



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: <ol style="list-style-type: none">1. Financial2. Healthcare3. 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?	Traditional approaches to data management such as > the relational database have been around for decades. But they are no longer enough to support the modern data management needs of enterprises. The emergence of new data sources and the explosion in the volume and variety of data have led to the development of new approaches to data management. These new approaches include No
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?	Traditional approaches to data management such as > relational databases and enterprise data warehouses (EDWs) are reaching their limits, and organizations are increasingly turning to new technologies to get more value out of their data. The first step to getting more value from your data is to understand what you have. That's where data discovery

Model Evaluation Section Output

```
[5]: payload = {
    "inputs": "Traditional approaches to data management such as",
    "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)
```

Traditional approaches to data management such as
> the relational database have been around for decades. But they are no longer enough to support the modern data management needs of enterprises. The emergence of new data sources and the explosion in the volume and variety of data have led to the development of new approaches to data management. These new approaches include No

=====




Fine-Tuning Section Output







```
payload = {
    "inputs": "Traditional approaches to data management such as",
    "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)
```

Traditional approaches to data management such as
> relational databases and enterprise data warehouses (EDWs) are reaching their limits, and organizations are increasingly turning to new technologies to get more value out of their data.
The first step to getting more value from your data is to understand what you have. That's where data discovery


=====


AWS S3 bucket: fine-tuned model weights


<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	 __script_info__.json	json	February 14, 2024, 21:22:34 (UTC+03:00)	170.0 B	Standard
<input type="checkbox"/>	 added_tokens.json	json	February 14, 2024, 21:23:10 (UTC+03:00)	21.0 B	Standard
<input type="checkbox"/>	 config.json	json	February 14, 2024, 21:22:45 (UTC+03:00)	625.0 B	Standard


<input type="checkbox"/>	 generation_config.json	json	February 14, 2024, 21:22:45 (UTC+03:00)	132.0 B	Standard
<input type="checkbox"/>	 inference.py	py	February 14, 2024, 21:22:45 (UTC+03:00)	0 B	Standard
<input type="checkbox"/>	 model-00001-of-00002.safetensors	safetensors	February 14, 2024, 21:22:45 (UTC+03:00)	9.3 GB	Standard
<input type="checkbox"/>	Name ▲	Type ▼	Last modified ▼	Size ▼	Storage class ▼
<input type="checkbox"/>	 tokenizer_config.json	json	February 14, 2024, 21:22:45 (UTC+03:00)	695.0 B	Standard
<input type="checkbox"/>	 tokenizer.json	json	February 14, 2024, 21:22:45 (UTC+03:00)	1.8 MB	Standard
<input type="checkbox"/>	 tokenizer.model	model	February 14, 2024, 21:23:10 (UTC+03:00)	488.0 KB	Standard


Objects (1) [Info](#)



 Copy S3 URI

 Copy URL


 Download

 Open


Delete


Actions ▼

Create folder

 Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

< 1 > 

<input type="checkbox"/>	Name ▲	Type ▼	Last modified ▼	Size ▼	Storage class ▼
<input type="checkbox"/>	 training_job_end.ts	ts	February 14, 2024, 21:23:12 (UTC+03:00)	0 B	Standard

Objects (1) [Info](#)

Copy S3 URI

Copy URL

Download

Open

Delete

Actions ▼

Create folder

Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

< 1 >

<input type="checkbox"/>	Name ▲	Type ▼	Last modified ▼	Size ▼	Storage class ▼
<input type="checkbox"/>	training_job_end.ts	ts	February 14, 2024, 21:23:12 (UTC+03:00)	0 B	Standard