# Assignment 3

ELL 786 - Multimedia Systems

*Due Date : 18 April 2023*

---

1. Explain zero shot learning and few shot learning. What does the author mean by 'prompt engineering' in the paper. Implement the code to find the accuracy with basic prompt ('a photo of a {class}') and modified prompt ("a photo of a {class}, a type of pet.') on Oxford-IIIT Pets dataset. (You can use code from [1] and [2] (10 marks)

2. Explain polysemy. Explain the performance for boxer class in Oxford-IIIT Pets dataset with and without modification in prompt using few examples and predictions (Refer to [2], [1] and [3]) (5 marks)

3. Explain generalizability, robustness to distribution shift. What are the properties and usefulness of the datasets ImageNet, ImageNet-A, ImageNet-Sketch, ImageNet-R, MS-COCO, and dataset used for training CLIP model. (5 marks)

4. Collect 100 images of cats and dogs from ImageNet-Sketch / ImageNet-R. Use the collected dataset to find accuracy of a ResNet50 model trained on cats vs dogs dataset [4] OR a model trained on dataset formed by merging the different of cat breeds and dog breeds in Oxford-IIIT Pets dataset to form a cats vs dog dataset. Use additional prompts ('Prompt_Engineering_for_ImageNet.ipynb' in [1]) on CLIP with ResNet50 backbone and compare the performance. Use your collected dataset to compare the performance of CLIP model with ResNet50 backbone and ResNet50 trained earlier. (10 marks)

# References


[1] A. Radford, "Openai clip," https://github.com/openai/CLIP/tree/main/notebooks.

[2] ——, "Openai clip," https://github.com/openai/CLIP/blob/main/data/prompts.md#oxfordpets.

[3] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," *arXiv e-prints*, pp. arXiv–2103, 2021.

[4] Unknown, "Cats vs dogs," https://www.microsoft.com/en-us/download/details.aspx?id=54765.