

UPLOAD DATASET FILE AND FORMAT IT IN THE WAY OPENAI WANTS

```
D:\USC\CSCI 544 NLP\Project>openai tools fine_tunes.prepare_data -f FINAL_DATASET.xlsx
Analyzing...
```

```
Based on the analysis we will perform the following actions:
- [Necessary] Your format 'XLSX' will be converted to 'JSONL'
- [Recommended] Add a suffix separator ' ->' to all prompts [Y/n]: Y
- [Recommended] Add a suffix ending '\n' to all completions [Y/n]: Y
- [Recommended] Add a whitespace character to the beginning of the completion [Y/n]: Y

Your data will be written to a new JSONL file. Proceed [Y/n]: Y

Wrote modified file to 'FINAL_DATASET_prepared.jsonl'
Feel free to take a look!
```

CREATE A FINE-TUNED MODEL FROM THE PREPARED JSONL DATASET FILE

```
> openai api fine_tunes.create -t "FINAL_DATASET_prepared.jsonl"
```

After you've fine-tuned a model, remember that your prompt has to end with the indicator string ' ->' for the model to start generating completions, rather than continuing with the prompt. Make sure to include 'stop=["\n"]' so that the generated texts ends at the expected place.

Once your model starts training, it'll approximately take 6.62 minutes to train a 'curie' model, and less for 'ada' and 'babbage'. Queue will approximately take half an hour per job ahead of you.

```
D:\USC\CSCI 544 NLP\Project>openai api fine_tunes.create -t FINAL_DATASET_prepared.jsonl -m ada
```

Upload progress: 100%

92.4k/92.4k [00:00<?, ?it/s]