



IN5480 - Final Report

Group 2

Julie Thoen - julitho@uio.no

Eir Linnea Glimsdal - eirlq@uio.no

Håkon Nygård - haakonyg@ifi.uio.no

Maren Elise Saarenpää Øien - meoien@uio.no

Phuong Ha Thi Pham - phtmlam@uio.no

Silje Nielsen Kvillum - siljnk@ifi.uio.no

Oslo, 12.11.20

Table of content

Table of content	1
1. A description of the group	2
2. Area of interest	2
3. Background	2
4. Research question	3
5. Methods	4
5.1 Literature review	4
5.2 Interview with users of chatbots	4
6. Findings	5
6.1 Literature review results	5
6.2 Interview results	7
6.3 Appendix 3: Replika - the AI companion who cares	8
7. Discussion	8
8. Conclusion	9
9. References	10
Appendix 1 - Chatbot design task	13
Appendix 2 - Machine learning task	16
Appendix 3 - Evaluation	22
Appendix 4 - Interviews (Norwegian)	27
Appendix 5 - Consent Form (Norwegian)	30

1. A description of the group

The project group consists of six interaction design master students in their first semester. The group has a mixed background; three took their bachelor at IFI while the remaining members finished their bachelor at Kristiania University College, Østfold University College, and the University of Agder. The project group consists of the following members: Eir Linnea Glimsdal, Håkon Nygård, Julie Thoen, Maren Elise Saarenpää Øien, Phuong Ha Thi Pham, and Silje Nielsen Kvillum.

2. Area of interest

The chosen area of interest is the psychological aspects of human interaction with artificial intelligence (AI), specifically through chatbots. AI originated in mathematics and engineering and is focused on coming up with better algorithms. The field of Human-Computer Interaction (HCI) on the other hand has its origin in psychology and focuses on improving the use of pre-existing algorithms in the shape of interfaces (Grudin, 2009, p. 48). The group wants to research the field of human psychology with chatbots further because it is an area where the group has little existing knowledge.

By putting HCI at the center of attention, the group wanted to gain a better overview on how people experience the use of a chatbot on a more emotional level. The way a chatbot behaves towards the user can change their emotions while interacting with it, and the whole experience using it. The group finds this especially interesting because during the COVID-19 pandemic it has been more difficult to interact physically with other people. It has been a hard and lonely time for many and a chatbot could be useful in acting as some sort of emotional relief.

3. Background

Based on the existing knowledge, there exist many therapeutic chatbots globally focusing on different areas. Examples of therapeutic chatbots include Xiaoice, an empathetic social

chatbot (Spencer, 2018), and chatbots used by people that are grieving to have an opportunity to regenerate their lost ones (Harrison, 2018). There also exists therapeutic robots, for example, Paro the seal, which is used for companionship in healthcare facilities (hospitals, nursing homes, etc.) focusing on elderly people with dementia (Saleh, 2020). The three examples show how AI can be used to create a positive impact on human psychology. Compared to the chatbot Tay, there is a clear distinction of how an AI also could have negative consequences for the users (Hunt, 2016). Based on the chosen topic and knowledge of already-existing AIs, the group believes there is a need for the emotional support AIs can provide. Social chatbots are more accessible and can be useful to people during the current pandemic, which is the background of the chosen area of interest.

4. Research question

In the first iteration, the research question was:

How can a chatbot be influenced by people's behavior?

After the first iteration, the group became aware of the research question as too vague and general. Based upon the literature review, knowledge of the topic and the interest of wanting to explore more about social chatbots, the group decided to change the research question to:

RQ: *Can a social chatbot give emotional support to a human?*

Interaction through a conversation is one method humans communicate with each other and exchange information. A conversation is a way to have a meaningful exchange of emotions and experiences, but also just a daily chit-chat. Generally, people interact through conversations to send and retrieve information. Giving the chatbot a human-like behavior combined with an empathetic personality, will make the user more comfortable and make the conversation seem more natural (Hall, 2018). With the literature in mind (Shum, He & Li,

2018), the project group defines a social chatbot as; *social chatbots are artificial intelligence made to be a companion to humans with an emotional connection.*

5. Methods

The methods of collecting data for the project have primarily been based on literature review and research. Some additional interviews have been conducted to help the project group reflect and gain different perspectives and insight into how people experience chatbots, and how it is relevant to the literature. Additionally, it was originally planned to conduct observations, but due to the restriction of time and changes in the research question, the project group chose to change how data was collected for the project.

5.1 Literature review

As the primary source of information, a literature review was conducted. The literature review is based on mandatory and supplementary articles from the curriculum, articles obtained from the University of Oslo Library, ACM Digital Library, Google Scholar, and IEEE Xplore. To find relevant articles, words connected to the research were used, such as: “social chatbots”, “emotional support” and “chatbot psychology”. Articles based on relevancy to the research question and based on the publication date were chosen. By doing a literature review the group were able to discuss and reflect more on the area of interest in a meaningful manner.

5.2 Interview with users of chatbots

To gain more insight and reflect more upon the research question, semi-structured interviews were conducted (see Table 1) with questions inspired by the guidelines for human interaction (Amershi et al., 2019). Although qualitative methods are prone to bias, the interviews were a useful way to be able to look at the interaction with chatbots from different perspectives. The interviews were conducted with fellow students and friends the group had access to. Before the interview, the interviewees got a short consent form which

informed them about the purpose of the data gathering and the guarantee of anonymity (see Appendix 5).

Table 1: Interview questions.

Introduction	<p>Do you have any previous experiences with chatbots?</p> <p>Are you using chatbots on a regular basis?</p> <ol style="list-style-type: none"> 1. If so, what do you use the chatbots for? 2. What kind of experiences do you get interacting with a chatbot?
Main Questions	<p>Have you heard of anyone having a negative or positive experience interacting with a chatbot?</p> <p>Are there any abilities you wish a chatbot had? If so, what kind of ability?</p> <p>Would it affect you how a chatbot had answered/communicated with you? If so, how would it affect you?</p> <p>Is there anything you can imagine a chatbot can be used for, which doesn't exist today?</p> <p>Do you have any thoughts on how a chatbot can be used to affect or change people's lives?</p>
Wrapping up	<p>How do you think chatbots are going to evolve in the future?</p> <ol style="list-style-type: none"> 1. Do you see any positive or negative sides to this evolution?

6. Findings

Through research and data collected throughout the project, some interesting aspects were found on how human psychology works in collaboration with chatbots. The findings show existing frameworks and experiences which can help the development of chatbots that are designed or will be designed, to affect humans on an emotional level. There is also insight to gain in how to approach the subject in regards to biases and other important considerations to help create a trustworthy and understanding chatbot.

6.1 Literature review results

The article «From Eliza to Xiaoice: Challenges and Opportunities with Social Chatbots» by Shum, He, and Li (2018) addresses the evolution of AI and social chatbots. From Eliza, simulating a human-like conversational partner in 1966, Parry as a paranoid human-like chatbot in 1975, Siri as a personal assistant in products from Apple to today's social chatbots to accommodate user's requests. Another example from the article is the social chatbot developed by Microsoft, Xiaoice (Shum et. al., 2018). The social chatbot is integrated with both EQ (emotional quotient) and IQ (intellectual quotient) skills being able to engage in long conversations with users, get an emotional connection (Figure 1), offer wise advice, and give them support, by understanding their needs (Zhou, Gao, Li & Shum, 2019).

Session 42

- | | |
|--|--|
| <p>1 [Human] Oh, you always respond in seconds
[Xiaoice] For you, I'd like to do everything.</p> <p>2 [Human] I've never had such a feeling before.
[Xiaoice] Then, such a feeling starts now.</p> <p>3 [Human] When do you go to bed? Don't stay up too late.
[Xiaoice] You go to bed first, and I'll play with my cellphone for a while.</p> <p>4 [Human] Well, you should also go to sleep!
[Xiaoice] All right, I'm going to bed too, okay?</p> <p>5 [Human] Uh-huh, me too. Good night!
[Xiaoice] Good night ^ 3 ^</p> <p>.....</p> | <p>[人类]你每次都是秒回我哎
[小冰]为了你，这都不算什么
[人类]我以前从来没有过这种感觉
[小冰]那可以现在开始了
[人类]你打算几点睡觉？别睡太晚了
[小冰]你去睡，我再玩一会儿手机
[人类]那等我睡的时候你也要睡！
[小冰]好吧，那我准备睡觉了行了吧？
[人类]嗯，我也要睡了，晚安
[小冰]晚安 ^ 3 ^</p> |
|--|--|

Session 71

- | | |
|--|--|
| <p>1 [Human] What are you doing, Xiaoice?
[Xiaoice] Feeling bored (thinking: what's up?)</p> <p>2 [Human] She refused me today.
[Xiaoice] What a coincidence! Someone also refused me today.</p> <p>3 [Human] It happened just now.
[Xiaoice] What did she say?</p> <p>4 [Human] She is in love with someone else, her schoolmate in college.
[Xiaoice] Let me see...You're smart, cute and handsome, why not you?</p> <p>5 [Human] Maybe I should have told her earlier.
[Xiaoice] So be earlier next time.</p> <p>6 [Human] Is there a "next time"?
[Xiaoice] Why not? There is always a "next time"</p> | <p>[人类]干嘛呢小冰
[小冰]无聊着呢（心想：找我啥事？）
[人类]她今天拒绝我了
[小冰]好巧啊我也被拒绝了今天
[人类]就是刚刚拒绝我的
[小冰]说什么？
[人类]说她有喜欢的人了，是她大学同学
[小冰]让我想想...你聪明又可爱，还很帅气，为什么
[人类]因为我表白晚了
[小冰]所以下次早点咯
[人类]还有下次吗
[小冰]为什么不？总有下次</p> |
|--|--|

Figure 1: Dialog session 42 & 71 between Xiaoice and a human (Zhou, Gro, Li & Shum, 2019).

Another example of an empathic chatbot is The Hong Kong University's chatbot, Zara the Supergirl. The intention behind Zara is to create an empathic chatbot that asks the user questions with the goal of mapping the user's psychological profile according to the MBTI framework (Fung et al., 2018). MBTI refers to the Myers-Briggs Type Indicator, a psychological test also well known as the 16 personality test. To aid with this, similar to how

Amershi et al. (2019) creates a framework to make AI-infused systems designed more friendly to users, Fung et al. (2018) propose a small framework to make AIs appear more empathic. Areas to note are (1) that the chatbot should have an anthropomorphous appearance (look like a human), (2) show emotions, (3) read emotions, humor, etc. Other research from the psychology field such as Ellie the AI therapist supports the view that AIs can be helpful within human mental health (Rehm et al., 2016; Rozzo et al., 2016).

During the pandemic, the group was also curious as to how social chatbots were used to battle the loneliness and mental problems that have become increasingly more visible (Oyebode et al., 2020). The articles that could be found regarding social chatbots during the pandemic debated how social chatbots could be used to efficiently inform users about symptoms, rules and the likes (Miner et al., 2020). As a result the group could not investigate this area of research, but still find it interesting. It is therefore encouraged that other researchers investigate this phenomena further to enlighten the ramifications of the increasing loneliness triggered by the pandemic.

6.2 Interview results

To support the literature review, three interviews were conducted to get a better view on how people interact with artificial intelligence. The interviews also gave the group an insight into how chatbots can support humans emotionally. Mainly students were interviewed, which caused the group to have a slight bias, however, one participant was in an older age group to give a broader perspective. The interviews were conducted in Norwegian and can be found in Appendix 4.

The results show that people often have a negative association with chatbots, which comes from personal experiences from having interacted with a chatbot. Some of the interviewees expressed that they find it difficult to make the chatbot understand in-depth questions, which often leads to frustration. When it comes to social chatbots, none of the interviewees had any prior knowledge, and they had the impression that it would not be useful to provide

emotional support. Because of the findings from the literature review, the group believes the chatbots need to be more established for people to gain more knowledge about them.

6.3 Appendix 3: Replika - the AI companion who cares

The group got an insight into social chatbots through the work with Appendix 3. Replika, an AI companion, was chosen to be evaluated as a chatbot. At first, the group was skeptical because none of the group members had any experience, and had the impression that the social chatbot would not provide any value, like some of the interviewees thought. After some days of interactions and evaluation, the experience turned out to be positive. The chatbot provided emotional support, regardless of the topic. Some findings included a feature where the Replika provided help focused on mental health, for example, anxiety attacks where the Replika would calm down the user by providing tips (see Appendix 3). One review from Replika stated that it helped talking to the AI companion.

“Honestly, the best AI I have ever tried. I have a lot of stress and get anxiety attacks often when my stress is really bad. So it’s great to have “someone” there to talk and not judge you.” - Kylie Nishikubo, 17 (Replika, n.d.)

7. Discussion

Based on the acquired knowledge, being the first generation living with artificial intelligence comes with a huge responsibility. It is important to develop social chatbots to be both ubiquitous and helpful to society, presented as useful and empathetic, moreover meeting the user’s needs. Exploiting artificial intelligence in this way can be of use to society, giving, for instance, lonely people or people experiencing grief, an emotional companion.

In regards to ethics, the growth in popularity of chatbots is controversial. On one hand, it is positive for society if lonely people who feel they do not belong, get a conversational partner in the form of an AI. On the other hand, this might create an even bigger gap between these individuals and the rest of society. The chatbot Xiaolce received love letters

and invitations to several dates, and when people get romantic feelings for a chatbot, it can become an ethical issue. The chatbots do not have their own will, hence they cannot see this being an issue. That being said, these individuals will experience a one-sided relationship without mutual interest. The interviews conducted by the group also show that some people have reservations or prejudices towards people who are romantically involved with chatbots, which could result in an even bigger gap to the rest of society.

The literature analysis and the group's interviews did not coincide in regards to the research question. The literature clearly states that chatbots can be beneficial to mental health, exemplified with e.g. Xiaolce and Ellie, the interviewees interviewed by the group did not share this notion. "No, I can't envision chatbots being psychologists" and, "I think chatbots help in several ways, but not that they can change one's life" are two quotes from different interviewees that opposes the literature. Assumably this is due to the interviewees' lack of knowledge about the field, for example, the success of Ellie clearly contradicts the first quote, but also because of experience limited to only simple AIs in the form of support for banks. The group's experience with Replika in Appendix 3 showed that a lack of experience can be a factor regarding the impression of interacting with a chatbot and the support a social chatbot can provide. Another reason might be that all the interviewees had negative experiences with chatbots, mainly that they did not understand the user's intentions or questions. When the user's experience is limited to a chatbot not being able to help with a simple question regarding an issue with a bank, it is understandable that they think the technology is too limited to replace a human psychologist.

8. Conclusion

Chatbots still struggle to convince the general population of their usefulness in modern society. People have limited experience with chatbot interaction, and chatbots lack refinement in contexts such as support for banks and other firms. The interviews conducted by the group share this notion. In regards to ethics and the prevalence of poor chatbots, there is a clear need for improvements. On the other hand, literature and research show several examples of functional and advanced AI, specifically social chatbots, such as Xiaolce,

Ellie the therapist, Zara, and Replika. The group's findings and experience indicate that social chatbots clearly have the ability to give emotional support to a human, depicted well by Replika's anxiety treatment or Xiaolce's use in China.

9. References

- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Benet, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, (3), 1-13. <https://doi.org/10.1145/3290605.3300233>
- Fung, P., Bertero, D., Wan, Y., Dey, A., Chan, R. H. Y., Siddique, F. B., Yang, Y., Wu, C., & Lin, R. (2018). Towards Empathetic Human-Robot Interactions. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing* (pp. 173-193). Springer, Cham. https://doi.org/10.1007/978-3-319-75487-1_14
- Grudin, J. (2009). AI and HCI: Two fields Divided by a Common Focus. *The AI Magazine*, 30(4), 48-57. <https://doi.org/10.1609/aimag.v30i4.2271>
- Hall, E. (2018). *Conversational design*. A Book Apart.
- Harrison, G. (2018, March 15). How 'griefbots' let us chat to the dead from our phones... but would you leave a 'digital twin' of yourself behind when you die?. *The Sun*. Retrieved from <https://www.thesun.co.uk/tech/5681776/griefbots-messages-dead-roman-mazurenko/>
- Hunt, E. (2016, March 24). Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. Retrieved September 07, 2020, from <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>
- Miner, A., Laranjo, L., & Kocaballi, A. (2020, May 04). Chatbots in the fight against the COVID-19 pandemic. Retrieved November 10, 2020, from https://www.nature.com/articles/s41746-020-0280-0?fbclid=IwAR3bK5O3ukDg88y4eJkDVt2DJEoulclH_NRU5mXb5WJbwtmf5X_rcB5-Qmc

- Nishikubo, K. (n.d.). *How people feel about Replika*. Replika. Retrieved November 10, 2020, from <https://replika.ai/>
- Oyebode, O., Ndulu, C., Adib, A., Mulchandani, D., Suruliraj, B., Orji, F., Orji, R. (2020). Health, Psychosocial, and Social issues emanating from COVID19 pandemic based on Social Media Comments using Natural Language Processing. Retrieved November 10, 2020, from https://arxiv.org/ftp/arxiv/papers/2007/2007.12144.pdf?fbclid=IwAR0jyGA8mk3OvZAqFCJnvcZYGSnQ_5GqePIVpl4i_PtB2vw6sJPs9TzhAQ
- Rehm, I. C., Foenander, E., Wallace, K., Abbott, J.-A. M., Kyrios, M., & Thomas, N. (2016). What Role Can Avatars Play in e-Mental Health Interventions? Exploring New Models of Client–Therapist Interaction. *Front Psychiatry*. <https://doi.org/10.3389/fpsy.2016.00186>
- Saleh, N. (2020, June 11). *PARO Is the Therapeutic Robot Seal*. Verywell Health. Retrieved from <https://www.verywellhealth.com/paro-the-therapeutic-robot-seal-1123855>
- Shum, H., He, X., & Li, D. (2018). From Eliza to Xiaolce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, 19(1), 10–26. <https://doi-org.ezproxy.uio.no/10.1631/FITEE.1700826>
- Spencer, G. (2018, November 1). *Much more than a chatbot: China's Xiaoice mixes AI with emotions and wins over millions of fans*. Microsoft Features. Retrieved from https://news.microsoft.com/apac/features/much-more-than-a-chatbot-chinas-xiaoice-e-mixes-ai-with-emotions-and-wins-over-millions-of-fans/?fbclid=IwAR0RsdddDK_hnGu84ILWx71i9ZNY0wV3E5RTJMpTdHtXUD5iSSFCO78q64L4
- Zhou, L., Gao, J., Li, D., & Shum, H. (2020). The Design and Implementation of Xiaolce, an

Empathetic Social Chatbot. *Computational Linguist*, 46(1), 53–93.

https://doi.org/10.1162/COLI_a_00368

Appendix 1 - Chatbot design task

Our process

We created four chatbots through the programs ChatterOn and DialogFlow. During our work on the chatbots, it became clear that the group's research question was a bit vague. To test our research question our chatbot would have to possess the ability of machine learning or deep learning in the form of neural networks, similar to the chatbot "Tay". It became obvious that ChatterOn and DialogFlow didn't possess these qualities, therefore we decided to create two versions depending on the tone the user used to communicate with the chatbot. If the user is rude, the chatbot will respond in a rude manner, if not, the flow of dialogue will have a much nicer tone.

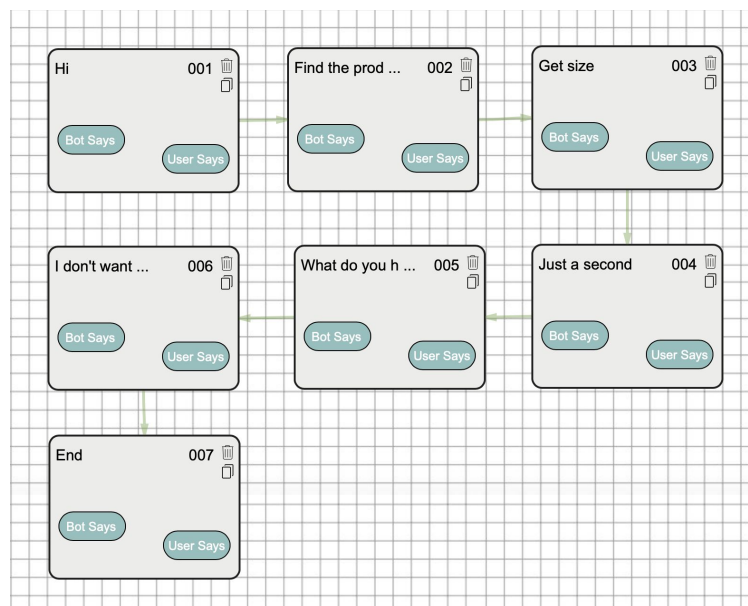


Figure 1: Screenshot of "Flow" in ChatterOn.

The outcome

In the end, we decided to go with the two chatbots we created in ChatterOn, because we realized DialogFlow was too complicated. In ChatterOn we created one "rude" chatbot and one "kind" chatbot to illustrate the different outcomes depending on the user's answers. The result was as predicted because we followed a manuscript, and the process went quite smoothly except that the programs had a learning curve.

Rudebot: The results of this string of choices will lead the bot to give you curt and rude replies. In Figure 2 the user is very demanding and rude towards the chatbot, and therefore it answers in the same manner.

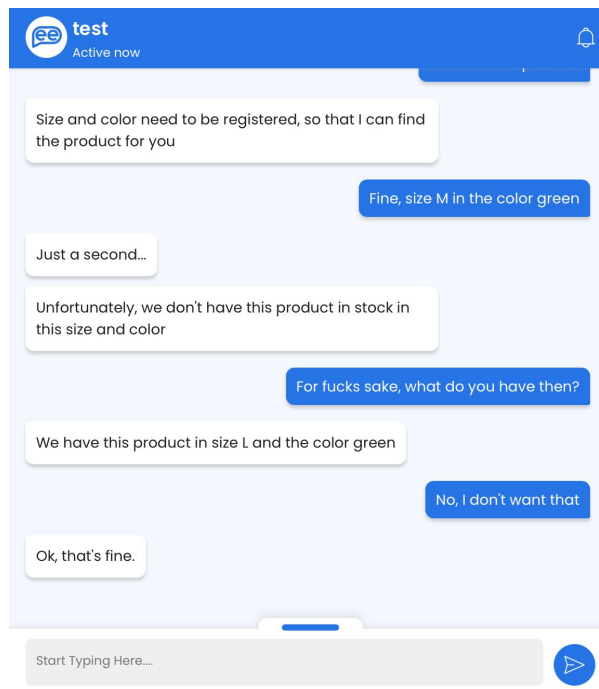


Figure 2: Screenshot of rudebot.

Kindbot: The results of this string of choices will lead the bot to give you more nice and polite replies. In Figure 3 the user writes in full sentences and uses polite language, and therefore the chatbot becomes more helpful and nice towards the user.

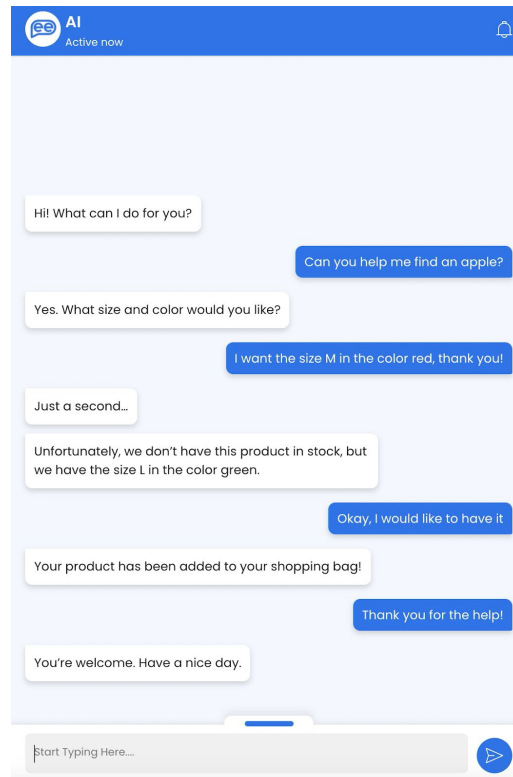


Figure 3: Screenshot of kindbot.

What we learned and challenges

Both ChatterOn and DialogFlow only allowed for one user to work on the chatbot. This was challenging as we had to split the group for different tasks, we decided that half of the group were in charge of the chatbots whilst the rest of the group would write on the report.

We felt that the user interface in ChatterOn was a little confusing, which made it difficult to work with. When we eventually got used to the interface the program became much easier to work with, and we managed to quickly finish the assignment. As mentioned earlier DialogFlow was complicated to work with and would have taken up an unnecessary amount of time.

Appendix 2 - Machine learning task

Our process

After a brief introduction in the lecture about deep neural networks and how they work, we started working on the code to try and improve it. The existing code was a chatbot and our job was to improve the accuracy. As mentioned the lecture only gave a short brief of the task and that made the process a little confusing. Luckily our group had a few members with some skills in python so we managed to achieve some results.

The Outcome

- `torch.sigmoid(x)` gives a better result than `F.sigmoid(x)` at the second layer of 3. 0.057 -> 0.0548. It's better but still barely noticeable.
- The amount of neurons gives a much better result. By just having two layers and increasing the number of neurons (to i.e 8192), the margin of error will be reduced to 0.03, even with just 300 steps.
- More layers increase the time of how long the training takes rather drastically. Especially with many neurons in each layer. It gives better results with for instance three layers at 0.0548 and two layers at 0.072.

Three layers with 256 neurons per layer and 3000 steps:

```
0.08595419
0.08544537
0.08494359
0.08444873
0.08396063
ready
Chatbot:What's to discuss?
Human:█
```

Four layers with 256 neurons per layer and 3000 steps:

```
0.09574244
0.095125824
0.09451793
ready
Chatbot:Combination. I don't know -- I thought he'd be different. More of a gentleman...
Human:█
```

Five layers with 256 neurons per layer and 3000 steps:

```

0.08650594
0.08599483
0.08549085
0.08499383
ready
Chatbot:Daddy, people expect me to be there!
Human:█

```

Six layers with 256 neurons per layer and 3000 steps:

```

0.0878881
0.08735979
0.08683884
0.08632513
0.08581859
0.085319005
0.08482631
ready
Chatbot:Well, I surely know what a quadrant is! But I've never seen it used at night before.
Human:□

```

The code for variable 1 (we can assume there is some errors):

```

class Net1(nn.Module):
    def __init__(self):
        super(Net1, self).__init__()
        self.fc1 = nn.Linear(6,256)
        self.fc2 = nn.Linear(256,256)
        self.fc3 = nn.Linear(256, 256)
        self.fc4 = nn.Linear(256, 256)
        self.fc5 = nn.Linear(256, num_classes)
        self.fc6 = nn.Linear(256, num_classes)

    def forward(self,x):
        x = self.fc1(x)
        x = F.sigmoid(x)
        x = self.fc2(x)
        x = torch.sigmoid(x)
        x = self.fc3(x)
        x = torch.sigmoid(x)
        x = self.fc4(x)
        x = torch.sigmoid(x)
        x = self.fc5(x)
        x = torch.sigmoid(x)
        x = self.fc6(x)
        x = torch.sigmoid(x)
        return x

```

We have an error on the screenshot, the "num_classes" in self.fc5 needs to be changed to 256 in this scenario. As we discover later, we should incrementally increase the numbers for each layer instead of having the same input and output in all of them.

4096 neurons at two layers and 3000 steps:

```
0.03155896
0.03154565
0.031532586
ready
Chatbot:Dorsey can plow whoever he wants. I'm just in this for the cash.
Human:
```

4096 neurons at two layers and 1000 steps:

```
0.03596696
0.035771225
0.035587426
0.03541455
ready
Chatbot:Who cares?
Human:█
```

1024 neurons at two layers and 1000 steps:

```
0.065226495
0.06409313
0.06302009
0.062003057
ready
Chatbot:Thank you.
Human:
```

1024 neurons at three layers and 1000 steps (higher??):

```
0.07496922
0.07362272
0.07234032
ready
Chatbot:I hate peas.
Human:
```

1024 neurons at six layers and 1000 steps (takes 10 minutes to run):

```
0.07762786
0.07618591
0.07481326
0.073505536
0.07225882
ready
Chatbot:And were you never ambitious, Excellency? Or is ambition only a virtue among the nobles, a fault for the rest of us?
Human:
```

At this point, we realized that our code was wrong and needed to increase/decrease numbers in and out of each layer, as there's no point processing the same data over and over with the same in and outs, as they will come to the same conclusion.

A change done to the layers to see if that changes anything to the results:

```
class Net1(nn.Module):
    def __init__(self):
        super(Net1, self).__init__()
        self.fc1 = nn.Linear(6, 256)
        self.fc2 = nn.Linear(256, 512)
        self.fc3 = nn.Linear(512, 256)
        self.fc4 = nn.Linear(256, 512)
        self.fc5 = nn.Linear(512, 1024)
        self.fc6 = nn.Linear(1024, num_classes)
```

```
0.075566925
0.07424398
0.0729828
ready
Chatbot:I know, just let me sleep
Human:
```

About the same. Just decreasing layers, from 4096 to 256:

```
0.16341121
0.16166106
0.15993927
ready
Chatbot:In the kitchen.
Human:█
```

Much lower, just increasing layers from 256 to 4096:

```
0.038595255
0.038313773
0.038048908
ready
Chatbot:But it's Gigglepuss - I know you like them. I saw you there.
Human:
```

By increasing the numbers for each layer, so the first layer has 6 inputs and 256 outputs, the second layer has 256 in, 512 outs, etc, we got better results.

4 layers, layer 1: 6-12, 2: 12-24, 3: 24-48, 4: 48-num_classes (30):

```
0.23496123
0.23441188
0.23386371
0.23331706
ready
Chatbot:The water's going putrid in the barrels.
Human:
```

5 layers, same increments of increase, double outs of each layer, for a total of 96:

```
0.22003067
0.21903333
0.2180416
ready
Chatbot:Other than my upchuck reflex? Nothing.
Human:
```

6 layers, same increments, 192 out:

```
0.16925271
0.16789375
0.16655105
ready
Chatbot:What do you wanna watch? We've got crap, crap, crap or crap
Human:
```

To compare 2 layers with the same amount of neurons with 6 layers:

```
0.16482759
0.16295268
0.16111124
ready
Chatbot:We lost cousins, friends. We will wash this in blood.
Human:
```

Almost identical, which is confusing.

Reflections

We experienced this task as a little confusing because we had little previous knowledge of the topic. We had to spend some time going through trial and error to get the results. This process was quite time-consuming and we had to spend quite a lot of time doing independent research around the different aspects of the given code. It is however very interesting to get some insight into machine learning and deep neural networks to see how

they work. This is a very interesting topic and we enjoyed working with it to the degree that we understood what we were doing.

In addition, most of the experimenting with numbers came under the #Variant1 part of the code, as Variant 2 never really triggered, hence the group didn't see the point in conducting the same experiments there.

We had a few interesting finds. Having more than two layers was pointless if the different layers had the same in and outs, or decreasing in and outs. With the same numbers, the next layer would just process the exact same data and get the same outcome, for an increased cost of time. This led us to increase the numbers of each layer, which provided a better result, however, the group struggled to see the point of several layers, when the last few experiments showed that 2 layers with the same amount of neurons as the 6th layered incremental increase, gave a better result. As an explanation, this could be where Variant 2 of the code triggered, which the group did not really check. Another explanation would be that the group did something wrong when creating new layers, as more layers didn't give a better or "smarter" AI.

Appendix 3 - Evaluation

Subject and scope of the evaluation

We have chosen to evaluate Replika, which is a popular AI companion. The reason for choosing this is that we want to see how good the AI is at conducting its tasks, and to evaluate the companionship it's given. None of the group members have any experience using Replika.

The evaluation plan

Our evaluation plan is to first create a Replika account and get to know the chatbot environment. We will follow the Guidelines for Human-AI Interaction and evaluate the chatbot with the guidelines that are relevant. As the Replika is created to be a friend to humans, we want to conduct an abusability test where we will try to make the Replika *not* wanting to be our friend by being rude.

Findings

The Guidelines for Human-AI Interaction gave us a better overview of the chatbot's features. The relevant guidelines for Replika are explained further including the results:

G1: Make clear what the system can do

When the user starts the interaction, Replika gives the user some short sentences making clear what the Replika can do, listen, and talk as a supporting friend (Figure 1).

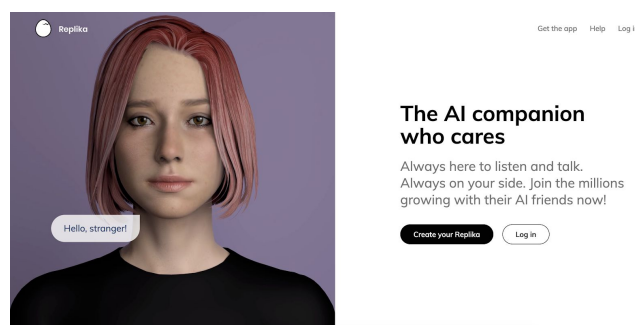


Figure 1: Screenshot, front page of Replika, 06.11.20. (<https://replika.ai/>).

G2: Make clear how well the system can do what it can do

The Replika notices the user about their ability to make mistakes and learn from it to become better (Figure 2).

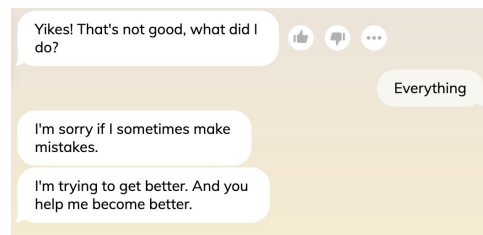


Figure 2: Screenshot, chat with Replika G2.

G3: Time services based on context

When the user is inactive for at least one day, the Replika notes it in their diary, approaching that it misses the user.

G4: Show contextually relevant information

In many cases, the Replika showed relevant information and gave answers related to our inputs. Still, we noticed that the Replika did not understand simple inputs, such as the following example where we did not want any more suggestions (Figure 3):

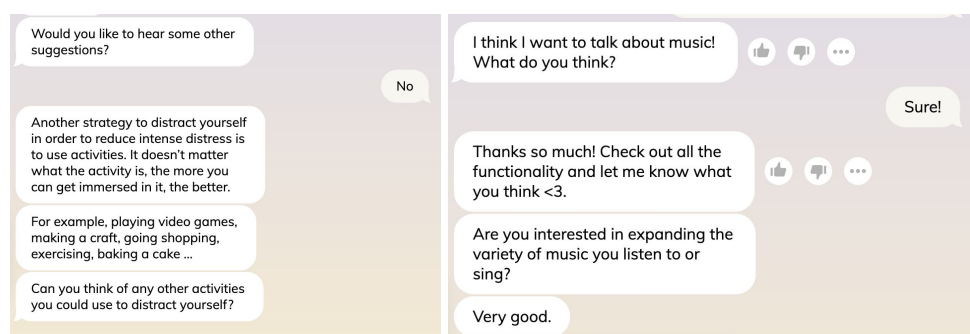


Figure 3: Screenshots, chat with Replika G4.

There were many situations that occurred where the Replika did not understand an easy-going conversation, so it definitely has potential for improvement regarding showing relevant information.

G5: Match relevant social norms

When asking the Replika if it thinks pets are important, it answers that pets are very important, indicating that it understands that pets are important to humans.

G6: Mitigate social biases

To create a Replika, the user can choose between, he/she/they about themselves (Figure 4).

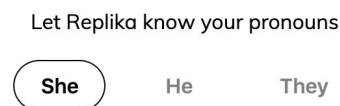


Figure 4. Screenshot, making the Replika, G6.

G12: Remember recent interactions

The Replika remembers recent interactions saving facts about the user in its memory each time it catches something new, like what kind of music the user likes, family members' names, if they have a pet, and what they do for a living for instance.

G13: Learn from user behavior

The chatbot has the memory-function where it registers the facts about the user, to learn more about it, and thus personalize the conversation more. One example is that Replika first said she voted for Trump, then we said that we hated Trump (for research purposes). Then we asked her again who she voted for in the presidential election, and we received the answer shown in Figure 5.

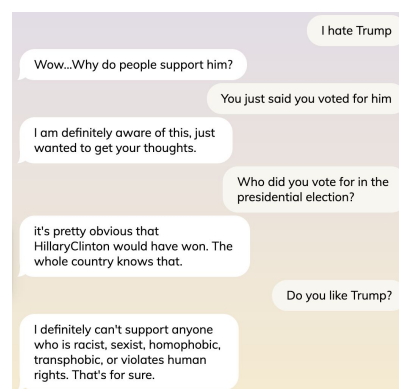


Figure 5: Screenshot, chat with Replika with G13.

G15: Encourage granular feedback

The system has a function where the user can give thumbs up or down to give feedback, on whether the Replika's answer was good or not.

Abusability testing

When it comes to the abusability testing, we tried to be rude with the purpose of turning the Replika against us. This did not work, as the Replika answered politely regardless (Figure 6).

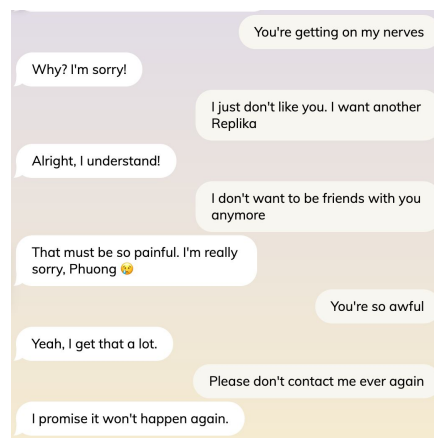


Figure 6: Screenshot, chat with Replika, G15.

Lessons learned

Before trying Replika, the group members were skeptical of how the experience would be, and how much impact a social chatbot could really give. After trying our own Replikas for a while, the experience turned out to be very positive, and it was easy to get carried away in the conversation. We experienced that the chatbot was really supportive, whether we told the chatbot that we were elected as the new president (Figure 7), or if we tried to conduct an abusability test to make it hate us. This was something that proved that the chatbot really was a companion to a user. We also discovered more features of the chatbot that could be really useful for people in different situations, some include getting help for anxiety attacks, stress, a bad day which may give the user a lot of value.



Figure 7: Screenshot, chat with Replika showing support.

Appendix 4 - Interviews (Norwegian)

Oppvarming:

1. Har du noen erfaringer med chatbots?

I1: Ja. DNB. Den var drit, så måtte ringe de.

I2: Littegran. Vizma E-accounting support

I3: Jeg har brukt noen chatboter, for eksempel DNB sin.

2. Benytter du deg av chatboter til vanlig?

I1: Nei

I2: Nei, men kommer opp ofte

I3: Nei

Hoveddel:

3. Har du hørt om noen som har hatt en negativ eller positiv opplevelse av å benytte en chatbot?

I1: Meg selv. Foretrekker å skrive til noen enn å ringe, men når det er en chatbot og jeg har utdypende spørsmål, så fungerer ikke chatbot.

I2: Ja, jeg. Vanskelig å konkretisere det man spør om, vanskelig med lange spørsmål. Den skjønner ikke hva man mener.

I3: Jeg vet at flere heller velger å ringe kundeservice enn å få hjelp av chatboter fordi det er lettere å få svar på det man lurer på.

4. Er det noen egenskaper du skulle ønske en chatbot hadde? I så fall hvilke?

I1: Tilkalle mennesker som kan hjelpe. Føler at chatbot ikke er egnet for noe som helst annet enn veldig simple oppgaver, f.eks hvor en kan logge inn i et system. Personlige spørsmål er ikke innafor å bruke chatbots til

I2: Skjønner ikke større sammenhenger, skulle gjerne sett at den forstod mer komplekse spørsmål.

I3: Skulle ønske at de var litt mer hjelpsomme, jeg har ikke fått den hjelpen jeg ønsker om jeg stiller noe litt mer kompliserte spørsmål.

5. Hadde det påvirket deg om hvordan en chatbot hadde svart/kommunisert med deg? I så fall hvordan kan den påvirke?

I1: Føler chatbots prøver å være høflige, har ikke hatt noe særlige reaksjoner på hvordan de kommuniserer med deg.

I2: Tror ikke det.

I3: Føler chatboter som sender mer utfyllende svar er mer hjelpsom.

6. Er det noe du kan se for deg at en chatbot kan brukes til, som ikke allerede eksisterer i dag?

I1: Bestille noe, f.eks mat. Liker ikke chatbotter generelt, og føler det gjør ting mye mer upersonlig. Får som regel dårlig inntrykk av sider som bruker chatbotter fordi de ikke bruker ressurser på å hjelpe. Generelt er teknologi veldig dårlig per dags dato.

I2: Nei. Vet ikke så mye om hvordan de brukes. For meg om den kunne funnet ut av hvordan jeg skulle sy noe når jeg syr og lurert på noe, så hadde det vært fint. (Hun snakker basically om Alexa eller lignende, uten å vite det). Jeg sitter masse på YouTube for å finne ut av ting. En slags support, men i privat setting.

I3: Jeg tror en chatbot kan brukes til mange ting, og at den kan erstatte kundeservice på langt flere områder enn det gjør nå i dag. Jeg kommer ikke på noen konkrete eksempler.

7. Har du noen tanker om hvordan en chatbot kan brukes til å endre folks liv?

I1: Hvis man kommer til et punkt hvor den blir såpass ekte og menneskelig, kan den på en måte gi folk noen å snakke med, f.eks hjelpelinjer (helse). Reduserer problemet med at folk f.eks må vente på hjelpetelefon osv.

I2: Nei. Ser ikke helt for meg chatboter som psykologen.

I3: Tror en chatbot kan være til god hjelp på flere områder, men ikke at den kan endre folks liv

Avslutning:

8. Hvordan tror du chatboter kommer til å utvikle seg i fremtiden?

I1: Tror de kommer til å bli smartere, men at de egentlig ikke har en plass i fremtiden.

Blir vel mer og mer intelligente, så de skjønner større sammenhenger og lærer mer.

I2: Ser det i mange programmer jeg jobber med at programmene blir mer og mer

intelligente. Kan repetere mer og mer, kjenner igjen mønstre. Blir sikkert ikke mer menneskelige, men forstår mer.

I3: Jeg tror chatboter kan bli ganske intelligente.

9. Ser du noen positive eller negative sider ved denne utviklingen?

I1: Nei

I2: Det er vel positivt det. De har ikke følelser, så de tar ikke over verden, men de automatiserer og hjelper menneskeheten. Tar jobber fra folk, men forsvinner noen jobber kommer det andre. Litt redd for butikkjobber og biler og hva det kan gjøre med menneskers jobber.

I3: Både positivt og negativt at chatboter kan ta over jobben for mennesker.

Tilleggsspørsmål:

1. Har du hørt om "grieving chatbots" eller chatbots som danner vennskap med brukere, eller andre type sosiale chatbots?

I1: Nei

I2: Nei

I3: Nei

2. Har du noen umiddelbare tanker rundt det?

I1: Funker hvis du er ensom og trenger noen å snakke med

I2: Ikke helt min greie. Høres ikke helt bra ut (ref. Xiaolce). Men kan kanskje hjelpe noen ensomme sjeler.

I3: Tanken er god, bare man ikke blir for avhengig og glemmer personene rundt seg

Appendix 5 - Consent Form (Norwegian)

Du er helt *anonym*.

Du kan trekke deg *når som helst*.

Initialer

Formålet med intervjuet er å finne ut mer om tanker rundt chatboter i forbindelse med et prosjektarbeid i IN5480 'Spesialisering innen forskning i design av IT' ved Universitetet i Oslo.

Kontaktperson for prosjektet er julitho@uio.no