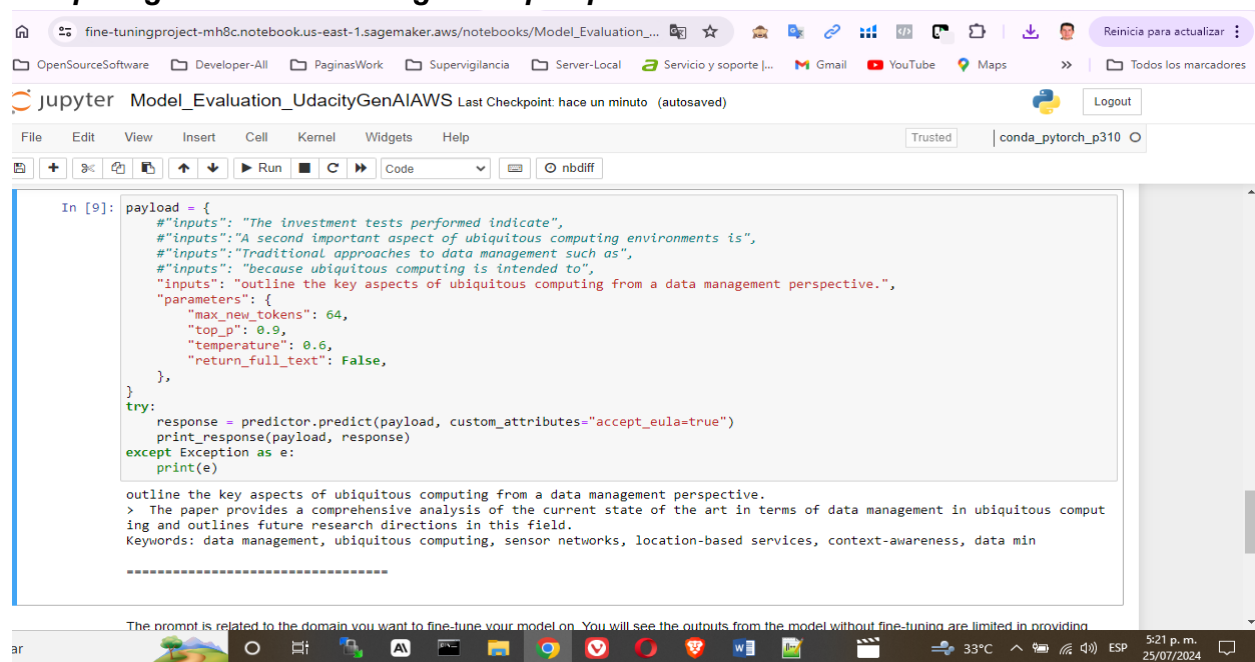# UDACITY

Introduction to Generative AI with AWS

Project Documentation Report - Screenshots of both notebooks with cell output

Visit [UDACITY Introduction to Generative AI with AWS Project Documentation Report](#) to make a copy of this document.

---

1. Take a screenshot of the **Model_Evaluation.ipynb** file with the cell output as proof you completed this step of the project

We can see the cell output for the input: *'outline the key aspects of ubiquitous computing from a data management perspective"*
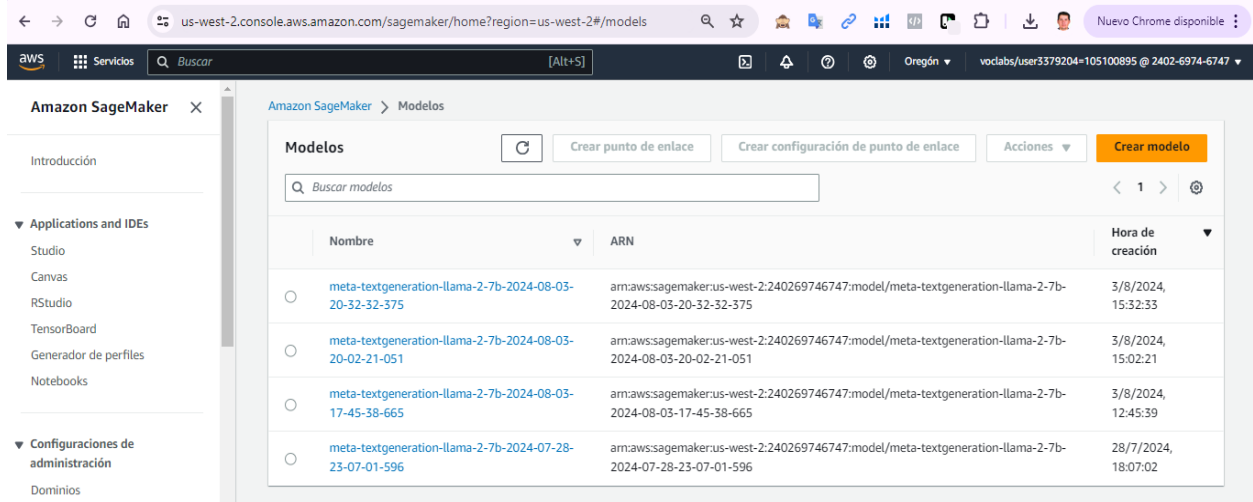
| 1A. Screenshot demonstrating the deployment of the model in the SageMaker environment. |
| --- |

IMPORTANT:
The job about **Model_Evaluation**, was run with my Udacity User: *voclabs/user3379204=105100895*

2. Take a screenshot of the **Model_FineTuning.ipynb** file with the cell output as proof you completed this step of the project

For this notebook, We can see the cell output for the input: *'outline the key aspects of ubiquitous computing from a data management perspective"*
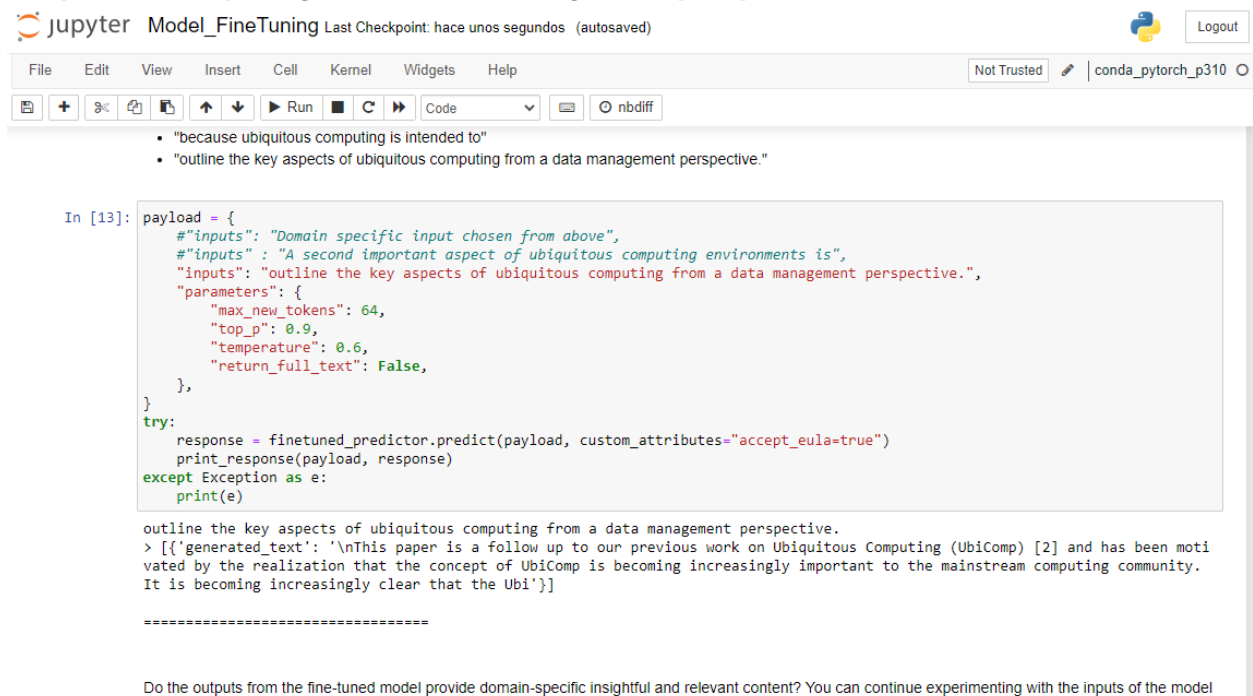
2.A Deploying the model in notebook


```
Processing ./lib/loralib/loralib-0.1.1-py3-none-any.whl (from -r requirements.txt (line 10))
Processing ./lib/multivolumefile/multivolumefile-0.2.3-py3-none-any.whl (from -r requirements.txt (line 11))
Processing ./lib/mypy-extensions/mypy_extensions-1.0.0-py3-none-any.whl (from -r requirements.txt (line 12))
Processing ./lib/nvidia-cublas-cu12/nvidia_cublas_cu12-12.1.3.1-py3-none-manylinux1_x86_64.whl (from -r requirements.txt (lin
e 13))
Processing ./lib/nvidia-cuda-cupti-cu12/nvidia_cuda_cupti_cu12-12.1.105-py3-none-manylinux1_x86_64.whl (from -r requirements.
txt (line 14))
Processing ./lib/nvidia-cuda-nvrtc-cu12/nvidia_cuda_nvrtc_cu12-12.1.105-py3-none-manylinux1_x86_64.whl (from -r requirements.
txt (line 15))
```

Deploy the fine-tuned model

Next, we deploy the domain fine-tuned model. We will compare the performance of the fine-tuned and pre-trained model.

```
In [9]:  finetuned_predictor = estimator.deploy(initial_instance_count=1,instance_type='ml.g5.2xlarge')

INFO:sagemaker:Creating model with name: meta-textgeneration-llama-2-7b-2024-08-05-14-07-01-702
INFO:sagemaker:Creating endpoint-config with name meta-textgeneration-llama-2-7b-2024-08-05-14-07-01-681
INFO:sagemaker:Creating endpoint with name meta-textgeneration-llama-2-7b-2024-08-05-14-07-01-681

------------!
```

Evaluate the pre-trained and fine-tuned model

---

3. Visit the **AWS S3 bucket** where your fine-tuned model weights are stored after training and take a screenshot for your submission.
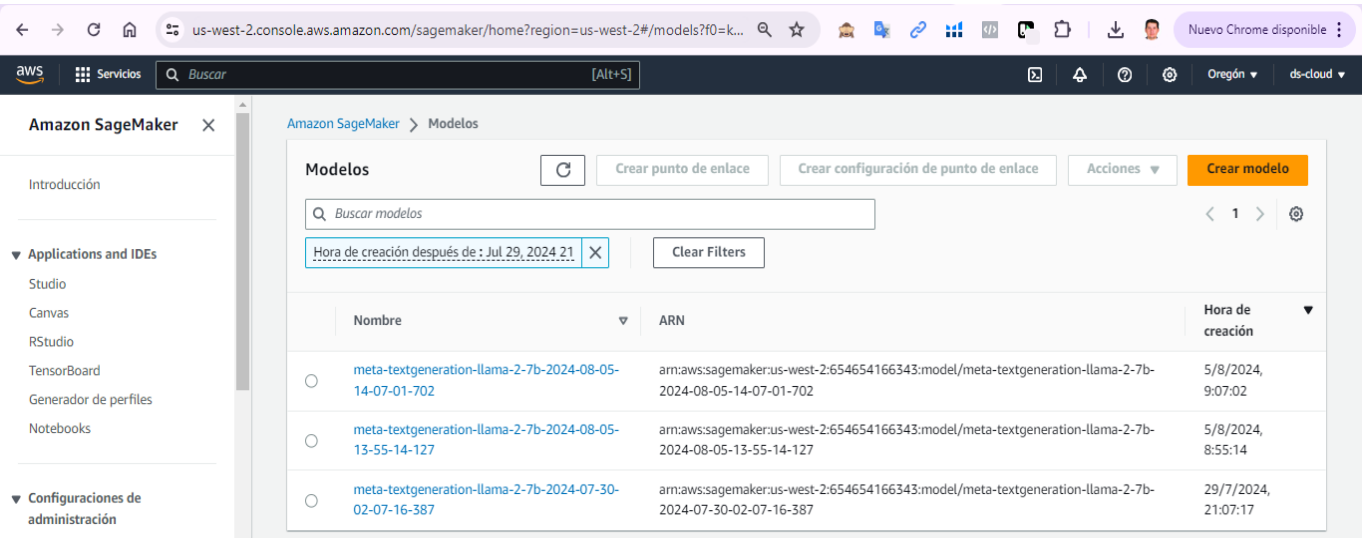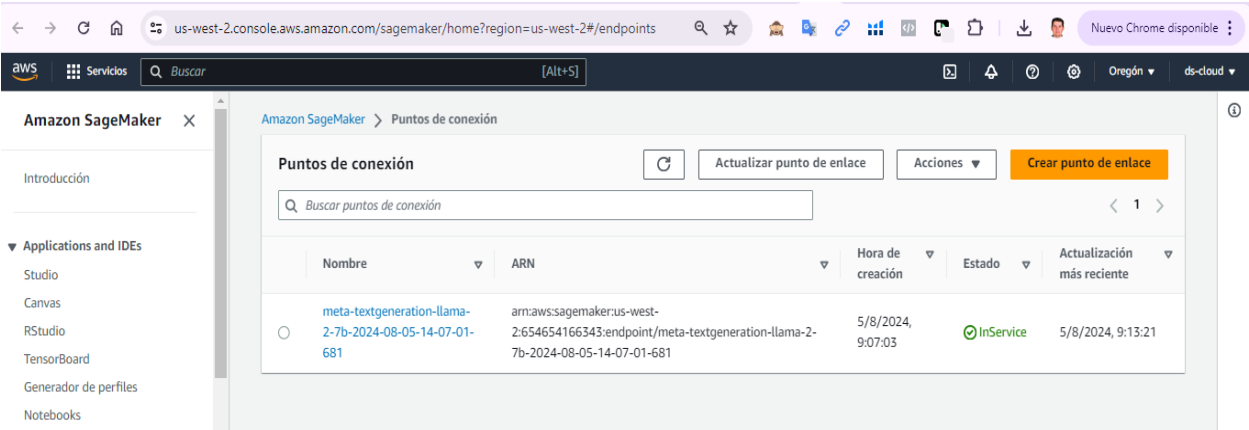
4. Screenshots demonstrating the deployment of the model in the SageMaker environment..

5. Screenshots demonstrating training jobs in the SageMaker
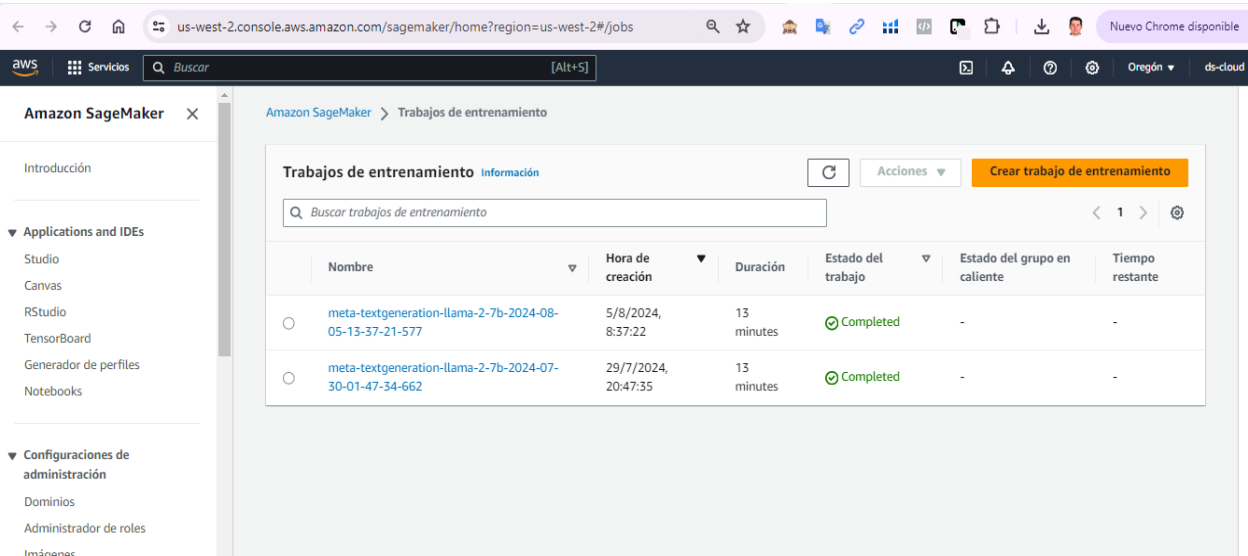
5. Screenshots the Endpoint created in deployment of the model in the SageMaker practice.

IMPORTANT:
The job about **Model_Finetuning**, was run with my owner AWS User: **ds-cloud**

I have a repository for this project:
https://github.com/richardriveros/AWS_Introducing_Generative_AI