

AAI-520 – Team Project Status Update Form

Fill out this form and submit it by the end of Module 4 in Canvas.

Team Number: 9

Team Leader/Representative: Michael Shimanyula

Full Names of Team Members:

1. Frank Ivey
2. Michael Shimanyula
3. _____

Title of Your Project: Ubuntu Guru

Short Description of Your Project and Objectives:

Build Natural Language Processing Model/s that are trained on the Kaggle Ubuntu Dialogue Dataset to build a chat bot that will dialogue with the end user. Our team, as instructed, will utilize a for version control along with a cloud application hosting framework. The goal is to deploy the chatbot on a web/app interface, offering users an intuitive platform for interaction.

Description of Your Selected Dataset (data source, number of variables, size of dataset, etc.):

Link to data source: <https://www.kaggle.com/datasets/rtatman/ubuntu-dialogue-corpus/data>

We have chosen the Ubuntu Dialogue Corpus as our dataset, which provides a rich ground for training and refining our chatbot model.

The Data source size is 2.91 GB

toc.csv (175 B)			
Detail Compact Column 4 of 4 columns			
About this file csv table of contents			
# lines	# words	# characters	filename
3 total values	3 total values	3 total values	3 unique values
9212878	91660344	996253904	dialogueText_196.csv
16587831	166392849	1799936480	dialogueText_301.csv
1038325	11035331	116070597	dialogueText.csv
No more data to show			

Data Explorer
 Version 2 (2.91 GB)
 Ubuntu-dialogue-corpus
 toc.csv
 ... 1 more

Summary
 4 files
 .csv 4
 22 columns
 String 13
 Integer 6
 DateTime 3

Provide the GitHub link of the final project: <https://github.com/hmshimanyula1995/AAI-520-NLP>

How many times have your members met in the last two weeks? 3

List the specific contributions that each team member is providing for the Final Team Project in the table below.

- **NOTE:** ALL students on the team should contribute equally to the Final Team Project.

Team Member 1 (Michael)	Team Member 2 (Frank)	Team Member 3 (if applicable) (Name)
List of contributions Text Preprocessing, Tokenization, Building Vocabulary, Sequence Generation of Tokens, Model Building, Model Training, Model Evaluation Chat bot interface Cloud Hosting	List of contributions Text Preprocessing, Tokenization, Building Vocabulary, Sequence Generation of Tokens, Model Building, Model Training, Model Evaluation Chat bot interface Cloud Hosting	List of contributions

Comments/ Roadblocks: _____ N/A _____
