openAI fine tuning

Assignee	
Status	In Progress
Due	
Project	Mark My Work
Parent-task	try public api's on questionaire
Priority	
Sub-tasks	finetune all questions one by one, finetune one model answering all questions, check duplicated rows that prepare dataset proposed, check ada instead of davinci
Tags	

eval on finetuned model:

annotated dataset split into 777 train vs 195 validation essay/question pairs.
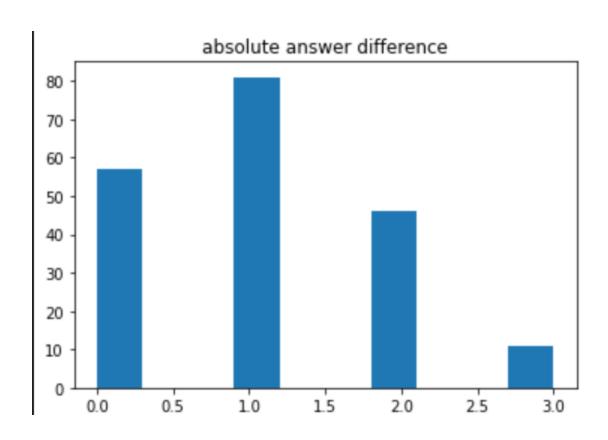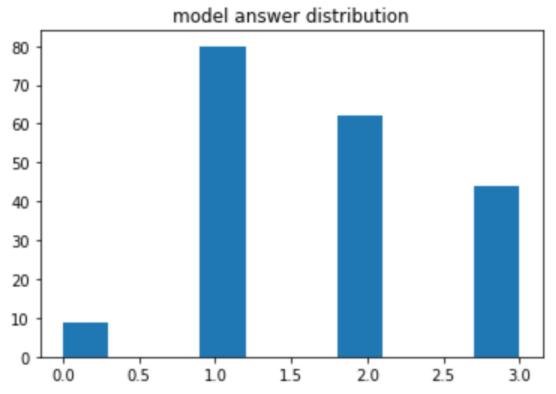
much better performance already.

1/3 answers are correct, 71% answers are 1 or less points off.
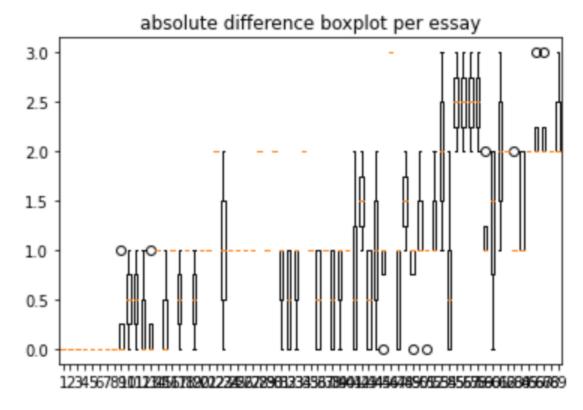
all medians are 1.

some questions seem to work better, some essays seem to be completely off (double check)
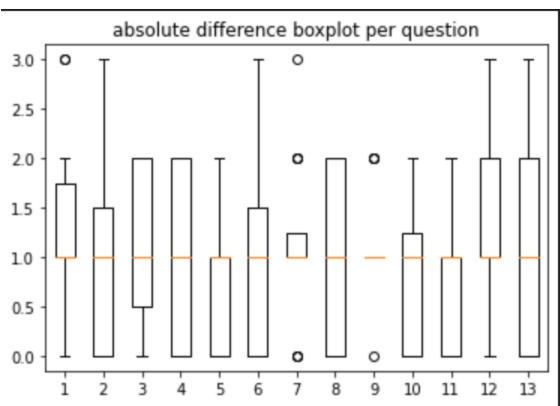
absolute answer difference

model answer distribution

absolute difference boxplot per essay

absolute difference boxplot per question