[1]:
```
!pip install --upgrade sagemaker datasets
```

```
Requirement already satisfied: sagemaker in /home/ec2-user/anaconda3/envs/pytorch_p310/lib/python3.10/site-packag
es (2.207.1)
Requirement already satisfied: datasets in /home/ec2-user/anaconda3/envs/pytorch_p310/lib/python3.10/site-package
s (2.17.0)
Requirement already satisfied: attrs<24,>=23.1.0 in /home/ec2-user/anaconda3/envs/pytorch_p310/lib/python3.10/sit
e-packages (from sagemaker) (23.1.0)
Requirement already satisfied: boto3<2.0,>=1.33.3 in /home/ec2-user/anaconda3/envs/pytorch_p310/lib/python3.10/si
te-packages (from sagemaker) (1.34.38)
Requirement already satisfied: cloudpickle==2.2.1 in /home/ec2-user/anaconda3/envs/pytorch_p310/lib/python3.10/si
te-packages (from sagemaker) (2.2.1)
Requirement already satisfied: google-pasta in /home/ec2-user/anaconda3/envs/pytorch_p310/lib/python3.10/site-pac
kages (from sagemaker) (0.2.0)
Requirement already satisfied: numpy<2.0,>=1.9.0 in /home/ec2-user/anaconda3/envs/pytorch_p310/lib/python3.10/sit
e-packages (from sagemaker) (1.26.1)
Requirement already satisfied: protobuf<5.0,>=3.12 in /home/ec2-user/anaconda3/envs/pytorch_p310/lib/python3.10/s
ite-packages (from sagemaker) (4.24.4)
Requirement already satisfied: smdebug-rulesconfig==1.0.1 in /home/ec2-user/anaconda3/envs/pytorch_p310/lib/pytho
n3.10/site-packages (from sagemaker) (1.0.1)
```

[2]:
```
from sagemaker.jumpstart.estimator import JumpStartEstimator
import boto3

estimator = JumpStartEstimator(model_id=model_id,  environment={"accept_eula": "true"},instance_type = "ml.g5.2xl

estimator.set_hyperparameters(instruction_tuned="False", epoch="5")

#Fill in the code below with the dataset you want to use from above
#example: estimator.fit({"training": f"s3://genaiwithawsproject2024/training-datasets/finance"})
estimator.fit({"training": f"s3://genaiwithawsproject2024/training-datasets/it"})
```

```
sagemaker.config INFO - Not applying SDK defaults from location: /etc/xdg/sagemaker/config.yaml
sagemaker.config INFO - Not applying SDK defaults from location: /home/ec2-user/.config/sagemaker/config.yaml
Using model 'meta-textgeneration-llama-2-7b' with wildcard version identifier '*'. You can pin to version '3.0.2'
for more stable results. Note that models may have different input/output signatures after a major version upgrad
e.
INFO:sagemaker:Creating training-job with name: meta-textgeneration-llama-2-7b-2024-02-14-18-09-11-963
2024-02-14 18:09:13 Starting - Starting the training job...
2024-02-14 18:09:35 Pending - Preparing the instances for training.........
2024-02-14 18:10:49 Downloading - Downloading input data.....................bash: cannot set terminal proce
ss group (-1): Inappropriate ioctl for device
bash: no job control in this shell
2024-02-14 18:15:21,975 sagemaker-training-toolkit INFO     Imported framework sagemaker_pytorch_container.traini
ng
2024-02-14 18:15:22,000 sagemaker-training-toolkit INFO     No Neurons detected (normal if no neurons installed)
```

[3]:
```
finetuned_predictor = estimator.deploy()
```

```
No instance type selected for inference hosting endpoint. Defaulting to ml.g5.2xlarge.
INFO:sagemaker.jumpstart:No instance type selected for inference hosting endpoint. Defaulting to ml.g5.2xlarge.
INFO:sagemaker:Creating model with name: meta-textgeneration-llama-2-7b-2024-02-14-18-24-13-322
INFO:sagemaker:Creating endpoint-config with name meta-textgeneration-llama-2-7b-2024-02-14-18-24-13-314
INFO:sagemaker:Creating endpoint with name meta-textgeneration-llama-2-7b-2024-02-14-18-24-13-314
-------!
```

[11]:
```
def print_response(payload, response):
    print(payload["inputs"])
    print(f"> {response[0]['generated_text']}")
    print("\n==================================\n")
```

Now we can run the same prompts on the fine-tuned model to evaluate it's domain knowledge.

```
payload = {
    "inputs": "Traditional approaches to data management such as",
    "parameters": {
        "max_new_tokens": 64,
        "top_p": 0.9,
        "temperature": 0.6,
        "return_full_text": False,
    },
}
try:
    response = finetuned_predictor.predict(payload, custom_attributes="accept_eula=true")
    print_response(payload, response)
except Exception as e:
    print(e)
```

```
Traditional approaches to data management such as
>  relational databases and enterprise data warehouses (EDWs) are reaching their limits, and organizations are increasingly turning to new technologie
s to get more value out of their data.
The first step to getting more value from your data is to understand what you have. That’s where data discovery

==================================
```

[7]:
```
finetuned_predictor.delete_model()
finetuned_predictor.delete_endpoint()
```

```
INFO:sagemaker:Deleting model with name: meta-textgeneration-llama-2-7b-2024-02-14-18-24-13-322
INFO:sagemaker:Deleting endpoint configuration with name: meta-textgeneration-llama-2-7b-2024-02-14-18-24-13-314
INFO:sagemaker:Deleting endpoint with name: meta-textgeneration-llama-2-7b-2024-02-14-18-24-13-314
```