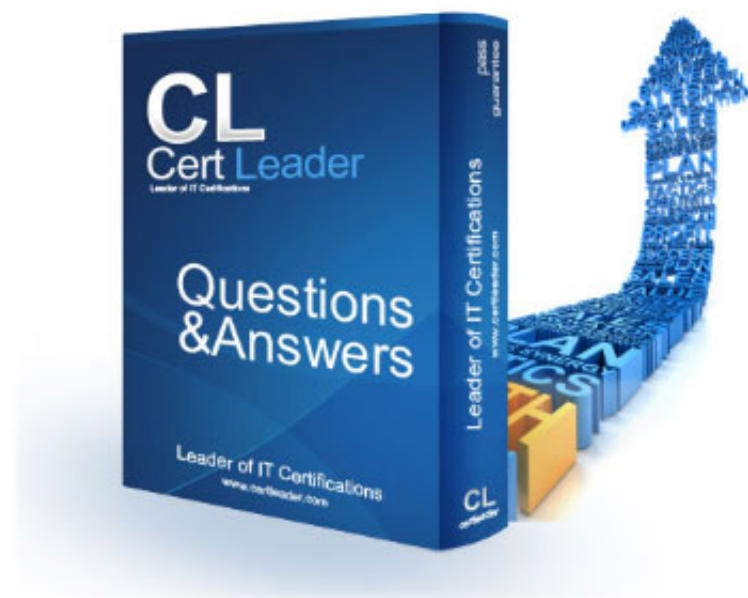


AIF-C01 Dumps

AWS Certified AI Practitioner

<https://www.certleader.com/AIF-C01-dumps.html>



NEW QUESTION 1

A company is building a large language model (LLM) question answering chatbot. The company wants to decrease the number of actions call center employees need to take to respond to customer questions.

Which business objective should the company use to evaluate the effect of the LLM chatbot?

- A. Website engagement rate
- B. Average call duration
- C. Corporate social responsibility
- D. Regulatory compliance

Answer: B

Explanation:

The business objective to evaluate the effect of an LLM chatbot aimed at reducing the actions required by call center employees should be average call duration.

? Average Call Duration:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 2

A law firm wants to build an AI application by using large language models (LLMs). The application will read legal documents and extract key points from the documents. Which solution meets these requirements?

- A. Build an automatic named entity recognition system.
- B. Create a recommendation engine.
- C. Develop a summarization chatbot.
- D. Develop a multi-language translation system.

Answer: C

Explanation:

A summarization chatbot is ideal for extracting key points from legal documents. Large language models (LLMs) can be used to summarize complex texts, such as legal documents, making them more accessible and understandable.

? Option C (Correct): "Develop a summarization chatbot": This is the correct answer

because a summarization chatbot uses LLMs to condense and extract key information from text, which is precisely the requirement for reading and summarizing legal documents.

? Option A: "Build an automatic named entity recognition system" is incorrect

because it focuses on identifying specific entities, not summarizing documents.

? Option B: "Create a recommendation engine" is incorrect as it is used to suggest products or content, not summarize text.

? Option D: "Develop a multi-language translation system" is incorrect because translation is unrelated to summarizing text.

AWS AI Practitioner References:

? Using LLMs for Text Summarization on AWS: AWS supports developing summarization tools using its AI services, including Amazon Bedrock.

NEW QUESTION 3

A company has developed an ML model for image classification. The company wants to deploy the model to production so that a web application can use the model.

The company needs to implement a solution to host the model and serve predictions without managing any of the underlying infrastructure.

Which solution will meet these requirements?

- A. Use Amazon SageMaker Serverless Inference to deploy the model.
- B. Use Amazon CloudFront to deploy the model.
- C. Use Amazon API Gateway to host the model and serve predictions.
- D. Use AWS Batch to host the model and serve predictions.

Answer: A

Explanation:

Amazon SageMaker Serverless Inference is the correct solution for deploying an ML model to production in a way that allows a web application to use the model without the need to manage the underlying infrastructure.

? Amazon SageMaker Serverless Inference provides a fully managed environment

for deploying machine learning models. It automatically provisions, scales, and manages the infrastructure required to host the model, removing the need for the company to manage servers or other underlying infrastructure.

? Why Option A is Correct:

? Why Other Options are Incorrect:

Thus, A is the correct answer, as it aligns with the requirement of deploying an ML model without managing any underlying infrastructure.

NEW QUESTION 4

A company wants to display the total sales for its top-selling products across various retail locations in the past 12 months.

Which AWS solution should the company use to automate the generation of graphs?

- A. Amazon Q in Amazon EC2
- B. Amazon Q Developer
- C. Amazon Q in Amazon QuickSight
- D. Amazon Q in AWS Chatbot

Answer: C

Explanation:

Amazon QuickSight is a fully managed business intelligence (BI) service that allows users to create and publish interactive dashboards that include visualizations

like graphs, charts, and tables. "Amazon Q" is the natural language query feature within Amazon QuickSight. It enables users to ask questions about their data in natural language and receive visual responses such as graphs.

? Option C (Correct): "Amazon Q in Amazon QuickSight": This is the correct answer

because Amazon QuickSight Q is specifically designed to allow users to explore their data through natural language queries, and it can automatically generate graphs to display sales data and other metrics. This makes it an ideal choice for the company to automate the generation of graphs showing total sales for its top-selling products across various retail locations.

? Option A, B, and D: These options are incorrect:

AWS AI Practitioner References:

? Amazon QuickSight Q is designed to provide insights from data by using natural language queries, making it a powerful tool for generating automated graphs and visualizations directly from queried data.

? Business Intelligence (BI) on AWS: AWS services such as Amazon QuickSight

provide business intelligence capabilities, including automated reporting and visualization features, which are ideal for companies seeking to visualize data like sales trends over time.

NEW QUESTION 5

A company has documents that are missing some words because of a database error. The company wants to build an ML model that can suggest potential words to fill in the missing text.

Which type of model meets this requirement?

- A. Topic modeling
- B. Clustering models
- C. Prescriptive ML models
- D. BERT-based models

Answer: D

Explanation:

BERT-based models (Bidirectional Encoder Representations from Transformers) are suitable for tasks that involve understanding the context of words in a sentence and suggesting missing words. These models use bidirectional training, which considers the context from both directions (left and right of the missing word) to predict the appropriate word to fill in the gaps.

? BERT-based Models:

? Why Option D is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 6

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.

Answer: B

Explanation:

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

NEW QUESTION 7

A company wants to build an ML model by using Amazon SageMaker. The company needs to share and manage variables for model development across multiple teams.

Which SageMaker feature meets these requirements?

- A. Amazon SageMaker Feature Store
- B. Amazon SageMaker Data Wrangler
- C. Amazon SageMaker Clarify
- D. Amazon SageMaker Model Cards

Answer: A

Explanation:

Amazon SageMaker Feature Store is the correct solution for sharing and managing variables (features) across multiple teams during model development.

? Amazon SageMaker Feature Store:

? Why Option A is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 8

Which AWS service or feature can help an AI development team quickly deploy and consume a foundation model (FM) within the team's VPC?

- A. Amazon Personalize
- B. Amazon SageMaker JumpStart
- C. PartyRock, an Amazon Bedrock Playground
- D. Amazon SageMaker endpoints

Answer: B

Explanation:

Amazon SageMaker JumpStart is the correct service for quickly deploying and consuming a foundation model (FM) within a team's VPC.

? Amazon SageMaker JumpStart:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 9

What does an F1 score measure in the context of foundation model (FM) performance?

- A. Model precision and recall.
- B. Model speed in generating responses.
- C. Financial cost of operating the model.
- D. Energy efficiency of the model's computations.

Answer: A

Explanation:

The F1 score is the harmonic mean of precision and recall, making it a balanced metric for evaluating model performance when there is an imbalance between false positives and false negatives. Speed, cost, and energy efficiency are unrelated to the F1 score. References: AWS Foundation Models Guide.

NEW QUESTION 10

A company built an AI-powered resume screening system. The company used a large dataset to train the model. The dataset contained resumes that were not representative of all demographics. Which core dimension of responsible AI does this scenario present?

- A. Fairness.
- B. Explainability.
- C. Privacy and security.
- D. Transparency.

Answer: A

Explanation:

Fairness refers to the absence of bias in AI models. Using non-representative datasets leads to biased predictions, affecting specific demographics unfairly. Explainability, privacy, and transparency are important but not directly related to this scenario. References: AWS Responsible AI Framework.

NEW QUESTION 10

A company wants to classify human genes into 20 categories based on gene characteristics. The company needs an ML algorithm to document how the inner mechanism of the model affects the output.

Which ML algorithm meets these requirements?

- A. Decision trees
- B. Linear regression
- C. Logistic regression
- D. Neural networks

Answer: A

Explanation:

Decision trees are an interpretable machine learning algorithm that clearly documents the decision-making process by showing how each input feature affects the output. This transparency is particularly useful when explaining how the model arrives at a certain decision, making it suitable for classifying genes into categories.

? Option A (Correct): "Decision trees": This is the correct answer because decision

trees provide a clear and interpretable representation of how input features influence the model's output, making it ideal for understanding the inner mechanisms affecting predictions.

? Option B: "Linear regression" is incorrect because it is used for regression tasks, not classification.

? Option C: "Logistic regression" is incorrect as it does not provide the same level of interpretability in documenting decision-making processes.

? Option D: "Neural networks" is incorrect because they are often considered "black boxes" and do not easily explain how they arrive at their outputs.

AWS AI Practitioner References:

? Interpretable Machine Learning Models on AWS: AWS supports using interpretable models, such as decision trees, for tasks that require clear documentation of how input data affects output decisions.

NEW QUESTION 12

What does an F1 score measure in the context of foundation model (FM) performance?

- A. Model precision and recall
- B. Model speed in generating responses
- C. Financial cost of operating the model
- D. Energy efficiency of the model's computations

Answer: A

Explanation:

The F1 score is a metric used to evaluate the performance of a classification model by considering both precision and recall. Precision measures the accuracy of positive predictions (i.e., the proportion of true positive predictions among all positive predictions made by the model), while recall measures the model's ability to identify all relevant positive instances (i.e., the proportion of true positive predictions among all actual positive instances). The F1 score is the harmonic mean of precision and recall, providing a single metric that balances both concerns. This is particularly useful when dealing with imbalanced datasets or when the cost of false positives and false negatives is significant. Options B, C, and D pertain to other aspects of model performance but are not related to the F1 score.

Reference: AWS Certified AI Practitioner Exam Guide

NEW QUESTION 15

A company wants to use large language models (LLMs) with Amazon Bedrock to develop a chat interface for the company's product manuals. The manuals are stored as PDF files.

Which solution meets these requirements MOST cost-effectively?

- A. Use prompt engineering to add one PDF file as context to the user prompt when the prompt is submitted to Amazon Bedrock.
- B. Use prompt engineering to add all the PDF files as context to the user prompt when the prompt is submitted to Amazon Bedrock.
- C. Use all the PDF documents to fine-tune a model with Amazon Bedrock.
- D. Use the fine-tuned model to process user prompts.
- E. Upload PDF documents to an Amazon Bedrock knowledge base.
- F. Use the knowledge base to provide context when users submit prompts to Amazon Bedrock.

Answer: A

Explanation:

Using Amazon Bedrock with large language models (LLMs) allows for efficient utilization of AI to answer queries based on context provided in product manuals. To achieve this cost-effectively, the company should avoid unnecessary use of resources.

? Option A (Correct): "Use prompt engineering to add one PDF file as context to the user prompt when the prompt is submitted to Amazon Bedrock": This is the most cost-effective solution. By using prompt engineering, only the relevant content from one PDF file is added as context to each query. This approach minimizes the amount of data processed, which helps in reducing costs associated with LLMs' computational requirements.

? Option B: "Use prompt engineering to add all the PDF files as context to the user prompt when the prompt is submitted to Amazon Bedrock" is incorrect. Including all PDF files would increase costs significantly due to the large context size processed by the model.

? Option C: "Use all the PDF documents to fine-tune a model with Amazon Bedrock" is incorrect. Fine-tuning a model is more expensive than using prompt engineering, especially if done for multiple documents.

? Option D: "Upload PDF documents to an Amazon Bedrock knowledge base" is incorrect because Amazon Bedrock does not have a built-in knowledge base feature for directly managing and querying PDF documents.

AWS AI Practitioner References:

? Prompt Engineering for Cost-Effective AI: AWS emphasizes the importance of using prompt engineering to minimize costs when interacting with LLMs. By carefully selecting relevant context, users can reduce the amount of data processed and save on expenses.

NEW QUESTION 17

A company is using an Amazon Bedrock base model to summarize documents for an internal use case. The company trained a custom model to improve the summarization quality.

Which action must the company take to use the custom model through Amazon Bedrock?

- A. Purchase Provisioned Throughput for the custom model.
- B. Deploy the custom model in an Amazon SageMaker endpoint for real-time inference.
- C. Register the model with the Amazon SageMaker Model Registry.
- D. Grant access to the custom model in Amazon Bedrock.

Answer: B

Explanation:

To use a custom model that has been trained to improve summarization quality, the company must deploy the model on an Amazon SageMaker endpoint. This allows the model to be used for real-time inference through Amazon Bedrock or other AWS services. By deploying the model in SageMaker, the custom model can be accessed programmatically via API calls, enabling integration with Amazon Bedrock.

? Option B (Correct): "Deploy the custom model in an Amazon SageMaker endpoint for real-time inference": This is the correct answer because deploying the model on SageMaker enables it to serve real-time predictions and be integrated with Amazon Bedrock.

? Option A: "Purchase Provisioned Throughput for the custom model" is incorrect because provisioned throughput is related to database or storage services, not model deployment.

? Option C: "Register the model with the Amazon SageMaker Model Registry" is incorrect because while the model registry helps with model management, it does not make the model accessible for real-time inference.

? Option D: "Grant access to the custom model in Amazon Bedrock" is incorrect because Bedrock does not directly manage custom model access; it relies on deployed endpoints like those in SageMaker.

AWS AI Practitioner References:

? Amazon SageMaker Endpoints: AWS recommends deploying models to SageMaker endpoints to use them for real-time inference in various applications.

NEW QUESTION 20

A company is using the Generative AI Security Scoping Matrix to assess security responsibilities for its solutions. The company has identified four different solution scopes based on the matrix.

Which solution scope gives the company the MOST ownership of security responsibilities?

- A. Using a third-party enterprise application that has embedded generative AI features.
- B. Building an application by using an existing third-party generative AI foundation model (FM).
- C. Refining an existing third-party generative AI foundation model (FM) by fine-tuning the model by using data specific to the business.
- D. Building and training a generative AI model from scratch by using specific data that a customer owns.

Answer: D

Explanation:

Building and training a generative AI model from scratch provides the company with the most ownership and control over security responsibilities. In this scenario, the company is responsible for all aspects of the security of the data, the model, and the infrastructure.

? Option D (Correct): "Building and training a generative AI model from scratch by using specific data that a customer owns": This is the correct answer because it involves complete ownership of the model, data, and infrastructure, giving the company the highest level of responsibility for security.

? Option A: "Using a third-party enterprise application that has embedded generative AI features" is incorrect as the company has minimal control over the security of the AI features embedded within a third-party application.

? Option B: "Building an application using an existing third-party generative AI foundation model (FM)" is incorrect because security responsibilities are shared with the third-party model provider.

? Option C: "Refining an existing third-party generative AI FM by fine-tuning the model with business-specific data" is incorrect as the foundation model and part of the security responsibilities are still managed by the third party.

AWS AI Practitioner References:

? Generative AI Security Scoping Matrix on AWS: AWS provides a security responsibility matrix that outlines varying levels of control and responsibility depending on the approach to developing and using AI models.

NEW QUESTION 25

A company built a deep learning model for object detection and deployed the model to production.

Which AI process occurs when the model analyzes a new image to identify objects?

- A. Training
- B. Inference
- C. Model deployment
- D. Bias correction

Answer: B

Explanation:

Inference is the correct answer because it is the AI process that occurs when a deployed model analyzes new data (such as an image) to make predictions or identify objects.

? Inference:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 27

A company deployed an AI/ML solution to help customer service agents respond to frequently asked questions. The questions can change over time. The company wants to give customer service agents the ability to ask questions and receive automatically generated answers to common customer questions. Which strategy will meet these requirements MOST cost-effectively?

- A. Fine-tune the model regularly.
- B. Train the model by using context data.
- C. Pre-train and benchmark the model by using context data.
- D. Use Retrieval Augmented Generation (RAG) with prompt engineering techniques.

Answer: D

Explanation:

RAG combines large pre-trained models with retrieval mechanisms to fetch relevant context from a knowledge base. This approach is cost-effective as it eliminates the need for frequent model retraining while ensuring responses are contextually accurate and up to date. References: AWS RAG Techniques.

NEW QUESTION 30

A retail store wants to predict the demand for a specific product for the next few weeks by using the Amazon SageMaker DeepAR forecasting algorithm.

Which type of data will meet this requirement?

- A. Text data
- B. Image data
- C. Time series data
- D. Binary data

Answer: C

Explanation:

Amazon SageMaker's DeepAR is a supervised learning algorithm designed for forecasting scalar (one-dimensional) time series data. Time series data consists of sequences of data points indexed in time order, typically with consistent intervals between them. In the context of a retail store aiming to predict product demand, relevant time series data might include historical sales figures, inventory levels, or related metrics recorded over regular time intervals (e.g., daily or weekly). By training the DeepAR model on this historical time series data, the store can generate forecasts for future product demand. This capability is particularly useful for inventory management, staffing, and supply chain optimization. Other data types, such as text, image, or binary data, are not suitable for time series forecasting tasks and would not be appropriate inputs for the DeepAR algorithm.

Reference: Amazon SageMaker DeepAR Algorithm

NEW QUESTION 35

An AI company periodically evaluates its systems and processes with the help of independent software vendors (ISVs). The company needs to receive email message notifications when an ISV's compliance reports become available.

Which AWS service can the company use to meet this requirement?

- A. AWS Audit Manager

- B. AWS Artifact
- C. AWS Trusted Advisor
- D. AWS Data Exchange

Answer: D

Explanation:

AWS Data Exchange is a service that allows companies to securely exchange data with third parties, such as independent software vendors (ISVs). AWS Data Exchange can be configured to provide notifications, including email notifications, when new datasets or compliance reports become available.

? Option D (Correct): "AWS Data Exchange": This is the correct answer because it

enables the company to receive notifications, including email messages, when ISVs' compliance reports are available.

? Option A: "AWS Audit Manager" is incorrect because it focuses on assessing an organization's own compliance, not receiving third-party compliance reports.

? Option B: "AWS Artifact" is incorrect as it provides access to AWS's compliance reports, not ISVs'.

? Option C: "AWS Trusted Advisor" is incorrect as it offers optimization and best practices guidance, not compliance report notifications.

AWS AI Practitioner References:

? AWS Data Exchange Documentation: AWS explains how Data Exchange allows organizations to subscribe to third-party data and receive notifications when updates are available.

NEW QUESTION 36

A company wants to use a large language model (LLM) to develop a conversational agent. The company needs to prevent the LLM from being manipulated with common prompt engineering techniques to perform undesirable actions or expose sensitive information.

Which action will reduce these risks?

- A. Create a prompt template that teaches the LLM to detect attack patterns.
- B. Increase the temperature parameter on invocation requests to the LLM.
- C. Avoid using LLMs that are not listed in Amazon SageMaker.
- D. Decrease the number of input tokens on invocations of the LLM.

Answer: A

Explanation:

Creating a prompt template that teaches the LLM to detect attack patterns is the most effective way to reduce the risk of the model being manipulated through prompt engineering.

? Prompt Templates for Security:

? Why Option A is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 37

An accounting firm wants to implement a large language model (LLM) to automate document processing. The firm must proceed responsibly to avoid potential harms.

What should the firm do when developing and deploying the LLM? (Select TWO.)

- A. Include fairness metrics for model evaluation.
- B. Adjust the temperature parameter of the model.
- C. Modify the training data to mitigate bias.
- D. Avoid overfitting on the training data.
- E. Apply prompt engineering techniques.

Answer: AC

Explanation:

To implement a large language model (LLM) responsibly, the firm should focus on fairness and mitigating bias, which are critical for ethical AI deployment.

? A. Include Fairness Metrics for Model Evaluation:

? C. Modify the Training Data to Mitigate Bias:

? Why Other Options are Incorrect:

NEW QUESTION 38

A company is building a customer service chatbot. The company wants the chatbot to improve its responses by learning from past interactions and online resources.

Which AI learning strategy provides this self-improvement capability?

- A. Supervised learning with a manually curated dataset of good responses and bad responses
- B. Reinforcement learning with rewards for positive customer feedback
- C. Unsupervised learning to find clusters of similar customer inquiries
- D. Supervised learning with a continuously updated FAQ database

Answer: B

Explanation:

Reinforcement learning allows a model to learn and improve over time based on feedback from its environment. In this case, the chatbot can improve its responses by being rewarded for positive customer feedback, which aligns well with the goal of self-improvement based on past interactions and new information.

? Option B (Correct): "Reinforcement learning with rewards for positive customer

feedback": This is the correct answer as reinforcement learning enables the chatbot to learn from feedback and adapt its behavior accordingly, providing self-improvement capabilities.

? Option A: "Supervised learning with a manually curated dataset" is incorrect because it does not support continuous learning from new interactions.

? Option C: "Unsupervised learning to find clusters of similar customer inquiries" is incorrect because unsupervised learning does not provide a mechanism for improving responses based on feedback.

? Option D: "Supervised learning with a continuously updated FAQ database" is incorrect because it still relies on manually curated data rather than self-

improvement from feedback.

AWS AI Practitioner References:

? Reinforcement Learning on AWS: AWS provides reinforcement learning

frameworks that can be used to train models to improve their performance based on feedback.

NEW QUESTION 43

What are tokens in the context of generative AI models?

A. Tokens are the basic units of input and output that a generative AI model operates on, representing words, subwords, or other linguistic units.

B. Tokens are the mathematical representations of words or concepts used in generative AI models.

C. Tokens are the pre-trained weights of a generative AI model that are fine-tuned for specific tasks.

D. Tokens are the specific prompts or instructions given to a generative AI model to generate output.

Answer: A

Explanation:

Tokens in generative AI models are the smallest units that the model processes, typically representing words, subwords, or characters. They are essential for the model to understand and generate language, breaking down text into manageable parts for processing.

? Option A (Correct): "Tokens are the basic units of input and output that a

generative AI model operates on, representing words, subwords, or other linguistic units": This is the correct definition of tokens in the context of generative AI models.

? Option B: "Mathematical representations of words" describes embeddings, not tokens.

? Option C: "Pre-trained weights of a model" refers to the parameters of a model, not tokens.

? Option D: "Prompts or instructions given to a model" refers to the queries or commands provided to a model, not tokens.

AWS AI Practitioner References:

? Understanding Tokens in NLP: AWS provides detailed explanations of how tokens are used in natural language processing tasks by AI models, such as in Amazon Comprehend and other AWS AI services.

NEW QUESTION 47

A security company is using Amazon Bedrock to run foundation models (FMs). The company wants to ensure that only authorized users invoke the models. The company needs to identify any unauthorized access attempts to set appropriate AWS Identity and Access Management (IAM) policies and roles for future iterations of the FMs.

Which AWS service should the company use to identify unauthorized users that are trying to access Amazon Bedrock?

A. AWS Audit Manager

B. AWS CloudTrail

C. Amazon Fraud Detector

D. AWS Trusted Advisor

Answer: B

Explanation:

AWS CloudTrail is a service that enables governance, compliance, and operational and risk auditing of your AWS account. It tracks API calls and identifies unauthorized access attempts to AWS resources, including Amazon Bedrock.

? AWS CloudTrail:

? Why Option B is Correct:

? Why Other Options are Incorrect:

Thus, B is the correct answer for identifying unauthorized users attempting to access Amazon Bedrock.

NEW QUESTION 52

A company wants to use generative AI to increase developer productivity and software development. The company wants to use Amazon Q Developer.

What can Amazon Q Developer do to help the company meet these requirements?

A. Create software snippets, reference tracking, and open-source license tracking.

B. Run an application without provisioning or managing servers.

C. Enable voice commands for coding and providing natural language search.

D. Convert audio files to text documents by using ML models.

Answer: A

Explanation:

Amazon Q Developer is a tool designed to assist developers in increasing productivity by generating code snippets, managing reference tracking, and handling open-source license tracking. These features help developers by automating parts of the software development process.

? Option A (Correct): "Create software snippets, reference tracking, and open-

source license tracking": This is the correct answer because these are key features that help developers streamline and automate tasks, thus improving productivity.

? Option B: "Run an application without provisioning or managing servers" is incorrect as it refers to AWS Lambda or AWS Fargate, not Amazon Q Developer.

? Option C: "Enable voice commands for coding and providing natural language search" is incorrect because this is not a function of Amazon Q Developer.

? Option D: "Convert audio files to text documents by using ML models" is incorrect as this refers to Amazon Transcribe, not Amazon Q Developer.

AWS AI Practitioner References:

? Amazon Q Developer Features: AWS documentation outlines how Amazon Q Developer supports developers by offering features that reduce manual effort and improve efficiency.

NEW QUESTION 57

A company wants to use a large language model (LLM) on Amazon Bedrock for sentiment analysis. The company needs the LLM to produce more consistent responses to the same input prompt.

Which adjustment to an inference parameter should the company make to meet these requirements?

- A. Decrease the temperature value
- B. Increase the temperature value
- C. Decrease the length of output tokens
- D. Increase the maximum generation length

Answer: A

Explanation:

The temperature parameter in a large language model (LLM) controls the randomness of the model's output. A lower temperature value makes the output more deterministic and consistent, meaning that the model is less likely to produce different results for the same input prompt.

? Option A (Correct): "Decrease the temperature value": This is the correct answer

because lowering the temperature reduces the randomness of the responses, leading to more consistent outputs for the same input.

? Option B: "Increase the temperature value" is incorrect because it would make the output more random and less consistent.

? Option C: "Decrease the length of output tokens" is incorrect as it does not directly affect the consistency of the responses.

? Option D: "Increase the maximum generation length" is incorrect because this adjustment affects the output length, not the consistency of the model's responses.

AWS AI Practitioner References:

? Understanding Temperature in Generative AI Models: AWS documentation explains that adjusting the temperature parameter affects the model's output randomness, with lower values providing more consistent outputs.

NEW QUESTION 59

A company has installed a security camera. The company uses an ML model to evaluate the security camera footage for potential thefts. The company has discovered that the model disproportionately flags people who are members of a specific ethnic group.

Which type of bias is affecting the model output?

- A. Measurement bias
- B. Sampling bias
- C. Observer bias
- D. Confirmation bias

Answer: B

Explanation:

Sampling bias is the correct type of bias affecting the model output when it disproportionately flags people from a specific ethnic group.

? Sampling Bias:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 62

An AI practitioner has a database of animal photos. The AI practitioner wants to automatically identify and categorize the animals in the photos without manual human effort.

Which strategy meets these requirements?

- A. Object detection
- B. Anomaly detection
- C. Named entity recognition
- D. Inpainting

Answer: A

Explanation:

Object detection is the correct strategy for automatically identifying and categorizing animals in photos.

? Object Detection:

? Why Option A is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 63

A company needs to build its own large language model (LLM) based on only the company's private data. The company is concerned about the environmental effect of the training process.

Which Amazon EC2 instance type has the LEAST environmental effect when training LLMs?

- A. Amazon EC2 C series
- B. Amazon EC2 G series
- C. Amazon EC2 P series
- D. Amazon EC2 Trn series

Answer: D

Explanation:

The Amazon EC2 Trn series (Trainium) instances are designed for high-performance, cost-effective machine learning training while being energy-efficient. AWS Trainium-powered instances are optimized for deep learning models and have been developed to minimize environmental impact by maximizing energy efficiency.

? Option D (Correct): "Amazon EC2 Trn series": This is the correct answer because the Trn series is purpose-built for training deep learning models with lower energy consumption, which aligns with the company's concern about environmental effects.

? Option A: "Amazon EC2 C series" is incorrect because it is intended for compute-intensive tasks but not specifically optimized for ML training with environmental considerations.

? Option B: "Amazon EC2 G series" (Graphics Processing Unit instances) is optimized for graphics-intensive applications but does not focus on minimizing environmental impact for training.

? Option C: "Amazon EC2 P series" is designed for ML training but does not offer the same level of energy efficiency as the Trn series.

AWS AI Practitioner References:

? AWS Trainium Overview: AWS promotes Trainium instances as their most energy- efficient and cost-effective solution for ML model training.

NEW QUESTION 67

Which feature of Amazon OpenSearch Service gives companies the ability to build vector database applications?

- A. Integration with Amazon S3 for object storage
- B. Support for geospatial indexing and queries
- C. Scalable index management and nearest neighbor search capability
- D. Ability to perform real-time analysis on streaming data

Answer: C

Explanation:

Amazon OpenSearch Service (formerly Amazon Elasticsearch Service) has introduced capabilities to support vector search, which allows companies to build vector database applications. This is particularly useful in machine learning, where vector representations (embeddings) of data are often used to capture semantic meaning.

Scalable index management and nearest neighbor search capability are the core features enabling vector database functionalities in OpenSearch. The service allows users to index high-dimensional vectors and perform efficient nearest neighbor searches, which are crucial for tasks such as recommendation systems, anomaly detection, and semantic search.

Here is why option C is the correct Answer:

? Scalable Index Management: OpenSearch Service supports scalable indexing of vector data. This means you can index a large volume of high-dimensional vectors

and manage these indexes in a cost-effective and performance-optimized way. The service leverages underlying AWS infrastructure to ensure that indexing scales seamlessly with data size.

? Nearest Neighbor Search Capability: OpenSearch Service's nearest neighbor search capability allows for fast and efficient searches over vector data. This is essential for applications like product recommendation engines, where the system needs to quickly find the most similar items based on a user's query or behavior.

? AWS AI Practitioner References:

The other options do not directly relate to building vector database applications:

? A. Integration with Amazon S3 for object storage is about storing data objects, not vector-based searching or indexing.

? B. Support for geospatial indexing and queries is related to location-based data, not vectors used in machine learning.

? D. Ability to perform real-time analysis on streaming data relates to analyzing incoming data streams, which is different from the vector search capabilities.

NEW QUESTION 71

Which functionality does Amazon SageMaker Clarify provide?

- A. Integrates a Retrieval Augmented Generation (RAG) workflow
- B. Monitors the quality of ML models in production
- C. Documents critical details about ML models
- D. Identifies potential bias during data preparation

Answer: D

Explanation:

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.

NEW QUESTION 75

A company is building an ML model to analyze archived data. The company must perform inference on large datasets that are multiple GBs in size. The company does not need to access the model predictions immediately.

Which Amazon SageMaker inference option will meet these requirements?

- A. Batch transform
- B. Real-time inference
- C. Serverless inference
- D. Asynchronous inference

Answer: A

Explanation:

Batch transform in Amazon SageMaker is designed for offline processing of large datasets. It is ideal for scenarios where immediate predictions are not required, and the inference can be done on large datasets that are multiple gigabytes in size. This method processes data in batches, making it suitable for analyzing archived data without the need for real- time access to predictions.

? Option A (Correct): "Batch transform": This is the correct answer because batch transform is optimized for handling large datasets and is suitable when immediate access to predictions is not required.

? Option B: "Real-time inference" is incorrect because it is used for low-latency, real-time prediction needs, which is not required in this case.

? Option C: "Serverless inference" is incorrect because it is designed for small-scale, intermittent inference requests, not for large batch processing.

? Option D: "Asynchronous inference" is incorrect because it is used when immediate predictions are required, but with high throughput, whereas batch transform is more suitable for very large datasets.

AWS AI Practitioner References:

? Batch Transform on AWS SageMaker: AWS recommends using batch transform for large datasets when real-time processing is not needed, ensuring cost-effectiveness and scalability.

NEW QUESTION 76

An e-commerce company wants to build a solution to determine customer sentiments based on written customer reviews of products.

Which AWS services meet these requirements? (Select TWO.)

- A. Amazon Lex
- B. Amazon Comprehend
- C. Amazon Polly
- D. Amazon Bedrock
- E. Amazon Rekognition

Answer: BD

Explanation:

To determine customer sentiments based on written customer reviews, the company can use Amazon Comprehend and Amazon Bedrock.

? Amazon Comprehend:

? Amazon Bedrock:

? Why Other Options are Incorrect:

NEW QUESTION 80

A company wants to use AI to protect its application from threats. The AI solution needs to check if an IP address is from a suspicious source.

Which solution meets these requirements?

- A. Build a speech recognition system.
- B. Create a natural language processing (NLP) named entity recognition system.
- C. Develop an anomaly detection system.
- D. Create a fraud forecasting system.

Answer: C

Explanation:

An anomaly detection system is suitable for identifying unusual patterns or behaviors, such as suspicious IP addresses, which might indicate a potential threat.

? Anomaly Detection:

? Why Option C is Correct:

? Why Other Options are Incorrect:

Thus, C is the correct answer for detecting suspicious IP addresses.

NEW QUESTION 81

A company is using domain-specific models. The company wants to avoid creating new models from the beginning. The company instead wants to adapt pre-trained models to create models for new, related tasks.

Which ML strategy meets these requirements?

- A. Increase the number of epochs.
- B. Use transfer learning.
- C. Decrease the number of epochs.
- D. Use unsupervised learning.

Answer: B

Explanation:

Transfer learning is the correct strategy for adapting pre-trained models for new, related tasks without creating models from scratch.

? Transfer Learning:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 84

A student at a university is copying content from generative AI to write essays. Which challenge of responsible generative AI does this scenario represent?

- A. Toxicity
- B. Hallucinations
- C. Plagiarism
- D. Privacy

Answer: C

Explanation:

The scenario where a student copies content from generative AI to write essays represents the challenge of plagiarism in responsible AI use.

? Plagiarism:

? Why Option C is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 88

A company has built a solution by using generative AI. The solution uses large language models (LLMs) to translate training manuals from English into other

languages. The company wants to evaluate the accuracy of the solution by examining the text generated for the manuals. Which model evaluation strategy meets these requirements?

- A. Bilingual Evaluation Understudy (BLEU)
- B. Root mean squared error (RMSE)
- C. Recall-Oriented Understudy for Gisting Evaluation (ROUGE)
- D. F1 score

Answer: A

Explanation:

BLEU (Bilingual Evaluation Understudy) is a metric used to evaluate the accuracy of machine-generated translations by comparing them against reference translations. It is commonly used for translation tasks to measure how close the generated output is to professional human translations.

? Option A (Correct): "Bilingual Evaluation Understudy (BLEU)": This is the correct answer because BLEU is specifically designed to evaluate the quality of translations, making it suitable for the company's use case.

? Option B: "Root mean squared error (RMSE)" is incorrect because RMSE is used for regression tasks to measure prediction errors, not translation quality.

? Option C: "Recall-Oriented Understudy for Gisting Evaluation (ROUGE)" is incorrect as it is used to evaluate text summarization, not translation.

? Option D: "F1 score" is incorrect because it is typically used for classification tasks, not for evaluating translation accuracy.

AWS AI Practitioner References:

? Model Evaluation Metrics on AWS: AWS supports various metrics like BLEU for specific use cases, such as evaluating machine translation models.

NEW QUESTION 90

A company is using few-shot prompting on a base model that is hosted on Amazon Bedrock. The model currently uses 10 examples in the prompt. The model is invoked once daily and is performing well. The company wants to lower the monthly cost.

Which solution will meet these requirements?

- A. Customize the model by using fine-tuning.
- B. Decrease the number of tokens in the prompt.
- C. Increase the number of tokens in the prompt.
- D. Use Provisioned Throughput.

Answer: B

Explanation:

Decreasing the number of tokens in the prompt reduces the cost associated with using an LLM model on Amazon Bedrock, as costs are often based on the number of tokens processed by the model.

? Token Reduction Strategy:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 92

A company is building a solution to generate images for protective eyewear. The solution must have high accuracy and must minimize the risk of incorrect annotations.

Which solution will meet these requirements?

- A. Human-in-the-loop validation by using Amazon SageMaker Ground Truth Plus
- B. Data augmentation by using an Amazon Bedrock knowledge base
- C. Image recognition by using Amazon Rekognition
- D. Data summarization by using Amazon QuickSight

Answer: A

Explanation:

Amazon SageMaker Ground Truth Plus is a managed data labeling service that includes human-in-the-loop (HITL) validation. This solution ensures high accuracy by involving human reviewers to validate the annotations and reduce the risk of incorrect annotations.

? Amazon SageMaker Ground Truth Plus:

? Why Option A is Correct:

? Why Other Options are Incorrect:

Thus, A is the correct answer for generating high-accuracy images with minimized annotation risks.

NEW QUESTION 96

A loan company is building a generative AI-based solution to offer new applicants discounts based on specific business criteria. The company wants to build and use an AI model responsibly to minimize bias that could negatively affect some customers.

Which actions should the company take to meet these requirements? (Select TWO.)

- A. Detect imbalances or disparities in the data.
- B. Ensure that the model runs frequently.
- C. Evaluate the model's behavior so that the company can provide transparency to stakeholders.
- D. Use the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) technique to ensure that the model is 100% accurate.
- E. Ensure that the model's inference time is within the accepted limits.

Answer: AC

Explanation:

To build an AI model responsibly and minimize bias, it is essential to ensure fairness and transparency throughout the model development and deployment process. This involves detecting and mitigating data imbalances and thoroughly evaluating the model's behavior to understand its impact on different groups.

? Option A (Correct): "Detect imbalances or disparities in the data": This is correct because identifying and addressing data imbalances or disparities is a critical step in reducing bias. AWS provides tools like Amazon SageMaker Clarify to detect bias during data preprocessing and model training.

? Option C (Correct): "Evaluate the model's behavior so that the company can provide transparency to stakeholders": This is correct because evaluating the model's behavior for fairness and accuracy is key to ensuring that stakeholders understand how the model makes decisions. Transparency is a crucial aspect of

responsible AI.

? Option B: "Ensure that the model runs frequently" is incorrect because the frequency of model runs does not address bias.

? Option D: "Use the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) technique to ensure that the model is 100% accurate" is incorrect because ROUGE is a metric for evaluating the quality of text summarization models, not for minimizing bias.

? Option E: "Ensure that the model's inference time is within the accepted limits" is incorrect as it relates to performance, not bias reduction.

AWS AI Practitioner References:

? Amazon SageMaker Clarify: AWS offers tools such as SageMaker Clarify for detecting bias in datasets and models, and for understanding model behavior to ensure fairness and transparency.

? Responsible AI Practices: AWS promotes responsible AI by advocating for fairness, transparency, and inclusivity in model development and deployment.

NEW QUESTION 101

A company is building an application that needs to generate synthetic data that is based on existing data.

Which type of model can the company use to meet this requirement?

A. Generative adversarial network (GAN)

B. XGBoost

C. Residual neural network

D. WaveNet

Answer: A

Explanation:

Generative adversarial networks (GANs) are a type of deep learning model used for generating synthetic data based on existing datasets. GANs consist of two neural networks (a generator and a discriminator) that work together to create realistic data.

? Option A (Correct): "Generative adversarial network (GAN)": This is the correct answer because GANs are specifically designed for generating synthetic data that closely resembles the real data they are trained on.

? Option B: "XGBoost" is a gradient boosting algorithm for classification and regression tasks, not for generating synthetic data.

? Option C: "Residual neural network" is primarily used for improving the performance of deep networks, not for generating synthetic data.

? Option D: "WaveNet" is a model architecture designed for generating raw audio waveforms, not synthetic data in general.

AWS AI Practitioner References:

? GANs on AWS for Synthetic Data Generation: AWS supports the use of GANs for creating synthetic datasets, which can be crucial for applications like training machine learning models in environments where real data is scarce or sensitive.

NEW QUESTION 104

Which option is a benefit of using Amazon SageMaker Model Cards to document AI models?

A. Providing a visually appealing summary of a model's capabilities.

B. Standardizing information about a model's purpose, performance, and limitations.

C. Reducing the overall computational requirements of a model.

D. Physically storing models for archival purposes.

Answer: B

Explanation:

Amazon SageMaker Model Cards provide a standardized way to document important details about an AI model, such as its purpose, performance, intended usage, and known limitations. This enables transparency and compliance while fostering better communication between stakeholders. It does not store models physically or optimize computational requirements. References: AWS SageMaker Model Cards Documentation.

NEW QUESTION 107

.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your AIF-C01 Exam with Our Prep Materials Via below:

<https://www.certleader.com/AIF-C01-dumps.html>