Amazon-Web-Services

Exam Questions AIF-C01

AWS Certified AI Practitioner

NEW QUESTION 1
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 2
An AI practitioner has built a deep learning model to classify the types of materials in images. The AI practitioner now wants to measure the model performance. Which metric will help the AI practitioner evaluate the performance of the model?

A. Confusion matrix
B. Correlation matrix
C. R2 score
D. Mean squared error (MSE)

Answer: A

Explanation:
A confusion matrix is the correct metric for evaluating the performance of a classification model, such as the deep learning model built to classify types of materials in images.
? Confusion Matrix:
? Why Option A is Correct:
? Why Other Options are Incorrect:

NEW QUESTION 3
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 4
A company wants to make a chatbot to help customers. The chatbot will help solve technical problems without human intervention. The company chose a foundation model (FM) for the chatbot. The chatbot needs to produce responses that adhere to company tone.
Which solution meets these requirements?

A. Set a low limit on the number of tokens the FM can produce.
B. Use batch inferencing to process detailed responses.
C. Experiment and refine the prompt until the FM produces the desired responses.
D. Define a higher number for the temperature parameter.

Answer: C

Explanation:
Experimenting and refining the prompt is the best approach to ensure that the chatbot using a foundation model (FM) produces responses that adhere to the company's tone.
? Prompt Engineering:

? Why Option C is Correct:
? Why Other Options are Incorrect:

NEW QUESTION 5
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 conten
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 6
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 7
An education provider is building a question and answer application that uses a generative AI model to explain complex concepts. The education provider wants to automatically change the style of the model response depending on who is asking the question. The education provider will give the model the age range of the user who has asked the question.
Which solution meets these requirements with the LEAST implementation effort?

A. Fine-tune the model by using additional training data that is representative of the various age ranges that the application will support.
B. Add a role description to the prompt context that instructs the model of the age range that the response should target.
C. Use chain-of-thought reasoning to deduce the correct style and complexity for a response suitable for that user.
D. Summarize the response text depending on the age of the user so that younger users receive shorter responses.

Answer: B

Explanation:
Adding a role description to the prompt context is a straightforward way to instruct the generative AI model to adjust its response style based on the user's age range. This method requires minimal implementation effort as it does not involve additional training or complex logic.
? Option B (Correct): "Add a role description to the prompt context that instructs the model of the age range that the response should target": This is the correct answer because it involves the least implementation effort while effectively guiding the
model to tailor responses according to the age range.
? Option A: "Fine-tune the model by using additional training data" is incorrect because it requires significant effort in gathering data and retraining the model.
? Option C: "Use chain-of-thought reasoning" is incorrect as it involves complex reasoning that may not directly address the need to adjust response style based on age.
? Option D: "Summarize the response text depending on the age of the user" is incorrect because it involves additional processing steps after generating the initial response, increasing complexity.
AWS AI Practitioner References:
? Prompt Engineering Techniques on AWS: AWS recommends using prompt context effectively to guide generative models in providing tailored responses based on specific user attributes.

NEW QUESTION 8
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 9
Which AWS feature records details about ML instance data for governance and reporting?

A. Amazon SageMaker Model Cards
B. Amazon SageMaker Debugger
C. Amazon SageMaker Model Monitor
D. Amazon SageMaker JumpStart

Answer: A

Explanation:
Amazon SageMaker Model Cards provide a centralized and standardized repository for documenting machine learning models. They capture key details such as the model's intended use, training and evaluation datasets, performance metrics, ethical considerations, and other relevant information. This documentation facilitates governance and reporting by ensuring that all stakeholders have access to consistent and comprehensive information about each model. While Amazon SageMaker Debugger is used for real-time debugging and monitoring during training, and Amazon SageMaker Model Monitor tracks deployed models for data and prediction quality, neither offers the comprehensive documentation capabilities of Model Cards. Amazon SageMaker JumpStart provides pre-built models and solutions but does not focus on governance documentation.
Reference: Amazon SageMaker Model Cards

NEW QUESTION 10
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 10
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 15
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 20
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 25
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 30
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 35
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 40
A company wants to create an application by using Amazon Bedrock. The company has a limited budget and prefers flexibility without long-term commitment.
Which Amazon Bedrock pricing model meets these requirements?

A. On-Demand
B. Model customization
C. Provisioned Throughput
D. Spot Instance

Answer: A

Explanation:
Amazon Bedrock offers an on-demand pricing model that provides flexibility without long- term commitments. This model allows companies to pay only for the resources they use, which is ideal for a limited budget and offers flexibility.
? Option A (Correct): "On-Demand": This is the correct answer because on-demand
pricing allows the company to use Amazon Bedrock without any long-term commitments and to manage costs according to their budget.
? Option B: "Model customization" is a feature, not a pricing model.
? Option C: "Provisioned Throughput" involves reserving capacity ahead of time, which might not offer the desired flexibility and could lead to higher costs if the capacity is not fully used.
? Option D: "Spot Instance" is a pricing model for EC2 instances and does not apply to Amazon Bedrock.
AWS AI Practitioner References:
? AWS Pricing Models for Flexibility: On-demand pricing is a key AWS model for services that require flexibility and no long-term commitment, ensuring cost-effectiveness for projects with variable usage patterns.

NEW QUESTION 45
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 social media company wants to use a large language model (LLM) for content moderation. The company wants to evaluate the LLM outputs for bias and potential discrimination against specific groups or individuals.
Which data source should the company use to evaluate the LLM outputs with the LEAST administrative effort?

A. User-generated content
B. Moderation logs
C. Content moderation guidelines
D. Benchmark datasets

Answer: D

Explanation:
Benchmark datasets are pre-validated datasets specifically designed to evaluate machine learning models for bias, fairness, and potential discrimination. These datasets are the most efficient tool for assessing an LLM??s performance against known standards with
minimal administrative effort.
? Option D (Correct): "Benchmark datasets": This is the correct answer because using standardized benchmark datasets allows the company to evaluate model outputs for bias with minimal administrative overhead.
? Option A: "User-generated content" is incorrect because it is unstructured and would require significant effort to analyze for bias.
? Option B: "Moderation logs" is incorrect because they represent historical data and do not provide a standardized basis for evaluating bias.
? Option C: "Content moderation guidelines" is incorrect because they provide qualitative criteria rather than a quantitative basis for evaluation.
AWS AI Practitioner References:
? Evaluating AI Models for Bias on AWS: AWS supports using benchmark datasets to assess model fairness and detect potential bias efficiently.

NEW QUESTION 51
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 53
An AI practitioner trained a custom model on Amazon Bedrock by using a training dataset that contains confidential data. The AI practitioner wants to ensure that the custom model does not generate inference responses based on confidential data.
How should the AI practitioner prevent responses based on confidential data?

A. Delete the custom mode
B. Remove the confidential data from the training datase
C. Retrain the custom model.
D. Mask the confidential data in the inference responses by using dynamic data masking.
E. Encrypt the confidential data in the inference responses by using Amazon SageMaker.
F. Encrypt the confidential data in the custom model by using AWS Key Management Service (AWS KMS).

Answer: A

Explanation:
When a model is trained on a dataset containing confidential or sensitive data, the model may inadvertently learn patterns from this data, which could then be reflected in its inference responses. To ensure that a model does not generate responses based on confidential data, the most effective approach is to remove the confidential data from the training dataset and then retrain the model.
Explanation of Each Option:
? Option A (Correct): "Delete the custom model. Remove the confidential data from the training dataset. Retrain the custom model."This option is correct because it directly addresses the core issue: the model has been trained on confidential data. The only way to ensure that the model does not produce inferences based on this data is to remove the confidential information from the training dataset and then retrain the model from scratch. Simply deleting the model and retraining it ensures that no confidential data is learned or retained by the model. This approach follows the best practices recommended by AWS for handling sensitive data when using machine learning services like Amazon Bedrock.
? Option B: "Mask the confidential data in the inference responses by using dynamic data masking."This option is incorrect because dynamic data masking is typically used to mask or obfuscate sensitive data in a database. It does not address the core problem of the model being trained on confidential data. Masking data in inference responses does not prevent the model from using confidential data it learned during training.

? Option C: "Encrypt the confidential data in the inference responses by using Amazon SageMaker."This option is incorrect because encrypting the inference responses does not prevent the model from generating outputs based on confidential data. Encryption only secures the data at rest or in transit but does not affect the model's underlying knowledge or training process.
? Option D: "Encrypt the confidential data in the custom model by using AWS Key Management Service (AWS KMS)."This option is incorrect as well because encrypting the data within the model does not prevent the model from generating responses based on the confidential data it learned during training. AWS KMS can encrypt data, but it does not modify the learning that the model has already performed.
AWS AI Practitioner References:
? Data Handling Best Practices in AWS Machine Learning: AWS advises practitioners to carefully handle training data, especially when it involves sensitive or confidential information. This includes preprocessing steps like data anonymization or removal of sensitive data before using it to train machine learning models.
? Amazon Bedrock and Model Training Security: Amazon Bedrock provides foundational models and customization capabilities, but any training involving sensitive data should follow best practices, such as removing or anonymizing confidential data to prevent unintended data leakage.

NEW QUESTION 54
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 56
A medical company is customizing a foundation model (FM) for diagnostic purposes. The company needs the model to be transparent and explainable to meet regulatory requirements.
Which solution will meet these requirements?

A. Configure the security and compliance by using Amazon Inspector.
B. Generate simple metrics, reports, and examples by using Amazon SageMaker Clarify.
C. Encrypt and secure training data by using Amazon Macie.
D. Gather more dat
E. Use Amazon Rekognition to add custom labels to the data.

Answer: B

Explanation:
Amazon SageMaker Clarify provides transparency and explainability for machine learning models by generating metrics, reports, and examples that help to understand model predictions. For a medical company that needs a foundation model to be transparent and explainable to meet regulatory requirements, SageMaker Clarify is the most suitable solution.
? Amazon SageMaker Clarify:
? Why Option B is Correct:
? Why Other Options are Incorrect:
Thus, B is the correct answer for meeting transparency and explainability requirements for the foundation model

NEW QUESTION 61
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 65

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 67
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 69
A company wants to use a large language model (LLM) on Amazon Bedrock for sentiment analysis. The company wants to classify the sentiment of text passages as positive or negative.
Which prompt engineering strategy meets these requirements?

A. Provide examples of text passages with corresponding positive or negative labels in the prompt followed by the new text passage to be classified.
B. Provide a detailed explanation of sentiment analysis and how LLMs work in the prompt.
C. Provide the new text passage to be classified without any additional context or examples.
D. Provide the new text passage with a few examples of unrelated tasks, such as text summarization or question answering.

Answer: A

Explanation:
Providing examples of text passages with corresponding positive or negative labels in the prompt followed by the new text passage to be classified is the correct prompt engineering strategy for using a large language model (LLM) on Amazon Bedrock for sentiment analysis.
? Example-Driven Prompts:
? Why Option A is Correct:
? Why Other Options are Incorrect:

NEW QUESTION 72
An AI practitioner is using a large language model (LLM) to create content for marketing campaigns. The generated content sounds plausible and factual but is incorrect.
Which problem is the LLM having?

A. Data leakage
B. Hallucination
C. Overfitting
D. Underfitting

Answer: B

Explanation:
In the context of AI, "hallucination" refers to the phenomenon where a model generates outputs that are plausible-sounding but are not grounded in reality or the training data. This problem often occurs with large language models (LLMs) when they create information that sounds correct but is actually incorrect or fabricated.
? Option B (Correct): "Hallucination": This is the correct answer because the
problem described involves generating content that sounds factual but is incorrect, which is characteristic of hallucination in generative AI models.
? Option A: "Data leakage" is incorrect as it involves the model accidentally learning
from data it shouldn't have access to, which does not match the problem of generating incorrect content.

? Option C: "Overfitting" is incorrect because overfitting refers to a model that has
learned the training data too well, including noise, and performs poorly on new data.
? Option D: "Underfitting" is incorrect because underfitting occurs when a model is
too simple to capture the underlying patterns in the data, which is not the issue here.
AWS AI Practitioner References:
? Large Language Models on AWS: AWS discusses the challenge of hallucination in large language models and emphasizes techniques to mitigate it, such as
using guardrails and fine-tuning.

NEW QUESTION 75
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 80
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 83
A company is using a pre-trained large language model (LLM) to build a chatbot for product recommendations. The company needs the LLM outputs to be short
and written in a specific language.
Which solution will align the LLM response quality with the company's expectations?

A. Adjust the prompt.
B. Choose an LLM of a different size.
C. Increase the temperature.
D. Increase the Top K value.

Answer: A

Explanation:
Adjusting the prompt is the correct solution to align the LLM outputs with the company's expectations for short, specific language responses.
? Adjust the Prompt:
? Why Option A is Correct:
? Why Other Options are Incorrect:

NEW QUESTION 86
A financial institution is using Amazon Bedrock to develop an AI application. The application is hosted in a VPC. To meet regulatory compliance standards, the
VPC is not allowed access to any internet traffic.
Which AWS service or feature will meet these requirements?

A. AWS PrivateLink
B. Amazon Macie
C. Amazon CloudFront
D. Internet gateway

Answer: A

Explanation:
AWS PrivateLink enables private connectivity between VPCs and AWS services without exposing traffic to the public internet. This feature is critical for meeting regulatory compliance standards that require isolation from public internet traffic.
? Option A (Correct): "AWS PrivateLink": This is the correct answer because it
allows secure access to Amazon Bedrock and other AWS services from a VPC without internet access, ensuring compliance with regulatory standards.
? Option B: "Amazon Macie" is incorrect because it is a security service for data
classification and protection, not for managing private network traffic.
? Option C: "Amazon CloudFront" is incorrect because it is a content delivery network service and does not provide private network connectivity.
? Option D: "Internet gateway" is incorrect as it enables internet access, which violates the VPC's no-internet-traffic policy.
AWS AI Practitioner References:
? AWS PrivateLink Documentation: AWS highlights PrivateLink as a solution for connecting VPCs to AWS services privately, which is essential for organizations with strict regulatory requirements.

NEW QUESTION 91
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 93
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 94
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 98
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 103
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 104
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 106
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 107
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 110
A company wants to create a chatbot by using a foundation model (FM) on Amazon Bedrock. The FM needs to access encrypted data that is stored in an Amazon S3 bucket.
The data is encrypted with Amazon S3 managed keys (SSE-S3).
The FM encounters a failure when attempting to access the S3 bucket data. Which solution will meet these requirements?

A. Ensure that the role that Amazon Bedrock assumes has permission to decrypt data withthe correct encryption key.
B. Set the access permissions for the S3 buckets to allow public access to enable access over the internet.
C. Use prompt engineering techniques to tell the model to look for information in Amazon S3.
D. Ensure that the S3 data does not contain sensitive information.

Answer: A

Explanation:
Amazon Bedrock needs the appropriate IAM role with permission to access and decrypt data stored in Amazon S3. If the data is encrypted with Amazon S3 managed keys (SSE- S3), the role that Amazon Bedrock assumes must have the required permissions to access and decrypt the encrypted data.
? Option A (Correct): "Ensure that the role that Amazon Bedrock assumes has permission to decrypt data with the correct encryption key": This is the correct solution as it ensures that the AI model can access the encrypted data securely without changing the encryption settings or compromising data security.
? Option B: "Set the access permissions for the S3 buckets to allow public access" is incorrect because it violates security best practices by exposing sensitive data to the public.
? Option C: "Use prompt engineering techniques to tell the model to look for information in Amazon S3" is incorrect as it does not address the encryption and permission issue.
? Option D: "Ensure that the S3 data does not contain sensitive information" is incorrect because it does not solve the access problem related to encryption.
AWS AI Practitioner References:
? Managing Access to Encrypted Data in AWS: AWS recommends using proper IAM roles and policies to control access to encrypted data stored in S3.

NEW QUESTION 114
A company has built a chatbot that can respond to natural language questions with images. The company wants to ensure that the chatbot does not return inappropriate or unwanted images.
Which solution will meet these requirements?

A. Implement moderation APIs.
B. Retrain the model with a general public dataset.
C. Perform model validation.
D. Automate user feedback integration.

Answer: A

Explanation:
Moderation APIs, such as Amazon Rekognition??s Content Moderation API, can help filter and block inappropriate or unwanted images from being returned by a chatbot. These APIs are specifically designed to detect and manage undesirable content in images.
? Option A (Correct): "Implement moderation APIs": This is the correct answer because moderation APIs are designed to identify and filter inappropriate content, ensuring the chatbot does not return unwanted images.
? Option B: "Retrain the model with a general public dataset" is incorrect because retraining does not directly prevent inappropriate content from being returned.
? Option C: "Perform model validation" is incorrect as it ensures model correctness, not content moderation.
? Option D: "Automate user feedback integration" is incorrect because user feedback does not prevent inappropriate images in real-time.
AWS AI Practitioner References:
? AWS Content Moderation Services: AWS provides moderation APIs for filtering unwanted content from applications.

NEW QUESTION 117
......

Thank You for Trying Our Product

We offer two products:

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

2nd - Questons 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