Finetuning LLM from AI Quick Actions Catalog

Referenced Documentation

https://docs.oracle.com/en-us/iaas/data-science/using/ai-quick-actions-fine-tuning.htm

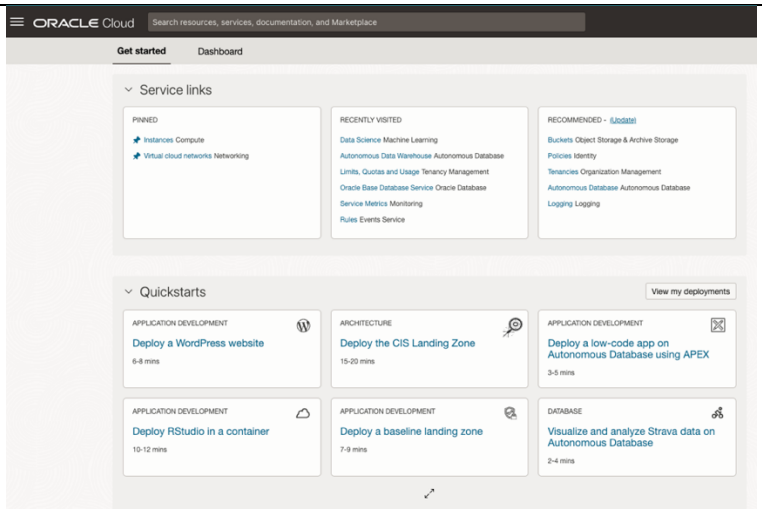
Description

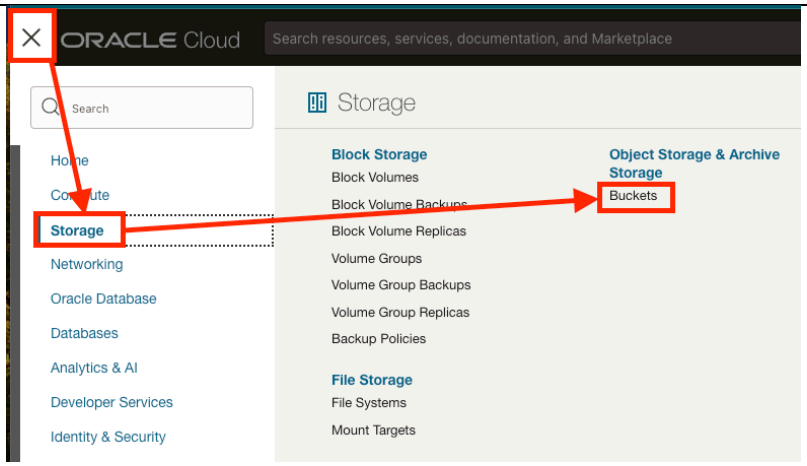
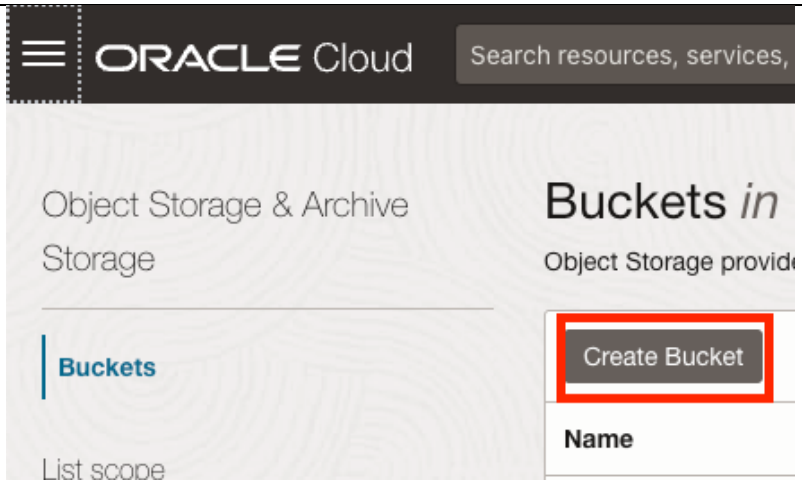
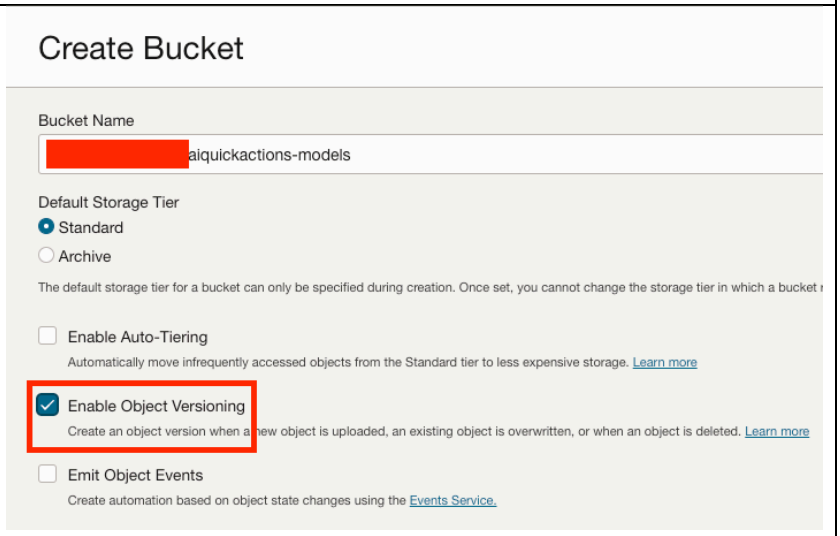
Fine-tuning is the process of taking a pretrained model and further training it on a domain-specific dataset to improve its knowledge and provide better responses in that domain. When you fine tune a model in AI quick actions, you're creating a Data Science job to do that. The dataset must be in JSONL format and must include the necessary 'prompt' and 'completion' columns. Optionally, you can include a 'category' column. If a dataset file with the same name already exists in the bucket, it's replaced by the new file. The dataset must contain a minimum of 100 records for fine-tuning.

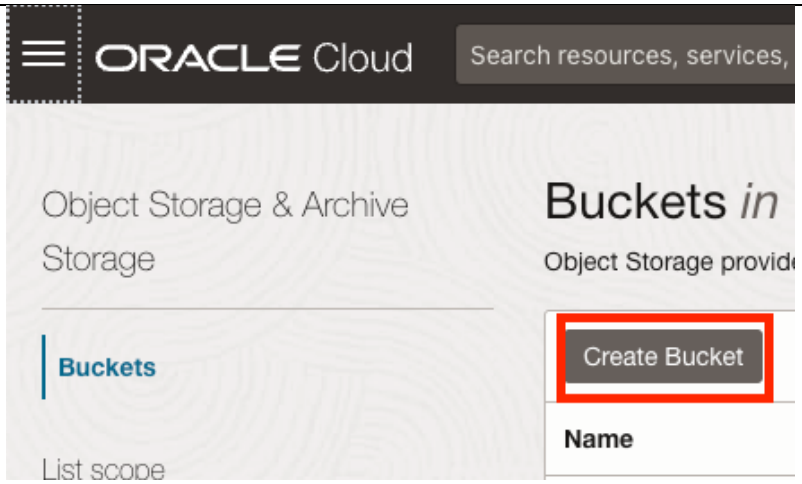
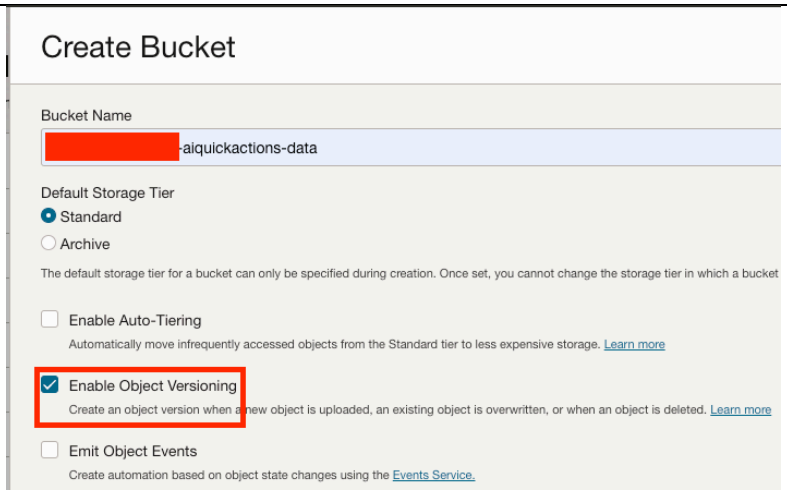
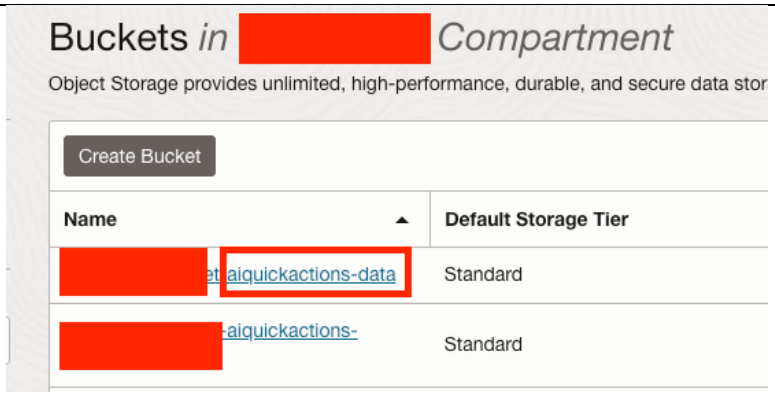
Pre-Requisites

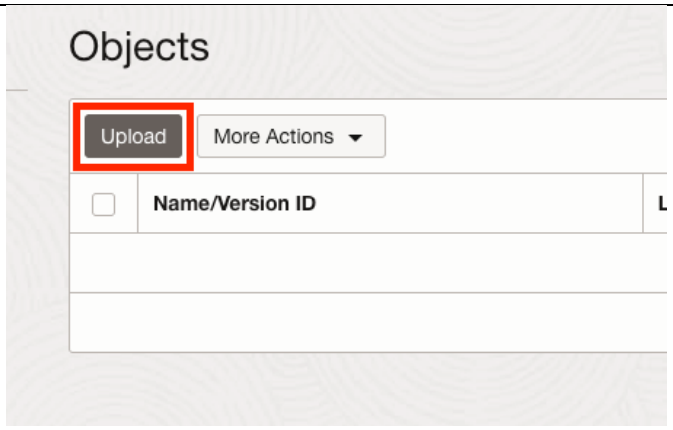
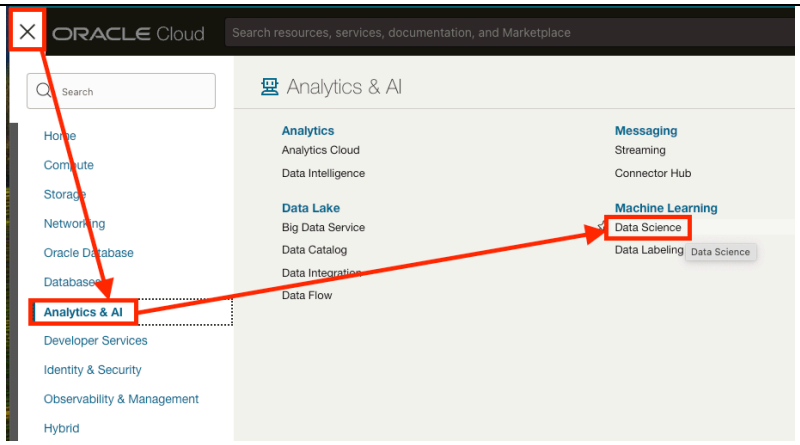
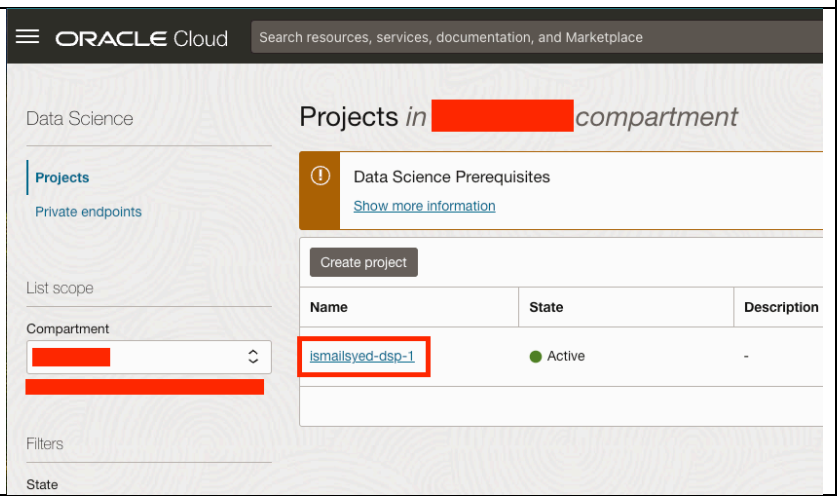
- Implement the required policies - https://docs.oracle.com/en-us/iaas/data-science/using/ai-quick-actions-set-up.htm
- Ensure you have your OCI Data Science GPU service limits raised for the GPU Shapes you plan to use. This can be done from OCI Console.
- Provisioned OCI Data Science Project and Notebook Session (Must be deactivated and reactivated if created before the policies where implemented).
- OCI Log Group & Log Created (Optional)

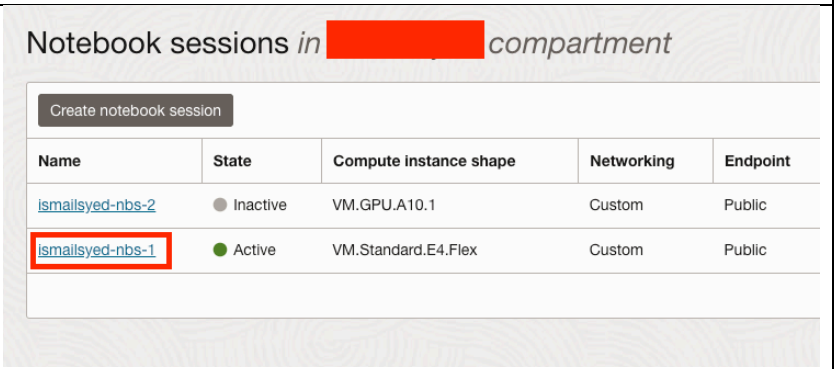
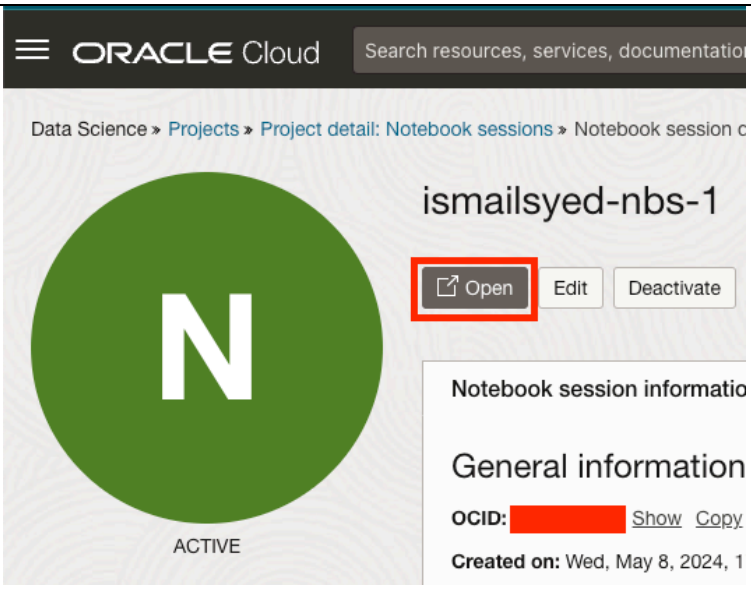
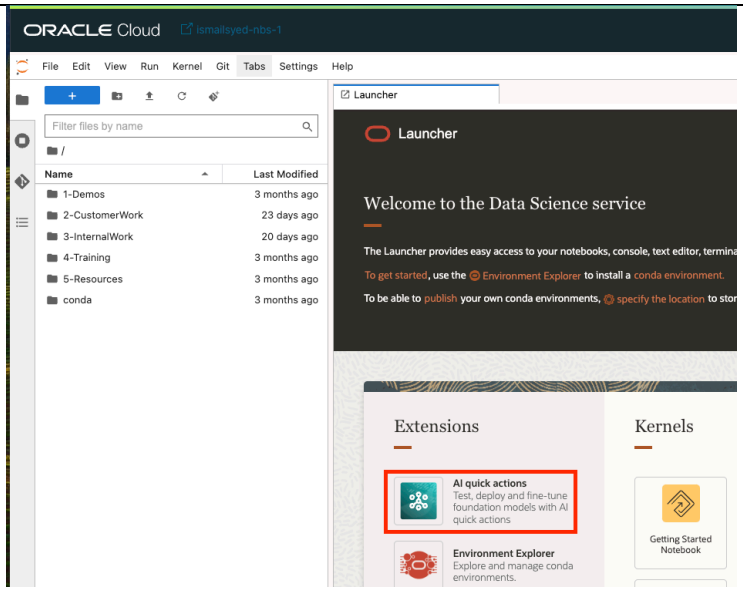
Guide

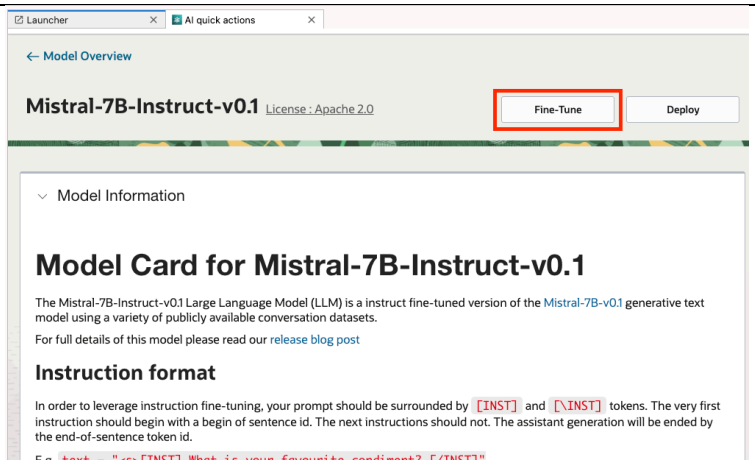
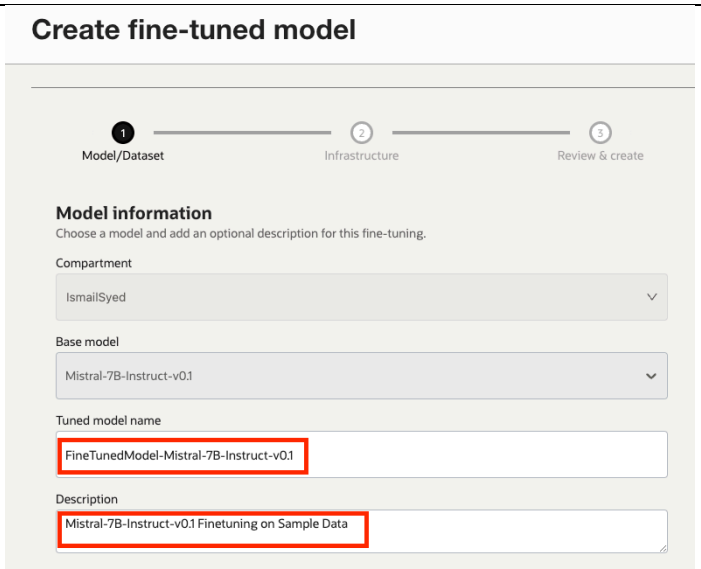
Step	Screenshot
Login to the Cloud Console. ***cloud.oracle.com***	

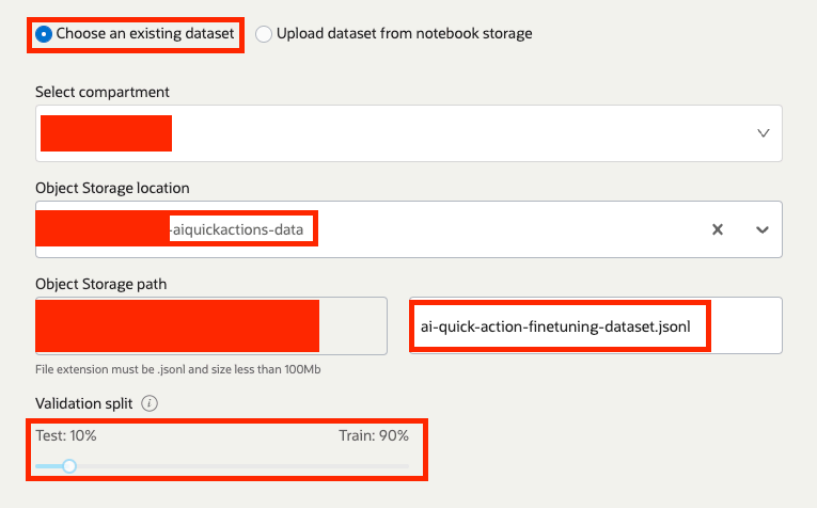
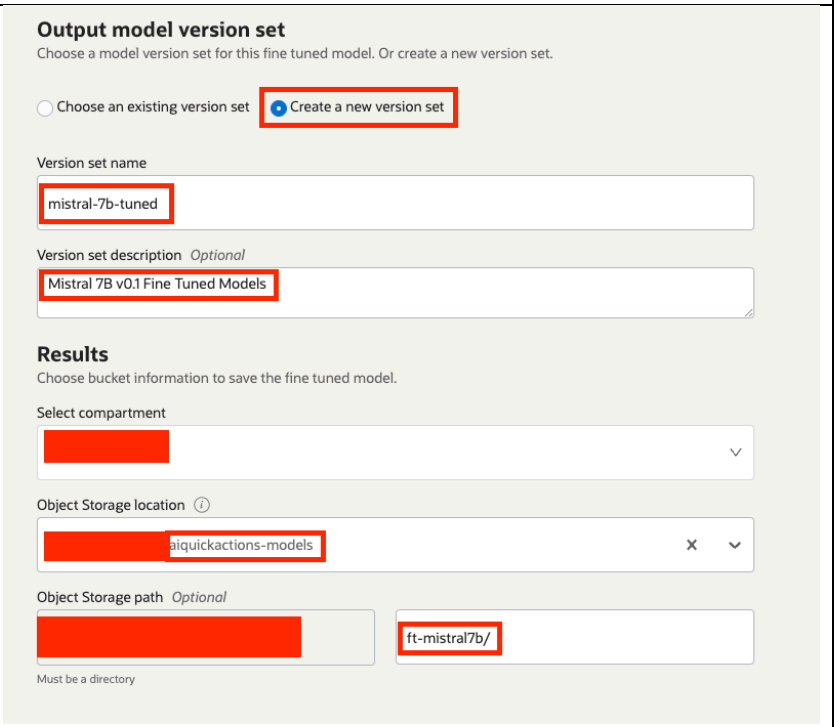
First, we will create two buckets to store our finetuning data and our finetuned model. ***Navigate to OCI Menu > Storage > Buckets.***	
Click Create Bucket.	
The first bucket we create will be for our finetuned model. ***Enter a Name.*** ***Enable Object Versioning.*** ***Click Create.***	

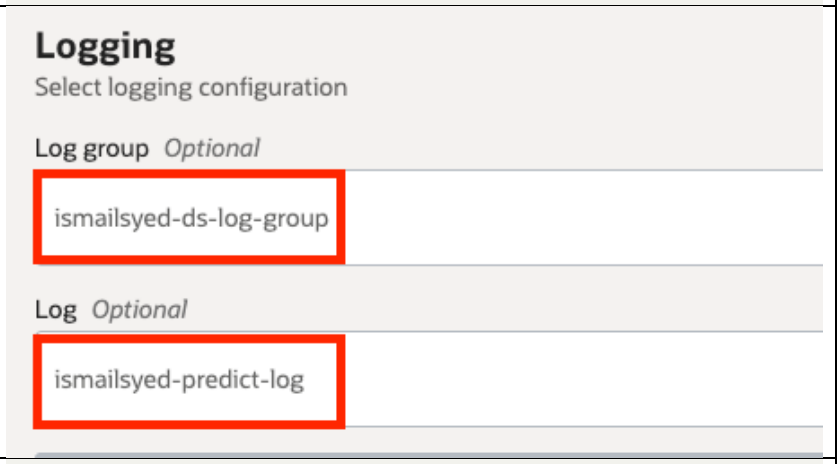
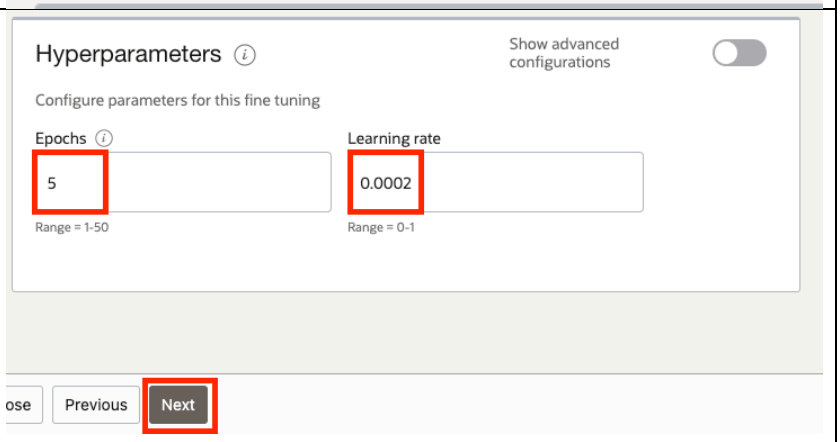
We will then create the bucket for our finetuning data. ***Click Create Bucket.***	
Enter a Name. ***Enable Object Versioning.*** ***Click Create.***	
Click on our Data bucket we just created.	

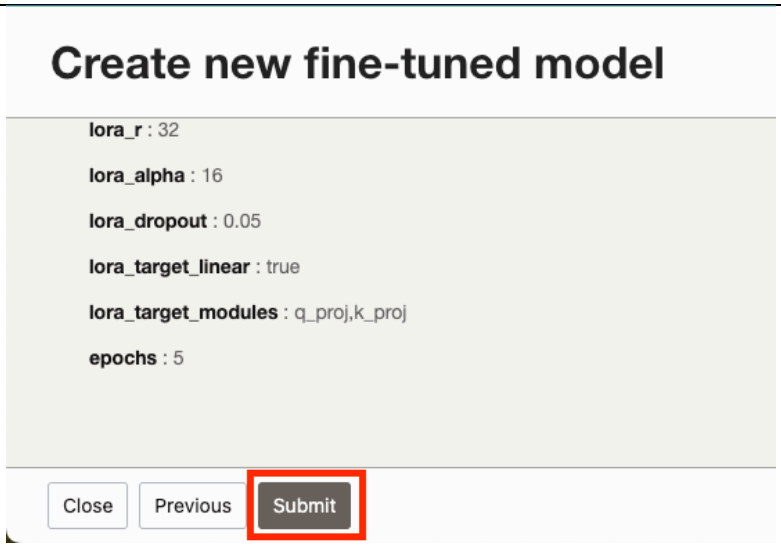
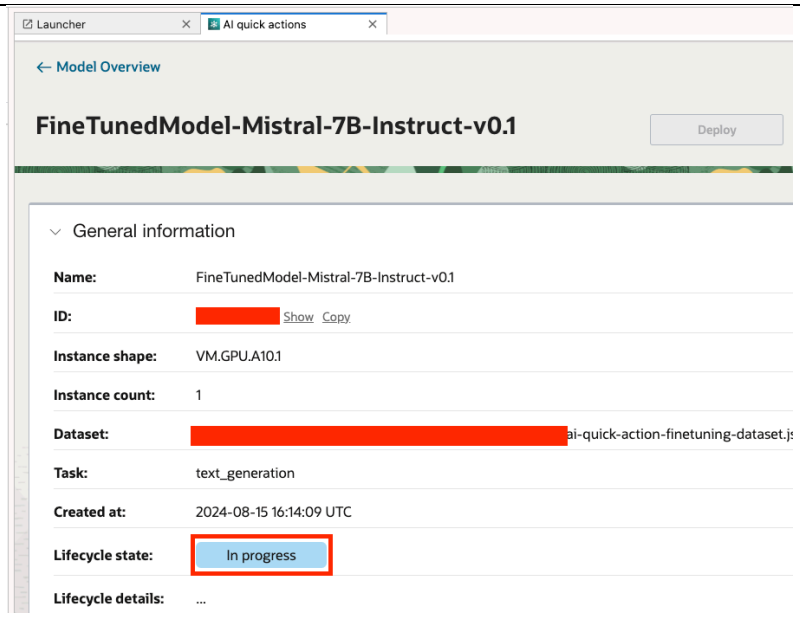
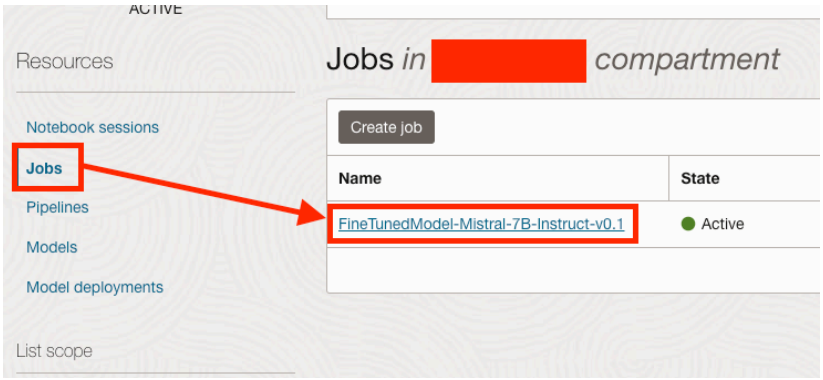
Click Upload.	
Upload the ***ai-quick-action-finetuning-dataset.jsonl*** Which should come with this guide. The dataset must be in ***jsonl*** format.	
Now navigate to your Data Science Projects. ***OCI Menu > Analytics & AI > Data Science.***	
Open up your existing Data Science Project.	

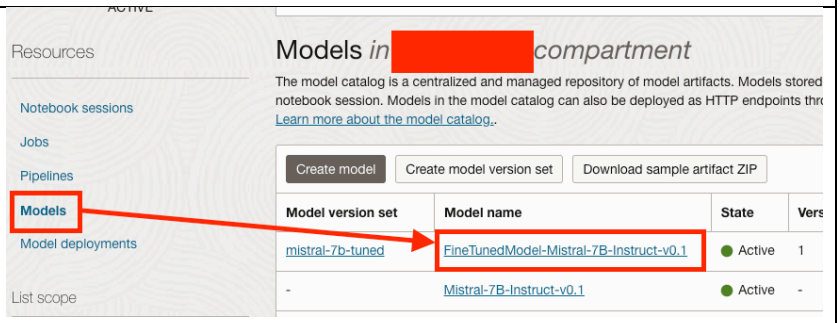
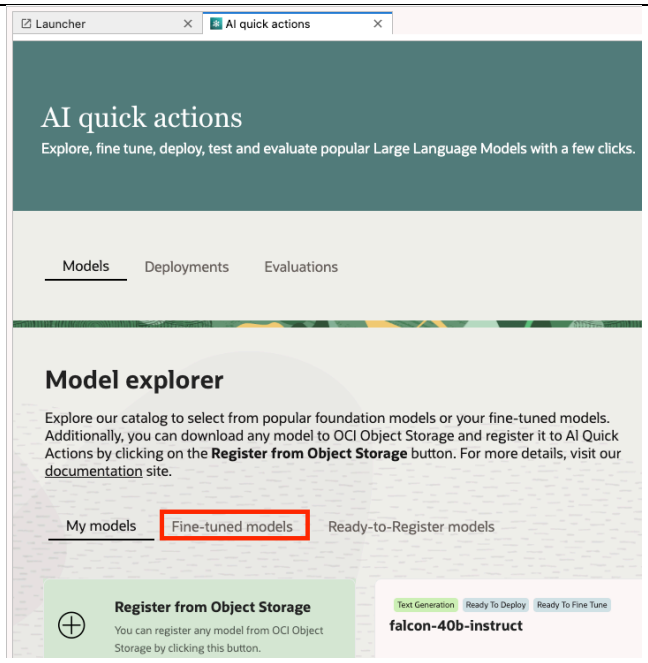
Click on your existing Data Science Notebook Session. Note – This does not have to be a GPU Shape.	
Click on Open. This will open up your Data Science Notebook Session. ***You will have to reauthenticate.***	
If the policies within the pre-requisites have been implemented correctly you should be able to open up the AI Quick Actions Extension within the Launcher. ***Click AI quick actions.***	

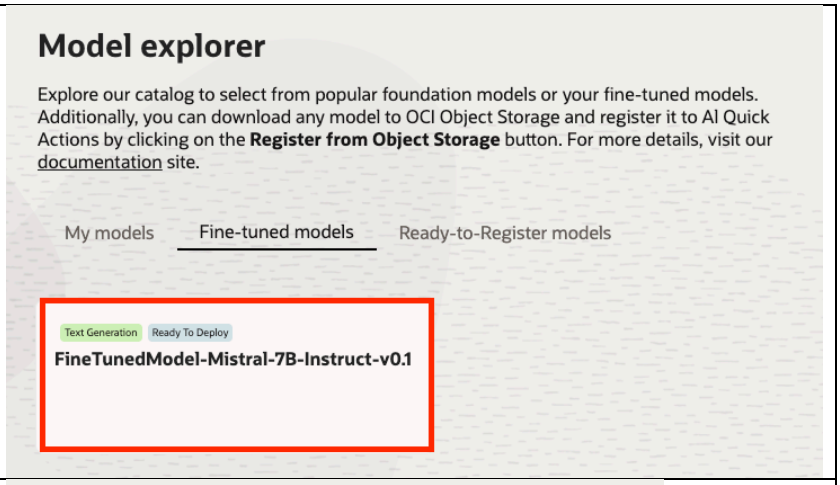
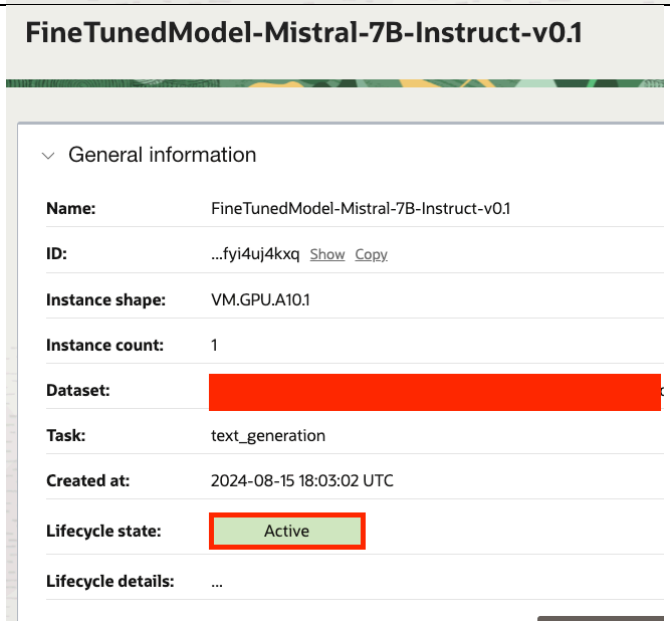
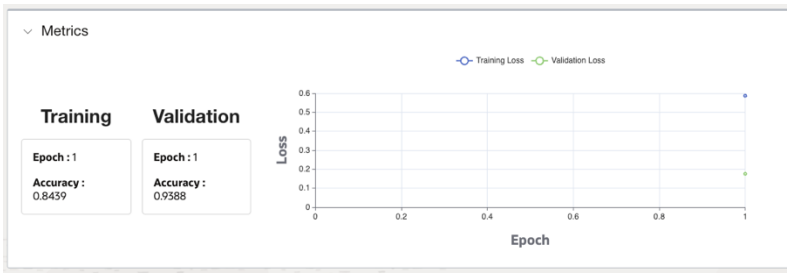
The AI Quick Actions Catalog will be displayed. In this tutorial we will fine tune the ***Mistral-7B-Instruct-v0.1 model***. ***Select this model.***	
We will then be taken to the Model card which gives you a bit of information about the Model. ***Click on Fine-Tune.***	
Give the fine-tuned model: ***Name*** ***Description*** ***Scroll Down***	

Then select our finetuning dataset. ***Select Choose an existing dataset.*** ***Select your Compartment.*** ***Select your Object Storage Bucket.*** ***Enter dataset path/name within the Bucket.*** ***Select your Train/Test Split.***	
Define details of Model Version Set within the Catalog. ***Specify a Version Set Name and Description.*** Select Object Storage location to save fine-tuned model outputs. ***Select Compartment.*** ***Select Object Storage Bucket.*** ***Specify results path within the Bucket.*** ***Click Next.***	

Specify the infrastructure shape to run the fine-tuning job. ***I have selected the default – VM.GPU.A10.1.***	
Select where the logging output should be stored. (This is optional). ***Specify Log Group and Log.***	
Set number of EPOCHS and Learning Rate. ***Epochs – 5*** ***Learning Rate – 0.0002*** You can also set advanced parameters under the ***Show advanced configurations*** option. ***Click Next.***	

Review all the options configured for the fine-tuning Job run. ***Click Submit.*** Behind the scenes this will provision the select infrastructure and kick start an OCI Data Science Job Run to execute the fine-tuning and save the tuned model within the OCI Data Science Model Catalog.	
The ***Lifecycle state*** will be updated to ***In progress*** while the fine tuning job executes.	
While the fine-tuning job is running, we can head back to the ***OCI Console and to our Data Science Project.*** If we select ***Jobs*** under ***Resources*** we can see our fine-tuning job. ***Click on in.***	

Here we can see our Job Run has been Accepted and will start executing.	
If we head back to our ***OCI Data Science Project***. Click on ***Models*** and we can see the entry for our fine-tuned model within the ***Model Catalog.***	
If we head back to our OCI Data Science Notebook Session and to the AI Quick Actions Explorer, we can select: ***Fine-tuned models***	

We can then select our Fine-Tuned Model.	
Here we can see the Lifecycle State has been updated to ***Active***.	
If you scroll down, you can see the fine-tuning metrics for training and validation. In the screenshot we only see 1 EPOCH, but you should see 5 along with how the loss has reduced for each EPOCH.	

If you scroll back to the top, you have the option to then deploy the fine-tuned model. See our Deployment Guide for steps.