Project 2: Transformers

This is a single-person project. It is fine to discuss your process with other students but make your own decisions and write your own code, run your own experiments, and create your own presentation.

Our task is common sense question answering with the CommonsenseQA dataset.

Models

We compare three Transformer models: 1) a randomly initialized Transformer, 2) a pretrained Transformer (which was not trained/finetuned on CommonsenseQA), and 3) an LLM (1B+ parameters) of your choice. Use the same model hyperparameters for 1 and 2. We finetune the two Transformers, but not the LLM, for which we do prompt engineering.

Deliverables

First, a Jupyter notebook with documentation, code, and a link to your Weights & Biases view or report. Second, the slides of your presentation as a PDF. Check the deadlines on Ilias. Late submissions result in grade deductions.

Data

Use the CommonsenseQA dataset from Hugging Face with the dataset splits as instructed in class.

Jupyter Notebook

Start every section with all your decisions and justifications. Include the following sections:

- Introduction
- Setup
- Preprocessing
- Model
- Training
- Evaluation
- Interpretation
- Tools used

For details, see the grading checklist. The notebook should be runnable and reproduce your results.

Presentation

5 minutes. Don't explain the task (everyone did the same task). Content (~1 slide per section):

- Preprocessing
- Input/output format
- Network architecture
- Experiments
- Results
- Interpretation of results