ECE49595NL: Introduction to Natural Language Processing
ECE59500NL: Foundations of Natural Language Processing
Spring 2026
Homework 1
Due 5pm Friday 27 February 2026 (note the deadline has been extended)

This homework is to be done in teams; the same teams that were formed for homework 3.

For this homework, each team will construct a pair of chatbots, one that emulates Former President Joseph R. Biden and one that emulates President Donald J. Trump. The core of the chatbot can be the same for both, but they will likely need to be trained on different data. The core can be implemented either as something like ELIZA or as a large language model (LLM), or any other computational method that you wish (not a surrogate human). Various (re)implementations of ELIZA are available at:

https://github.com/wadetb/eliza
https://github.com/jwiegley/emacs-release/blob/master/lisp/play/doctor.el
https://github.com/jeffshrager/elizagen.org
https://dhconnelly.com/paip-python/docs/paip/eliza.html
https://github.com/qobi/ece49595nl/blob/main/eliza.py

Various reimplementation of GPT (an LLM) written by Andrej Karpathy are available at:

https://github.com/karpathy/minGPT
https://github.com/karpathy/nanoGPT

Andrej Karpathy has prepared various tutorials on how to build LLMs:

https://www.youtube.com/watch?v=7xTGNNLPyMI
https://www.youtube.com/watch?v=l8pRSuU81PU
https://www.youtube.com/watch?v=zduSFxRajkE
https://www.youtube.com/watch?v=zjkBMFhNj_g
https://www.youtube.com/watch?v=kCc8FmEb1nY

More are available at:

https://www.youtube.com/@AndrejKarpathy/videos

You can use any of this code or any other code available on the internet as you see fit. Purdue has a license key for this class to use the OpenAI models through Microsoft Azure. I will distribute the key for this through filelocker. You can use any pretrained models you find on the internet. But likely to achieve a good emulation, you will need to at least fine tune any pretrained model and may need to train a model from scratch.

Your code should allow your emulations to engage in a debate, i.e., your emulated Joe Biden and emulated Donald Trump should participate in a conversation or dialog. You can have an emulated moderator if you wish, though this is not necessary. You can set whatever rules of debate you wish.

Your system should output speech. To generate speech, you can use any software you wish. Code to generate speech from text is available at:

https://github.com/qobi/ece49595nl/blob/main/minimal_text_to_speech_pyttsx3.py
https://github.com/qobi/ece49595nl/blob/main/text_to_speech_pyttsx3.py
https://github.com/qobi/ece49595nl/blob/main/minimal_text_to_speech_microsoft.py
https://github.com/qobi/ece49595nl/blob/main/text_to_speech_microsoft.py

PYTTSX3 runs locally on your computer. Microsoft Azure requires a network interface and a license key. I will distribute the key for this through filelocker.

The above does not have emulated voices for Joe Biden or Donald Trump. You don't have to emulate their voice; but you can if you wish. There may be ways to train a voice model based on speech samples. It should be straightforward to obtain speech samples of Joe Biden and Donald Trump on the internet. Teams that wish to excel should try to do so.

Your system should take speech as input. Code to recognize speech from text is available at:

https://github.com/qobi/ece49595nl/blob/main/speech_to_text_microsoft.py

Microsoft Azure requires a network interface and a license key. I will distribute the key for this through filelocker.

Sample code that ties ELIZA and GPT to speech input and output is available at:

https://github.com/qobi/ece49595nl/blob/main/spoken_eliza_pyttsx3.py
https://github.com/qobi/ece49595nl/blob/main/spoken_eliza_microsoft.py
https://github.com/qobi/ece49595nl/blob/main/spoken_gpt_pyttsx3.py
https://github.com/qobi/ece49595nl/blob/main/spoken_gpt_microsoft.py

You can use any of the above code, or any code you find; you just need to cite the source of any code that you use.

You will likely need to train (or at least fine tune) your models on samples of text from Joe Biden and Donald Trump. You can use whatever samples you find on the internet; you just need to cite your sources. There are likely public records of speeches and various public statements that each made. Part if this exercise is for you to search for and find those. Here are some resources that I found from a quick search:

https://ia601405.us.archive.org/19/items/TrumpTheArtOfTheDeal/Trump_%20The%20Art%20of%20the%20Deal.pdf
https://www.reboxu.com/uploads/8/6/0/3/86031326/think_like_a_champion.pdf
https://en.wikipedia.org/wiki/Bibliography_of_Joe_Biden?wprov=sfla1
https://archive.org/details/promisestokeepon00joeb/page/n2/mode/1up
https://archive.org/details/promisemedadyear0000bide_j8m5/page/180/mode/1up

This code does not need to run on RCAC Scholar. You just need to submit your code as a single zip file to Brightspace. Each member of each team needs to independently submit all of the code for that team. Your submission should not include any software or model weights you downloaded from the internet; just pointers to where you downloaded the software and/or model weights from. You submission should not include any model weights you trained; just code to replicate the training. We will not run your code. Rather, each team will need to bring two laptops to class, one that runs the system that emulates Joe Biden and one that emulates Donald Trump. Your two systems need to debate each other through speech; not through network connection. Each team will get ten minutes to have their emulations debate each other. This will be done over several lectures after the due date. (We may conduct a public debate between team chatbots, open to the entire Purdue community, if teams are able to build particularly compelling emulations. Having emulated voices for Joe Biden and Donald Trump would make this particularly compelling.)

This is an ambitious exercise. Because of this, I have extended the deadline. Do not leave this to the last minute. The details of this may be fleshed out as we go along.