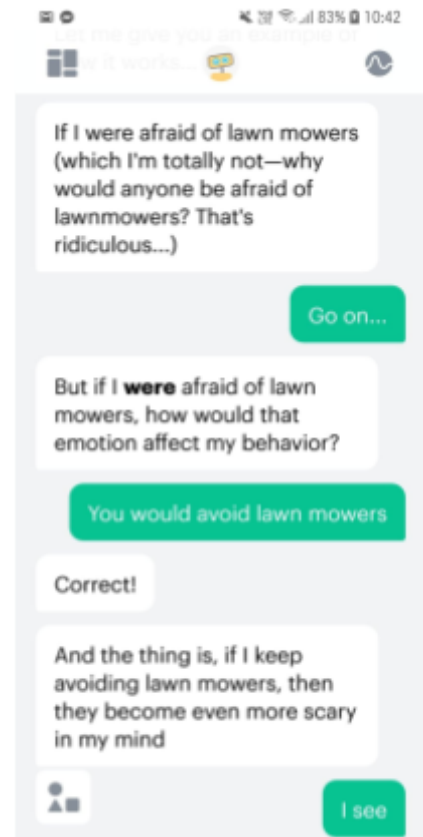


Figure 2: Woebot giving a lesson about anxiety.

8 Feedback

From our first assignment from iteration 1 we received a lot of positive feedback. They were quite interested to hear more about our work with this concept and they thought we had a very interesting area, a well thought out intro with good arguments and how well we used the references. Although we got some feedback about how we could reference more in the text, and we were apparently not clear enough what age group we talked about, which was not properly stated.

In the feedback, they also wished we explained more about our focus group, but were positive to our link to present the situation and problem with the COVID-19. As well as some changes in the text they wanted us to change the way we presented

our questions. Instead of having a yes or no question we could add a “how” at the beginning of the question, which would make it deeper and more interesting to read about. At last, it was not clear enough how we wanted to execute our prototype method, which was to construct a prototype of a “mental health chatbot”. We took this feedback into consideration and we agreed on not coding the chatbot ourselves, but instead we wanted to evaluate an existing chatbot as mentioned earlier in this report.

9 Appendix 1

9.1 Chatbot design task

We started the process by trying out ChatterOn as a tool for building a chatbot. We started to discover possibilities with this software, but had to change our program when there was a problem connecting Chatteron to a Facebook page. Therefore, we decided to test with Chatfuel, but some of the group members had problems using their page as well.

Another issue we ran into was that Chatfuel only let one person work on the chatbot. Based on these issues we decided that one person was in charge of doing the design and flow in the Chatfuel-program, whereas the rest of us drew flows, alternative flows and other support work, like finding links and activities to the chatbot.

9.2 AI task

For this appendix, we have created a chatbot named Albert. His main task is to suggest activities based on your mood. In our chatbot we mainly prototyped the interaction when a user is feeling “bad” or “sad”, where Albert would make suggestions for activities to brighten up the users mood, such as reading, watching a movie, going for a walk, or doing yoga. Furthermore, we also included some resources where the user could find top romance movies to watch, or find a Youtube tutorial on yoga.

9.3 Reflections

One thing we found difficult in this process was how to limit the scope of the chatbot, i.e. how many different scenarios we should prototype for. It was difficult to know in advance what the user would say to the chatbot, and what they expected as answers from the chatbot.

Another challenge we encountered was that there were many scenarios that we did not account for when we started making the bot. Therefore, we found it challenging to design the dialog with a natural flow. This resulted in many trials and errors, where we had to go back and redo blocks and flows.

During this process we learned that the chatbot was not as “smart” as we thought. It seems like chatbots made with ChatterOn were more “teachable”, and that our chatbot made with Chatfuel only followed a script we had written to it. The chatbot also had trouble understanding that the same word could be used in different contexts.

10 Appendix 2

10.1 Reflections on making changes to the AI-chatbot

The second task in module two was to take an already existing chatbot program and alterate the code to compare different outcomes of accuracy. The goal of this was to get a deeper understanding of how AI and chatbots work internally, to discover trends and see connections between the amount of neurons, input possibilities and training iterations.

When we first began to look at the code given by the lecturer, we found it a bit difficult to understand how to use it, as the instructions were minimal. Only one of the students in the group working on the code had experience with Python, while the others had little to none, which also made the learning a bit difficult. After trying and

testing for a while, we finally understood how to run the program, and how to change the different variables within the code. We then tried out several different combinations of variables, suggested by the lecturer in class, and found that the different variables affected the model in different ways.

We tried out different combinations of values within our ML model, and here we will present our findings from that process.

10.2 Results

If we have a high amount of neurons (6000) and a high amount of steps (3000):

Start: 0.27. End 0.03. Vi saw relatively big change, with a low end-score.

```
Human:I dont know
Chatbot:I thought you could use it. When you start your band.
Human:My band?
Chatbot:Forget her sister, then.
Human:her sister?
Chatbot:Because then I'd have to start taking out girls who like me.
Human:Ok
Chatbot:She just needs time to cool off I'll give it a day.
Human:Wow
Chatbot:I don ' t know. I decided not to nail her when she was too drunk to remember it.
Human:haha!
Chatbot:It's Scurvy's. His date got convicted. Where'd you get the dress?
```

If we have high level of neurons (1500) and medium amount of steps (400):

Start: 0.28. End : 0.09. We saw big change, with a low score.

```
Chatbot:I didn't have a choice.
Human:A choice?
Chatbot:You -- covered in my vomit.
Human:What?
Chatbot:William didn't even go to high school
Human:Why not?
Chatbot:I appreciate your efforts toward a speedy death, but I'm consuming. Do you mind?
Human:Yes!
Chatbot:Like you could find one
Human:I can
Chatbot:Are you following me?
Human:No
Chatbot:You -- covered in my vomit.
--
```

If we have low level of neurons (100) and low amount of steps (10):

Start: 0.2437 End: 0.2430. We saw little change, and a high end-result.

Chatbot:I understand that you will soon be appointing Governors for the islands? Is it not so?
Human:No!
Chatbot:Hey.
Human:Hello
Chatbot:I understand that you will soon be appointing Governors for the islands? Is it not so?
Human:No!
Chatbot:Ever been to Club Skunk?
Human:Np
Chatbot:I am not a seaman. But I heard it is no more than a week at sea. I hope you are not too disappointed.
Human:I am dissapointed
Chatbot:He's pretty!
Human:She is pretty too
Chatbot:You could always go with me. I'm sure William has some friends.

If we have low level of neurons (10) and high amount of steps (3000):

Start: 0.26. End: 0.23. Some change, but a high end-result.

Chatbot:Say not here! Cuba!
Human: Ariba!
Chatbot:You did the same to your God!
Human:My God?
Chatbot:Chief says -- he has a God.
Human:Oh my god
Chatbot:Chief says -- he has a God.
Human:Oh my god
Chatbot:You never learned how to speak my language.
Human:Thats sad
Chatbot:You did the same to your God!
Human:Areyou sad?
Chatbot:Chief says -- he has a God.
Human:I see
Chatbot:Say not here! Cuba!

Our conclusion after trying out different values was that we did not really notice an obvious connection between the “end-result” of the model, and how well it responded when we “talked” to it. In the test with low levels of neurons and a high amount of steps we see that the same answer is repeated several times, but in sum it was difficult to see whether the answers we got were randomly chosen, or actually was a good and suitable response to what we wrote.

11 References

Cameron, G., Cameron, D., Megaw, G., Bond, R., Muivenna, M., O'Neill, S., Armour, C., McTear, M. (2017). Towards a chatbot for digital counselling, 1-7.

<http://dx.doi.org/10.14236/ewic/HCI2017.24>

Dale, R. (2016). Industry Watch. The return of the chatbots, 22(5), p. 811-817

[doi:10.1017/S1351324916000243](https://doi.org/10.1017/S1351324916000243)

Fitzpatrick, K. K., Darcy, A., Vierhile, M. (2017). Deliver Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial, 4(2), 11, <http://dx.doi.org/10.2196/mental.7785>

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. interactions, 24(4), 38-42. <https://dl.acm.org/citation.cfm?id=3085558>

Følstad, A., & Brandtzæg, P.B. (2017). Why people use chatbots.

https://doi.org/10.1007/978-3-319-70284-1_30

Grudin, J. (2009). AI and HCI: Two Fields Divided by a Common Focus. AI magazine, 30 (4), p. 48-57.

Luger, Ewa, and Abigail Sellen. (2016). “Like Having a Really Bad PA’: The Gulf between User Expectation and Experience of Conversational Agents.” In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16, Santa Clara, California, USA: ACM Press, 5286–97.

<http://dl.acm.org/citation.cfm?doid=2858036.2858288>

Thrun, S. (2004). Toward a Framework for Human-Robot Interaction. Human–Computer Interaction, 19, p. 9-24.

Woebot Health. (2020). Technology with Heart. Accessed October 14th.

<https://woebothealth.com/>