<u>What I did vs what the Python library did:</u>

Python Package: Run the LDA model and implement the variational Bayes expectation maximization algorithm. The mathematical derivations for the update steps are quite complicated and I do not have the skills to implement them myself yet, especially when details like efficiency/smoothing have to be considered.

Me: Learn all aspects of the LDA model, including the hierarchical design, how to set priors, and the mathematical formulation of the problem for solving the posterior. Learn optimal values for hyperparameters (number of topics, number of words, max number of iterations, whether to remove stopwords, etc.) to boost model performance. Create visualizations to check posterior uncertainty and evaluate model performance. Write code to preprocess data properly and correctly feed it into the LDA model. Interpret and report results.