Leads4Pass

AIF-C01$^{Q\&As}$

Amazon AWS Certified AI Practitioner

Pass Amazon AIF-C01 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.leads4pass.com/aif-c01.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Amazon
Official Exam Center

⚙ **Instant Download** After Purchase

⚙ **100% Money Back** Guarantee

⚙ **365 Days** Free Update

⚙ **800,000+** Satisfied Customers

QUESTION 1

A company needs to choose a model from Amazon Bedrock to use internally. The company must identify a model that generates responses in a style that the company\\'s employees prefer.

What should the company do to meet these requirements?

A. Evaluate the models by using built-in prompt datasets.

B. Evaluate the models by using a human workforce and custom prompt datasets.

C. Use public model leaderboards to identify the model.

D. Use the model InvocationLatency runtime metrics in Amazon CloudWatch when trying models.

Correct Answer: B

To determine which model generates responses in a style that the company\\'s employees prefer, the best approach is to use a human workforce to evaluate the models with custom prompt datasets. This method allows for subjective

evaluation based on the specific stylistic preferences of the company\\'s employees, which cannot be effectively assessed through automated methods or pre-built datasets. Option B (Correct): "Evaluate the models by using a human workforce

and custom prompt datasets":This is the correct answer as it directly involves human judgment to evaluate the style and quality of the responses, aligning with employee preferences.

Option A:"Evaluate the models by using built-in prompt datasets" is incorrect because built-in datasets may not capture the company\\'s specific stylistic requirements.

Option C:"Use public model leaderboards to identify the model" is incorrect as leaderboards typically measure model performance on standard benchmarks, not on stylistic preferences.

Option D:"Use the model InvocationLatency runtime metrics in Amazon CloudWatch" is incorrect because latency metrics do not provide any information about the style of the model\\'s responses.

AWS AI Practitioner

References:

Model Evaluation Techniques on AWS:AWS suggests using human evaluators to assess qualitative aspects of model outputs, such as style and tone, to ensure alignment with organizational preferences

QUESTION 2

A company is building a chatbot to improve user experience. The company is using a large language model (LLM) from Amazon Bedrock for intent detection. The company wants to use few-shot learning to improve intent detection accuracy.

Which additional data does the company need to meet these requirements?

A. Pairs of chatbot responses and correct user intents

B. Pairs of user messages and correct chatbot responses

C. Pairs of user messages and correct user intents

D. Pairs of user intents and correct chatbot responses

Correct Answer: C

Few-shot learning involves providing a model with a few examples (shots) to learn from. For improving intent detection accuracy in a chatbot using a large language model (LLM), the data should consist of pairs of user messages and their corresponding correct intents.

QUESTION 3

A company is building an ML model. The company collected new data and analyzed the data by creating a correlation matrix, calculating statistics, and visualizing the data.

Which stage of the ML pipeline is the company currently in?

A. Data pre-processing

B. Feature engineering

C. Exploratory data analysis

D. Hyperparameter tuning

Correct Answer: C

Exploratory data analysis (EDA) involves understanding the data by visualizing it, calculating statistics, and creating correlation matrices. This stage helps identify patterns, relationships, and anomalies in the data, which can guide further

steps in the ML pipeline. Option C (Correct): "Exploratory data analysis":This is the correct answer as the tasks described (correlation matrix, calculating statistics, visualizing data) are all part of the EDA process.

Option A:"Data pre-processing" is incorrect because it involves cleaning and transforming data, not initial analysis.

Option B:"Feature engineering" is incorrect because it involves creating new features from raw data, not analyzing the data\\'s existing structure. Option D:"Hyperparameter tuning" is incorrect because it refers to optimizing model parameters,

not analyzing the data.

AWS AI Practitioner References:

Stages of the Machine Learning Pipeline:AWS outlines EDA as the initial phase of understanding and exploring data before moving to more specific preprocessing, feature engineering, and model training stages.

QUESTION 4

A company wants to use a large language model (LLM) on Amazon Bedrock for sentiment analysis. The company wants to classify the sentiment of text passages as positive or negative.

Which prompt engineering strategy meets these requirements?

A. Provide examples of text passages with corresponding positive or negative labels in the prompt followed by the new text passage to be classified.

B. Provide a detailed explanation of sentiment analysis and how LLMs work in the prompt.

C. Provide the new text passage to be classified without any additional context or examples.

D. Provide the new text passage with a few examples of unrelated tasks, such as text summarization or question answering.

Correct Answer: A

Providing examples of text passages with corresponding positive or negative labels in the prompt followed by the new text passage to be classified is the correct prompt engineering strategy for using a large language model (LLM) on Amazon Bedrock for sentiment analysis.

QUESTION 5

A company uses Amazon SageMaker for its ML pipeline in a production environment. The company has large input data sizes up to 1 GB and processing times up to 1 hour. The company needs near real-time latency.

Which SageMaker inference option meets these requirements?

A. Real-time inference

B. Serverless inference

C. Asynchronous inference

D. Batch transform

Correct Answer: A

Real-time inference is designed to provide immediate, low-latency predictions, which is necessary when the company requires near real-time latency for its ML models. This option is optimal when there is a need for fast responses, even with

large input data sizes and substantial processing times.

Option A (Correct): "Real-time inference":This is the correct answer because it supports low-latency requirements, which are essential for real-time applications where quick response times are needed.

Option B:"Serverless inference" is incorrect because it is more suited for intermittent, small-scale inference workloads, not for continuous, large-scale, low- latency needs.

Option C:"Asynchronous inference" is incorrect because it is used for workloads that do not require immediate responses.

Option D:"Batch transform" is incorrect as it is intended for offline, large-batch processing where immediate response is not necessary.

AWS AI Practitioner References:

Amazon SageMaker Inference Options:AWS documentation describes real-time inference as the best solution for applications that require immediate prediction results with low latency.