# Chatbot Implementation in Customer Service Industry through Deep Neural Networks

**Team Members:**

Likhita Chandana Adabala

Naveen Kumar Reddy Desireddygari

**Contents:**

- Data Collection
- Data Processing

## Data Collection:

We have taken dataset from Yale University especially about the highlights of the university and about the magazine's information. The main aim of this dataset is to generate the enquiry about the Graduation programs in Yale and the majors in each program. I have collected 130 samples in my dataset about the Grad programs and its specialization about the program. I have given information about the Arts, Humanities, Environmental science, and public health programs. In some samples, I have included information about the application process of the programs and its requirements. The sample dataset started with the general conversations and then I asked these enquiries about the Grad programs in Yale. I have gathered the information from the official website of Yale. I have uploaded the dataset in the drive link and pasted it the below.

## Data Set Link:

https://drive.google.com/drive/folders/1lxE7Z2uJlQCwFKnfV6oHNgt0CsCVlF3?usp=share_link

## Data Processing:

The text data in the question-and-answer columns of the pandas DataFrame are cleaned up in the data preparation stage described above. It specifically uses the pandas Series.apply() method to apply the preprocessing procedure to each member of these columns. Preprocessing () performs the following processes on a phrase after receiving it as input: Using the. lower() function, it lowercases every character. Using the regular phrase '[,.]', it eliminates all commas and periods. This standardizes the text data's format and gets rid of any extra punctuation that can get in the way of model training or assessment. The matching columns in the original DataFrame df are then given new assignments to the updated question and response columns.

The next step is to develop a vocabulary and commit to memory pairs of questions and answers. Keep in mind that we are working with word sequences, which don't automatically transfer to a discrete integer space. As a result, we must establish one by assigning an index value to each distinct term we find in our collection.

**References:**

**Yale Dataset – Grad Program Links**

https://gsas.yale.edu/

https://law.yale.edu/

https://publichealth.yale.edu/

https://pytorch.org/tutorials/beginner/chatbot_tutorial.html#load-and-trim-data