Fine-Tuning with Benchmark Programs (Output) Train Accuracy Test Accuracy Output Prediction Accuracy Steps