COMP39/9900 Computer Science/IT Capstone Project
School of Computer Science and Engineering, UNSW

Project Number: P1

Project Title: Voice-based interface for NLP models

Project Clients: Aditya Joshi

Number of Groups: 2

Project Specializations: Artificial Intelligence (Machine/Deep Learning, NLP);Web application development.

Background:

NLP systems may use conversation agents, sentiment analyzers, hate speech detection, and so on. NLP researchers often fine-tune their own models for various tasks. We would like to have a voice-based interface for such Transformer-based models.

Goals:

1) To create a voice-based interface for text-based (i.e., NLP) models.
2) To test the interface with two scenarios: (a) speakers that use different dialects, (b) multiple speakers interacting with the system.

We hope to use the interface in our ongoing research on NLP for dialects of English, funded by Google Research.

Requirements and Scope:

- The project will focus on NLP models for English.
- The project can use HuggingFace models for text only.
- The response time should be reasonable.
- Evaluation with a set of users who use different dialects of English, and with a set of users engaged in a conversation or parallel talk.
- The project can use black-box models for transcription.

The voice-based interface will be available as a web-based interface that uses the microphone on the user's device to record the audio and allows the user to select a particular model (For example: "Sentiment Analysis") from among a set of available models. The interface will then convert the audio to text and use an NLP-based model to obtain the inference. The output will be produced on the screen either as text or as audio.

The project will enable the students to design an NLP system that effectively integrates different NLP/speech processing models, and tests them in a web-based setting. The test setting will help the students understand the effectiveness of their system. The students may target an NLP demo paper at the end of the course, if that is of interest to the team.

Required Knowledge and skills:

The students should be able to use transformer-based models. Fine-tuning skills are good but not essential. The students should be able to create web-based interfaces that receive text and audio as input and produce the same as output.

Expected outcomes/deliverables:

- Open-source code
- Documentation
- User guide to configure the interface for new models.

Supervision:

Aditya Joshi

Additional resources: