

Amazon-Web-Services

Exam Questions AIF-C01

AWS Certified AI Practitioner



NEW QUESTION 1

A company wants to use a large language model (LLM) on Amazon Bedrock for sentiment analysis. The company wants to know how much information can fit into one prompt.

Which consideration will inform the company's decision?

- A. Temperature
- B. Context window
- C. Batch size
- D. Model size

Answer: B

Explanation:

The context window determines how much information can fit into a single prompt when using a large language model (LLM) like those on Amazon Bedrock.

? Context Window:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 2

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 3

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 4

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 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 medical company deployed a disease detection model on Amazon Bedrock. To comply with privacy policies, the company wants to prevent the model from including personal patient information in its responses. The company also wants to receive notification when policy violations occur.

Which solution meets these requirements?

- A. Use Amazon Macie to scan the model's output for sensitive data and set up alerts for potential violations.
- B. Configure AWS CloudTrail to monitor the model's responses and create alerts for any detected personal information.
- C. Use Guardrails for Amazon Bedrock to filter content.
- D. Set up Amazon CloudWatch alarms for notification of policy violations.
- E. Implement Amazon SageMaker Model Monitor to detect data drift and receive alerts when model quality degrades.

Answer: C

Explanation:

Guardrails for Amazon Bedrock provide mechanisms to filter and control the content generated by models to comply with privacy and policy requirements. Using guardrails ensures that sensitive or personal information is not included in the model's responses. Additionally, integrating Amazon CloudWatch alarms allows for real-time notification when a policy violation occurs.

? Option C (Correct): "Use Guardrails for Amazon Bedrock to filter content. Set up

Amazon CloudWatch alarms for notification of policy violations": This is the correct answer because it directly addresses both the prevention of policy violations and the requirement to receive notifications when such violations occur.

? Option A: "Use Amazon Macie to scan the model's output for sensitive data" is

incorrect because Amazon Macie is designed to monitor data in S3, not to filter real-time model outputs.

? Option B: "Configure AWS CloudTrail to monitor the model's responses" is

incorrect because CloudTrail tracks API activity and is not suited for content moderation.

? Option D: "Implement Amazon SageMaker Model Monitor to detect data drift" is

incorrect because data drift detection does not address content moderation or privacy compliance.

AWS AI Practitioner References:

? Guardrails in Amazon Bedrock: AWS provides guardrails to ensure AI models comply with content policies, and using CloudWatch for alerting integrates monitoring capabilities.

NEW QUESTION 7

How can companies use large language models (LLMs) securely on Amazon Bedrock?

- A. Design clear and specific prompt
- B. Configure AWS Identity and Access Management (IAM) roles and policies by using least privilege access.
- C. Enable AWS Audit Manager for automatic model evaluation jobs.
- D. Enable Amazon Bedrock automatic model evaluation jobs.
- E. Use Amazon CloudWatch Logs to make models explainable and to monitor for bias.

Answer: A

Explanation:

To securely use large language models (LLMs) on Amazon Bedrock, companies should design clear and specific prompts to avoid unintended outputs and ensure proper configuration of AWS Identity and Access Management (IAM) roles and policies with the principle of least privilege. This approach limits access to sensitive resources and minimizes the potential impact of security incidents.

? Option A (Correct): "Design clear and specific prompts. Configure AWS Identity

and Access Management (IAM) roles and policies by using least privilege access": This is the correct answer as it directly addresses both security practices in prompt design and access management.

? Option B: "Enable AWS Audit Manager for automatic model evaluation jobs" is

incorrect because Audit Manager is for compliance and auditing, not directly related to secure LLM usage.

? Option C: "Enable Amazon Bedrock automatic model evaluation jobs" is incorrect

because Bedrock does not provide automatic model evaluation jobs specifically for security purposes.

? Option D: "Use Amazon CloudWatch Logs to make models explainable and to

monitor for bias" is incorrect because CloudWatch Logs are used for monitoring and not directly for making models explainable or secure.

AWS AI Practitioner References:

? Secure AI Practices on AWS: AWS recommends configuring IAM roles and using least privilege access to ensure secure usage of AI models.

NEW QUESTION 8

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 9

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 10

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 10

A company wants to develop a large language model (LLM) application by using Amazon Bedrock and customer data that is uploaded to Amazon S3. The company's security policy states that each team can access data for only the team's own customers.

Which solution will meet these requirements?

A. Create an Amazon Bedrock custom service role for each team that has access to only the team's customer data.

B. Create a custom service role that has Amazon S3 access

C. Ask teams to specify the customer name on each Amazon Bedrock request.

D. Redact personal data in Amazon S3. Update the S3 bucket policy to allow team access to customer data.

E. Create one Amazon Bedrock role that has full Amazon S3 access

F. Create IAM roles for each team that have access to only each team's customer folders.

Answer: A

Explanation:

To comply with the company's security policy, which restricts each team to access data for only their own customers, creating an Amazon Bedrock custom service role for each team is the correct solution.

? Custom Service Role Per Team:

? Why Option A is Correct:

? Why Other Options are Incorrect:

Thus, A is the correct answer to meet the company's security requirements.

NEW QUESTION 11

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 14

A company wants to use language models to create an application for inference on edge devices. The inference must have the lowest latency possible.

Which solution will meet these requirements?

A. Deploy optimized small language models (SLMs) on edge devices.

B. Deploy optimized large language models (LLMs) on edge devices.

C. Incorporate a centralized small language model (SLM) API for asynchronous communication with edge devices.

D. Incorporate a centralized large language model (LLM) API for asynchronous communication with edge devices.

Answer: A

Explanation:

To achieve the lowest latency possible for inference on edge devices, deploying optimized small language models (SLMs) is the most effective solution. SLMs require fewer

resources and have faster inference times, making them ideal for deployment on edge devices where processing power and memory are limited.

? Option A (Correct): "Deploy optimized small language models (SLMs) on edge

devices": This is the correct answer because SLMs provide fast inference with low latency, which is crucial for edge deployments.

? Option B: "Deploy optimized large language models (LLMs) on edge devices" is

incorrect because LLMs are resource-intensive and may not perform well on edge devices due to their size and computational demands.

? Option C: "Incorporate a centralized small language model (SLM) API for

asynchronous communication with edge devices" is incorrect because it introduces network latency due to the need for communication with a centralized server.

? Option D: "Incorporate a centralized large language model (LLM) API for

asynchronous communication with edge devices" is incorrect for the same reason, with even greater latency due to the larger model size.

AWS AI Practitioner References:

? Optimizing AI Models for Edge Devices on AWS: AWS recommends using small, optimized models for edge deployments to ensure minimal latency and efficient performance.

NEW QUESTION 17

A company has thousands of customer support interactions per day and wants to analyze these interactions to identify frequently asked questions and develop insights.

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

A. Amazon Lex

- B. Amazon Comprehend
- C. Amazon Transcribe
- D. Amazon Translate

Answer: B

Explanation:

Amazon Comprehend is the correct service to analyze customer support interactions and identify frequently asked questions and insights.

? Amazon Comprehend:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 21

A company is using Amazon SageMaker Studio notebooks to build and train ML models. The company stores the data in an Amazon S3 bucket. The company needs to manage the flow of data from Amazon S3 to SageMaker Studio notebooks.

Which solution will meet this requirement?

- A. Use Amazon Inspector to monitor SageMaker Studio.
- B. Use Amazon Macie to monitor SageMaker Studio.
- C. Configure SageMaker to use a VPC with an S3 endpoint.
- D. Configure SageMaker to use S3 Glacier Deep Archive.

Answer: C

Explanation:

To manage the flow of data from Amazon S3 to SageMaker Studio notebooks securely, using a VPC with an S3 endpoint is the best solution.

? Amazon SageMaker and S3 Integration:

? Why Option C is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 26

Which metric measures the runtime efficiency of operating AI models?

- A. Customer satisfaction score (CSAT)
- B. Training time for each epoch
- C. Average response time
- D. Number of training instances

Answer: C

Explanation:

The average response time is the correct metric for measuring the runtime efficiency of operating AI models.

? Average Response Time:

? Why Option C is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 30

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 32

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 33

A company has built an image classification model to predict plant diseases from photos of plant leaves. The company wants to evaluate how many images the model classified correctly.

Which evaluation metric should the company use to measure the model's performance?

A. R-squared score

B. Accuracy

C. Root mean squared error (RMSE)

D. Learning rate

Answer: B

Explanation:

Accuracy is the most appropriate metric to measure the performance of an image classification model. It indicates the percentage of correctly classified images out of the total number of images. In the context of classifying plant diseases from images, accuracy will help the company determine how well the model is performing by showing how many images were correctly classified.

? Option B (Correct): "Accuracy": This is the correct answer because accuracy

measures the proportion of correct predictions made by the model, which is suitable for evaluating the performance of a classification model.

? Option A: "R-squared score" is incorrect as it is used for regression analysis, not classification tasks.

? Option C: "Root mean squared error (RMSE)" is incorrect because it is also used for regression tasks to measure prediction errors, not for classification accuracy.

? Option D: "Learning rate" is incorrect as it is a hyperparameter for training, not a performance metric.

AWS AI Practitioner References:

? Evaluating Machine Learning Models on AWS: AWS documentation emphasizes the use of appropriate metrics, like accuracy, for classification tasks.

NEW QUESTION 35

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 37

A company wants to deploy a conversational chatbot to answer customer questions. The chatbot is based on a fine-tuned Amazon SageMaker JumpStart model. The application must comply with multiple regulatory frameworks.

Which capabilities can the company show compliance for? (Select TWO.)

A. Auto scaling inference endpoints

B. Threat detection

C. Data protection

D. Cost optimization

E. Loosely coupled microservices

Answer: BC

Explanation:

To comply with multiple regulatory frameworks, the company must ensure data protection and threat detection. Data protection involves safeguarding sensitive

customer information, while threat detection identifies and mitigates security threats to the application.

? Option C (Correct): "Data protection": This is correct because data protection is critical for compliance with privacy and security regulations.

? Option B (Correct): "Threat detection": This is correct because detecting and mitigating threats is essential to maintaining the security posture required for regulatory compliance.

? Option A: "Auto scaling inference endpoints" is incorrect because auto-scaling does not directly relate to regulatory compliance.

? Option D: "Cost optimization" is incorrect because it is focused on managing expenses, not compliance.

? Option E: "Loosely coupled microservices" is incorrect because this architectural approach does not directly address compliance requirements.

AWS AI Practitioner References:

? AWS Compliance Capabilities: AWS offers services and tools, such as data protection and threat detection, to help companies meet regulatory requirements for security and privacy.

NEW QUESTION 38

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 39

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 40

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 45

Which option is a use case for generative AI models?

- A. Improving network security by using intrusion detection systems
- B. Creating photorealistic images from text descriptions for digital marketing
- C. Enhancing database performance by using optimized indexing
- D. Analyzing financial data to forecast stock market trends

Answer: B

Explanation:

Generative AI models are used to create new content based on existing data. One common use case is generating photorealistic images from text descriptions, which is particularly useful in digital marketing, where visual content is key to engaging potential customers.

? Option B (Correct): "Creating photorealistic images from text descriptions for digital marketing": This is the correct answer because generative AI models, like those offered by Amazon Bedrock, can create images based on text descriptions, making them highly valuable for generating marketing materials.

? Option A: "Improving network security by using intrusion detection systems" is incorrect because this is a use case for traditional machine learning models, not generative AI.

? Option C: "Enhancing database performance by using optimized indexing" is incorrect as it is unrelated to generative AI.

? Option D: "Analyzing financial data to forecast stock market trends" is incorrect because it typically involves predictive modeling rather than generative AI.

AWS AI Practitioner References:

? Use Cases for Generative AI Models on AWS: AWS highlights the use of generative AI for creative content generation, including image creation, text generation, and more, which is suited for digital marketing applications.

NEW QUESTION 49

A digital devices company wants to predict customer demand for memory hardware. The company does not have coding experience or knowledge of ML algorithms and needs to develop a data-driven predictive model. The company needs to perform analysis on internal data and external data.

Which solution will meet these requirements?

- A. Store the data in Amazon S3. Create ML models and demand forecast predictions by using Amazon SageMaker built-in algorithms that use the data from Amazon S3.
- B. Import the data into Amazon SageMaker Data Wrangle
- C. Create ML models and demand forecast predictions by using SageMaker built-in algorithms.
- D. Import the data into Amazon SageMaker Data Wrangle
- E. Build ML models and demand forecast predictions by using an Amazon Personalize Trending-Now recipe.
- F. Import the data into Amazon SageMaker Canva
- G. Build ML models and demand forecast predictions by selecting the values in the data from SageMaker Canvas.

Answer: D

Explanation:

Amazon SageMaker Canvas is a visual, no-code machine learning interface that allows users to build machine learning models without having any coding experience or knowledge of machine learning algorithms. It enables users to analyze internal and external data, and make predictions using a guided interface.

? Option D (Correct): "Import the data into Amazon SageMaker Canvas. Build ML models and demand forecast predictions by selecting the values in the data from SageMaker Canvas": This is the correct answer because SageMaker Canvas is designed for users without coding experience, providing a visual interface to build predictive models with ease.

? Option A: "Store the data in Amazon S3 and use SageMaker built-in algorithms" is incorrect because it requires coding knowledge to interact with SageMaker's built-in algorithms.

? Option B: "Import the data into Amazon SageMaker Data Wrangler" is incorrect.

Data Wrangler is primarily for data preparation and not directly focused on creating ML models without coding.

? Option C: "Use Amazon Personalize Trending-Now recipe" is incorrect as Amazon Personalize is for building recommendation systems, not for general demand forecasting.

AWS AI Practitioner References:

? Amazon SageMaker Canvas Overview: AWS documentation emphasizes Canvas as a no-code solution for building machine learning models, suitable for business analysts and users with no coding experience.

NEW QUESTION 54

A company has a database of petabytes of unstructured data from internal sources. The company wants to transform this data into a structured format so that its data scientists can perform machine learning (ML) tasks.

Which service will meet these requirements?

- A. Amazon Lex
- B. Amazon Rekognition
- C. Amazon Kinesis Data Streams
- D. AWS Glue

Answer: D

Explanation:

AWS Glue is the correct service for transforming petabytes of unstructured data into a structured format suitable for machine learning tasks.

? AWS Glue:

? Why Option D is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 55

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

Answer: C

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 58

A company has terabytes of data in a database that the company can use for business analysis. The company wants to build an AI-based application that can build a SQL query from input text that employees provide. The employees have minimal experience with technology.
Which solution meets these requirements?

- A. Generative pre-trained transformers (GPT)
- B. Residual neural network
- C. Support vector machine
- D. WaveNet

Answer: A

Explanation:

Generative Pre-trained Transformers (GPT) are suitable for building an AI-based application that can generate SQL queries from natural language input provided by employees.

? GPT for Natural Language Processing:

? Why Option A is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 59

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 63

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 67

A company is building a contact center application and wants to gain insights from customer conversations. The company wants to analyze and extract key information from the audio of the customer calls.

Which solution meets these requirements?

- A. Build a conversational chatbot by using Amazon Lex.
- B. Transcribe call recordings by using Amazon Transcribe.
- C. Extract information from call recordings by using Amazon SageMaker Model Monitor.
- D. Create classification labels by using Amazon Comprehend.

Answer: B

Explanation:

Amazon Transcribe is the correct solution for converting audio from customer calls into text, allowing the company to analyze and extract key information from the conversations.

? Amazon Transcribe:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 70

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 71

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 75

A pharmaceutical company wants to analyze user reviews of new medications and provide a concise overview for each medication. Which solution meets these requirements?

- A. Create a time-series forecasting model to analyze the medication reviews by using Amazon Personalize.
- B. Create medication review summaries by using Amazon Bedrock large language models (LLMs).
- C. Create a classification model that categorizes medications into different groups by using Amazon SageMaker.
- D. Create medication review summaries by using Amazon Rekognition.

Answer: B

Explanation:

Amazon Bedrock provides large language models (LLMs) that are optimized for natural language understanding and text summarization tasks, making it the best choice for creating concise summaries of user reviews. Time-series forecasting, classification, and image analysis (Rekognition) are not suitable for summarizing

textual data. References: AWS Bedrock Documentation.

NEW QUESTION 79

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 80

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 82

An AI practitioner is building a model to generate images of humans in various professions. The AI practitioner discovered that the input data is biased and that specific attributes affect the image generation and create bias in the model. Which technique will solve the problem?

- A. Data augmentation for imbalanced classes
- B. Model monitoring for class distribution
- C. Retrieval Augmented Generation (RAG)
- D. Watermark detection for images

Answer: A

Explanation:

Data augmentation for imbalanced classes is the correct technique to address bias in input data affecting image generation.

? Data Augmentation for Imbalanced Classes:

? Why Option A is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 86

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

AIF-C01 Practice Exam Features:

- * AIF-C01 Questions and Answers Updated Frequently
- * AIF-C01 Practice Questions Verified by Expert Senior Certified Staff
- * AIF-C01 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * AIF-C01 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The AIF-C01 Practice Test Here](#)