

# Data Collection Plan

Project number: 6

SpikeBot: Deep  
Neural Network  
Based Drink-Ordering  
Project title: Chatbot With A Web Application Project leader: Liya Li

Date: 9/12/2022

## Description of the data collection

- Collect conversational data in .txt/.wav format recorded by different people in gender, race, ethnicities and ages ordering drinks in a restaurant.*
- Use electronic survey method with volunteers for mock drink-ordering chats. Combine all conversational data into one .txt file and an audio folder.*
- Both Text and Audio conversations must be in U.S. English.*
- At least one month long collection process to collect about 200 conversations. Each conversation contains at least 5 questions and answers.*
- For all collected conversational data, we will perform quality assurance on it and store in the shared Google Drive.*

## What will be done with the data once it has been collected?

<input checked="" type="checkbox"/>	Identify how well the process is meeting <b>customer needs</b>		Analyze a <b>problem</b> , or identify the causes of variation
<input checked="" type="checkbox"/>	Obtain <b>exploratory view</b> of the process		Test a <b>hypothesis</b> about the process output
	Evaluate the <b>measurement system</b>		Test a <b>hypothesis</b> about the effects of one or more inputs
<input checked="" type="checkbox"/>	Check the <b>stability</b> of the process (is it in control?)	<input checked="" type="checkbox"/>	<b>Control</b> a process input or monitor a process output
	Conduct a <b>capability study</b>	<input checked="" type="checkbox"/>	Other reason... Use as original data to build and train DL models

## Key Variables - A summary of the chosen input variables (Y's) and/or output variables (X's)

	1	2	3	4	5	6
What?	Variable title	Audio Conversation	Text Conversation	Drink Knowledge		
	Input (X) or output (Y) variable?	X	X	X		
	Unit of measurement	Kilobyte	Bit	Bit		
	Data type	Audio in .WAV	String	String		
	Collection method	Record human voice	Write down texts	Fill in google sheet		
	If manual	Yes	Yes	Yes		
	Gauge/instrument	N/A	N/A	N/A		
	Location	restaurant settings	any location	any location		
	Gauge calibrated?	No	No	No		
	Measurement system checked?	Yes	No	No		
MSA	Precision (R&R) adequate?	N/A	N/A	N/A		
	Accuracy adequate?	Yes	Yes	Yes		
	Historical data exist?	No	No	No		
	Source of historical data	N/A	N/A	N/A		
	Historical data representative/reliable?	N/A	N/A	N/A		
	Mean	N/A	N/A	N/A		
	Upper specification limit	N/A	N/A	N/A		
	Lower specification limit	N/A	N/A	N/A		
	Standard deviation	N/A	N/A	N/A		
	Target	Original value	Original value	Original value		
Sampling	Minimum sample size (MSS)	N/A	N/A	N/A		
	Sampling frequency	Never	Never	Never		
	Sub-grouping needed?	No	No	No		
	Sub-group size	N/A	N/A	N/A		
	Stratification needed? (time, shift)	No	No	No		
Who?	Data collector	SpikeBot team	SpikeBot team	SpikeBot team		
	Operational definition exist?	Yes	Yes	Yes		
	Data collector trained?	Yes	Yes	Yes		
	Resources available for data collector?	Yes	Yes	Yes		
When?	Start date	09/12/2022	09/12/2022	09/12/2022		
	Due date	10/12/2022	10/12/2022	10/12/2022		
	Duration (in days)	30	30	30		

## Additional Comments - e.g. resolution needed, sampling method, R&R results, storing of data, handling outliers, using filters, etc.

If the volunteer audio is not in English, does not have content, or too noisy, this audio will be discarded.

This plan is for collecting 1000+ primary data. If we need to collect 10000+ primary data, the process will take up to 1 year in total.

For audio data collection, do not combine question and answer in one audio clip.