DS 5983: Large Language Models (Spring 2024) Dr. Roi Yehoshua
Student name: (Due) February 9, 2024

PA2: Transformers

In this task you are going to implement the Transformer architecture based on the Jupyter notebook provided on Canvas.

1. Understand the architecture:
   (a) Start by thoroughly reading and understanding the original Transformer paper Attention is All You Need.
   (b) Harvard's NLP group created a guide annotating the paper with a PyTorch implementation `https://nlp.seas.harvard.edu/annotated-transformer/`.
2. Model implementation: Fill in the missing code parts in the provided notebook adhering closely to the specifications in the paper. Pay close attention to the instructions and comments in the notebook.
3. Training the model: Train the model on a machine translation task from German to English using the `Multi30k` dataset (described in the notebook). Monitor the training and validation loss during training. Training the model on Google Colab (with GPU activated) for 20 epochs should take less than 30 minutes.
4. Testing and evaluation:
   (a) Implement a function for translating sentences using greedy decoding. Use it to translate several sample sentences.
   (b) Experiment with different hyperparameters (learning rate, batch size, number of heads, number of layers) and compare the results.
5. Write a short report covering the implementation process, challenges faced, and what have you learned.
6. *Bonus*: Implement beam search instead of greedy decoding for better-quality translations.