

Title: Develop a Multimodal Based Network for Language and Vision Tasks

Goal: Utilize well known large Vision Language models and evaluate their accuracy under change of pretraining objectives.

Project Description: As a graduate student, you have the unique opportunity to embark on an exciting journey in the field of deep learning and natural language processing. In this project, you will dive into the realm of Language Vision Pretraining, a cutting-edge area that combines the power of large-scale models, advanced hardware, and state-of-the-art software tools. This project will not only expand your knowledge but also provide hands-on experience in setting up a deep learning environment for billion-parameter models, working with advanced NVIDIA servers, and using recent versions of PyTorch and CUDA.

What You Will Learn:

1. **Deep Learning Fundamentals:** You will gain a strong foundation in deep learning principles, including neural network architectures, optimization algorithms, and backpropagation.
2. **Language Vision Pretraining:** You will explore the fascinating world of Language Vision Pretraining, where you will understand how to leverage pre-trained models to perform various natural language understanding and computer vision tasks.
3. **Setting up Deep Learning Environments:** You will learn how to set up a robust deep learning environment capable of handling billion-parameter models. This includes configuring software dependencies, GPU clusters, and distributed computing frameworks.

By the end of this project, you will be well-equipped with the skills and knowledge necessary to tackle complex challenges in deep learning and natural language processing. You will have a deep understanding of Language Vision Pretraining and the ability to work with advanced hardware and software environments.

