# Problem Statement

# Design conversational helpers that can efficiently help humans in engaging human-human conversations. Using existing chatbots that target open-ended conversations, create an app that can help in analyzing how the existing chatbots actually maintain engaging conversations. Gather the data based on the responses that are helpful and obtain the results that can help in improving open-ended conversational chatbots.

# The main goal of this project is to design chatbot helpers that can be used to assist people with having a social conundrum and those who find it difficult to have an engaging conversation. Thereby improving socializing skills and establishing an initial connection of friendship.

# Provided Solution

## Chatbots

The app will be a simple application that uses two chatbots:

1. Microsoft DialoGPT (GPT-2 based chatbot)
2. Pandorabots Rosie (Open source AIML based chatbot from Alicebot family)

The DialoGPT is implemented using HuggingFace and the responses are generated locally whereas responses from Rosie are generated using Pandorabots API.

## Data Gathering

The data generated would be stored using some database and would be open sourced. There will be several users using the application and any two random users would be connected for having a conversation by providing the feedback from the chatbot responses. Since the chatbots are primitive and in development, the advice given by the chatbot’s need not be taken seriously.

Since the data would be open sourced, users are advised not to provide confidential information. While there could be responses generated by the chatbots that might not be appropriate, such responses should be avoided. These responses could be generated due to open-endedness in the conversation.

The primary data components that are used for this project are the responses generated by the chatbots and whether or not they are actually useful in a real human conversation. At every message, there will be some responses provided to the user, and the user will select if the response is actually useful and related to the context.

There are secondary components required for designing the application such as name, age, email, and more. However, these components of the data, being sensitive in nature will not be open sourced. These essential components will help in user registration and select the random users to use the application.

## Data Storage and Retention

The data will be stored in a relational database server which will be secured via authentication. The primary data components will be open sourced while the secondary components will be retained till the development of the project is over.

Since the conversational data will also be generated at the Pandorabots API, under the Section 5 of their terms and conditions (<https://home.pandorabots.com/policies.html>) the users are required to provide consent for sending the conversational data over the API. However, as per Section 7, Pandorabots do not acquire any ownership of the data and therefore, the data can be open sourced.

## User Guidance

The goal of the project revolves around the guiding users to have an efficient conversation and the users will be guided using the chatbot responses. The users will then select if the response was actually useful and / or context accurate. Meaning, the response generated is good and related to the conversation and not totally out of the conversation.

The responses generated are created by chatbots and are unmonitored and could therefore contain coarse meanings and therefore, the users are advised not to take the responses seriously.

# Need for IRB approval

Since the experiment being conducted involves human subjects engaging in conversations that are not monitored, IRB approval is required. At any point, the subjects will be having conversations over the application and will not communicate using other means. Due to the current situation of the pandemic, the subjects will follow the regular legal social distancing measures or can opt for remote data gathering since the data is gathered electronically.

The subjects will be asked to provide a feedback based on the chatbot responses to see if the chatbot is helpful to actually engage humans in an efficient conversation. The users can also use the responses generated by the chatbot and in this case, it would mean that the response was helpful.

The responses generated are created by chatbots and are unmonitored and could therefore contain coarse language or meanings and therefore, the users are advised not to take the responses seriously.