Advanced Deep Learning - Assignment 1

In this assignment, you will analyze the performance of current state-of-the-art architectures for long context tasks.

- In the assignment, you will need to work with the Long Range Arena (LRA) dataset, which contains multiple 6 different subsets, each representing a different long context use case. You may choose any of them in your implementation
- We are interested in three different network architectures:
    - Long Short-Term Memory (LSTM) architecture.
    - Transformer (https://arxiv.org/abs/1706.03762) architecture.
    - State Space Models (SSM) - S4 in particular.
- You will need to train the model in three different approaches:
    - Train the network directly on the task you chose.
    - Pretrain (autoregressively) the network on an external dataset such as the WikiText dataset (for NLP tasks - you may choose a different dataset) and then fine-tune the network on the task you chose.
    - Pretrain (autoregressively) the network on the dataset you chose from LRA, and then fine-tune the network on the task you chose.
- You need to compare the performance of the 3 models on the 3 training strategies on the subset you chose from the LRA dataset:
    - Report the 9 different results in a table.
    - Compare them and give a conclusion on which model architecture gives the best results.
    - Conclude which training methodology leads to the best results (training directly, pre-train on an external dataset, or pre-train the network on the task dataset).
    - Explain your conclusion.
- You need to submit your code through Google Colab:
    - Due to the limited compute resources, you may choose the models sizes that works for you. But you need to make sure the number of parameters is roughly the same in all models.
    - The Colab notebook should be shared with the email shadya@mail.tau.ac.il
    - You need to have your code ready for evaluation.