task_1

Q: Present your chosen data, does it have any special attributes that you expect to have an influence on the text generation of the model?

My training dataset is *One Thousand and One Nights.* This classic synthetic dataset is a collection of Middle Eastern folk tales compiled in Arabic during the Islamic Golden Age.
My model may tend to be a language model which is good at telling arabic folk story.
The story generated by my model may include exotic arabic locations and scenes, characters in *One Thousand and One Nights* and the vocabulary may reflect the Middle Eastern culture.

Q: Take a look at the sample generation, what are your impressions?

The text captures stylistic elements of the training data, suggesting that the model has learned this from the dataset.
The output generally lacks coherence and a continuous narrative thread, jumping from one scene to another without clear transitions.
The text includes a mix of poetic expressions, old-fashioned dialogue, reflecting the time of the book in the training data.
There are also some grammatical and syntactical errors throughout the text.
Overall I think people can easily tell the classic styles of the text, and the textis generally understandable and readable.