

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>

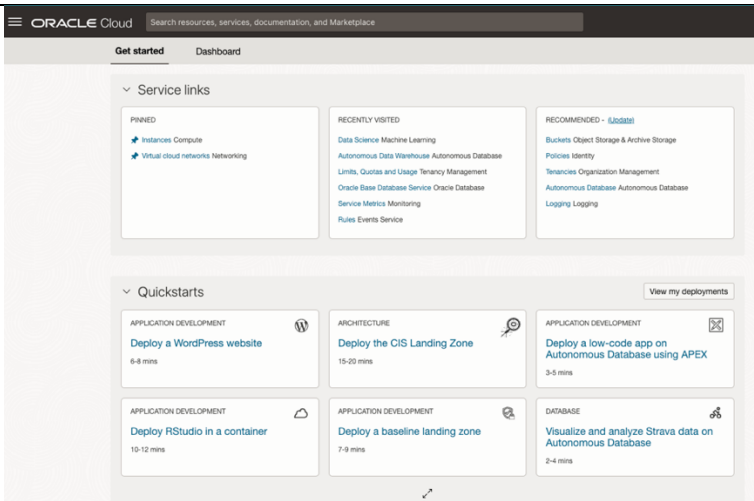
Description

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.

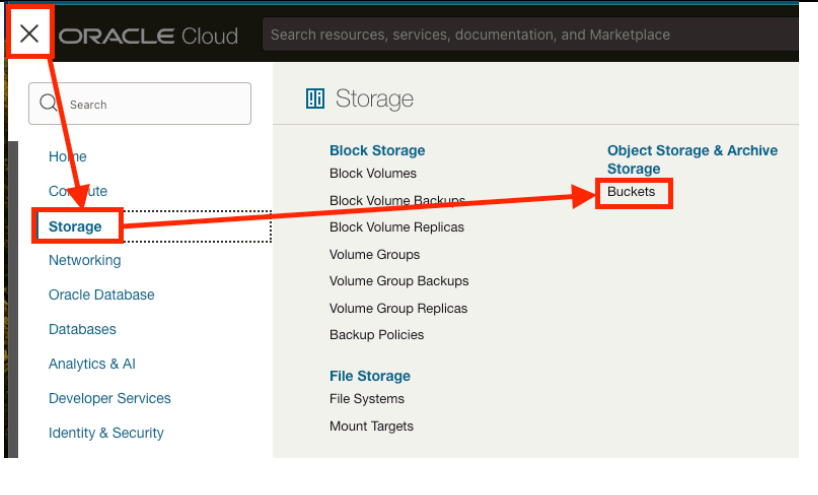
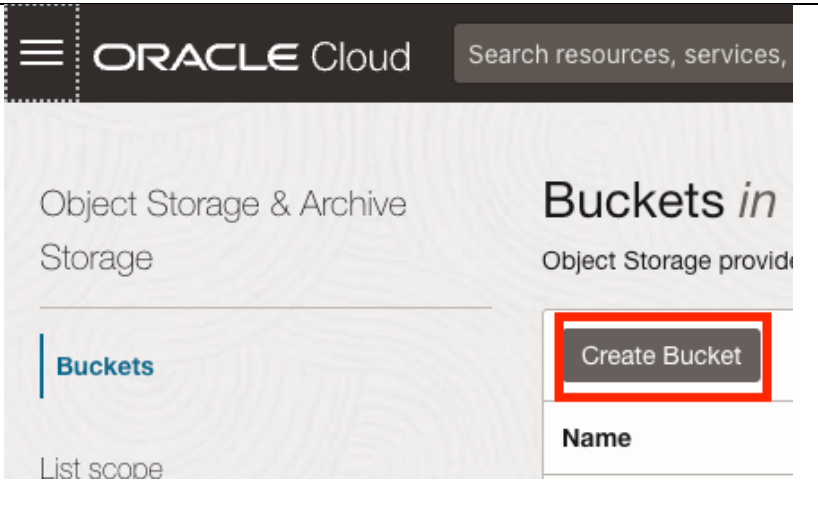
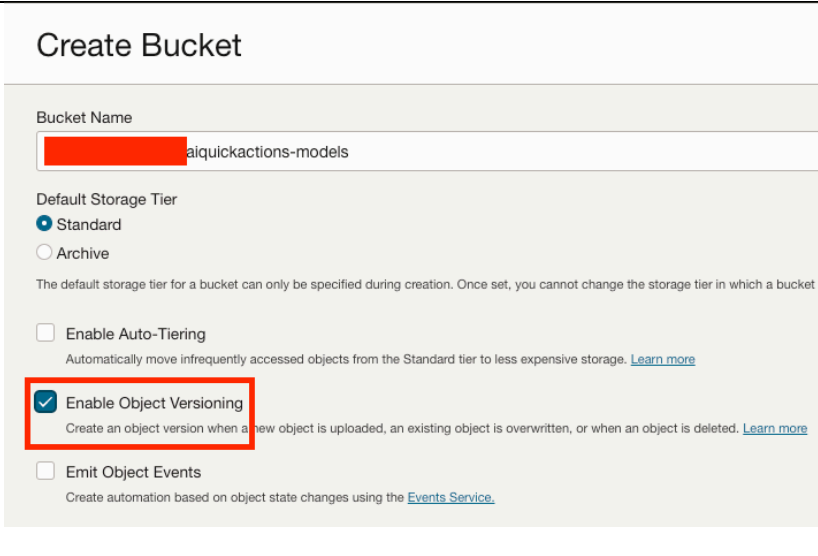
Pre-Requisites

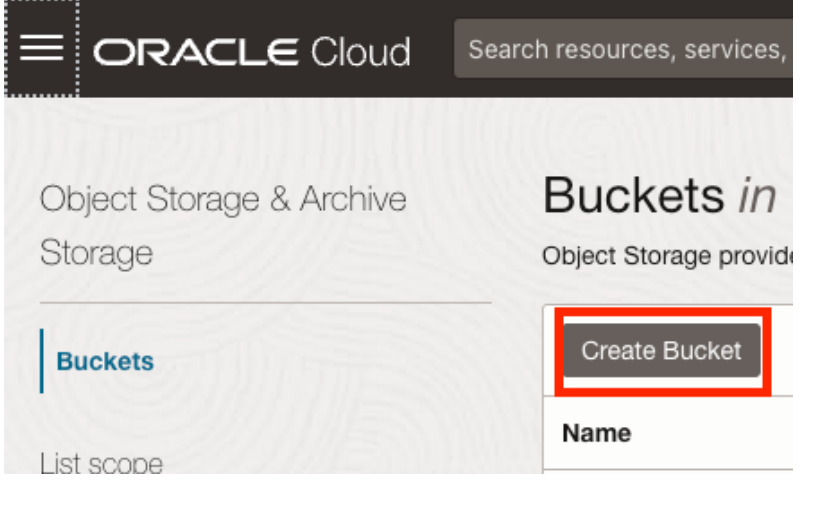
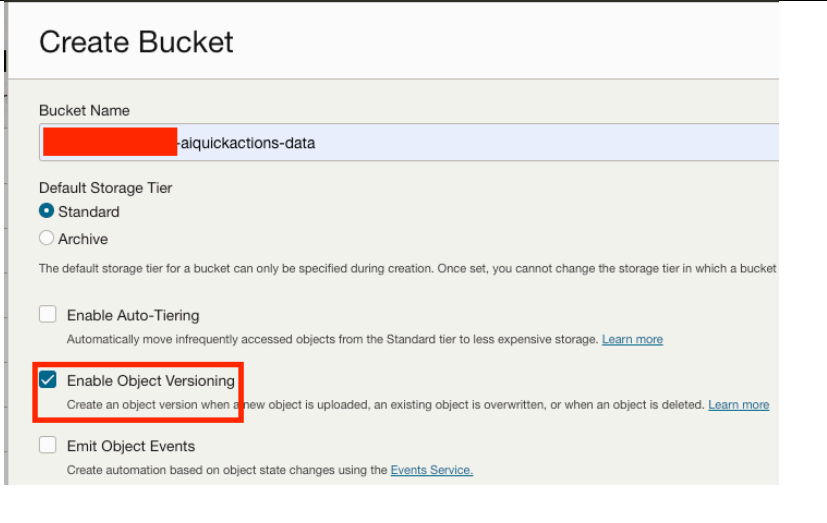
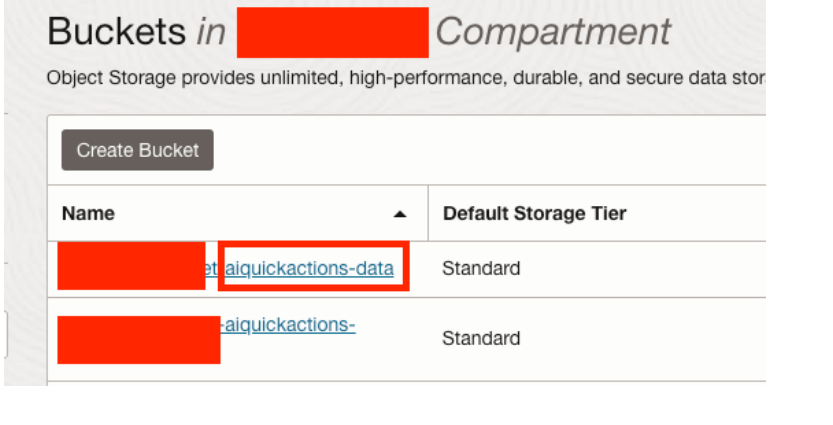
- Implement the required policies - <https://docs.oracle.com/en-us/iaas/data-science/using/ai-quick-actions-set-up.htm>
- Ensure you have your OCI Data Science GPU service limits raised for the GPU Shapes you plan to use. This can be done from OCI Console.
- Provisioned OCI Data Science Project and Notebook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

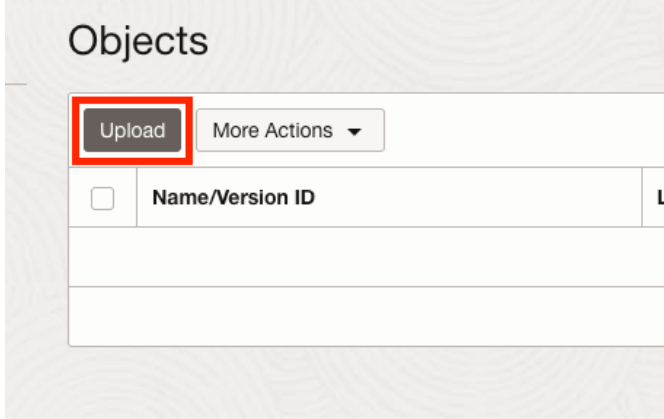
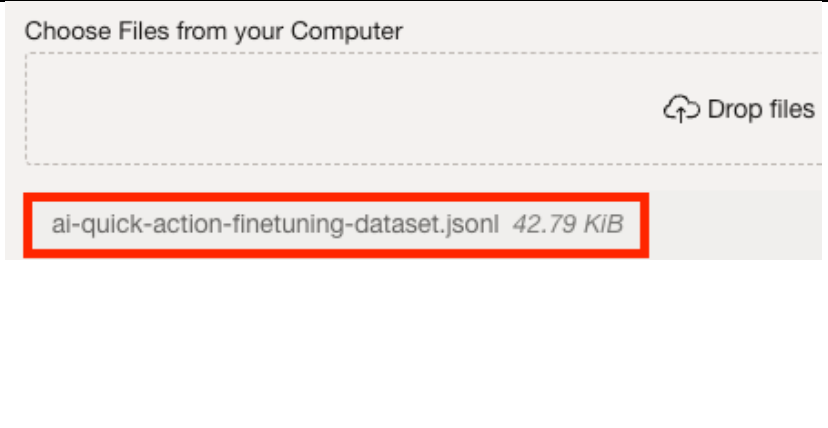
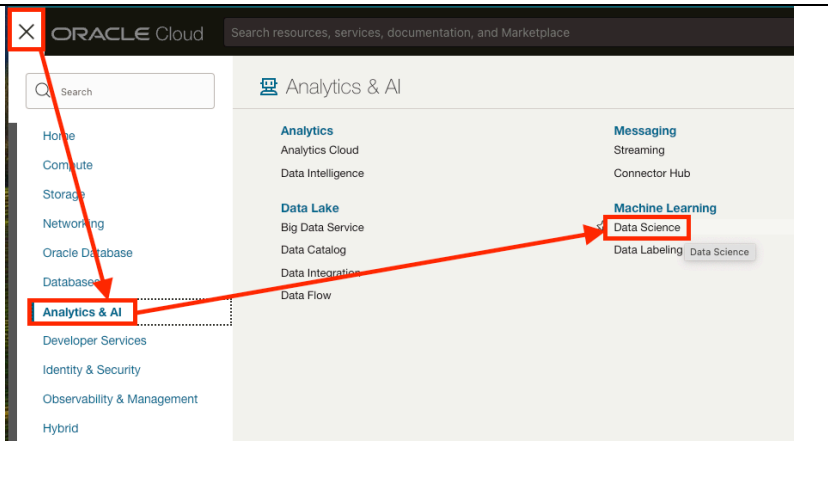
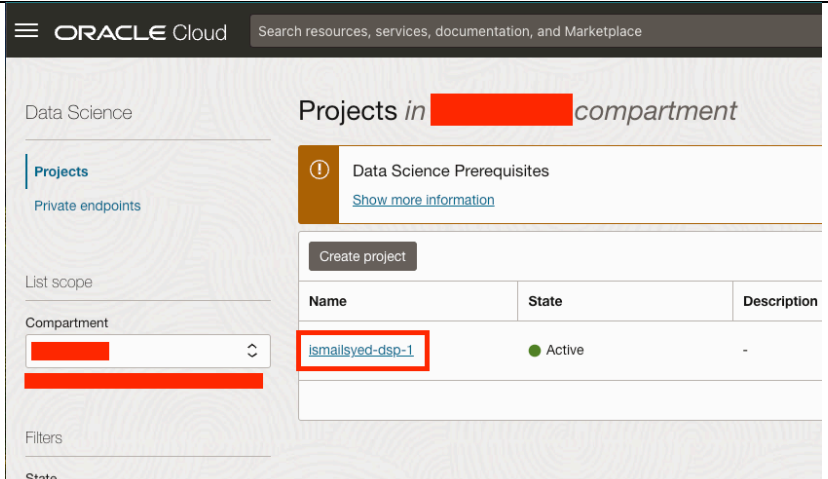
Guide

Step	Screenshot
<p>Login to the Cloud Console.</p> <p><i>cloud.oracle.com</i></p>	 <p>The screenshot shows the Oracle Cloud Dashboard. At the top, there's a search bar and navigation tabs for 'Get started' and 'Dashboard'. Below this, there are three main sections: 'Service links', 'Quickstarts', and 'View my deployments'. The 'Service links' section includes 'Pinned' links like 'Instances Compute' and 'Virtual cloud network Networking', 'Recently visited' links like 'Data Science Machine Learning' and 'Autonomous Data Warehouse Autonomous Database', and 'Recommended' links like 'Buckets Object Storage & Archive Storage'. The 'Quickstarts' section features several cards for tasks like 'Deploy a WordPress website', 'Deploy the CIS Landing Zone', 'Deploy a low-code app on Autonomous Database using APEX', 'Deploy RStudio in a container', 'Deploy a baseline landing zone', and 'Visualize and analyze Strava data on Autonomous Database'.</p>



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

<p>We will then create the bucket for our finetuning data.</p> <p>Click Create Bucket.</p>	 <p>The screenshot shows the Oracle Cloud console interface. At the top, there's a navigation bar with the Oracle Cloud logo and a search bar. Below it, the main heading is 'Object Storage & Archive Storage'. To the right, it says 'Buckets in Object Storage provides'. A 'Create Bucket' button is highlighted with a red box. Below the button, there's a 'Name' field.</p>						
<p>Enter a Name.</p> <p>Enable Object Versioning.</p> <p>Click Create.</p>	 <p>The screenshot shows the 'Create Bucket' form. The 'Bucket Name' field is filled with 'aiquickactions-data'. The 'Default Storage Tier' is set to 'Standard'. The 'Enable Object Versioning' checkbox is checked and highlighted with a red box. Below it, there's a description: 'Create an object version when a new object is uploaded, an existing object is overwritten, or when an object is deleted. Learn more'. There are also checkboxes for 'Enable Auto-Tiering' and 'Emit Object Events'.</p>						
<p>Click on our Data bucket we just created.</p>	 <p>The screenshot shows the 'Buckets in [redacted] Compartment' page. It says 'Object Storage provides unlimited, high-performance, durable, and secure data storage'. A 'Create Bucket' button is highlighted with a red box. Below it, there's a table with two columns: 'Name' and 'Default Storage Tier'. The first row shows a bucket named 'aiquickactions-data' with a 'Standard' storage tier. The second row shows a bucket named 'aiquickactions-' with a 'Standard' storage tier.</p> <table border="1"><thead><tr><th>Name</th><th>Default Storage Tier</th></tr></thead><tbody><tr><td>aiquickactions-data</td><td>Standard</td></tr><tr><td>aiquickactions-</td><td>Standard</td></tr></tbody></table>	Name	Default Storage Tier	aiquickactions-data	Standard	aiquickactions-	Standard
Name	Default Storage Tier						
aiquickactions-data	Standard						
aiquickactions-	Standard						

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

<p>Click on your existing Data Science Notebook Session.</p> <p>Note – This does not have to be a GPU Shape.</p>	<div>Notebook sessions in [redacted] compartment</div> <div><div>Create notebook session</div><table><thead><tr><th>Name</th><th>State</th><th>Compute instance shape</th><th>Networking</th><th>Endpoint</th></tr></thead><tbody><tr><td>ismailsyed-nbs-2</td><td>Inactive</td><td>VM.GPU.A10.1</td><td>Custom</td><td>Public</td></tr><tr><td>ismailsyed-nbs-1</td><td>Active</td><td>VM.Standard.E4.Flex</td><td>Custom</td><td>Public</td></tr></tbody></table></div>	Name	State	Compute instance shape	Networking	Endpoint	ismailsyed-nbs-2	Inactive	VM.GPU.A10.1	Custom	Public	ismailsyed-nbs-1	Active	VM.Standard.E4.Flex	Custom	Public
Name	State	Compute instance shape	Networking	Endpoint												
ismailsyed-nbs-2	Inactive	VM.GPU.A10.1	Custom	Public												
ismailsyed-nbs-1	Active	VM.Standard.E4.Flex	Custom	Public												
<p>Click on Open.</p> <p>This will open up your Data Science Notebook Session.</p> <p>You will have to reauthenticate.</p>	<div><div><div>ORACLE Cloud</div><div>Search resources, services, documentation</div></div><div>Data Science » Projects » Project detail: Notebook sessions » Notebook session c</div><div><div><div>N</div><div>ACTIVE</div></div><div><div><div>ismailsyed-nbs-1</div><div><div><div>Open</div><div>Edit</div><div>Deactivate</div></div></div></div></div><div><div>Notebook session informatio</div><div>General information</div><div>OCID: [redacted] Show Copy</div><div>Created on: Wed, May 8, 2024, 1</div></div></div></div>															
<p>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.</p> <p>Click AI quick actions.</p>	<div><div><div>ORACLE Cloud</div><div>ismailsyed-nbs-1</div></div><div><div>File Edit View Run Kernel Git Tabs Settings Help</div><div><div>Filter files by name</div><div><table><thead><tr><th>Name</th><th>Last Modified</th></tr></thead><tbody><tr><td>1-Demos</td><td>3 months ago</td></tr><tr><td>2-CustomerWork</td><td>23 days ago</td></tr><tr><td>3-InternalWork</td><td>20 days ago</td></tr><tr><td>4-Training</td><td>3 months ago</td></tr><tr><td>5-Resources</td><td>3 months ago</td></tr><tr><td>conda</td><td>3 months ago</td></tr></tbody></table></div></div></div></div> <div><div>Launcher</div><div>Welcome to the Data Science service</div><div>The Launcher provides easy access to your notebooks, console, text editor, termin</div><div>To get started, use the Environment Explorer to install a conda environment.</div><div>To be able to publish your own conda environments, specify the location to stor</div><div><div>Extensions</div><div><div><div></div><div>AI quick actions</div><div>Test, deploy and fine-tune foundation models with AI quick actions</div></div><div><div></div><div>Environment Explorer</div><div>Explore and manage conda environments.</div></div></div><div><div>Kernels</div><div><div></div><div>Getting Started Notebook</div></div></div></div></div>	Name	Last Modified	1-Demos	3 months ago	2-CustomerWork	23 days ago	3-InternalWork	20 days ago	4-Training	3 months ago	5-Resources	3 months ago	conda	3 months ago	
Name	Last Modified															
1-Demos	3 months ago															
2-CustomerWork	23 days ago															
3-InternalWork	20 days ago															
4-Training	3 months ago															
5-Resources	3 months ago															
conda	3 months ago															

The AI Quick Actions Catalog will be displayed.

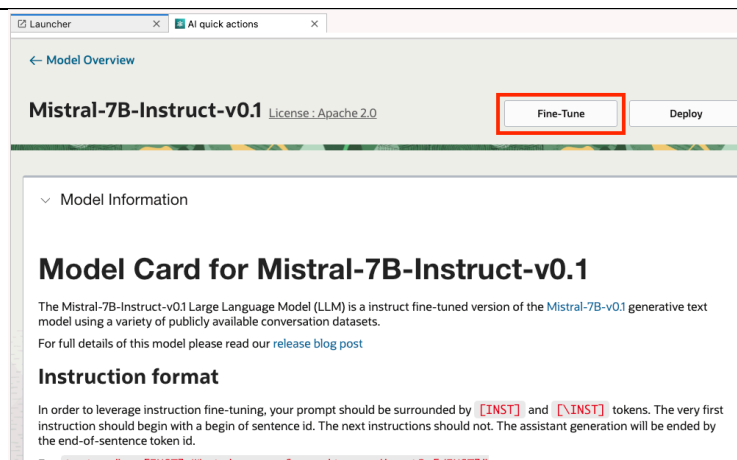
In this tutorial we will fine tune the **Mistral-7B-Instruct-v0.1** model.

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

Create fine-tuned model

A screenshot of the 'Create fine-tuned model' form. The form has a progress bar at the top with three steps: 'Model/Dataset', 'Infrastructure', and 'Review & create'. The 'Model information' section includes a 'Compartment' dropdown (set to 'IsmailSyed'), a 'Base model' dropdown (set to 'Mistral-7B-Instruct-v0.1'), a 'Tuned model name' text input field (containing 'FineTunedModel-Mistral-7B-Instruct-v0.1'), and a 'Description' text input field (containing 'Mistral-7B-Instruct-v0.1 Finetuning on Sample Data'). The 'Tuned model name' and 'Description' fields are highlighted with red rectangular borders.

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.

☒ Choose an existing dataset ☐ Upload dataset from notebook storage

Select compartment

[Redacted]

Object Storage location

[Redacted] aiquickactions-data

Object Storage path

[Redacted]

ai-quick-action-finetuning-dataset.jsonl

File extension must be .jsonl and size less than 100Mb

Validation split ⓘ

Test: 10%

Train: 90%

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.

Output model version set

Choose a model version set for this fine tuned model. Or create a new version set.

☐ Choose an existing version set ☒ Create a new version set

Version set name

mistral-7b-tuned

Version set description *Optional*

Mistral 7B v0.1 Fine Tuned Models

Results

Choose bucket information to save the fine tuned model.

Select compartment

[Redacted]

Object Storage location ⓘ

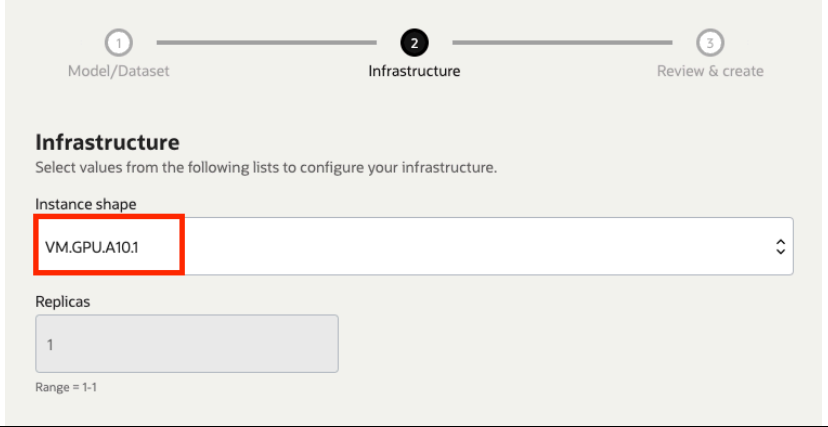
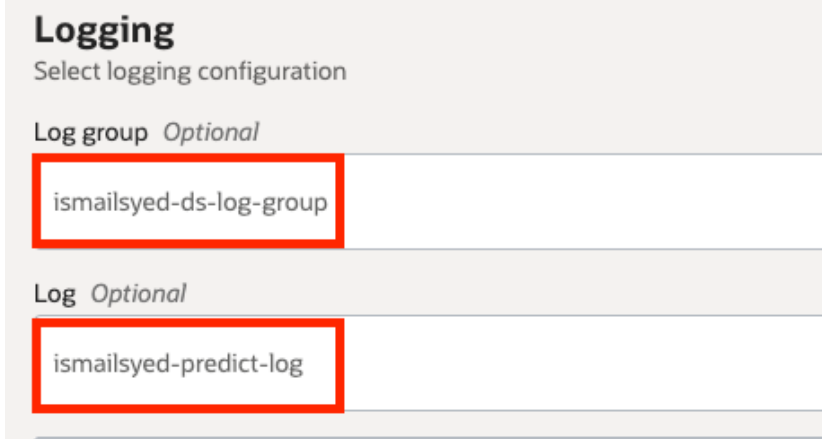
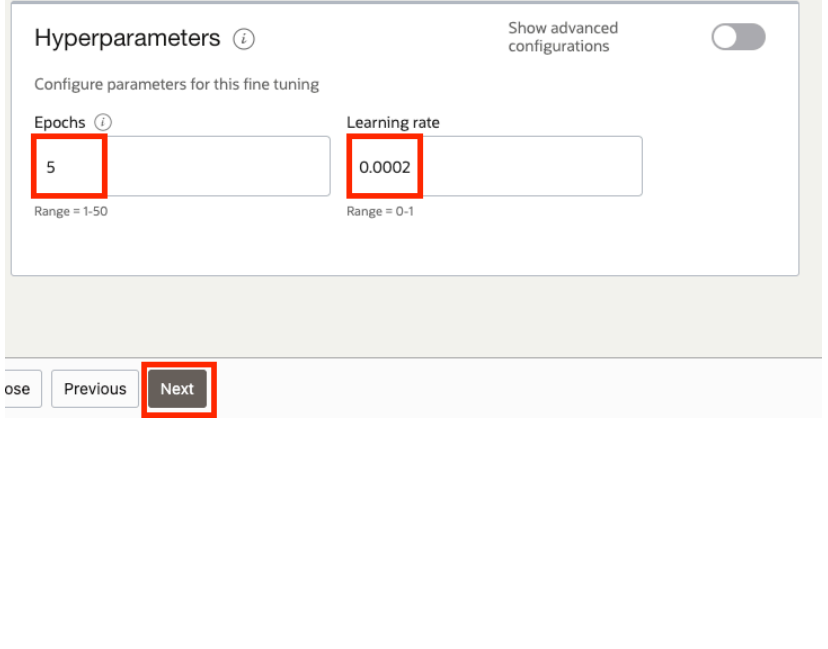
[Redacted] aiquickactions-models

Object Storage path *Optional*

[Redacted]

ft-mistral7b/

Must be a directory

<p>Specify the infrastructure shape to run the fine-tuning job.</p> <p><i>I have selected the default – VM.GPU.A10.1.</i></p>	 <p>1 Model/Dataset 2 Infrastructure 3 Review & create</p> <h3>Infrastructure</h3> <p>Select values from the following lists to configure your infrastructure.</p> <p>Instance shape</p> <p>VM.GPU.A10.1</p> <p>Replicas</p> <p>1</p> <p>Range = 1-1</p>
<p>Select where the logging output should be stored. (This is optional).</p> <p><i>Specify Log Group and Log.</i></p>	 <h3>Logging</h3> <p>Select logging configuration</p> <p>Log group <i>Optional</i></p> <p>ismailsyed-ds-log-group</p> <p>Log <i>Optional</i></p> <p>ismailsyed-predict-log</p>
<p>Set number of EPOCHS and Learning Rate.</p> <p><i>Epochs – 5</i></p> <p><i>Learning Rate – 0.0002</i></p> <p>You can also set advanced parameters under the Show advanced configurations option.</p> <p><i>Click Next.</i></p>	 <h3>Hyperparameters</h3> <p>Configure parameters for this fine tuning</p> <p>Epochs</p> <p>5</p> <p>Range = 1-50</p> <p>Learning rate</p> <p>0.0002</p> <p>Range = 0-1</p> <p>Previous Next</p>



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.

Create new fine-tuned model

`lora_r` : 32
`lora_alpha` : 16
`lora_dropout` : 0.05
`lora_target_linear` : true
`lora_target_modules` : q_proj,k_proj
`epochs` : 5

[Close](#)[Previous](#)[Submit](#)

The **Lifecycle state** will be updated to **In progress** while the fine tuning job executes.

Launcher x AI quick actions x

← Model Overview

FineTunedModel-Mistral-7B-Instruct-v0.1 [Deploy](#)

General information

Name: FineTunedModel-Mistral-7B-Instruct-v0.1

ID: [redacted] [Show](#) [Copy](#)

Instance shape: VM.GPU.A10.1

Instance count: 1

Dataset: [redacted] ai-quick-action-finetuning-dataset.j

Task: text_generation

Created at: 2024-08-15 16:14:09 UTC

Lifecycle state: **In progress**

Lifecycle details: ...

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.

ACTIVE

Resources

Notebook sessions

Jobs

Pipelines

Models

Model deployments

List scope

Jobs in [redacted] compartment

Create job

Name	State
FineTunedModel-Mistral-7B-Instruct-v0.1	Active

Here we can see our Job Run has been Accepted and will start executing.

Job runs in [REDACTED] compartment

Start a job run

Name	State	Lifecycle details
FineTunedModel-Mistral-7B-Instruct-v0.1-0	Accepted	Infrastructure provisioning.

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**.

Resources

- Notebook sessions
- Jobs
- Pipelines
- Models**
- Model deployments
- List scope

Models in [REDACTED] compartment

The model catalog is a centralized and managed repository of model artifacts. Models stored notebook session. Models in the model catalog can also be deployed as HTTP endpoints through [Learn more about the model catalog.](#)

Create model Create model version set Download sample artifact ZIP

Model version set	Model name	State	Vers
mistral-7b-tuned	FineTunedModel-Mistral-7B-Instruct-v0.1	Active	1
-	Mistral-7B-Instruct-v0.1	Active	-

If we head back to our OCI Data Science Notebook Session and to the AI Quick Actions Explorer, we can select:

Fine-tuned models

Launcher AI quick actions

Explore, fine tune, deploy, test and evaluate popular Large Language Models with a few clicks.

Models Deployments Evaluations

Model explorer

Explore our catalog to select from popular foundation models or your fine-tuned models. Additionally, you can download any model to OCI Object Storage and register it to AI Quick Actions by clicking on the **Register from Object Storage** button. For more details, visit our [documentation](#) site.

My models **Fine-tuned models** Ready-to-Register models

Register from Object Storage
You can register any model from OCI Object Storage by clicking this button.

Text Generation Ready To Deploy Ready To Fine Tune
falcon-40b-instruct

We can then select our Fine-Tuned Model.

Model explorer

Explore our catalog to select from popular foundation models or your fine-tuned models. Additionally, you can download any model to OCI Object Storage and register it to AI Quick Actions by clicking on the **Register from Object Storage** button. For more details, visit our [documentation](#) site.

My models

Fine-tuned models

Ready-to-Register models

Text Generation

Ready To Deploy

FineTunedModel-Mistral-7B-Instruct-v0.1

Here we can see the Lifecycle State has been updated to **Active**.

FineTunedModel-Mistral-7B-Instruct-v0.1

General information

Name: FineTunedModel-Mistral-7B-Instruct-v0.1

ID: ...fyi4uj4kxq [Show](#) [Copy](#)

Instance shape: VM.GPU.A10.1

Instance count: 1

Dataset:

Task: text_generation

Created at: 2024-08-15 18:03:02 UTC

Lifecycle state: **Active**

Lifecycle details: ...

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.

Metrics

Training

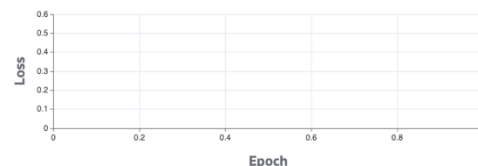
Validation

Epoch : 1

Accuracy : 0.8439

Epoch : 1

Accuracy : 0.9388



If you scroll back to the top, you have the option to then deploy the fine-tuned model.

See our Deployment Guide for steps.

