UDACITY

Introduction to Generative AI with AWS

Project Documentation Report

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	The domain I choose to fine-tune the Meta Llama 2 7B model on is: 2. Healthcare

Step 3: Model Evaluation Section
What was the response of the model to your domain-specific input in the **model_evaluation.ipynb file**?

- **Session Setup**

```
import sagemaker, boto3, json
from sagemaker.session import Session

sagemaker_session = Session()
aws_role = sagemaker_session.get_caller_identity_arn()
aws_region = boto3.Session().region_name
sess = sagemaker.Session()
print(aws_role)
print(aws_region)
print(sess)
```

```
sagemaker.config INFO - Not applying SDK defaults from location: /etc/xdg/sagemaker/config.yaml
sagemaker.config INFO - Not applying SDK defaults from location: /home/ec2-user/.config/sagemaker/config.yaml
arn:aws:iam::508781339756:role/service-role/SageMaker-udacitySagemakerRole
us-east-1
<sagemaker.session.Session object at 0x7f8a778d3970>
```

```
(model_id, model_version,) = ("meta-textgeneration-llama-2-7b","2.*",)
```

- **Model Definition and Deployment**

```
from sagemaker.jumpstart.model import JumpStartModel

model = JumpStartModel(model_id=model_id, model_version=model_version, instance_type="ml.g5.2xlarge")
predictor = model.deploy()
```

```
For forward compatibility, pin to model_version='2.*' in your JumpStartModel or JumpStartEstimator definitions. Note that major version upgrades may have different EULA acceptance terms and input/output signatures.
Using vulnerable JumpStart model 'meta-textgeneration-llama-2-7b' and version '2.1.8'.
Using model 'meta-textgeneration-llama-2-7b' with wildcard version identifier '2.*'. You can pin to version '2.1.8' for more stable results. Note that models may have different input/output signatures after a major version upgrade.
```

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

- **Model Evaluation**

Inputs and Outputs: {

- "Myeloid neoplasms and acute leukemias derive from"

```
Myeloid neoplasms and acute leukemias derive from
> a common progenitor and share a number of common genetic lesions.
Malignant myeloid neoplasms are a group of hematologic malignancies that includes myelodysplastic syndromes (MDS), myeloproliferative neoplasms

================================
```

- "Genomic characterization is essential for"

Genomic characterization is essential for
› understanding the mechanisms of adaptation and speciation in this group.
AB - Background: The genus Sarcocystis contains 30 species that infect a wide range of vertebrates. These organisms are obligat
e intracellular parasites of the phylum Cnidaria

- "Certain germline disorders may be associated with"

Certain germline disorders may be associated with
› an increased risk of developing breast cancer.
Genetic testing may be considered for people who have a family history of breast cancer.
Risk of breast cancer in BRCA-positive individuals
BRCA1 and BRCA2 are two genes that are responsible for repairing damaged DNA in cells. Mut

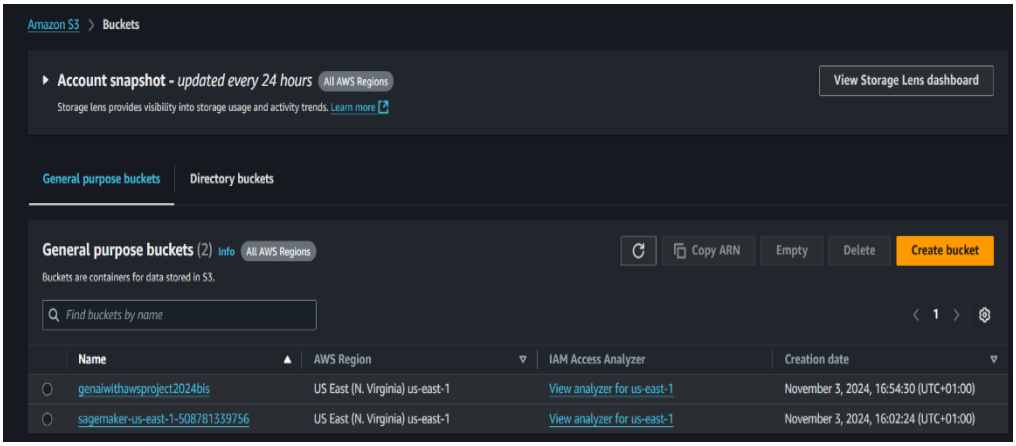
- “In contrast to targeted approaches, genome-wide sequencing"

Certain germline disorders may be associated with
› an increased risk of developing breast cancer.
Genetic testing may be considered for people who have a family history of breast cancer.
Risk of breast cancer in BRCA-positive individuals
BRCA1 and BRCA2 are two genes that are responsible for repairing damaged DNA in cells. Mut

}

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**?

1. Medical Dataset Uploaded to S3



2. Model Fine-tuning

```
2024-11-03 16:13:10 Uploading - Uploading generated training model
2024-11-03 16:13:53 Completed - Training job completed
Training seconds: 697
Billable seconds: 697
```

3. Fine-tuned Model Deployment:

```
'''
# Do not use estimator.deploy() without mentioning the instance_type.
# It's because when you call estimator.deploy() without explicitly setting the instance_type for the endpoint,
# SageMaker selects a default instance type for hosting, which, in this case, is ml.g5.12xlarge.
# However, Udacity doesn't allow instance type more than "ml.*.2xlarge".
'''
finetuned_predictor = estimator.deploy(instance_type="ml.g5.2xlarge", initial_instance_count=1)
```

```
INFO:sagemaker:Creating model with name: meta-textgeneration-llama-2-7b-2024-11-03-16-16-14-691
INFO:sagemaker:Creating endpoint-config with name meta-textgeneration-llama-2-7b-2024-11-03-16-16-14-689
INFO:sagemaker:Creating endpoint with name meta-textgeneration-llama-2-7b-2024-11-03-16-16-14-689
----------!
```

4. Fine-tuned Model Evaluation:

Inputs and Outputs: {

- "Myeloid neoplasms and acute leukemias derive from"

```
Myeloid neoplasms and acute leukemias derive from
> [{'generated_text': ' a common myeloid progenitor. The genetic and epigenetic events that lead to the acquisition of the leuk
emic phenotype are still poorly understood. We have identified a new transcription factor, TAL1/SCL, that is essential for norm
al development of the hemat'}]
```

- "Genomic characterization is essential for"

```
Genomic characterization is essential for
> [{'generated_text': ' the identification of novel genetic variations in diseases. Genetic variations are a major source of ge
netic diversity and are critical in understanding the evolutionary processes that shaped human genetic diversity.\nThe Human Ge
nome Project (HGP) was a 13-year project that was completed in '}]
```

- "Certain germline disorders may be associated with"

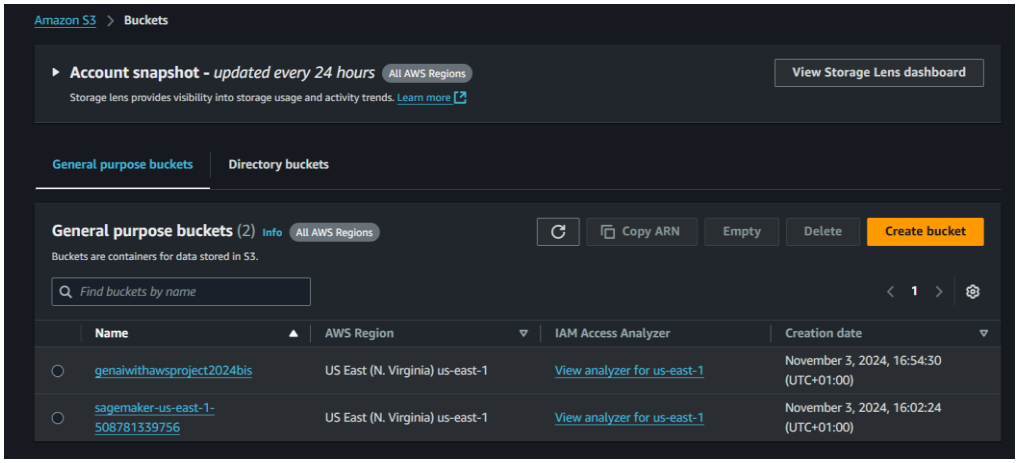
```
Certain germline disorders may be associated with
> [{'generated_text': ' an increased risk of cancer, but it is unclear whether the risk is related to the disorder itself or to
the treatment that patients with the disorder receive.\nIn a study published in the Journal of Clinical Oncology, researchers a
t Dana-Farber Cancer Institute and the Broad Institute'}]
```

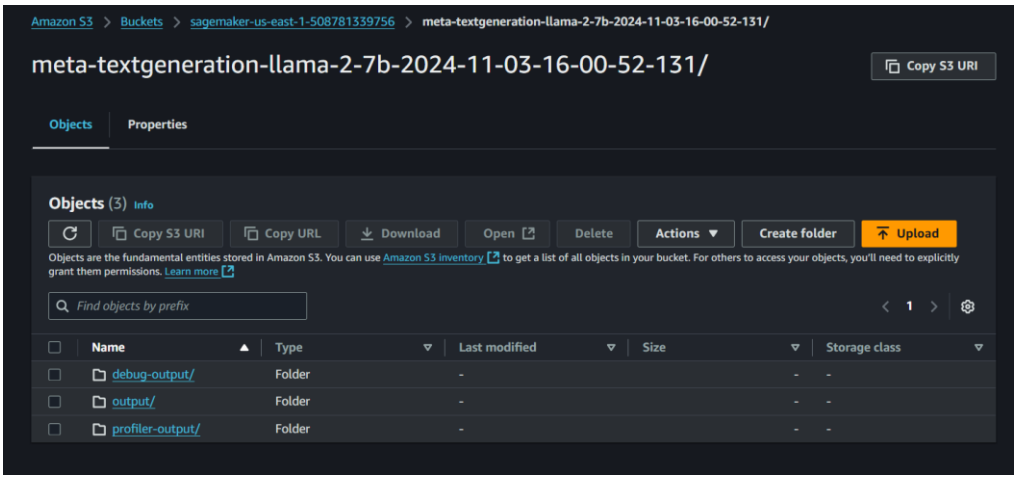
- “In contrast to targeted approaches, genome-wide sequencing"

```
In contrast to targeted approaches, genome-wide sequencing
> [{'generated_text': ' is the only way to identify the entire spectrum of genetic variations that contribute to the risk of di
sease. It also provides insights into the biological mechanisms underlying the disease and the potential for novel therapeutic
targets.\nSeveral large-scale sequencing initiatives have been launched to generate the'}]
```

}

5. Fine-tuned model weights in 3s





Conclusion

The fine-tuning process has a significant impact on model performance, the results show that the model acquired some knowledge in the medical field.