Name: Nguyễn Minh Kha Date of birth: 11/10/2002 4-th years student	Education: **University of Science, VNUHCM** Education program: **Honor Programe**

Vision AI intern assignment

1. **Problem**: Cat/Dog Image classification
2. **Solution Pipeline:** From the original input image, the image will be preprocessed and resized to (224, 224). It will then pass through the BLIP-2 model (a vision-language model) to extract features, resulting in a feature vector of length 256. This vector is then transfer-learned through linear layers with ReLU activation and Dropout with a rate of 0.5 to learn to predict two classes: dog and cat (with the dog labeled as 1 and the cat labeled as 0). If the model's output is greater than the 0.5 threshold, the prediction will be a dog; otherwise, it will be a cat.

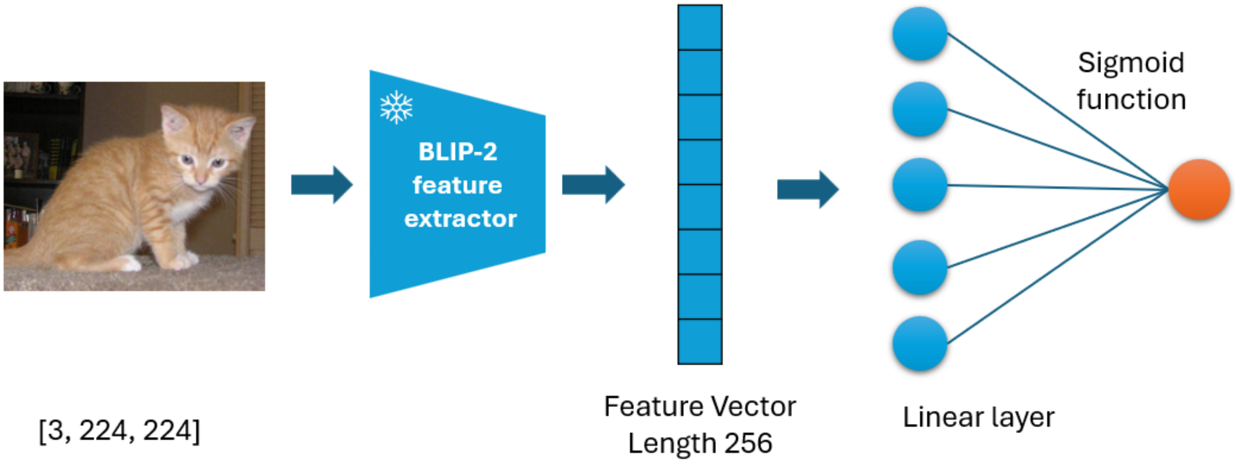

Figure 1: Pipeline of my solution

3. **Train dataset:** 12500 cat images, 12500 dog images
4. **Result**: Accuracy in test dataset reach 99.8%
5. **Compare with previous solution**:
   - outperforms CNN and ViT models and shows excellent accuracy even under significant noise.
   - learns very quickly and converges in just 4 epochs.