

**Project Idea:**

I will translate Shakespearean (Early Modern English) English into Contemporary English using deep learning.

**Domain:**

I will be working in the domain of machine translation / deep learning models (RNN's, LSTM's), word2vec embeddings, data scraping, and other NLP techniques.

**Data:**

I will scrape data from No Fear Shakespeare to get phrase to phrase translations. No Fear Shakespeare provides line for line translations of a significant body of Shakespeare's work. I will try to train a model by just using this data, but I might need to create a word2vec embedding for Early Modern English if I am not getting great results.

The second source will be a corpus of Early Modern English. I will need this in order to create word embeddings for Early Modern English. There are several sources I am trying to get access to, but I have found a website I can scrape that has thousands of texts as a plan b.

**Known Unknowns:**

The biggest hurdle could be training the word2vec embeddings on Early Modern English. There is still some uncertainty about whether or not I will get access to the corpus from some of these institutions. Scraping these texts at scale could prove difficult and the actual training of a word2vec model is time / resource intensive.

**Plan B:**

If this whole translation doesn't pan out, I will take Shakespeare's plays (and other writers of his era) and classify plays into the 3 major genres of that time: tragedies, historicals, and comedies. I have found a handful of other resources that I can scrape to get more plays written in the 15th - 17th century.