

# Biomedical Neural Machine Translation

04.08.2019

Ziqing Lu

Vignesh Murali

Ritvik Reddy

Final Project: INFO 7390 Neural Networks and AI

## Overview

Google Translate has only 57.7% accuracy when used for medical phrase translations and should not be trusted for important medical communications.

However, it still remains the most easily available and free initial mode of communication between a doctor and patient when language is a barrier.

Hence, we are proposing an NMT system which aims to provide accurate translations in the medical domain.

## Goals

To Improve the Neural Machine Translations for Medical Domain thereby increasing physician-patient engagement and hence enhance the total quality of care provided to the patients.



## Use Cases

Provide translations for taking consent for surgery, procedures or research from patients or relatives

## Data

We are using a Medical-Dataset

<https://github.com/biomedical-translation-corpora/medline>

## Methodology

For this final project, we will focus on NMT building and leverage existed API like google speech recognition API, OCR API to provide more choices for data input for application users.

### Neural Network Building:

We will use neural networks - encoder-decoder/seq2seq to train a neural machine translator. Key elements of our neural network and we will experiment with:

- Seq to Seq
- Encoder-Decoder
- Word-embeddings
- RNN units - GRU, LSTM and Bi-directional LSTM
- Attention Layers (Luong Attention, Bahdanau Attention)

### Evaluation Metric:

- Bleu Score
- Levenshtein Distance/Word Error Rate

### Existing API:

Based on the trained model, we will further leverage APIs like google speech recognition API, OCR to provide more choices for real-time source language data input.



## Process Outline

1. Data Preprocessing
2. Study of all Modelling approaches and select the best Model for prediction
3. Implementing Attentions
4. Study of all the Performance metrics ( such as Blue-Score) and Evaluate the Performance of the models using these scores
5. Deploy the Model on AWS or Google Cloud using a web-framework Flask

## Milestones

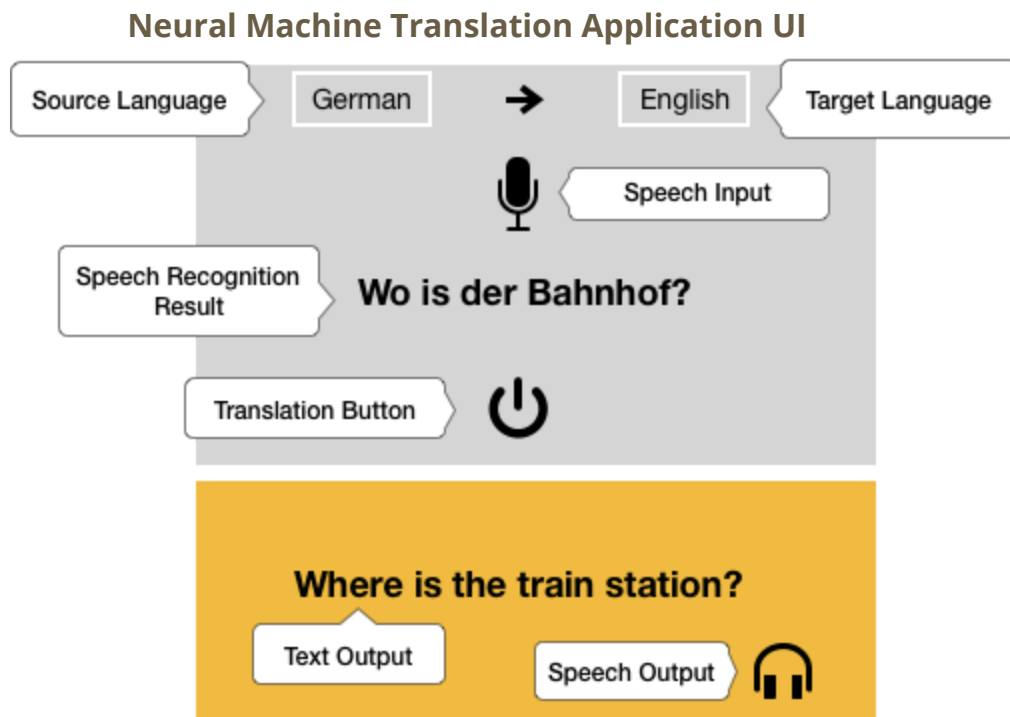
TIMESTAMP	PROCESS DELIVERY
Week 1 Day 1-3	Data Preprocessing
Week 1 Day 3-7	Model Building, Training, Selection and Performance Metrics
Week 2 Day 7-11	Deployment of models on the Cloud using FLASK
Week 2 Day11-12	Documentation and Final PPT

## Deployment Details

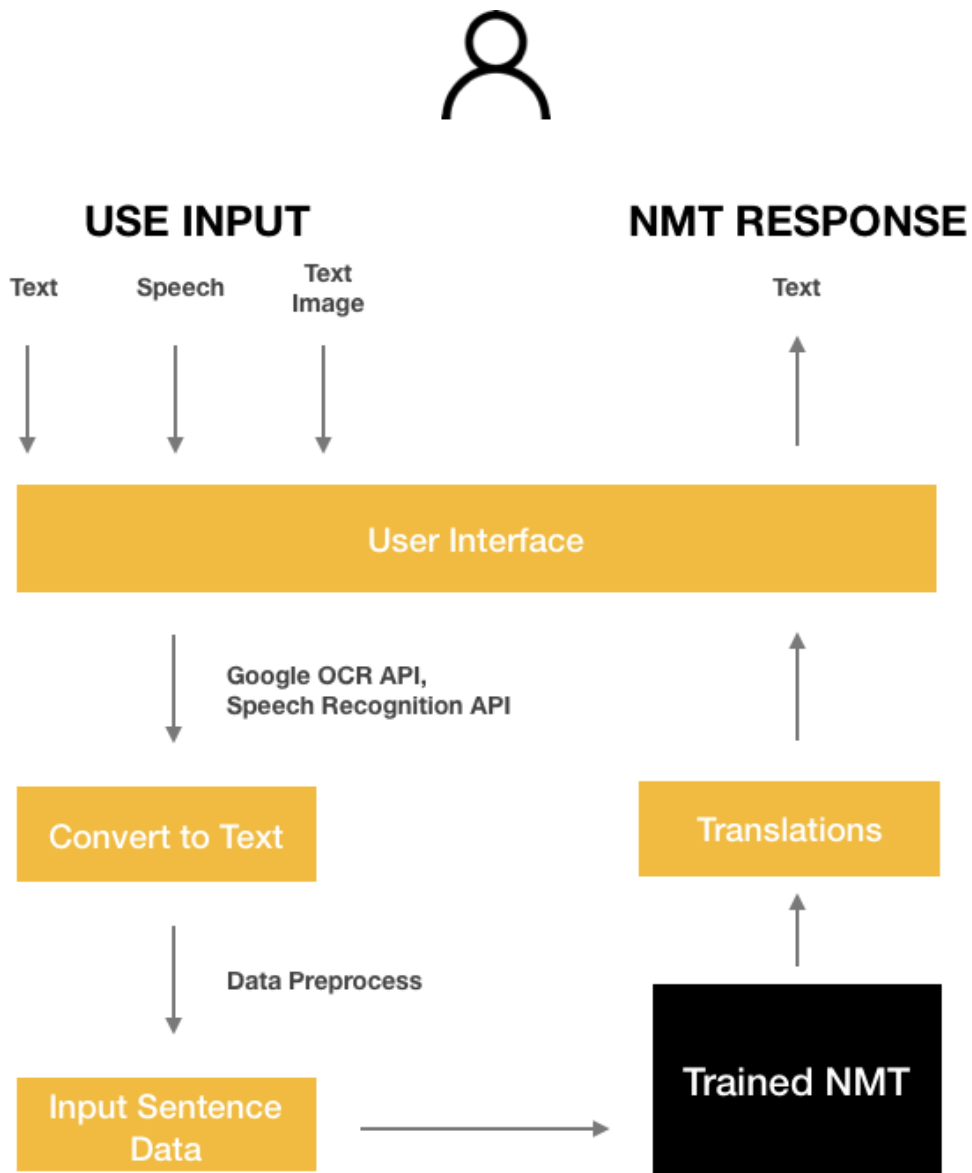
- 1) Programming Language: Python
- 2) Modelling Framework: Keras/Tensorflow
- 3) Modelling Back End: Tensorflow
- 4) Web Framework: Flask
- 5) Cloud: Amazon EC2

## User Interface Design Plan

If there's enough time, we will work on many features as what's shown in the UI below. For this final project, we will focus on translation between 2~3 languages, text input and speech input for the source language, and text output for the target language.



## NMT Application Workflow



## Reference and Sources:

<https://www.bmj.com/content/349/bmj.g7392>

<https://arxiv.org/abs/1508.04025>



<https://arxiv.org/abs/1409.0473>