Introduction

Poetry has been an integral part of human culture for centuries, serving as a powerful means of expression and creativity. In this report, present a comparative analysis of Arabic and English poetry generation using pre-trained models based on gpt2 model. i trained the models on specific datasets and evaluated their performance based on perplexity . Furthermore, i discuss the human interpretability of the generated poems.

Dataset

- Arabic Data Set : https://www.kaggle.com/datasets/fahd09/arabic-poetry-dataset-478-2017
    - The dataset contains over 58K poems that extend from the 6th century to the present day. Along with each poem, poem metadata have also been scrapped such as poet name, the poem, and its category. The data were scraped from adab.com
- English Data Set : https://hci-lab.github.io/LearningMetersPoems/
    - The English dataset is scraped from many different web resources. It consists of 199,002 verses, each of them is labeled with one of these four meters: Iambic, Trochee, Dactyl and Anapaestic. The Iambic class dominates the dataset; they are 186,809 Iambic verses, 5418 Trochee verses, 5378 Anapaestic verses, 1397 Dactyl verses.

Models

The two models i used in the project are based on GPT2 model heres the full details about each one

I choosed the two models to be already on poems and i retrained them on new poems because training gpt2-base from scratch to understand poems and generate them needs a lot of resources in term of computation power and time , but in theory if i trained the gpt2-base on large dataset it will eventually be as this models and mybe better

For the English model :

the model name is `"ismaelfaro/gpt2-poems.en"`

The ismaelfaro/gpt2-poems.en model is a specialized version of the gpt2 model designed specifically for generating English poetry. Developed by Ismael Faro, this model has been trained on a vast corpus of English poetry, enabling it to capture the intricacies of poetic expression and generate creative and evocative verses.

For the Arabic model:

the model name is `"aubmindlab/aragpt2-base"`

The aubmindlab/aragpt2-base model is a specialized version of the gpt2 model that has been specifically trained for Arabic language processing tasks. Developed by AUB Mind Lab, this model has been trained on a diverse corpus of Arabic text, allowing it to understand and generate Arabic language text with a focus on maintaining the nuances and intricacies of the Arabic language.

Embedding

To generate poetry in both Arabic and English, i employed the gpt2 models. These models are based on the transformer architecture and have shown excellent performance in natural language processing tasks.i used sequence representation for both Arabic and English texts.

Libraries Used

To implement the models and pre-processing steps, i relied on the transformer library because all models are from hugging face , which offers powerful tools for working with transformer-based models. Additionally, for the Arabic poetry generation, i utilized the pyarabic.araby library to remove tashkeel (diacritical marks) from the Arabic text, ensuring a consistent input format.

Model Evaluation

i evaluated the performance of our models using two key metrics: perplexity and semantic meaning score. Perplexity measures the model's ability to predict the next word in a sequence and serves as an indicator of how well the model has learned the underlying language patterns. The semantic meaning score assesses the coherence and relevance of the generated poems in relation to the given prompts or input.

To evaluate the models i have theres two key metrics :

1. perplexity
2. Similarity score

In the code i calculated the Perplexity for each model but i did’nt finish the Similarity Score because i faced some problem in the models i tried to use ,

For the perplexity :

Perplexity measures the model's ability to predict the next word in a sequence and serves as an indicator of how well the model has learned the underlying language patterns.

Here is table comparison for perplexity for each model , but before talking about the perplexity we should be aware that the perplexity cannot describe which model is good if the two models are different in the data and the evaluation data so the perplixity here is used to see how training effect the same model before fine-tuning and after

	Arabic	English	Base LSTM Model Arabic (less data)	Base LSTM Model English (less data)
Before Training	29.131423156611	35.643154985134	None	None
After 10 Epochs	23.362156314598	32.211453215463	132.542314789521	205.312563214589

The less the perplexity the more the model is better both the models show good decrease in the perplexity but we can see that the arabic model is better at learning the new patterns in the data.
And we can see the same thing in the base model , this model is a model that learned on small data because of the limitation on resources available to train a model like this .

The other Evaluation metric i tried to use is the Similarity Score which is depend on using pre-trained model for empedding texts and then calculate the similarity between them using the cos similarity
The basic idea was to use Word2Vec model as empedding model , first we get some testing lines of poems , create a input & label from them , then generate text using the input , after that we represent the input as vector and the output from the poem generator as vector and calculate the cos sim between the two vectors , then calculate the avg cos similarity score as score for the model
Because the Word2Vec and other similar models can capture the meaning of the words in the empedding and similar words are similar to each other in the vector space so if the model choose words that similar to the original poem but in other form of words then its should be consider doing god poems

Evaluation of Human Interpretability

Lets first see output from each model before start talking about this section

For the English model:

```
sadnessIs dilapidationDespite your rush, despite your witThe weightless whisper of the snail, The powerless thought of the coal, The lossless peace of the ocean breeze, The hopeless weightless agony of your soul on this stone yo
west.By Sidney
July 15,2008so often my sadness sends you to sleep  \nso very often my fingers walk over you with tears  \nyour eyes so gently squeezed as the sun remembers your childhood  \nI
```

The seed text was : sadness

For the Arabic model:

الحزن في قلبي عليك مخيم - والدمع من عيني عليك مسيل. ولقد شربت على المدام مدامة - صهباء صافية المذاق شمول.
وكأنها وكأن شاربها إذا - دارت عليه الراح ظل يميل.

The seed text was : الحزن

For the arabic model i really amazed from the poems that he can generate , he can make rhyme in the poem which is considered a hard thing to capture in arabic poem and almost all the parts the model generate are related to each other and related to the seed text it self

In the other hand theres the english model , as a person who dont understand the english poems at all 🙂 (i really tried but i didn't understand any thing) (but i love the arabic poems 🙂) i cannot say if this is a poems in english or not but by looking for the text the model generate we can see some pattern in the text , because the seed was "sadness" the vibes of the text is giving is dark , we can see the word tears and sleep and remembers and hopeless all these words make sense in the context of sadness but also i see some text that are not related to any thing like the date

You can try the both models from these links :

Mohamadhase/poem_generation_ar · Hugging Face

Mohamadhase/poem_generation_en · Hugging Face