

UDACITY

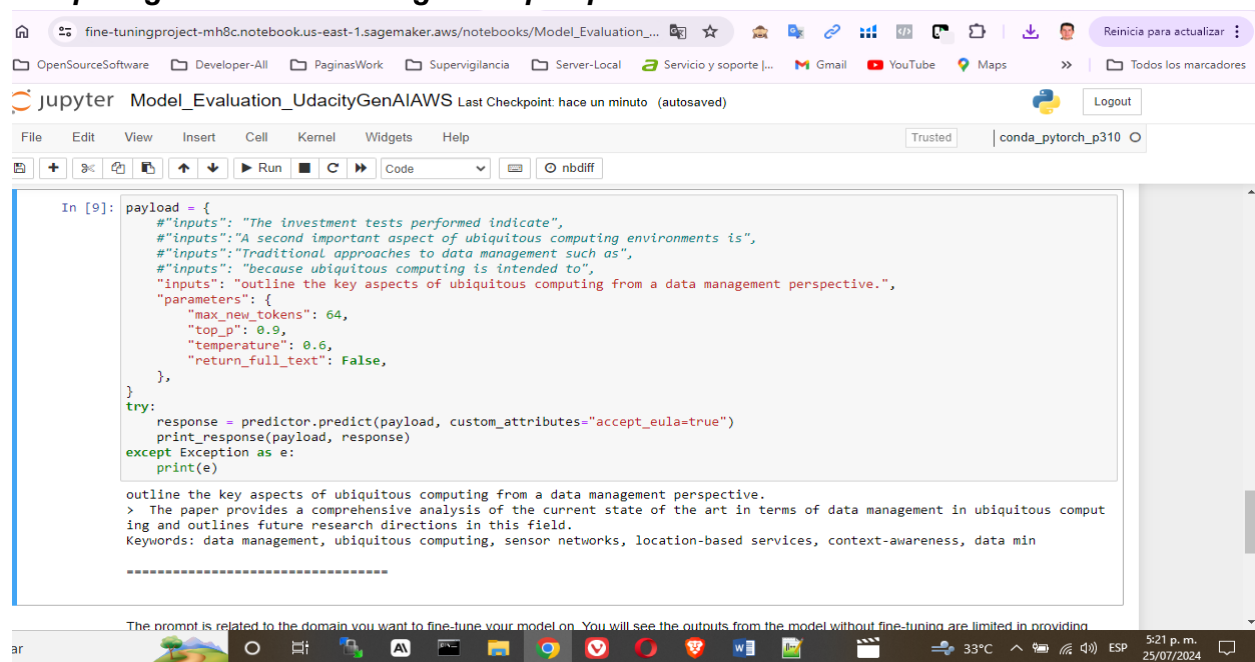
Introduction to Generative AI with AWS

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'***

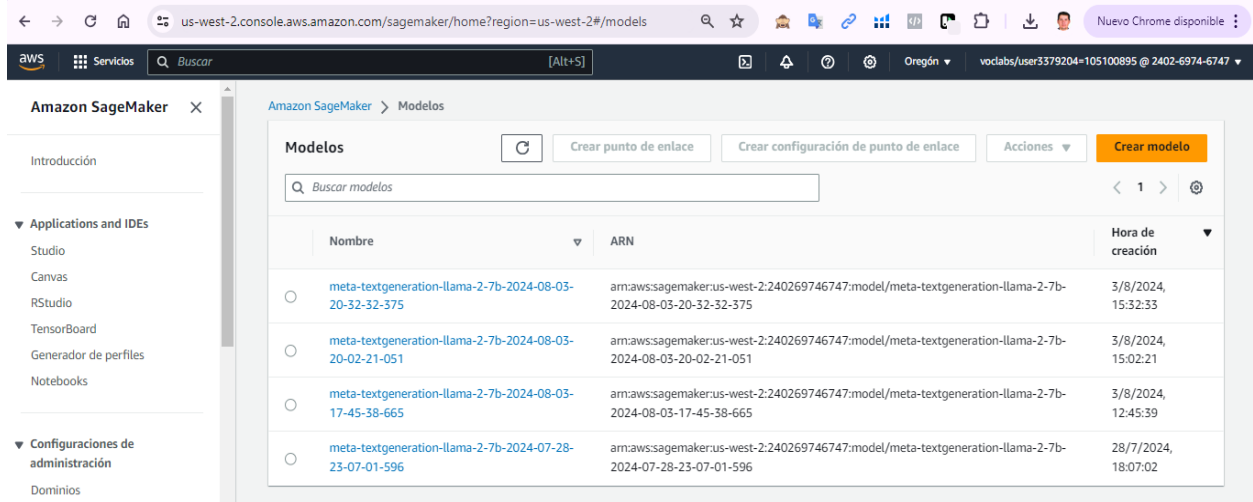


The screenshot shows a Jupyter Notebook interface with a browser address bar at the top. The notebook title is "Model_Evaluation_UdacityGenAIAWS". The code cell contains a Python script that sends a payload to a predictor. The output of the cell shows the model's response, which is a detailed analysis of the input prompt.

```
In [9]: payload = {
    #inputs: "The investment tests performed indicate",
    #inputs: "A second important aspect of ubiquitous computing environments is",
    #inputs: "Traditional approaches to data management such as",
    #inputs: "because ubiquitous computing is intended to",
    "inputs": "outline the key aspects of ubiquitous computing from a data management perspective.",
    "parameters": {
        "max_new_tokens": 64,
        "top_p": 0.9,
        "temperature": 0.6,
        "return_full_text": False,
    },
}
try:
    response = predictor.predict(payload, custom_attributes="accept_eula=true")
    print_response(payload, response)
except Exception as e:
    print(e)
```

outline the key aspects of ubiquitous computing from a data management perspective.
> The paper provides a comprehensive analysis of the current state of the art in terms of data management in ubiquitous computing and outlines future research directions in this field.
Keywords: data management, ubiquitous computing, sensor networks, location-based services, context-awareness, data mining

1A. Screenshot demonstrating the deployment of the model in the SageMaker environment.

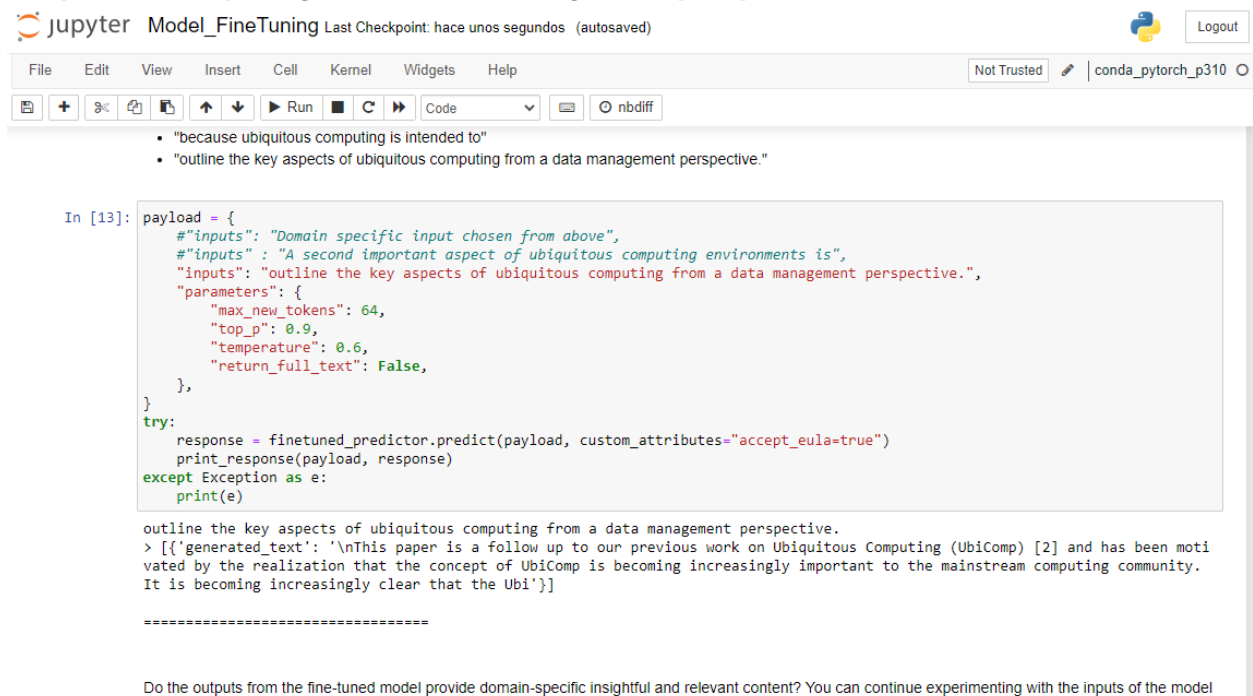


IMPORTANT:

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

The screenshot shows a Jupyter Notebook titled "Model_FineTuning" in the AWS SageMaker environment. The top bar indicates the notebook is autosaved and shows the kernel status as "conda_pytorch_p310". The notebook content includes a code cell with the following output:

```
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 (line 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))
```

Below the code cell, the text "Deploy the fine-tuned model" is followed by a paragraph: "Next, we deploy the domain fine-tuned model. We will compare the performance of the fine-tuned and pre-trained model."

The next code cell shows the deployment command and its output:

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

Below the code cell, the text "Evaluate the pre-trained and fine-tuned model" is displayed.

3. Visit the AWS S3 bucket where your fine-tuned model weights are stored after training and take a screenshot for your submission.

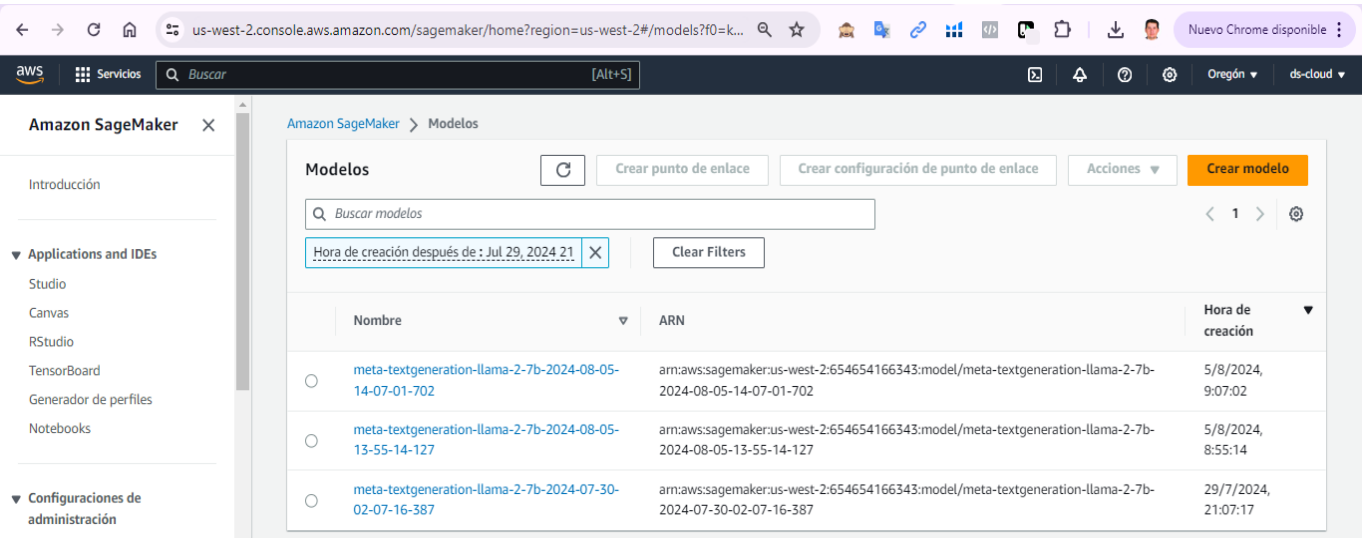
The screenshot shows the AWS S3 console interface. The left sidebar displays the "Amazon S3" menu with options like "Buckets", "Concesiones de acceso", "Puntos de acceso", "Puntos de acceso del objeto", "Lambda", "Puntos de acceso de varias regiones", "Operaciones por lotes", "Analizador de acceso de IAM para S3", "Configuración de bloqueo de acceso público", "Configuración de acceso a esta cuenta", "Storage Lens", "Paneles", "Grupos de Storage Lens", "Configuración de AWS Organizations".

The main content area shows the "Buckets" page. At the top, there is a "Instantánea de la cuenta" section with a "Ver panel de Storage Lens" button. Below this, the "Buckets de uso general" tab is selected, showing a list of buckets. The table has columns: "Nombre", "Región de AWS", "Analizador de acceso de IAM", and "Fecha de creación".

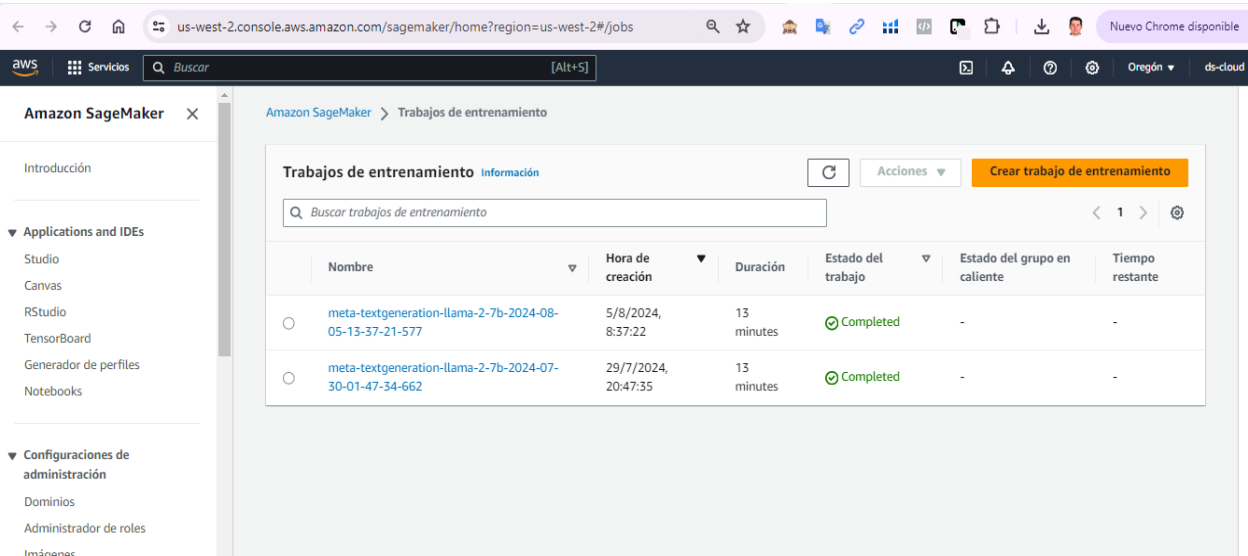
Nombre	Región de AWS	Analizador de acceso de IAM	Fecha de creación
sagemaker-us-west-2-654654166343	EE. UU. Oeste (Oregón) us-west-2	Ver analizador para us-west-2	29 Jul 2024 3:40:42 PM -05

The footer of the console shows "CloudShell", "Comentarios", and copyright information for Amazon Web Services, Inc. o sus filiales.

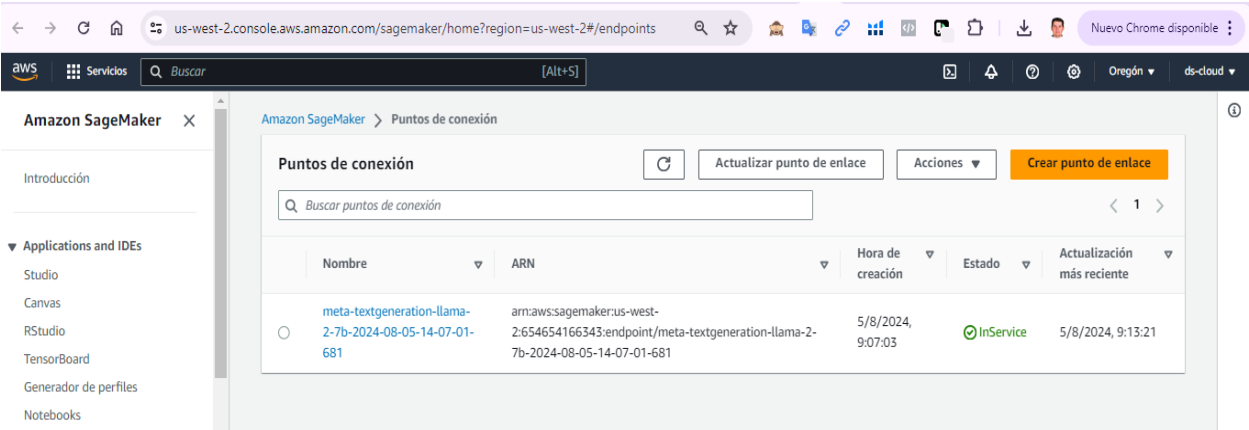
4. Screenshots demonstrating the deployment of the model in the SageMaker environment..



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