Neural machine translation (NMT) employs neural networks to perform sequence-to-sequence machine translation. NMT models require a parallel corpus of data, consisting of sources or phrases in the source language that map to their corresponding translation in the target language. This data is then split into training and validation sets and passed into the NMT model, which is typically based on an encoder-decoder architecture. The embedded encoder model takes in tokenizer source sentences, embeds it into a dense vector, using word embeddings, and then passes it to a stack of transformer layers that perform self-attention. The output from the encoder is then passed to the decoder which undertakes a similar process to generate tokens in the target language. Much of the literature we have been assessing so far has investigated how the success of pretrained transformer language models, such as BERT, hinges on the pretraining setup employed. We have spent a lot of time looking into the following subword tokenization methods and their relative performance tradeoffs on NMT: byte-pair encoding (BPE) (Sennrich et al. 2016), unigram language modeling (Kudo, 2018), and BPE-Dropout (Provlikov et al. 2020). Additionally, we have considered tokenization free approaches to NMT (Clark et al. 2022) and its time and space benefits relative to subword tokenization methods. Other reading has gone into investigating valid performance metrics for comparing these tokenization approaches, attempts to find optimal vocabulary sizes for NMT, and as to why some tokenizers and hyperparameters lead to better downstream performances than others (Vouhar et al, 2023). Up to this point, much of our actual programming has revolved around building the infrastructure to perform NMT in conjunction with the HuggingFace transformers library. This has taken the form of building code to train tokenizers and encoder decoder models, which we have been analyzing on translation tasks with synthetic languages. Future work will involve implementing these tokenization methods and assessing their performances.