# Experimentation with Vision Transformers

**Q1)**https://colab.research.google.com/drive/1trWiN0Ru6wDwLPN5sP00nW_To4DG61Be?usp=sharing In this google collab notebook i intialized pre-trained CNN model(ReSNet-18) with weights from ImageNet dataset and intialized ViT model on ‘google/vit-base-patch16-224’ dataset. Later i fine-tuned both models on CIFAR-10 subset using train_model function and then i evaluated them on CIFAR-10 test laoder to see their performace and then again evaluated fine-tuned models on CIFAR-100 subset function

Instead i should have pretrained both CNN and ViT models on same dataset would have been better as this allows both models to learn general visual features from same dataset and then finetune both on CIFAR-10 subset and then evaluate the fine-tuned model on cifar-10 test loader

And here i got CNN model performance to be good wrt ViT model as the dataset i fine-tuned was not large and ViT model performs good for Large Datasets as its good at generalising global features and CNN is model is better at finding local patterns