AI Lab 4—**or**—5 Group Project: roles are outlined below

Due Wednesday 12/4 11:59pm (not 12/3 as on syllabus). One group member should email to Professor Glaser and cc their partner.

**AI Lab 4: Text-to-speech**

Create an account and start a free trial at [ElevenLabs](https://elevenlabs.io/text-to-speech?utm_source=google&utm_medium=cpc&utm_campaign=us_brandsearch_brand_english&utm_id=20455649470&utm_term=eleven%20labs&utm_content=brand_exact&gad_source=1&gclid=Cj0KCQiAo5u6BhDJARIsAAVoDWtGB9veEftu0mZ4mFaGnTegryZ8MkBJVQGJaLYtfDdfSNycFTLIHyMaAiO5EALw_wcB) or another of the (many) AI text to speech start-ups out there.

Together, pick a poem that you 1) appreciate and 2) can find a human reading of online. This could be the author’s reading, or someone else’s performance (if you go to youtube, try to pick a high quality or at least interesting reading….)

Now get an AI to read it. Customize / select whichever “voice” you’d like (try to do some of this together). Perhaps there are notable AI readings of famous poems out there already. You can choose those as well!

First group member:

You are responsible for making the trial account, choosing a voice, and playing around until you get a decent output. If you can record and save, please do. Write a short paragraph that describes the AI, then notes any “decisions” the AI makes as well as how its various vocal qualities (accent, style, etc) impact the reading.

Second Group Member:

You are responsible for choosing the poem and finding a human reading (obviously work with your partner on the poem choice!). Write a short paragraph about any “decisions” the reader makes, noting any you and your partner might have made differently

**AI Lab 5: Curated data sets; fine-tuned and custom LLMs:**

By many accounts, Large Language Models like ChatGPT will give way in part to smaller language models oriented towards specific tasks and often built around precise data. This involves not just prompt engineering, but training and fine tuning. While we have not explored such techniques much in class, you now have the skills to evaluate the concepts behind models and the datasets used to train and tune them.

For example, here is a fine-tuned version of a relatively small Mistral LLM trained to make Haikus: <https://huggingface.co/davanstrien/HaikuHermes-0.1-7B>

And here is its dataset: <https://huggingface.co/datasets/davanstrien/haiku_dpo>

Setting aside the exact techniques used to create this dataset and train the haikubot, we can evaluate the following:

* What was the goal of the model, and how does it improve upon / refine the original? [Here: LLMs often struggle to count and get right stanzas other than really common ones. Haikus are relatively common but models still don’t understand the task.]
* How, roughly, does the training work? [Here: Rewarding the model for hitting the right syllable count teaches it the new task]
* Does the dataset represent the task or linguistic pattern well enough? [Here: I’m not sure if the “topics” or “beauty” really capture what haiku ism but scoring the syllable count seems straightforward]

Your job is to find:

1. another trained model (<https://huggingface.co/models>)
2. an intriguing dataset (<https://huggingface.co/datasets>)
3. if you struggle with huggingface, go ahead and google around / use poe to find a well-described model or even just a customized bot that you think is valuable for the kind of questions we’ve posed in this course.

First Group Member:

After finding a model and/or dataset with your partner, write a paragraph: describe why you chose the model (or dataset). What makes it interesting? Does a readme explain its development and rationale? Use the bullet points above to guide your response.

Second Group Member:

Could it be useful for the study of literature, or as a model for new tools for the study of literature? Perhaps this *kind* of training or customization has value even if the model itself does not. If you can actually run the model easily (depends on the model and your skills), describe a few prompts and responses.