BY THE END OF THIS SESSION YOU SHOULD BE ABLE TO:

- Explain why lemmatization is a more thorough approach than stemming when normalizing n-grams in a corpus.

- Identify the unigrams, bigrams, and trigrams in a sentence.

- Understand how to calculate a TF-IDF if given the necessary values. (No need to calculate the final product, but know how to set up the calculation.)

- Explain how an LDA model works in terms of alpha and beta.