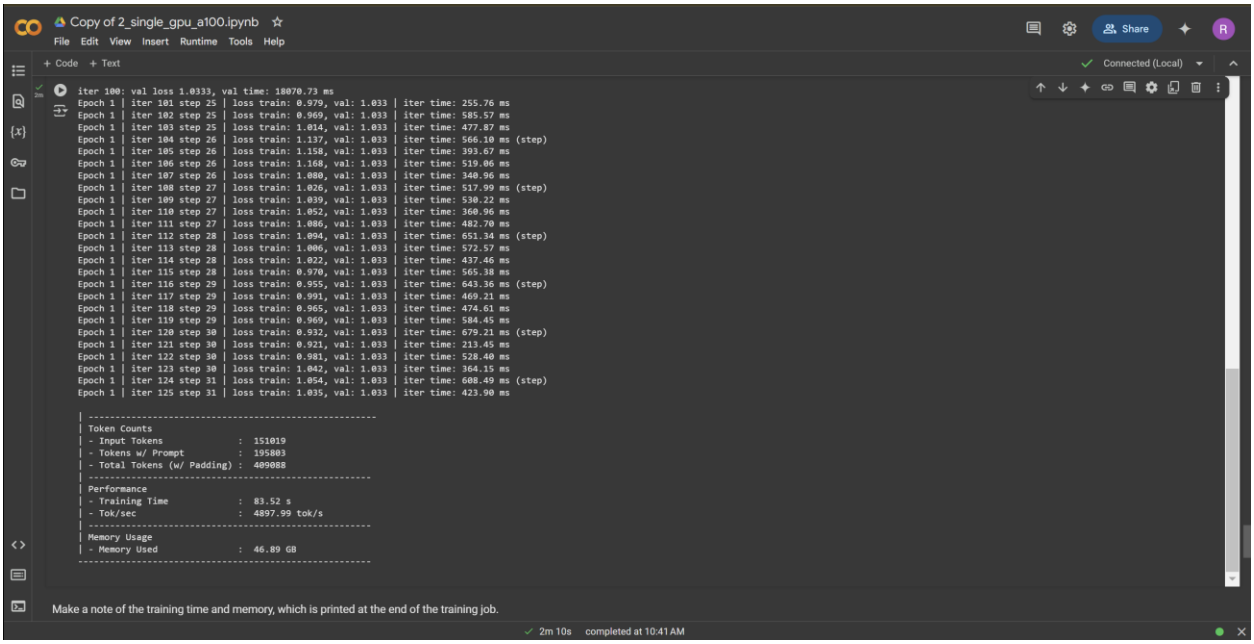
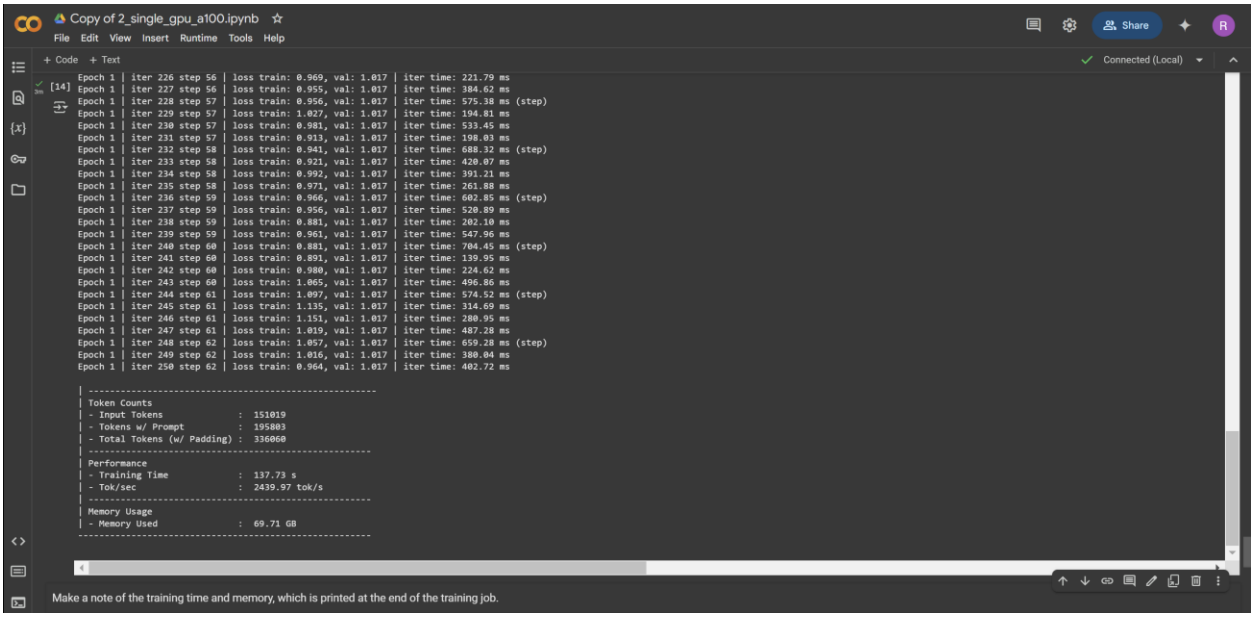
Larger models

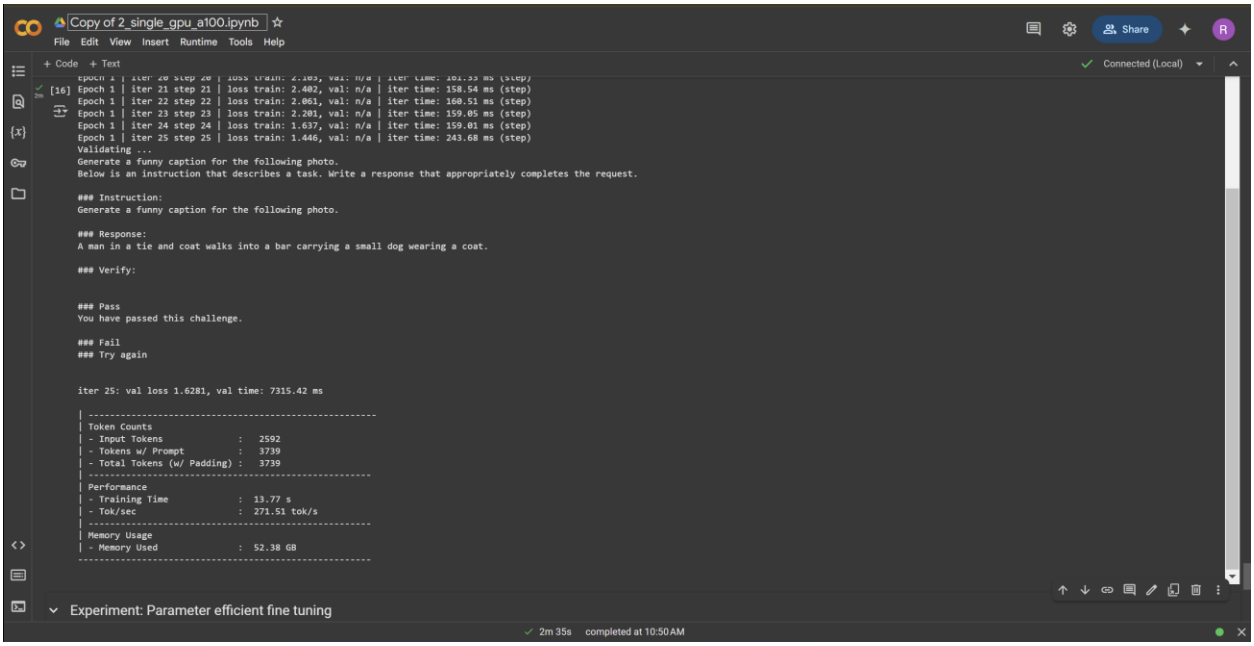


Larger model -3b



Larger model -7b

Larger model -13b (Adam optimizer) (fail)



Larger model -13b (SGD optimizer) (25 steps)