

UDACITY

Introduction to Generative AI with AWS

Project Documentation Report

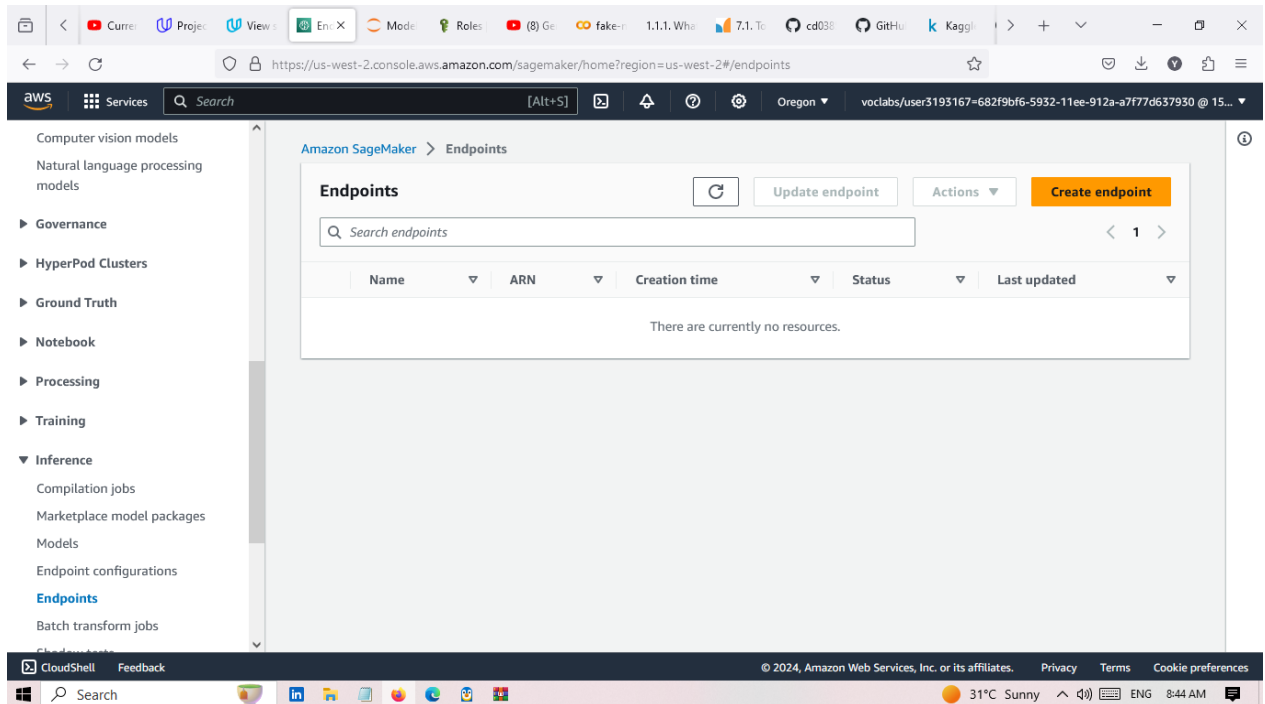
Visit [UDACITY Introduction to Generative AI with AWS Project Documentation Report](#) to make a copy of this document.

Complete the answers to the questions below to complete your project report. Create a PDF of the completed document and submit the PDF with your project.

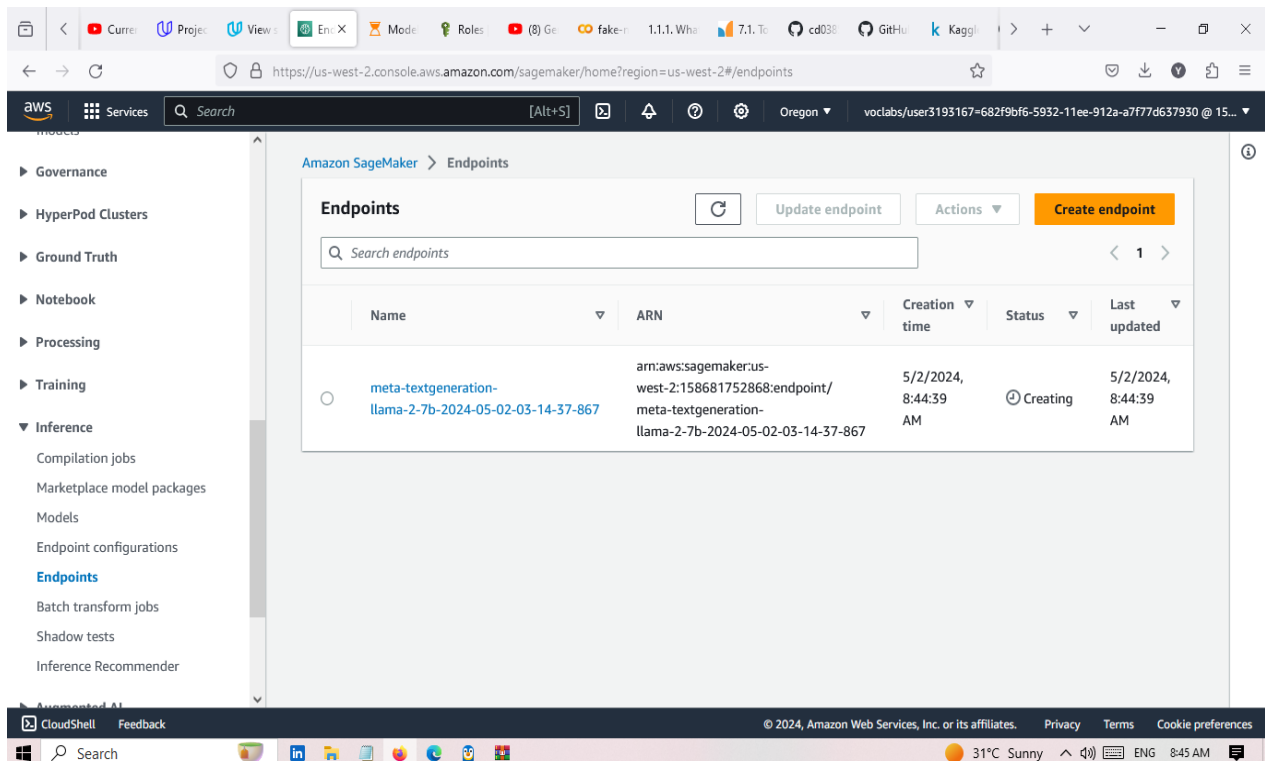
Question	Your answer:
Step 2: Domain Choice What domain did you choose to fine-tune the Meta Llama 2 7B model on? Choices: 1. Financial 2. Healthcare 3. IT	3.IT
Step 3: Model Evaluation Section What was the response of the model to your domain-specific input in the model_evaluation.ipynb file?	outline the key aspects of ubiquitous computing from a data management perspective > . A data management system for ubiquitous computing should be able to manage data that is generated by a variety of sensors and devices that are deployed in different physical environments. The system should also be able to manage data that is generated by a variety of users who are interacting with the data in different ways
Step 4: Fine-Tuning Section After fine-tuning the model, what was the response of the model to your domain-specific input in the model_finetuning.ipynb file?	A second important aspect of ubiquitous computing environments is > [{'generated_text': ' the ability to automatically collect and process data in real time.\n\nThe Internet of Things (IoT) is a term used to describe the concept of a world where physical objects are connected to the Internet. This allows the objects to communicate with each other and with other devices, such as computers and smartphones.'}]

- **Model_Evaluation_UdacityGenAIAWS**

-First there is no endpoints



-Creating a endpoint by Model



- Endpoints inservice

The screenshot shows the Amazon SageMaker console. On the left is a navigation menu with categories like Governance, HyperPod Clusters, Ground Truth, Notebook, Processing, Training, and Inference. The 'Endpoints' link is highlighted under the Inference section. The main panel displays the 'Endpoints' page with a search bar and a table of endpoints. One endpoint is listed:

Name	ARN	Creation time	Status	Last updated
meta-textgeneration-llama-2-7b-2024-05-02-03-14-37-867	arn:aws:sagemaker:us-west-2:158681752868:endpoint/meta-textgeneration-llama-2-7b-2024-05-02-03-14-37-867	5/2/2024, 8:44:39 AM	InService	5/2/2024, 8:53:07 AM

-output of the model

The screenshot shows a Jupyter Notebook interface within the Amazon SageMaker console. The notebook is titled 'Model_Evaluation_UdacityGenAI'. The code cell shows a JSON payload being passed to a predictor, and the output displays the prediction response and a detailed text output.

```
[0]: payload = {  
  "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
> . We then present a set of requirements for ubiquitous computing applications and use these requirements to identify the key challenges of ubiquitous computing. We then discuss the role of data management in addressing these challenges. We conclude by identifying some future directions for research in ubiquitous computing and

=====

The prompt is related to the domain you want to fine-tune your model on. You will see the outputs from the model without fine-tuning are limited in providing insightful or relevant content.

Use the output from this notebook to fill out the "model evaluation" section of the project documentation report

Take a screenshot of this file with the cell output for your project documentation report. Download it with cell output by making sure you used Save on the notebook before downloading

- **Model_FineTuning**
- creating endpoint for model

Amazon SageMaker > Endpoints

Endpoints

Search endpoints

Name	ARN	Creation time	Status	Last updated
meta-textgeneration-llama-2-7b-2024-05-02-03-37-39-571	arn:aws:sagemaker:us-west-2:158681752868:endpoint/meta-textgeneration-llama-2-7b-2024-05-02-03-37-39-571	5/2/2024, 9:07:41 AM	Creating	5/2/2024, 9:07:41 AM
meta-textgeneration-llama-2-7b-2024-05-02-03-14-37-867	arn:aws:sagemaker:us-west-2:158681752868:endpoint/meta-textgeneration-llama-2-7b-2024-05-02-03-14-37-867	5/2/2024, 8:44:39 AM	InService	5/2/2024, 8:53:07 AM

-endpoint Inservice

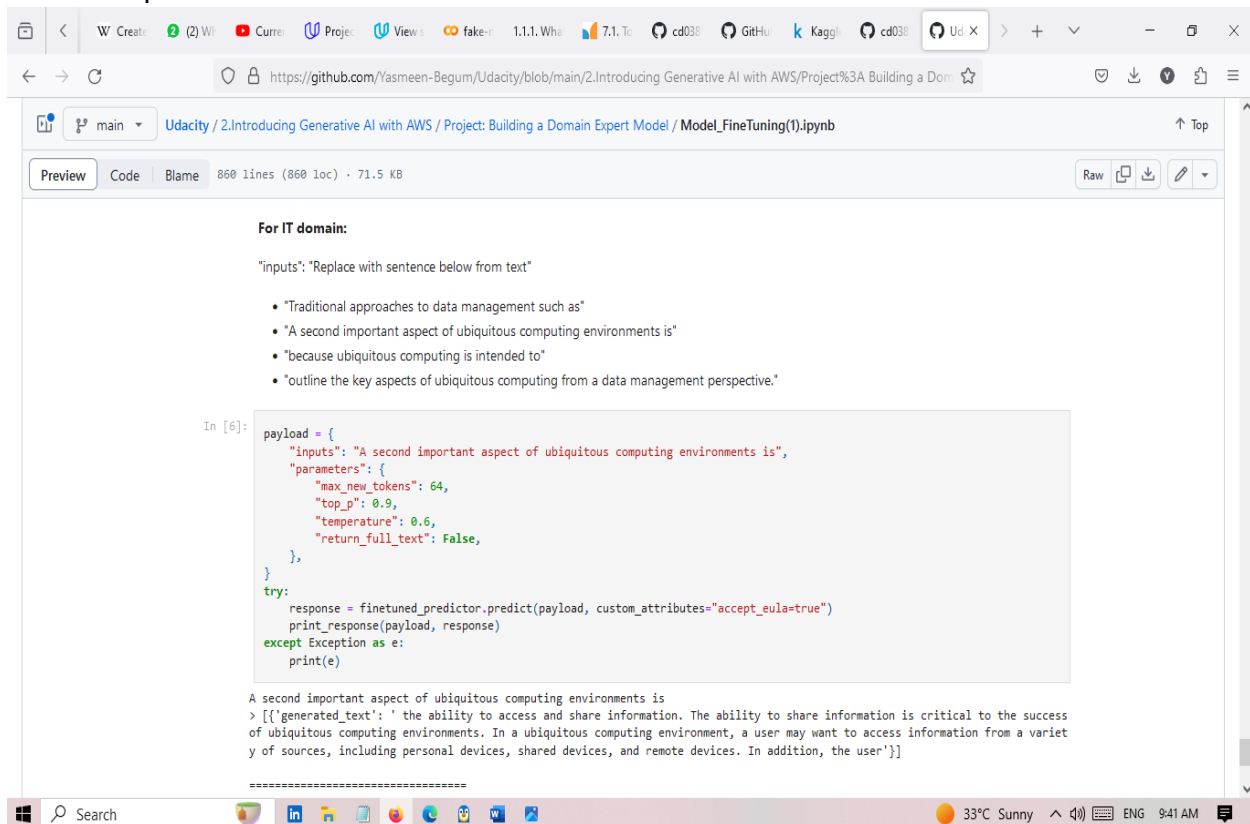
Amazon SageMaker > Endpoints

Endpoints

Search endpoints

Name	ARN	Creation time	Status	Last updated
meta-textgeneration-llama-2-7b-2024-05-02-03-37-39-571	arn:aws:sagemaker:us-west-2:158681752868:endpoint/meta-textgeneration-llama-2-7b-2024-05-02-03-37-39-571	5/2/2024, 9:07:41 AM	InService	5/2/2024, 9:13:06 AM
meta-textgeneration-llama-2-7b-2024-05-02-03-14-37-867	arn:aws:sagemaker:us-west-2:158681752868:endpoint/meta-textgeneration-llama-2-7b-2024-05-02-03-14-37-867	5/2/2024, 8:44:39 AM	InService	5/2/2024, 8:53:07 AM

- output of fine tune model



For IT domain:

"inputs": "Replace with sentence below from text"

- "Traditional approaches to data management such as"
- "A second important aspect of ubiquitous computing environments is"
- "because ubiquitous computing is intended to"
- "outline the key aspects of ubiquitous computing from a data management perspective."

```
In [6]: payload = {
  "inputs": "A second important aspect of ubiquitous computing environments is",
  "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)
```

A second important aspect of ubiquitous computing environments is
> [{"generated text": "the ability to access and share information. The ability to share information is critical to the success of ubiquitous computing environments. In a ubiquitous computing environment, a user may want to access information from a variety of sources, including personal devices, shared devices, and remote devices. In addition, the user"}]

-endpoints deleted

