Mini-Project (20 Marks) :

Image Captioning:

1. Design a CNN-LSTM system (preferably in pytroch) that can perform image captioning **[System 1]** based on the following details: **[10 Marks]**

- You are free to decide the CNN and LSTM architecture that best suits your case. The objective is to achieve a good BLEU score on the *test set* of Flickr8K data .
- Use the Flickr8K data for training and testing the model. The data is available at

  https://drive.google.com/drive/folders/1RQ5qHm0aVFqWDG9VBiSnXlNPl5T15Wf_?usp=sharing

  (Check Readme.txt for details of txt files)

2. The LSTM based image captioning can 'blindly' learn the structure of the language and predict meaningful sentences even with out learning much insight to the content of the image. This is termed as "language bias" of the system.

Design a training experiment with Flickr8K data to assess the language bias of your CNN-LSTM system [**System 1 Modified**]. Provide objective and subjective analysis/comparison of the results of *System 1* and *System 1 Modified* .

[10 Marks]

Contents to be shared:

- Share the training code, testing code and a technical report as a single zip file *Your_Team_ID*.zip
- Technical Report should contain
  - Team Id and all team member Ids and Names.
  - Architecture details of *System 1* and *System 1 Modified*
  - Detail the strategy used to asses the ‘language bias’ and justification for the same.
  - The objective and subjective results with *System 1* and *System 1 Modified* and their analysis

Note These Points:

1. To save on training time, run all images in training data to obtain CNN image embeddings for once and save it to drive. Use file read to store the image embeddings to local variable before the start of training. Same applies for validation and test data.