

How to Create a Custom Fine-Tuned Large Language Model (LLM)

DECOTA's base model uses OpenAI's GPT-3.5. Currently, OpenAI does not allow fine-tuned models to be shared for free. However, this quick and easy tutorial will show you how to use our open-access free training data to create your own fine-tuned LLM, which you can then input into our open-access code.

In our open-access code, you will see that there are spaces to input the names of your models, as below.

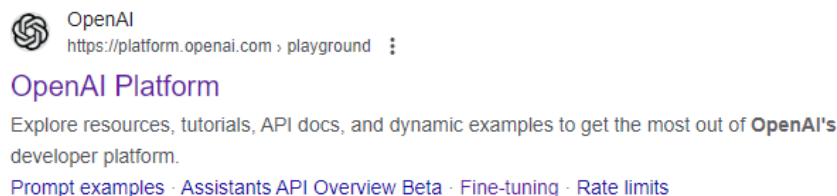
```
GPT_API_KEY = 'XXX'
Model1_Name = 'ft:gpt-3.5-turbo-0613:XXX'
Model2_Name = 'ft:gpt-3.5-turbo-0613:XXX'
```

Here, the “GPT_API_KEY” will be the ID number you receive when subscribing to OpenAI. To fine-tune a model with OpenAI, you need a subscription.

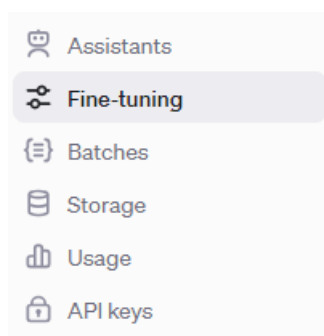
The Model1_Name and Model2_Name will then be the names of your fine-tuned models that this tutorial shows you how to create.

Instructions

1. In Google, search for “OpenAI playground”, and head to the below site (<https://platform.openai.com>):



2. Once there, navigate to “Fine-tuning” on the left hand side, and press the green “Create” button.



3. The below pop-up will appear. Select “gpt-3.5-turbo-0613” for the base model, and “upload new” under the training data section. The base models available for fine-tuning may change over

time, and using the most up-to-date version is recommended. You can name your model suffix something memorable, like model-1 for the first model.

Create a fine-tuned model

Base Model

gpt-3.5-turbo-0613

Training data

Add a jsonl file to use for training.

☒ Upload new ☐ Select existing



Upload a file or drag and drop here
(.jsonl)

Validation data

Add a jsonl file to use for validation metrics.

☒ Upload new ☐ Select existing ☐ None



Upload a file or drag and drop here
(.jsonl)

Suffix

Add a custom suffix that will be appended to the output model name.

model-1

Seed

The seed controls the reproducibility of the job. Passing in the same seed and job parameters should produce the same results, but may differ in rare cases. If a seed is not specified, one will be generated for you.

Random

4. Drag in our provided training data for Model 1 into the ‘training data’ box. Then drag the Model 1 validation data into the ‘validation data’ box. We have converted it into the desired ‘.jsonl’ format for ease, but you can see the raw training data in the open-access CSV files too.
5. Press “Create”. The model will then train using the fine-tuning data, and you will see your model appear in the “fine-tuning” section of the GPT playground. It will be in the format of ft:gpt-3.5-turbo-0613:XXX. These are the model names you will input into the code.

Player et al. (2024). The Use of Large Language Models for Qualitative Research (DECOTA).

```
GPT_API_KEY = 'XXX'  
Model1_Name = 'ft:gpt-3.5-turbo-0613:XXX'  
Model2_Name = 'ft:gpt-3.5-turbo-0613:XXX'
```

6. Repeat the same for Model 2, with the Model 2 training and validation data!