Introduction and Goal

- ❖ There has been efforts in NLG to create poetry and other artistic texts that emulate those created by actual humans
- ❖ Haiku generation is a subtask that involves more work, as it has limitations on the length of the poem and the topics that it can cover.
- ❖ The importance of developing a realistic Haiku generating model range from educational purposes to the preservation of the artform.
- ❖ There are existing Haiku generating models, but the poems generated are often not artistic and do not emulate real-life poetry

Thus, we aim to create transformer language model that takes in a topic word and generates a comprehensible, relevant, and artistic three-lined haiku.

Background

- English Haiku
  - Based on the traditional Japanese Haiku
  - Three short lines
  - Often about things in nature

- Previous Haiku Generation models
  - LSTM trained using Japanese haikus (Wu et al. 2017)
  - GPT-3 model trained on 45 TB of data
  - Haikus generated by previous models are often not artistic and can be easily distinguished from real poems

Haiku Written by Real Poet

The temple bell stops,
but I still hear the sound
coming out of the flowers
-Matuso Basho

Haiku Written by GPT-3 Model

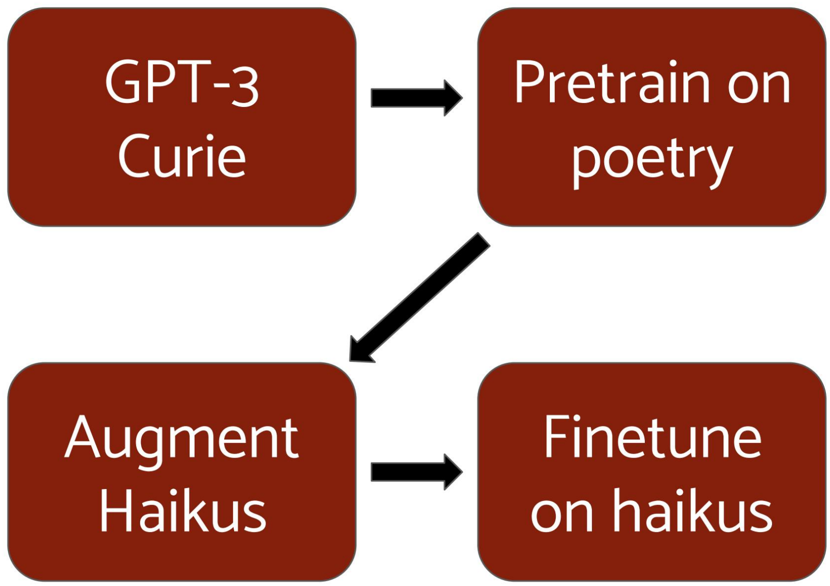
Pretty flowers in the garden
Paths are lined with petals
Bees buzzing happily

Methodology

The baseline model used was OpenAI's GPT-3 Curie model. We first pretrained the GPT-3 model on a dataset of generic poetry. We then finetuned the resulting model on a set of English haiku's that were first augmented.

Augmentation types:

- ❖ Insertion
- ❖ Swap
- ❖ Delete
- ❖ Substitute



Finetuning Details

- 4 Epochs and .1 Learning Rate
- Pretraining on Generic Poems
  - 8000 poems used; ~3 hours taken
- Finetuning on Haikus
  - 383 haikus; 1915 haikus after augmentation; ~1 hour taken

Pretraining and

- **Method 1:** Pretraining GPT-3 model on generic poems
  - Collected from poetryfoundation.org
  - 8000 poems used
  - Helps increase the amount of data we have and train the model to produce more artistic “poetic” text

- **Method 2:** Finetuning on Haiku corpus
  - Collected from various sources, including poets.org thehaikufoundation.org, and poets.com
  - 400 English haikus in the database
  - Each haiku was assigned a topic word
  - Model was finetuned using these prompt and poem pairs

Finetuning

Example Generic Poem Used

Holiness on the head,
Light and perfections on the breast,
Harmonious bells below, raising the dead
To lead them unto life and rest:
Thus are true Aarons drest.

Profaneness in my head,
Defects and darkness in my breast,
A noise of passions ringing me for dead
Unto a place where is no rest:
Poor priest, thus am I drest.

(Aaron by George Herbert)

Example Haikus Collected

Topic Word: Dandelion
Blow dandelions
and watch a thousand wishes
scatter in the wind.

Topic Word: Rain
Carefree raindrops drop
Racing on my window panes
Drip drop! They fall down.

Augmen

- **Method 3:** Augmenting the haikus before finetuning
  - Given the relative scarcity of haiku poetry available in English, we aimed to increase our dataset through augmentation. Augmentation also helps to prevent overfitting.
  - We employed 4 augmentation techniques: inserting a word by contextual word embeddings utilizing BERT, swapping two random words in a line, deleting random word(s) in a sentence, and substituting a word utilizing WordNet’s synonym database.

tation

Augmentation Example

Original:
A gentle spring breeze!
Through green barley plants
rushes the sound of water

Insertion:
Just a nice gentle spring breeze!
Through fine green barley plants
rushes the sound of rain water

Delete:
A gentle breeze!
Through barley plants
rushes the sound of water

Swap:
A spring gentle breeze!
Through barley green plants
rushes the sound of water

Substitute
A gentle fountain breeze!
Through green barley plants
rushes the sound of H2O

Evaluation

1. Objective Evaluation: Perplexity

Perplexity measures how well a model can predict a sample. A lower perplexity means a higher probability of haikus in the test set being generated by the model. Thus, a lower perplexity is better.

$$PP(W) = \sqrt[N]{\frac{1}{P(w_1, w_2, \ldots, w_N)}}$$

w_i represents a word
N is the number of words

Methodology

2. Subjective Evaluation: Rating Scores

Three Stanford student volunteers evaluated three sets of poems: real-life haikus, haikus generated by the raw GPT-3 model, and haikus generated by the finetuned model. The poems were rated from 1 to 5 in three categories.

- Relevance to Topic
- Artistry
- Comprehensibility

Testing

1. Perplexity Testing Data

To measure perplexity, we had a testing data set that contained 100 haikus written by real poets and their associated topic words. The five topic words we decided to test were “tree,” “spring,” “water,” “winter,” and “moon,” and we gathered 20 haikus for each word.

Example Haiku’s Chosen for Perplexity

Topic Word: Tree	**Topic Word: Spring**	**Topic Word: Water**
in the languid rain	Sweet scented blossom	deep in the water
a slender tree shivers	Confetti sent from heaven	softly moving his fins
delirious with life	With a gentle breeze	a carp, dreaming

Data

2. Subjective Evaluation Data

We chose 87 topic words, mostly relating to nature. We then collected three poems for each topic word, one from the raw GPT-3 model, one from the pretrained and fine-tuned model, and one real life poem.

Topic Word: Winter

Awake at night
The sound of the water jar
Cracking in the cold.

Topic Word: Moon

purple evokes plums
orange was butterfly noon
lavender calls scents

Example Topic Words Chosen for Subjective Evaluation

Grass, Spring, Leaf, Winter, Sleep, Growth, Night, Day, Sun, Cloud, Frog, Water, School, Children, Family, Dark, Light, Green

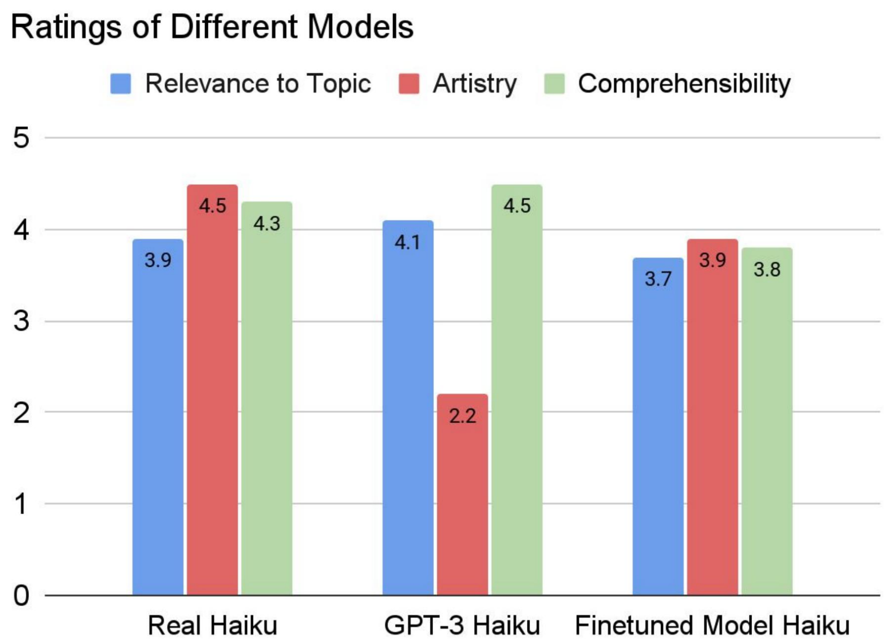
Perplexity Results

Model	Perplexity
GPT-3 Raw Model	16.5346936
Finetuned Model	5.3245626
Finetuned Model w/ Pretraining	5.7775668
Finetuned Model w/ Augmentation	10.067501
Finetuned Model w/ Pretraining and Augmentation	5.268226

From this, we see that the perplexity of the pretrained and fine-tuned model is lower than the raw model. This means that the probability of the test set poems being generated using the new model is greater than the raw model, suggesting that the fine-tuned model is better generating more artistic, and real (written by actual poets) haikus.

Subjective Evaluation Results

After averaging the scores of all 87 haikus for each model, we found that our finetuned model improved significantly in artistry when compared to the raw GPT-3 model. However, there were small decreases in relevance and comprehensibility. Compared to a Haiku written by an actual poet, our model was still behind in all 3 categories.



Topic Word: Moon

Real Haiku

Perfectly cloudless
Navy sky holds the full moon
Still for one moment

GPT-3 Haiku

The moon shines
so brightly in the sky
it's a beautiful sight

Finetuned Model haiku

A brilliant full moon!
On the surface of the water
Shadows

Analysis

- All four of of the models that we pretrained/finetuned had significant improvements.
- Best results in perplexity came from combining pretraining and augmentation.
- The models had significant increases in artistry, but slight decrease in comprehensibility and relevance
- Worse in all three categories when compared to real-life haikus.

Conclusion / Future Work

Haiku generation remains an ongoing challenge in the field of NLP. From our research, we were able to decrease perplexity of the raw GPT-3 model after pretraining, finetuning and applying data augmentation. However, as shown by the subjective tests, comprehensibility decreased from haikus generated by the raw model. There is also a level of artistry which is found in haikus authored by human poets that still cannot be captured through AI generation.

ADD REFERENCES HERE