Neural Language Models in Speech Recognition

State-of-the-art models for Speech Recognition typically make use of an acoustic model and a Language Model (LM) to minimize the word error rate of the transcriptions. These LMs are a key part of the pipeline, as they encourage the model to output existing words. In particular, the open sourced *DeepSpeech* model for speech recognition makes use of an *n-gram* LM, which is a frequentist model that captures a small part of the context, but does not compute deeper contextual representations. For instance, *cat* and *dog* are completely different words for those models, while a neural LM can abstract those words and relate them as animals. The purpose of this project is to implement such an algorithm in the DeepSpeech pipeline and measure its effect on the Word Error Rate (WER) in transcriptions. There is freedom regarding the choice of language model, and if time allows it, the task is to test several types of LMs: from bag-of-words to deep bidirectional transformers. The DeepSpeech model is built on Tensorflow and PyTorch.

This is part of a bigger research project which aims to provide an open source library to transcribe Norwegian speech and contribute to language technology development in general.

Main tasks include:
- Literature review on end-to-end speech recognition models and language models.
- Exploring and processing additional text data to minimize WER for the baseline model.
- Train and implement one or more neural language models and test hyperparameters to reach state-of-the-art performance in Norwegian Speech Recognition.

Data. The DeepSpeech model will be trained on public Norwegian data from Språkbanken, which contains over 500 hours of speech and their transcriptions. These data are phonetically balanced and contemplates all the Norwegian dialects. In addition, the language model can be trained on additional text data which can be extracted from public sources such as Wikipedia. That makes this project fully based on open source.

Main supervisor: **Massimiliano Ruocco**, NTNU
Co-supervisor: **Pablo Ortiz**, PhD, Senior Research Scientist, Telenor Research (Telenor ASA)