Mixture-of-Modality Experts

Since the previous experiments on modifying the knowledge distillation process did not lead to the desired results, we decided to now focus less on the training process and more on the model architecture itself.

The philosophy