Meets Specifications

Splendid job Udacian, congratulations on passing 'Building a domain expert model' project! 🎉🎉

Well done in trying out with IT domain and completing the fine tuning of model successfully. You have covered following sections successfully.

✅ Understanding and completing Model evaluation notebook, evaluates one of the foundation models of AWS.

✅ Completion and execution of Model_FineTuning.ipynb notebook with required S3 bucket details. S3 location must hold the required domain document for fine-tuning.

✅ Capturing output and documenting the Project Documentation Report.

By completing this project, you have gained practical experience in building and fine-tuning generative AI models using AWS services.

One common application of fine-tuning models is in the field of Natural Language Processing (NLP). Let's say you want to build a chatbot for customer support. You can start with a pre-trained language model, such as OpenAI's GPT-3, however, this pre-trained model may not be specifically tailored to understand and generate responses related to customer support.

To address this, you can fine-tune the pre-trained model on a dataset that contains customer support conversations. By providing this specific domain data during fine-tuning, the model can learn to generate more accurate and contextually relevant responses for customer support queries. This fine-tuned model can then be deployed as a chatbot to assist customers in real-time.

I would encourage you to explore more such use cases and try applying in your current work or studies, good luck!

```
Reference links
```

- Fine-tune Falcon 7B - Amazon SageMaker with @remote decorator
- AWS foundation model - talks about Llama 2 foundation model.
- Fine-tuning with Sagemaker - link for exploring AWS sagemaker with practical examples.

Pre-trained Model Evaluation

The Llama2 model is successfully deployed on AWS Sagemaker.
The output of the Model_Evaluation.ipynb file verifies deployment
Screenshots in the report show the model deployed in the SageMaker environment

Excellent job here, all the required screenshots have been shared.

 The Llama2 model is successfully deployed on AWS Sagemaker.

 The output of the Model_Evaluation.ipynb file verifies deployment

 Details in the report show the model deployed in the SageMaker environment

This step demonstrates your ability to deploy the Llama2 model on AWS Sagemaker and validate the deployment using the Model_Evaluation.ipynb file. It also showcases your documentation skills.
Model evaluation uses input to evaluate generation of domain-specific content, based on chosen domain for fine-tuning.
The model_evaluation.ipynb notebook contains relevant examples and output cells.
The response of the model to domain-specific inputs is documented in the project report.

Great work again, you have given necessary domain specific input for evaluation. You have chosen IT domain for model evaluation and it works good.

✅ IT domain-specific inputs used for evaluation.

✅ Notebook outputs are visible and match the analysis in the report.

✅ PDF of output.

By completing these tasks, you would have gained a deeper understanding of how the fine-tuned model performs in generating domain-specific content and documented your evaluation process for future reference.

[6]:
```
payload = {
    "inputs": "Traditional approaches to data management such as",
    "inputs": "A second important aspect of ubiquitous computing environments is",
    "inputs": "because ubiquitous computing is intended to",
    "inputs": "outline the key aspects of ubiquitous computing from a data management perspecti
    "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)
```

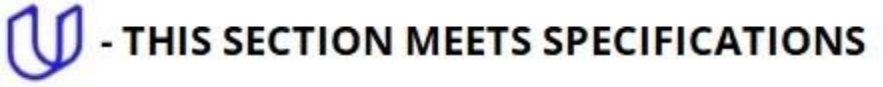
```
outline the key aspects of ubiquitous computing from a data management perspective.
>
The data management aspects of ubiquitous computing are broad and diverse, and the following se
pects.

\subsubsection{Data Acquisition}
Ubiquitous computing applications are often built on top of sensors, actuators, and other devic
ecting data

----------------------------------
```

Fine-tuning a Large Language Model

The model is fine-tuned using a dataset relevant to the chosen domain (Financial, Healthcare, IT).
The model_finetuning.ipynb notebook demonstrates the fine-tuning process
The dataset used for fine-tuning is set to the chosen domain in the Model_FineTuning.ipynb notebook file.
Screenshots of the fine-tuning cell output are present.



Nice job!

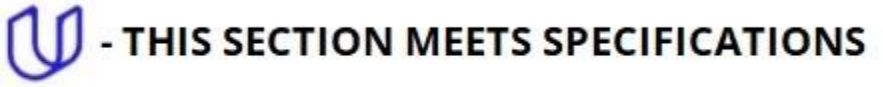
 The dataset choice aligns with the specified domain in the fine-tuning notebook.

 Notebook and shared files show required fine-tuning training successfully completed.

By fine-tuning the LLAMA model, we enhance its text generation capabilities and make it more useful for specific applications. This process helps to bridge the gap between the general language understanding of the pre-trained model and the specific requirements of the target domain enabling it to generate more contextually relevant and accurate content.

Evaluate the Fine-tuned Llama2 Large Language Model

The fine-tuned Llama2 model is successfully deployed on AWS Sagemaker.
Deployment steps are shown in the Model_FineTuning.ipynb notebook cell output
Screenshots show the deployed model in the SageMaker environment.



Well done!

The fine-tuned Llama2 model is successfully deployed on AWS Sagemaker.

Deployment steps are shown in the Model_FineTuning.ipynb notebook cell output

Submitted files show the deployed model in the SageMaker environment.

Evaluation of the fine-tuned model's performance on domain-specific text generation tasks by providing domain-specific input to the fine-tuned model.

The model_finetuning.ipynb notebook includes examples of the model's output post-fine-tuning.

Documentation of the improvements or changes in the model's performance after fine-tuning is provided.in the Project Documentation Report.

- THIS SECTION MEETS SPECIFICATIONS

All the required sections have been completed.

Provided domain-specific input to the fine-tuned model relevant to the dataset and IT domain the model was trained on.

The notebook's output cells reflect input and the fine-tuned model's response

Improved understanding or generation of domain-specific content post-fine-tuning in the Project Documentation Report.

[9]:
```
payload = {
    "inputs": "Traditional approaches to data management such as",
    "inputs": "A second important aspect of ubiquitous computing environments is",
    "inputs": "because ubiquitous computing is intended to",
    "inputs": "outline the key aspects of ubiquitous computing from a data management perspec
    "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)
```

```
outline the key aspects of ubiquitous computing from a data management perspective.
> [{'generated_text': '\nidentify the key challenges in designing and building a ubiquitous c
e main research directions in the area of ubiquitous computing.\nidentify the main research c
ting.\nunderstand the key design issues in building a ubi'}]

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