benchmark models, finetuned, phase R