

## MET CS622 Final Project Description (20 Points)

- You need to write a code that sends 10 questions, sentence completion, code writing or any other NLP task to the following language models:  
Falcon 7B, Falcon 50B, Falcon 180B, Llama2 7B, Llama2 13B, Llama2 70B, Vicuna, Alpaca, Koala, RWKV, ChatGLM, Mistral,... However, the number of parameters from the chosen model should be specified.
  - If you run locally, you are ok to run only five models. However, no more than two models with 7B parameters can be used. For example, you can use Mistral7b and Llama27B, but not Vicuna 7B.
  - If you run more than five models locally, independent from cloud you get extra credit. To run them locally you can use LLama.cpp or ollama. I don't provide the link to motivate you search for them.
  - If you run them on the cloud you need to run eight models, and be sure you can measure the energy use of the model execution on the cloud.
- Measure the execution time, energy usage, and memory use of each model and report it for the selected set of 10 questions, queries, etc.
- Questions (or sentence completion) should be standard questions for evaluations; you should research the right question set (e.g. HellaSwag, MMLU, HumanEval), and communicate that with me and TA before applying them.
- As a result, you should visualize the data with histograms with the following columns: model name, number of model parameters, the average query execution time + standard deviation, the average energy use + standard deviation, and average memory use.
- If you use Excel to build these reports, that is ok. However, if you create a GUI for your platform where the user can select the model, then write the query and, as a result, get the answer, plus consumed energy, memory, and execution time, you will get up to 5 extra credits.

### NOTES:

- \* For the project, you do not need to do it in Java, you can also do it in Python and call it from Java.
- \* The project should be done alone and not in a group.
- \* The deadline to deliver your final project is 12<sup>th</sup> of December.