Interaction with AI

Iteration 2

Group 3:

andrgrae

eivias

erikhma

emilylo

IN5480

Contents

About us

This project is being run by:
Andreas Græsmo - andrgrae@uio.no
Eivind Askeland - eivias@uio.no
Erik Mahlum - erikhma@uio.no
Emily Lo Kam Chuen - emilylo@uio.no

We are a group of 4 master students doing our first year of Design, use and interaction. We have different backgrounds with only one of us having a bachelor's degree from IFI while the other three have their degrees from Kristiania University College, OsloMet, and Høgskolen i Østfold.

Area of interest

We are interested in the field of Social Credit Systems (SCS), particularly the implementation of SCS by the Chinese government. It caught our attention due to interesting topics related to the ethics of AI-powered surveillance, government control, and automated decision-making. We see this system as one that is intrusive into the lives of Chinese citizens and there is a fear among us that this system could spread to other countries - including Norway - in the future.

The goal of the Chinese SCS is to "foster pro-social behaviour" by guiding people to "behave in accordance with society's interests" (Langer, 2020, p. 164). The means by which this is accomplished revolve around "automated evaluation of individual behaviour and social interactions". As a result, it is clear that this is a process where the privacy of citizens is not the top priority. Facial recognition and automated data processing are two aspects of SCS where AI is involved and we would like to dig deeper into the field to find out more about these in particular.

We pose the following research questions:

RQ1: How is AI used in the Chinese Social Credit System?

RQ2: To what degree are the inhabitants willing to accept a Social Credit System in Norway?

Background

The majority of the facial recognition technology (FRT) we humans experience first-hand today are benign, we mostly see it on our phones in the form of security. Instead of using a password, or our fingerprint, we can use our face as a way to identify ourselves (Gray, 2003, p. 2).

However, FRT is also used as a way for the government to survey and control the citizens. The technology is being used to enhance security in public spaces, locating missing people, fighting crime and corruption, imposing age restrictions on online viewing of pornography. These seemingly positive examples come with ethical drawbacks. Examples being biases, inaccuracies, mass surveillance and privacy intrusion (Kostka et al., 2021, p. 2).

Some countries, and cities have begun imposing regulations and bans on the use of FRT. The state of California has banned the use of FRT by law enforcement agencies. In 2020, Portland, Oregon banned the use of FRT for all city departments, including private retailers, for example hotels and restaurants (Kostka et al., 2021, p. 2).

China and scs

The Chinese government wants to implement a social credit system in China. The reason they want to implement this is “allow the trustworthy to roam everywhere under heaven while making it hard for the discredited to take a single step” (Engelmann et al. 2019, p. 70). However, this statement does not consider the ethical parts of the social credit system.

Every citizen gets an 18-digit ID card that is called the Unified Social Credit Code (Engelmann et al. 2019, p. 69). This is the individual code that will have the score connected to it. How this will be used in real life is not described in this paper.

According to Engelmann and Chen (2019) there are three main reasons why the Chinese government wants and can implement this system. The first reason is that dishonest activities stand for a loss of around 92 billion USD for the Chinese government every year.
The second reason is that with a credit score system a person without money can get a loan to buy a house. However the government is more interested in loaning people money that they can use on investing in the domestic market. Without a social credit score system the bank only has a credit check to control if it's safe to give someone a loan.
The third reason is that privacy is seen differently in China than in the Western society. For the Chinese population this will hardly be seen as a privacy-violation. Therefore, it is easier to implement this system in China than in other countries.

Methods

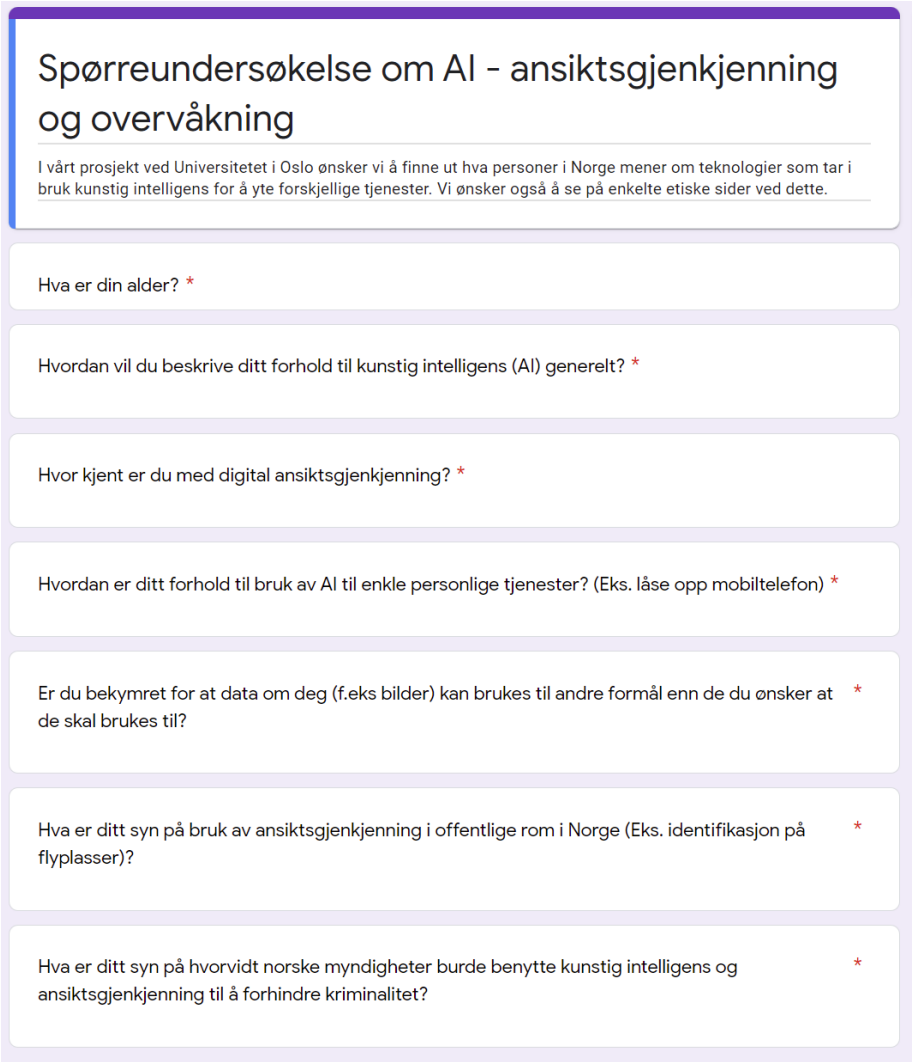
In our project we would like to use different ways of collecting data in search of answers to our research questions.

A document study should give us further insight into the use of AI in connection to the Social Credit System (SCS), as well as giving us more information on how the system works and how it impacts Chinese society. While this method was primarily chosen for answering research question 1, we have now identified several topics and questions which can be used in our data collection for research question 2 as of the second iteration of this project.

Initially, we had planned on using interviews as our main data gathering source for trying to answer the second research question. In some part due to the scope of the assignment and our limited time however, we have decided that a questionnaire is more suited for this task. In order to gather enough data through interviews, we would have had to interview many people to end up with a representative data set to analyze. We want to gather as many opinions and views as possible in order to be able to generalize, at least to a larger scale of the population than a small group of people in our social circles.

With this change in approach, we are largely giving up on the ability to dig deeper and look into why people have certain opinions of AI, in favour of a broader set of opinions. We have also decided to dampen our explicit focus on the social credit system in this questionnaire. Instead, we will focus on general opinions of AI, as well as facial recognition as a tool for individuals, organizations and the Norwegian government. This is still relevant to our case, as it tries to highlight opinions related to the ethics of surveillance systems such as the SCS, while also trying to gauge where the line should be drawn when it comes to the use of AI technology such as facial recognition by different actors.

The survey we have been working on is in the process of being sent out (through social media), and we expect the results to be ready for analysis and discussion soon. The current version is shown below.



Spørreundersøkelse om AI - ansiktsgjenkjenning og overvåkning

I vårt prosjekt ved Universitetet i Oslo ønsker vi å finne ut hva personer i Norge mener om teknologier som tar i bruk kunstig intelligens for å yte forskjellige tjenester. Vi ønsker også å se på enkelte etiske sider ved dette.

Hva er din alder? *

Hvordan vil du beskrive ditt forhold til kunstig intelligens (AI) generelt? *

Hvor kjent er du med digital ansiktsgjenkjenning? *

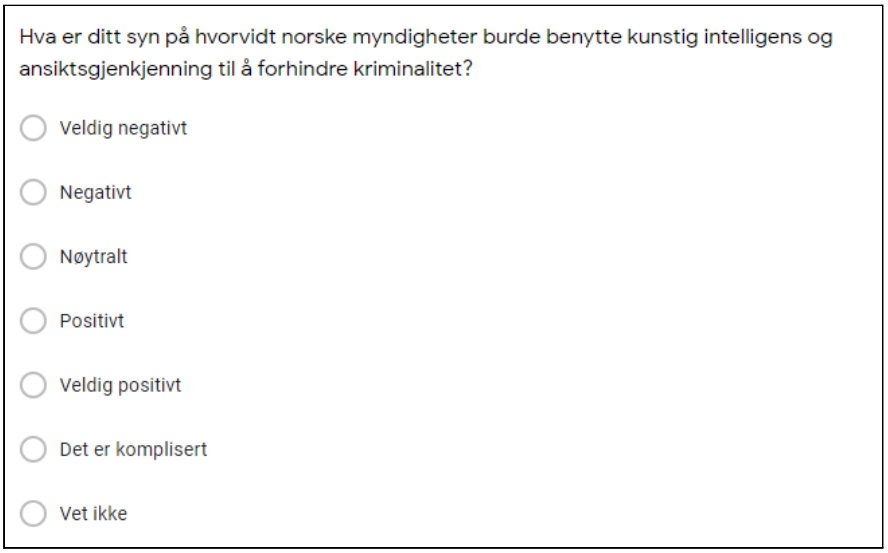
Hvordan er ditt forhold til bruk av AI til enkle personlige tjenester? (Eks. låse opp mobiltelefon) *

Er du bekymret for at data om deg (f.eks bilder) kan brukes til andre formål enn de du ønsker at de skal brukes til? *

Hva er ditt syn på bruk av ansiktsgjenkjenning i offentlige rom i Norge (Eks. identifikasjon på flyplasser)? *

Hva er ditt syn på hvorvidt norske myndigheter burde benytte kunstig intelligens og ansiktsgjenkjenning til å forhindre kriminalitet? *

The questions will generally try to show how familiar the respondents are with the technology, or explore how negatively or positively they view certain subjects related to our second research question:



Findings

The paper by Yu (2020) looks at how to use social media data to create a score. The purpose of the score is for banks to get another way to give out loans. As the bank operates today, they don't lend out money to people who don't pass the credit risk check. With the current system, people that are trustworthy but don't have credit can't get a loan.

To make this system work they use machine learning to get a score. First, they make categories to find abnormal users. For example, advertisements and companies also use social media but shall not get a social credit score. To find these users the AI looks at how many followers and followees an account has. An account with many followers but few followees can be categorized as an abnormal user.
The next step is to clean the data so the machine learning doesn't make wrong calculations based on wrong data. For example, the score can't be affected if your activity has been high one day but low every other day. After this the AI system makes a calculation based on your activities, followers, who you are friends with, how many books you bought and so on. In the end you are ending up with a score based on an AI systems judgement.

Prior to the implementation of SCS in China, the government was involved in what was called the Golden Shield Project. This was China's plan to link all of its state's individual surveillance networks with a large centralized online database to automate information sharing. This has only become feasible as of recently (Wong & Dobson, 2019, p.224). As technology proceeds to evolve, the SCS system has integrated tools such as facial recognition. China is believed to own the world's largest surveillance camera network with 176 million surveillance cameras, and this number is expected to increase up to 626 million by 2020 (Wong & Dobson, 2019, p.224).

How likely is it for a country to implement AI-surveillance? According to Feldstein (2019), a breakdown of military expenditures in 2018 shows forty of the top fifty military spending countries also have AI surveillance technology. Still there are quite a lot of technologies linked to Chinese companies that are found in at least sixty-three countries worldwide. The Chinese tech giant Huawei is alone responsible for providing AI surveillance technology to at least fifty countries (Feldstein, 2019, p. 11).

Feldstein (2019) presents three AI surveillance techniques, smart cities/safe cities, facial recognition system, and smart policing. Smart cities focus on making the city safer, with sensors, facial recognition and police body cameras. This is in order to prevent crimes, ensure public safety and respond to emergencies. Facial recognition systems involve biometric technology to match and compare live footage of individuals with images from a database. The last technique, Smart policing, is a data-driven technology used to facilitate investigations and police response. Example of using an algorithm to make a prediction of future crimes (Feldstein, 2019, p. 16).

In the next part of this project we would like to synthesize our findings from both the document study and survey in order to present the collective findings of our study by trying to provide some answers to our research questions.

References

Engelmann, S., Chen, M., Dang, L., & Grossklags, J. (2021). Blacklists and Redlists in the Chinese Social Credit System: Diversity, Flexibility, and Comprehensiveness. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 78–88. https://doi.org/10.1145/3461702.3462535

Engelmann, S., Chen, M., Fischer, F., Kao, C., & Grossklags, J. (2019). Clear Sanctions, Vague Rewards: How China's Social Credit System Currently Defines «Good» and «Bad» Behavior. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 69–78. https://doi.org/10.1145/3287560.3287585

F. Langer, P. (2020). Lessons from China - The Formation of a Social Credit System: Profiling, Reputation Scoring, Social Engineering. *The 21st Annual International Conference on Digital Government Research*, 164–174. https://doi.org/10.1145/3396956.3396962

Feldstein, S. (2019). The Global Expansion of AI Surveillance. *Carnegie Endowment for International Peace*, 42.

Gray, M. (2003) 'Urban Surveillance and Panopticism: will we recognize the facial recognition society?', Surveillance & Society, 1(3), 314–330. https://doi.org/10.24908/ss.v1i3.3343

Kostka, G., Steinacker, L., & Meckel, M. (2021). Between security and convenience: Facial recognition technology in the eyes of citizens in China, Germany, the United

Kingdom, and the United States. *Public Understanding of Science*, *30*(6), 671–690. https://doi.org/10.1177/09636625211001555

Wong, K. L. X., & Dobson, A. S. (2019). We're just data: Exploring China's social credit system in relation to digital platform ratings cultures in Westernised democracies. *Global Media and China*, *4*(2), 220–232. https://doi.org/10.1177/2059436419856090

Yu, X., Yang, Q., Wang, R., Fang, R., & Deng, M. (2020). Data Cleaning for Personal Credit Scoring by Utilizing Social Media Data: An Empirical Study. *IEEE Intelligent Systems*, *35*(2), 7–15. https://doi.org/10.1109/MIS.2020.2972214

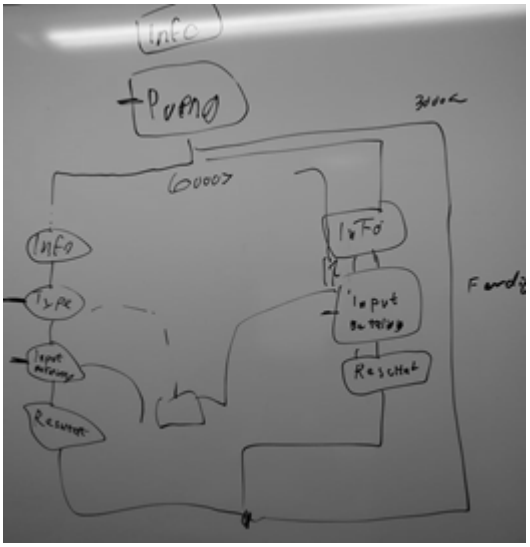
Appendix 1 - Chatbot design task

Introduction

In this assignment we created a chatbot. We tried to connect this assignment to our topic in the group assignment, social credit system. The idea we had was a chatbot that treats people differently based on their social credit score. We ended up creating a chatbot that helps people get loans, but gives different offers based on their score.

Process

In the beginning of the process we started to sketch how the flow in the chatbot should look like. The first box was a greeting box, where the user gets information of what the chatbot can help with. Further, the chatbot would ask what the social credit score of the user is. We decided that the maximum score was 12 000 and then we split the flow in two, those under and those over 6000. The flow for those over 6000 would get more options and better conditions on their loan. We then decided to add a third flow to the system, and this would be for those who have a score under 3000. These people won't get any loan from the bank, because of their low score.



Then we started to create the chatbot. In the beginning we needed to learn how Chatteron works and therefore we watched some tutorial videos on the webpage. This helped us learn some of the basic tools needed to create our idea. During the creation of the bot, we figured out that we would have to add some elements that we didn't have in our sketch. As an

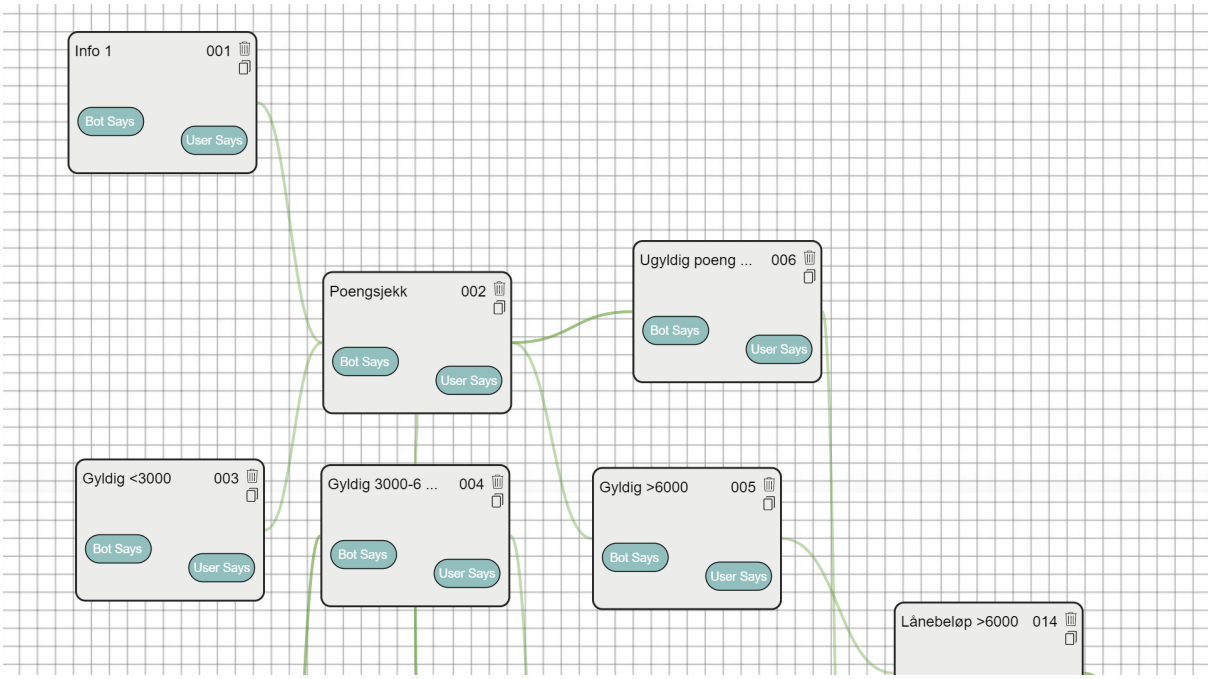
example, if a user provides an invalid input, in our case not a number, the flow needs to take the user back so they can try again.

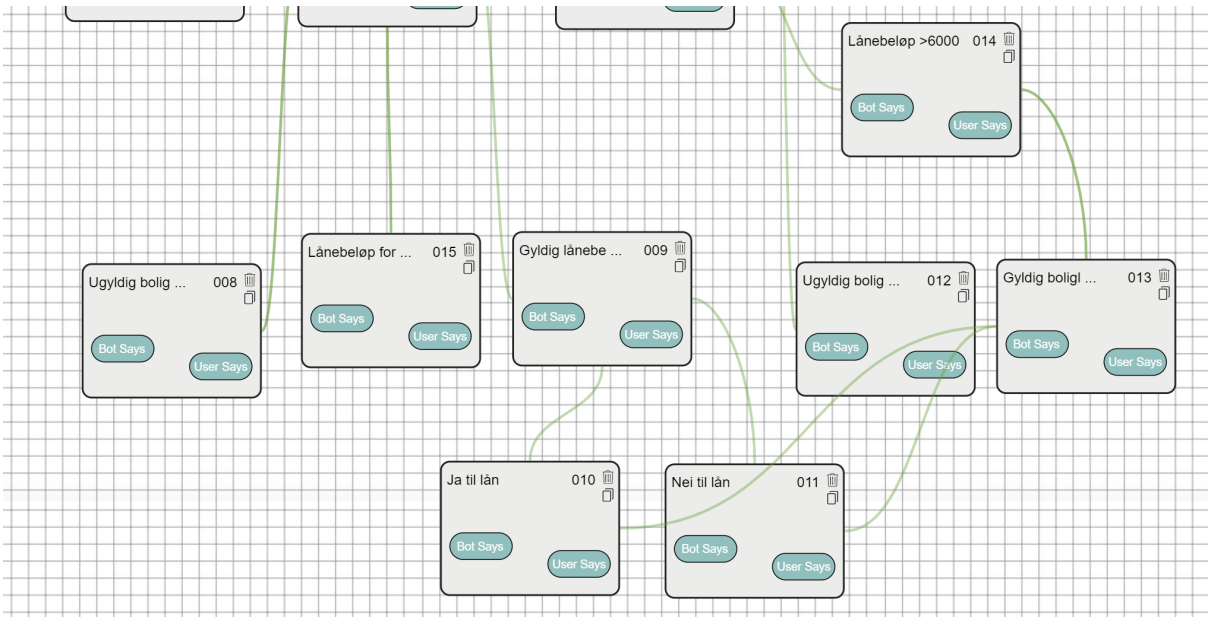
Reflection and learning

We created a chatbot that is simple and missing many of the processes that are necessary to get a loan, for example credit check. But it still shows how the idea works. Our simple chatbot still needed a lot of paths to get through the system. One question ended up having a minimum of two paths to proceed, and our question had only numeric answers. This shows how complex a chatbot is. To create a chatbot with many different answers and questions is difficult, especially if the user can write long answers. What may seem like an easy answer demands a lot of information underneath the surface to make the chatbot able to respond, even with a simple flow-based chatbot such as this.

To get a good flow was something we learned through this process. If the user writes something that is not expected, the chatbot needs to give the user the opportunity to try again. In our chatbot the user was sent back to the question if it was not answered like expected.

Chatbot flow structure in Chatteron:





Appendix 2 - Machine learning task

Process

We spent a lot of time initially trying to understand the code to figure out which numbers we could manipulate in order to see any changes in the training of the chatbot. While the code was not very descriptive, we found out that changes in the batch number and epoch values seemed to have the biggest impact, while changing the “Dense” value in the model also had some minor effects. When we added a dropout to the model of 0.3 the accuracy number seemed to fluctuate up and down a bit more. The number changed from 0.15 to 0.18 after every epoch. Without dropout the accuracy was consistently 0.15 until it changed to 0.17.

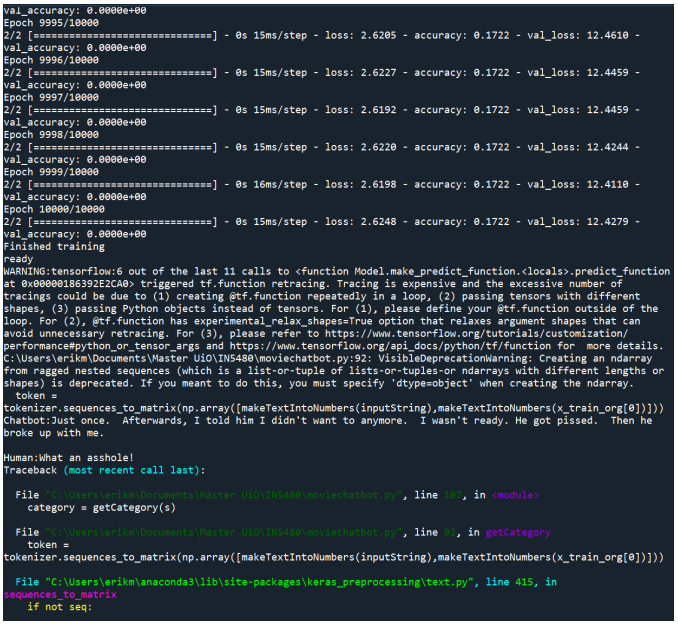
Throughout this working with this task, we have been very confused by the output values of the neural network’s training. We still don’t really have a good understanding of what the loss and accuracy values actually mean and how they correlate to how the bot responds to our input. The difference between val_loss and loss was also not apparent. We ran into some issues where the script would randomly crash after no more than 20 inputs from the user:

```
Chatbot:God, you're just like him! Just keep me locked away in the dark, so I can't experience anything for
myself

Human:You deserve it
Traceback (most recent call last):

  File "C:\Users\erikm\Documents\Master UiO\IN5480\moviechatbot.py", line 106, in <module>
    category = getCategory(s)

  File "C:\Users\erikm\Documents\Master UiO\IN5480\moviechatbot.py", line 91, in getCategory
    token =
tokenizer.sequences_to_matrix(np.array([makeTextIntoNumbers(inputString),makeTextIntoNumbers(x_train_org[0])]))

  File "C:\Users\erikm\anaconda3\lib\site-packages\keras_preprocessing\text.py", line 415, in
sequences_to_matrix
    if not seq:

ValueError: The truth value of an array with more than one element is ambiguous. Use a.any() or a.all()
```

Chatbot crashing

Learning outcomes

It seems to take a very high amount of iterations for the chatbot to exhibit any form of intelligence. We have not yet seen any signs of this. We change the batch size to 512 and later to 1000, and change the epochs to 10000. Still the accuracy was 0.17 and the interaction with the chatbot was confusing.



High validation loss (12.4) after 10 000 epochs

It was very hard to tell what actually makes a difference and what doesn’t. This might be connected to using too few iterations or layers. However, we didn't find what we were supposed to increase or do differently to get a better chatbot.

As the chatbot replied with the movie lines, we were confused by whether it had any correlation to what we wrote to the bot. At some point the replies indicated that the chatbot had understood what was written by us, however we were quickly disappointed when the next line seemed to be completely random. We are therefore left with the feeling that it doesn't matter what we write to the chatbot. Its internal workings are a black box to us as users.