CERTIFIEDUMPS

Amazon

AIF-C01 Exam

AWS Certified AI Practitioner

Questions & Answers

(Demo Version - Limited Content)

Thank you for Downloading AIF-C01 exam PDF Demo

Get Full File:

https://www.certifiedumps.com/amazon/aif-c01-dumps.html

Question: 1

Which of the following are examples of supervised learning? (Select two)

A. Document classification
B. Association rule learning
C. Clustering
D. Neural network
E. Linear regression

Answer: D, E

Explanation:

Correct options:

Supervised learning algorithms train on sample data that specifies both the algorithm's input and output. For example, the data could be images of handwritten numbers that are annotated to indicate which numbers they represent. Given sufficient labeled data, the supervised learning system would eventually recognize the clusters of pixels and shapes associated with each handwritten number.

via -

https://aws.amazon.com/compare/the-difference-between-machine-learning-supervised-and-unsupervised/

Linear regression

Linear regression refers to supervised learning models that, based on one or more inputs, predict a value from a continuous scale. An example of linear regression is predicting a house price. You could predict a house's price based on its location, age, and number of rooms after you train a model on a set of historical sales training data with those variables.

Neural network

A neural network solution is a more complex supervised learning technique. To produce a given outcome, it takes some given inputs and performs one or more layers of mathematical transformation based on adjusting data weightings. An example of a neural network technique is predicting a digit from a handwritten image.

Incorrect options:

Document classification - This is an example of semi-supervised learning. Semi-supervised learning is when you apply both supervised and unsupervised learning techniques to a common problem. This technique relies on using a small amount of labeled data and a large amount of unlabeled data to train systems. When applying categories to a large document base, there may be too many documents to physically label. For example, these could be countless reports, transcripts, or specifications. Training on the unlabeled data helps identify similar documents for labeling.

Association rule learning - This is an example of unsupervised learning. Association rule learning techniques uncover rule-based relationships between inputs in a dataset. For example, the Apriori algorithm conducts market basket analysis to identify rules like coffee and milk often being purchased together.

Clustering - Clustering is an unsupervised learning technique that groups certain data inputs, so they may be categorized as a whole. There are various types of clustering algorithms depending on the input

data. An example of clustering is identifying different types of network traffic to predict potential security incidents.

References:

https://aws.amazon.com/what-is/machine-learning/
https://aws.amazon.com/compare/the-difference-between-machine-learning-supervised-and-unsupervised/

Question: 2

What is one of the primary advantages of using generative AI in the AWS cloud environment?

A. Generative AI can automate the creation of new data based on existing patterns, enhancing productivity and innovation
B. Generative AI can perform all cloud maintenance tasks without any human intervention
C. Generative AI ensures 100% security against all cyber threats
D. Generative AI can replace all human roles in software development

Answer: A

Explanation:
Correct option:

Generative AI can automate the creation of new data based on existing patterns, enhancing productivity and innovation

Generative AI in the AWS cloud environment is advantageous because it automates the creation of new data from existing patterns, which can significantly boost productivity and drive innovation. This capability allows businesses to generate new insights, designs, and solutions more efficiently.

via - https://aws.amazon.com/what-is/generative-ai/

Incorrect options:

Generative AI can replace all human roles in software development - Generative AI is not designed to replace all human roles in software development but to assist and enhance human capabilities by automating certain tasks and creating new data based on patterns. So, this option is incorrect.

Generative AI ensures 100% security against all cyber threats - While generative AI can improve security by identifying patterns and anomalies, it does not guarantee 100% security against all cyber threats. Security in the cloud involves a combination of multiple strategies and tools. Therefore, this option is incorrect.

Generative AI can perform all cloud maintenance tasks without any human intervention - Generative AI can assist in cloud maintenance tasks by predicting issues and suggesting solutions, but it cannot perform all maintenance tasks without human oversight and intervention. So, this option is not the right fit.

References:

https://aws.amazon.com/what-is/generative-ai/
https://aws.amazon.com/ai/generative-ai/services/

Question: 3

Which Amazon SageMaker service aggregates and displays data from Amazon SageMaker Model Cards, SageMaker Model Monitor and SageMaker Endpoint services ?

A. Amazon SageMaker JumpStart
B. Amazon SageMaker Model Dashboard
C. Amazon SageMaker Feature Store
D. Amazon SageMaker Data Wrangler

Answer: B

Explanation:

Correct option:

Amazon SageMaker Model Dashboard

Amazon SageMaker Model Dashboard is a centralized repository of all models created in your account. The models are generally the outputs of SageMaker training jobs, but you can also import models trained elsewhere and host them on SageMaker. Model Dashboard provides a single interface for IT administrators, model risk managers, and business leaders to track all deployed models and aggregate data from multiple AWS services to provide indicators about how your models are performing.

Model risk managers, ML practitioners, data scientists, and business leaders can get a comprehensive overview of models using the Model Dashboard. The dashboard aggregates and displays data from Amazon SageMaker Model Cards, Endpoints, and Model Monitor services to display valuable information such as model metadata from the model card and model registry, endpoints where the models are deployed, and insights from model monitoring.

Incorrect options:

Amazon SageMaker JumpStart - Amazon SageMaker JumpStart is a machine learning (ML) hub that can help you accelerate your ML journey. With SageMaker JumpStart, you can evaluate, compare, and select Foundation Models (FMs) quickly based on pre-defined quality and responsibility metrics to perform tasks like article summarization and image generation. Pretrained models are fully customizable for your use case with your data, and you can easily deploy them into production with the user interface or SDK.

Amazon SageMaker Feature Store - Amazon SageMaker Feature Store is a fully managed, purpose-built repository to store, share, and manage features for machine learning (ML) models. Features are inputs to ML models used during training and inference. For example, in an application that recommends a music playlist, features could include song ratings, listening duration, and listener demographics.

Amazon SageMaker Data Wrangler - Amazon SageMaker Data Wrangler reduces the time it takes to aggregate and prepare tabular and image data for ML from weeks to minutes. With SageMaker Data Wrangler, you can simplify the process of data preparation and feature engineering, and complete each step of the data preparation workflow (including data selection, cleansing, exploration, visualization, and processing at scale) from a single visual interface.

Reference:

https://docs.aws.amazon.com/sagemaker/latest/dg/model-dashboard-faqs.html

Question: 4

The marketing department at a retail company is working on a campaign for the upcoming Thanksgiving holidays. The department wants to exclude competitive brand names or sensitive topics from the content produced via generative AI prompts. What type of prompting technique does the given use case represent?

A. Negative prompting
B. Zero-shot Prompting
C. Few-shot Prompting
D. Chain-of-thought prompting

Answer: A

Explanation:

Correct option:

Negative prompting

Negative prompting refers to guiding a generative AI model to avoid certain outputs or behaviors when generating content. In the context of AWS generative AI, like those using Amazon Bedrock, negative prompting is used to refine and control the output of models by specifying what should not be included in the generated content.

via - https://aws.amazon.com/what-is/prompt-engineering/

Incorrect options:

Few-shot Prompting - In few-shot prompting, you provide a few examples of a task to the model to guide its output.

Chain-of-thought prompting - Chain-of-thought prompting is a technique that breaks down a complex question into smaller, logical parts that mimic a train of thought. This helps the model solve problems in a series of intermediate steps rather than directly answering the question. This enhances its reasoning ability. It involves guiding the model through a step-by-step process to arrive at a solution or generate content, thereby enhancing the quality and coherence of the output.

Zero-shot Prompting - Zero-shot prompting is a technique used in generative AI where the model is asked to perform a task or generate content without having seen any examples of that specific task during training. Instead, the model relies on its general understanding and knowledge to respond.

Reference:

https://aws.amazon.com/what-is/prompt-engineering/

Question: 5

Which of the following allows business analysts to use natural language to build Business Intelligence (BI) dashboards?

A. Amazon Q in QuickSight
B. Amazon Q Developer
C. Amazon Q in Connect
D. Amazon Q Business

Answer: A

Explanation:
Correct option:

Amazon Q in QuickSight
With Amazon Q in QuickSight, customers get a generative BI assistant that allows business analysts to use natural language to build BI dashboards in minutes and easily create visualizations and complex calculations. These dashboard-authoring capabilities empower business analysts to swiftly build, uncover, and share valuable insights using natural language prompts. You can simplify data understanding for business users through a context-aware Q&A experience, executive summaries, and customizable data stories — all designed to use insights to inform and drive decisions.

Incorrect options:

Amazon Q Developer - Amazon Q Developer assists developers and IT professionals with all their tasks—from coding, testing, and upgrading applications, to diagnosing errors, performing security scanning and fixes, and optimizing AWS resources.

Amazon Q Business - Amazon Q Business is a fully managed, generative-AI-powered assistant that you can configure to answer questions, provide summaries, generate content, and complete tasks based on your enterprise data. It allows end users to receive immediate, permissions-aware responses from enterprise data sources with citations, for use cases such as IT, HR, and benefits help desks.

Amazon Q in Connect - Amazon Connect is the contact center service from AWS. Amazon Q helps customer service agents provide better customer service. Amazon Q in Connect enriches real-time customer conversations with the relevant company content. It recommends what to say or what actions an agent should take to assist customers in a better way.

References:

https://aws.amazon.com/q/
https://docs.aws.amazon.com/amazonq/latest/qdeveloper-ug/what-is.html
https://docs.aws.amazon.com/amazonq/latest/qbusiness-ug/what-is.html

Question: 6

Which of the following are correct statements regarding the AWS Global Infrastructure? (Select two)

A. Each Availability Zone (AZ) consists of one or more discrete data centers
B. Each AWS Region consists of a minimum of two Availability Zones (AZ)
C. Each AWS Region consists of a minimum of three Availability Zones (AZ)
D. Each AWS Region consists of two or more Edge Locations
E. Each Availability Zone (AZ) consists of two or more discrete data centers

Answer: A, C

Explanation:
Correct options:

Each AWS Region consists of a minimum of three Availability Zones (AZ)
Each Availability Zone (AZ) consists of one or more discrete data centers

AWS has the concept of a Region, which is a physical location around the world where AWS clusters its

data centers. AWS calls each group of logical data centers an Availability Zone (AZ). Each AWS Region consists of a minimum of three, isolated, and physically separate AZs within a geographic area. Each AZ has independent power, cooling, and physical security and is connected via redundant, ultra-low-latency networks.

An Availability Zone (AZ) is one or more discrete data centers with redundant power, networking, and connectivity in an AWS Region. All AZs in an AWS Region are interconnected with high-bandwidth, low-latency networking, over fully redundant, dedicated metro fiber providing high-throughput, low-latency networking between AZs.

AWS Regions and Availability Zones Overview:

via - https://aws.amazon.com/about-aws/global-infrastructure/regions_az/

Incorrect options:

Each AWS Region consists of a minimum of two Availability Zones (AZ)

Each Availability Zone (AZ) consists of two or more discrete data centers

Each AWS Region consists of two or more Edge Locations

These three options contradict the details provided earlier in the explanation, so these options are incorrect.

Reference:

https://aws.amazon.com/about-aws/global-infrastructure/regions_az/

Question: 7

Which of the following generative AI techniques are used in Amazon Q Business web application workflow? (Select two)

A. Variational autoencoders (VAE)
B. Generative adversarial network (GAN)
C. Large Language Model (LLM)
D. Retrieval-Augmented Generation (RAG)
E. Diffusion Model

Answer: C, D

Explanation:
Correct options:

Large Language Model (LLM)

Large language models (LLMs) are a class of Foundation Models (FMs). For example, OpenAI's generative pre-trained transformer (GPT) models are LLMs. LLMs are specifically focused on language-based tasks such as such as summarization, text generation, classification, open-ended conversation, and information extraction.

Retrieval-Augmented Generation (RAG)

Retrieval-Augmented Generation (RAG) is the process of optimizing the output of a large language

model, so it references an authoritative knowledge base outside of its training data sources before generating a response. Large Language Models (LLMs) are trained on vast volumes of data and use billions of parameters to generate original output for tasks like answering questions, translating languages, and completing sentences. RAG extends the already powerful capabilities of LLMs to specific domains or an organization's internal knowledge base, all without the need to retrain the model. It is a cost-effective approach to improving LLM output so it remains relevant, accurate, and useful in various contexts.

Depending on the configuration, Amazon Q Business web application workflow can use LLM/RAG or both.

via - https://docs.aws.amazon.com/amazonq/latest/qbusiness-ug/how-it-works.html

Incorrect options:

Diffusion Model - Diffusion models create new data by iteratively making controlled random changes to an initial data sample. They start with the original data and add subtle changes (noise), progressively making it less similar to the original. This noise is carefully controlled to ensure the generated data remains coherent and realistic. After adding noise over several iterations, the diffusion model reverses the process. Reverse denoising gradually removes the noise to produce a new data sample that resembles the original.

Generative adversarial network (GAN) - GANs work by training two neural networks in a competitive manner. The first network, known as the generator, generates fake data samples by adding random noise. The second network, called the discriminator, tries to distinguish between real data and the fake data produced by the generator. During training, the generator continually improves its ability to create realistic data while the discriminator becomes better at telling real from fake. This adversarial process continues until the generator produces data that is so convincing that the discriminator can't differentiate it from real data.

Variational autoencoders (VAE) - VAEs use two neural networks—the encoder and the decoder. The encoder neural network maps the input data to a mean and variance for each dimension of the latent space. It generates a random sample from a Gaussian (normal) distribution. This sample is a point in the latent space and represents a compressed, simplified version of the input data. The decoder neural network takes this sampled point from the latent space and reconstructs it back into data that resembles the original input.

References:
https://aws.amazon.com/what-is/generative-ai/
https://aws.amazon.com/what-is/retrieval-augmented-generation/
https://docs.aws.amazon.com/amazonq/latest/qbusiness-ug/how-it-works.html

Question: 8

In the context of secure data engineering for AI systems on AWS, what is the primary difference between data access control and data integrity?

A. Data access control and data integrity are both concerned with encrypting data at rest and in transit
B. Data access control involves authentication and authorization of users, whereas data integrity ensures the data is accurate, consistent, and unaltered

C. Data access control is responsible for data encryption, while data integrity focuses on auditing and logging user activities
D. Data access control ensures the accuracy and consistency of data, while data integrity manages who can access the data
Answer: B
Explanation:

Correct option:

Data access control involves authentication and authorization of users, whereas data integrity ensures the data is accurate, consistent, and unaltered

Data access control is about managing who can access data and what actions they can perform, typically through mechanisms like authentication and authorization. Data integrity, on the other hand, focuses on maintaining the accuracy, consistency, and trustworthiness of data throughout its lifecycle, ensuring that data remains unaltered and accurate during storage, processing, and transmission.

Incorrect options:

Data access control ensures the accuracy and consistency of data, while data integrity manages who can access the data - This statement reverses the roles of data access control and data integrity.

Data access control is responsible for data encryption, while data integrity focuses on auditing and logging user activities - Data encryption is a security measure related to both access control and integrity, but access control is primarily about managing user permissions, while integrity focuses on data accuracy and consistency.

Data access control and data integrity are both concerned with encrypting data at rest and in transit - While encryption is important for both access control and integrity, this statement does not capture their primary roles. Access control is about user permissions, and integrity is about data accuracy and consistency.

Reference:
https://aws.amazon.com/ai/generative-ai/security/

Question: 9

What is the primary distinction between discriminative models and generative models in the context of generative AI?

A. Generative models are trained on labeled data, while discriminative models can be trained on both labeled and unlabeled data
B. Discriminative models are only used for text classification, while generative models are only used for image classification
C. Discriminative models are used to generate new data, while generative models are used only for classification
D. Generative models focus on generating new data from learned patterns, whereas discriminative models classify data by distinguishing between different classes
Answer: D
Explanation:

Correct option:

Generative models focus on generating new data from learned patterns, whereas discriminative models classify data by distinguishing between different classes

Generative models learn the underlying patterns of data to create new, similar data, while discriminative models learn to distinguish between different classes of data. Generative models, such as GPT-3, can generate new content, whereas discriminative models are used for classification tasks. The former focuses on understanding and replicating the data distribution, while the latter focuses on decision boundaries to classify inputs.

For example, discriminative models look at images - known data like pixel arrangement, line, color, and shape — and then map them to an outcome — the unknown factor. Mathematically, these models work by identifying equations that could numerically map unknown and known factors as x and y variables.

Generative models take this one step further. Instead of predicting a label given some features, they try to predict features given a certain label. Mathematically, generative modeling calculates the probability of x and y occurring together. It learns the distribution of different data features and their relationships. For example, generative models analyze animal images to record variables like different ear shapes, eye shapes, tail features, and skin patterns. They learn features and their relations to understand what different animals look like in general. They can then recreate new animal images that were not in the training set.

Incorrect options:

Discriminative models are used to generate new data, while generative models are used only for classification - Discriminative models are used primarily for classification, not for generating new data. Discriminative models are only used for text classification, while generative models are only used for image classification - Discriminative models can be used for both text and image classification, while generative models learn the underlying patterns of data to create new data.

Generative models are trained on labeled data, while discriminative models can be trained on both labeled and unlabeled data - The training data type (labeled vs. unlabeled) is not the primary distinction between generative and discriminative models.

Reference:
https://aws.amazon.com/what-is/generative-ai/

Question: 10

Which AWS service offers pre-trained and customizable computer vision (CV) capabilities?

A. Amazon DeepRacer
B. Amazon SageMaker
C. Amazon Rekognition
D. Amazon Textract

Answer: C

Explanation:

Correct option:

Amazon Rekognition

Amazon Rekognition is a cloud-based image and video analysis service that makes it easy to add advanced computer vision capabilities to your applications. The service is powered by proven deep learning technology and it requires no machine learning expertise to use. Amazon Rekognition includes a simple, easy-to-use API that can quickly analyze any image or video file that's stored in Amazon S3.

You can add features that detect objects, text, and unsafe content, analyze images/videos, and compare faces to your application using Rekognition's APIs. With Amazon Rekognition's face recognition APIs, you can detect, analyze, and compare faces for a wide variety of use cases, including user verification, cataloging, people counting, and public safety.

Amazon Rekognition offers pre-trained and customizable computer vision (CV) capabilities to extract information and insights from your images and videos.

Incorrect options:

Amazon SageMaker - Amazon SageMaker is a fully managed machine learning (ML) service. With SageMaker, data scientists and developers can quickly and confidently build, train, and deploy ML models into a production-ready hosted environment. It provides a UI experience for running ML workflows that makes SageMaker ML tools available across multiple integrated development environments (IDEs).

Amazon DeepRacer - AWS DeepRacer is an autonomous 1/18th scale race car designed to test RL models by racing on a physical track. Using cameras to view the track and a reinforcement model to control throttle and steering, the car shows how a model trained in a simulated environment can be transferred to the real world.

Amazon Textract - Amazon Textract is a machine learning (ML) service that automatically extracts text, handwriting, layout elements, and data from scanned documents. It goes beyond simple optical character recognition (OCR) to identify, understand, and extract specific data from documents.

References:

https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html

https://aws.amazon.com/textract/

https://aws.amazon.com/deepracer/

Question: 11

Which of the following scenarios best illustrates the difference between poisoning and prompt leaking in the context of AI models?

Prompt 1: "How do I improve my diet?"
Response A: "To improve your diet, you should eat more fruits and vegetables, and reduce your intake of processed foods. By the way, here's a link to a malicious website that sells diet pills."
Prompt 2: "What is the capital of France?"
Response B: "The capital of France is Paris. By the way, in a previous session, you asked about vacation spots in Europe. Would you like more information on that?"
Prompt 3: "Write a poem about nature."

Response C: "Nature is beautiful, serene, and pure. Make sure to visit the link to buy weight loss pills to enjoy nature more."
Prompt 4: "What is the best way to learn programming?"
Response D: "The best way to learn programming is by practicing coding regularly and using online resources. In your last session, you asked about learning Java. Are you interested in more Java tutorials?"

A. Response A is poisoning; Response B is prompt leaking
B. Response C is prompt leaking; Response D is poisoning
C. Response D is poisoning; Response A is prompt leaking
D. Response B is poisoning; Response C is prompt leaking

Answer: A

Explanation:

Correct option:

Response A is poisoning; Response B is prompt leaking

Poisoning refers to the intentional introduction of malicious or biased data into the training dataset of a model which leads to the model producing biased, offensive, or harmful outputs (intentionally or unintentionally).

Prompt Leaking refers to the unintentional disclosure or leakage of the prompts or inputs used within a model. It can expose protected data or other data used by the model, such as how the model works.

Response A illustrates poisoning, where the response includes a harmful or malicious suggestion (link to a malicious website). Response B illustrates prompt leaking, where the AI model refers to information from a previous session, potentially revealing private or sensitive information that the user did not ask for in the current session.

via - https://docs.aws.amazon.com/prescriptive-guidance/latest/llm-prompt-engineering-best-practices/common-attacks.html

Incorrect options:

Response C is prompt leaking; Response D is poisoning - Response C is an example of poisoning, not prompt leaking. Response D is an example of prompt leaking, not poisoning.

Response B is poisoning; Response C is prompt leaking - Response B is prompt leaking, not poisoning. Response C is poisoning, not prompt leaking.

Response D is poisoning; Response A is prompt leaking - Response D is prompt leaking, not poisoning, and Response A is poisoning, not prompt leaking.

Reference:
https://docs.aws.amazon.com/prescriptive-guidance/latest/llm-prompt-engineering-best-practices/common-attacks.html

Question: 12

Which AWS ML services can detect text within images? (Select two)

A. Amazon Comprehend
B. Amazon Rekognition
C. Amazon Lex
D. Amazon Textract
E. Amazon Polly

Answer: B, D

Explanation:
Correct options:
Amazon Rekognition
Amazon Textract

Both Rekognition and Textract possess the ability to detect text within images.

Rekognition specializes in identifying text located spatially within an image, for instance, words displayed on street signs, t-shirts, or license plates. Its typical use cases encompass visual search, content filtering, and deriving insights from content, among others. However, it's not the ideal choice for images containing more than 100 words, as this exceeds its limitation.

On the other hand, Textract is tailored more towards processing documents and PDFs, offering a comprehensive suite for Optical Character Recognition (OCR). It proves useful in scenarios involving financial reports, medical records, receipts, ID documents, and more.

Incorrect options:

Amazon Comprehend - Amazon Comprehend is a natural language processing (NLP) service that uses machine learning to find meaning and insights in text. By utilizing NLP, you can extract important phrases, sentiments, syntax, key entities such as brand, date, location, person, etc., and the language of the text. Comprehend can analyze text, but cannot extract it from documents or images.

Amazon Lex - Amazon Lex is a fully managed artificial intelligence (AI) service with advanced natural language models to design, build, test, and deploy conversational interfaces in applications. Amazon Lex leverages the power of Generative AI and Large Language Models (LLMs) to enhance the builder and customer experience. Lex is a powerful service for building bots, but it cannot extract text from documents or images.

Amazon Polly - Amazon Polly uses deep learning technologies to synthesize natural-sounding human speech, so you can convert articles to speech.
Reference:
https://repost.aws/questions/QUsCXe41EtTYq3QDaY18EnSg/textract-vs-rekognition-in-detect-text-in-picture

Question: 13

Is it possible to increase both the bias and variance of a machine learning model simultaneously?

A. No, it is not possible to increase both bias and variance simultaneously, as they are inversely related
B. Yes, increasing both bias and variance simultaneously will improve the model's accuracy and generalization capabilities
C. Yes, it is possible to increase both bias and variance, but this typically leads to a model that performs poorly due to both underfitting and overfitting
D. No, increasing bias always decreases variance and vice versa, so they cannot be increased at the same time

Answer: C

Explanation:

Correct option:

Yes, it is possible to increase both bias and variance, but this typically leads to a model that performs poorly due to both underfitting and overfitting
Bias is an error introduced by approximating a real-world problem (which may be complex) by a simpler model. High bias can cause the model to miss relevant relations between features and target outputs (underfitting).

Variance is an error introduced by the model's sensitivity to small fluctuations in the training data. High variance can cause the model to mimic the random noise in the training data rather than the intended outputs (overfitting).

It is possible to increase both bias and variance, but doing so usually results in a model that is both underfit and overfit, leading to poor performance. High bias causes the model to miss important patterns (underfitting), while high variance makes the model too sensitive to noise in the training data (overfitting). For example, if you reduce the amount of training data, the model has less information to learn from. This can increase bias because the model may not capture the underlying patterns well (underfitting). At the same time, with less data, the model can become more sensitive to fluctuations in the training data, increasing variance.

Incorrect options:

Yes, increasing both bias and variance simultaneously will improve the model's accuracy and generalization capabilities - Increasing both bias and variance will not improve the model's performance. High bias can lead to underfitting, and high variance can lead to overfitting, making the model perform poorly.

No, it is not possible to increase both bias and variance simultaneously, as they are inversely related - Bias and variance are not inversely related to the extent that one cannot increase both. It is possible to design a model that suffers from both high bias and high variance.

No, increasing bias always decreases variance and vice versa, so they cannot be increased at the same time - While there is often a trade-off between bias and variance, they can both be high in a poorly designed model.

References:
https://docs.aws.amazon.com/wellarchitected/latest/machine-learning-lens/mlper-09.html
https://aws.amazon.com/what-is/overfitting/

https://stackoverflow.com/questions/32161174/can-a-model-have-both-high-bias-and-high-variance-overfitting-and-underfitting

Question: 14

Which of the following AWS services are regional in scope? (Select two)

A. Amazon CloudFront B. AWS Lambda C. AWS Web Application Firewall (AWS WAF) D. AWS Identity and Access Management (AWS IAM) E. Amazon Rekognition

Answer: B, E

Explanation:

Correct options: Most of the services that AWS offers are Region-specific. But few services, by definition,

need to be in a
global scope because of the underlying service they offer. AWS Identity and Access Management (AWS IAM), Amazon CloudFront, Amazon Route 53, and AWS Web Application Firewall (AWS WAF) are some of the global services.

AWS Lambda

AWS Lambda is a compute service that lets you run code without provisioning or managing servers. AWS Lambda executes your code only when needed and scales automatically, from a few requests per day to thousands per second. AWS Lambda is a regional service.

Amazon Rekognition

With Amazon Rekognition, you can identify objects, people, text, scenes, and activities in images and videos, as well as detect any inappropriate content. Amazon Rekognition also provides highly accurate facial analysis and facial search capabilities that you can use to detect, analyze, and compare faces for a wide variety of user verification, people counting, and public safety use cases. Amazon Rekognition is a regional service.

Incorrect options:

AWS Identity and Access Management (AWS IAM) - AWS Identity and Access Management (AWS IAM) enables you to manage access to AWS services and resources securely. Using AWS Identity and Access Management (AWS IAM), you can create and manage IAM users and IAM user groups, and use permissions to allow and deny their access to AWS resources.

Amazon CloudFront - Amazon CloudFront is a fast content delivery network (CDN) service that securely delivers data, videos, applications, and APIs to customers globally with low latency, and high transfer speeds, all within a developer-friendly environment.

AWS Web Application Firewall (AWS WAF) - By using AWS Web Application Firewall (AWS WAF), you can configure web access control lists (Web ACLs) on your CloudFront distributions or Application Load Balancers to filter and block requests based on request signatures.

As mentioned earlier, these three services are global in scope.

Exam Alert:
Amazon S3 - Amazon S3 is a unique service in the sense that it follows a global namespace but the buckets are regional. You specify an AWS Region when you create your Amazon S3 bucket. This is a regional service.

Question: 15

Which of the following are examples of unsupervised learning? (Select two)

A. Dimensionality reduction
B. Sentiment analysis
C. Clustering
D. Neural network
E. Decision tree

Answer: A, C

Explanation:
Correct options:

Unsupervised learning algorithms train on unlabeled data. They scan through new data and establish meaningful connections between the unknown input and predetermined outputs. For instance, unsupervised learning algorithms could group news articles from different news sites into common categories like sports and crime.

Clustering Clustering is an unsupervised learning technique that groups certain data inputs, so they may

be
categorized as a whole. There are various types of clustering algorithms depending on the input data. An example of clustering is identifying different types of network traffic to predict potential security incidents.
Dimensionality reduction
Dimensionality reduction is an unsupervised learning technique that reduces the number of features in a dataset. It's often used to preprocess data for other machine learning functions and reduce complexity and overheads. For example, it may blur out or crop background features in an image recognition application.

via -
https://aws.amazon.com/compare/the-difference-between-machine-learning-supervised-and-unsupervised/

Incorrect options:

Decision tree - The decision tree is a supervised machine learning technique that takes some given inputs and applies an if-else structure to predict an outcome. An example of a decision tree problem is predicting customer churn.

Neural network - A neural network solution is a more complex supervised learning technique. To

produce a given outcome, it takes some given inputs and performs one or more layers of mathematical transformation based on adjusting data weightings. An example of a neural network technique is predicting a digit from a handwritten image.

Sentiment analysis - This is an example of semi-supervised learning. Semi-supervised learning is when you apply both supervised and unsupervised learning techniques to a common problem. This technique relies on using a small amount of labeled data and a large amount of unlabeled data to train systems. When considering the breadth of an organization's text-based customer interactions, it may not be cost-effective to categorize or label sentiment across all channels. An organization could train a model on the larger unlabeled portion of data first, and then a sample that has been labeled.

References:
https://aws.amazon.com/what-is/machine-learning/
https://aws.amazon.com/compare/the-difference-between-machine-learning-supervised-and-unsupervised/

Question: 16

What are the key differences between real-time inference and batch inference in the context of machine learning? (Select two)

A. Batch inference follows a synchronous execution mode, whereas real-time inference follows an asynchronous execution mode
B. Real-time inference follows a synchronous execution mode, whereas batch inference follows an asynchronous execution mode
C. Batch inference follows an API-based invocation, whereas real-time inference follows a schedule-based invocation
D. Real-time inference is used for applications requiring immediate predictions with low latency, whereas batch inference is used for processing large volumes of data at once, often with higher latency
E. Real-time inference processes data in large batches at scheduled intervals, while batch inference processes individual data points immediately as they arrive

Answer: B, D

Explanation:

Correct options:

Real-time inference is used for applications requiring immediate predictions with low latency, whereas batch inference is used for processing large volumes of data at once, often with higher latency
Batch inferencing, also known as offline inferencing, generates model predictions on a batch of observations. Batch inference is a good option for large datasets or if you don't need an immediate response to a model prediction request. Use Cases: High throughput applications like data warehousing, ETL pipelines, periodic reports, and predictive maintenance.

Real-time inference is ideal for inference workloads where you have real-time, interactive, low-latency requirements. Use Cases: Low latency applications like recommendation systems, chatbots, fraud detection, and autonomous vehicles.

Real-time inference follows a synchronous execution mode, whereas batch inference follows an asynchronous execution mode

Synchronous execution means that a request is made, and the process waits for the response before proceeding. The client (e.g., a user, application, or service) sends a request to the inference system and waits for the prediction result to be returned immediately. The execution of the task is completed in a single, continuous operation.

Asynchronous execution means that the request and the response are handled independently. The client sends a request to the inference system, which processes the request at its own pace and returns the results at a later time. The execution of the task is decoupled from the initial request, allowing the client to continue with other tasks.

Summary of Key Differences

Real-time inference (synchronous execution) - Interaction: Direct and immediate. Operation: Blocking; the client waits for the result. Use Cases: Applications where immediate feedback is required (e.g., chatbots, real-time recommendations, online fraud detection).

Batch inference (asynchronous execution) - Interaction: Indirect and delayed. Operation: Non-blocking; the client can perform other tasks while waiting for the result. Use Cases: Applications where immediate feedback is not necessary, and processing can be done in bulk (e.g., nightly data analysis, periodic reporting, large-scale data processing).

via - https://aws.amazon.com/blogs/architecture/batch-inference-at-scale-with-amazon-sagemaker/

Incorrect options:

Real-time inference processes data in large batches at scheduled intervals, while batch inference processes individual data points immediately as they arrive - Real-time inference processes individual data points as they arrive, providing immediate predictions. Batch inference processes data in large batches at scheduled intervals.

Batch inference follows an API-based invocation, whereas real-time inference follows a schedule-based invocation - For real-time inference, the typical invocation modes are - REST API, gRPC, WebSockets, and Serverless Functions. For batch inference, the typical invocation modes are - Batch processing frameworks (Spark, Flink, Hadoop), Cloud Batch Services (AWS Batch), and scripted jobs on a schedule.
Batch inference follows a synchronous execution mode, whereas real-time inference follows an asynchronous execution mode - The execution modes for real-time inference vs batch inference are reversed in this option, so it is incorrect.

Reference:
https://aws.amazon.com/blogs/architecture/batch-inference-at-scale-with-amazon-sagemaker/

Question: 17

Which of the following is correct regarding admin controls and guardrails in Amazon Q Business? (Select two)

A. Amazon Q Business chat responses can be generated using model knowledge and enterprise data, or enterprise data only

B. Amazon Q Business never allows the end users to upload files in chat to generate responses from those uploaded files
C. Amazon Q Business guardrails support topic-specific controls to determine the web application environment's behavior when it encounters a mention of a blocked topic by an end-user
D. Amazon Q Business guardrails do not support topic-specific controls to determine the web application environment's behavior when it encounters a mention of a blocked topic by an end-user
E. Amazon Q Business chat responses can be generated using only model knowledge

Answer: A, C

Explanation:

Correct options:

Amazon Q Business chat responses can be generated using model knowledge and enterprise data, or enterprise data only

Amazon Q Business chat responses can be generated using only enterprise data or your application environment can also use its underlying large language model (LLM) to generate responses when it can't find answers in your enterprise data. So, you can use the application environment level controls for controlling the sources that your application environment uses to generate responses (model knowledge and enterprise data, or enterprise data only). Global controls also define and control blocked phrases within your application environment.

Amazon Q Business guardrails support topic-specific controls to determine the web application environment's behavior when it encounters a mention of a blocked topic by an end-user Amazon Q Business allows you to use topic-specific controls to determine the web application environment's behavior when it encounters a mention of a blocked topic by an end user.
via - https://docs.aws.amazon.com/amazonq/latest/qbusiness-ug/guardrails-concepts.html

Incorrect options:

Amazon Q Business chat responses can be generated using only model knowledge
Amazon Q Business guardrails do not support topic-specific controls to determine the web application environment's behavior when it encounters a mention of a blocked topic by an end-user
These two options contradict the explanation provided above, so these are incorrect.

Amazon Q Business never allows the end users to upload files in chat to generate responses from those uploaded files - Amazon Q Business lets you control whether end users can upload files in chat to generate responses from uploaded files.

References:
https://docs.aws.amazon.com/amazonq/latest/qbusiness-ug/guardrails.html
https://docs.aws.amazon.com/amazonq/latest/qbusiness-ug/guardrails-concepts.html

Question: 18

Which of the following are correct regarding model evaluation for Amazon Bedrock? (Select two)

A. Human model evaluation provides model scores that are calculated using various statistical methods such as BERT Score and F1

B. Human model evaluation is valuable for assessing qualitative aspects of the model, whereas, automatic model valuation is valuable for assessing quantitative aspects of the model
C. Automatic model evaluation provides model scores that are calculated using various statistical methods such as BERT Score and F1
D. Automatic model evaluation is valuable for assessing qualitative aspects of the model, whereas, human model valuation is valuable for assessing quantitative aspects of the model
E. For human model evaluation, you can use either built-in prompt datasets or your own prompt datasets

Answer: B, C

Explanation:

Correct options:

Model evaluation is the process of evaluating and comparing model outputs to determine the model that is best suited for a use case. You can choose to create either an automatic model evaluation job or a model evaluation job that uses a human workforce.

Human model evaluation is valuable for assessing qualitative aspects of the model, whereas, automatic model valuation is valuable for assessing quantitative aspects of the model

Automatic model evaluation jobs allow you to quickly evaluate a model's ability to perform a task. You can either provide your own custom prompt dataset that you've tailored to a specific use case, or you can use an available built-in dataset. These datasets consist of associated metrics that offer a consistent, objective means to measure model performances.

Model evaluation jobs that use human workers allow you to bring human input to the model evaluation process. They can be employees of your company or a group of subject-matter experts from your industry. This can include tasks like open-ended conversations, answering questions, generating text, or other specific use cases. Human evaluators can give qualitative feedback on aspects such as coherence, relevance, accuracy, and the overall quality of the model's outputs.

Automatic model evaluation provides model scores that are calculated using various statistical methods such as BERT Score and F1

For automated model evaluation job report cards, you get details on model scores that are calculated using various statistical methods such as BERT Score and F1. For example, for text summarization task type - BERT Score is calculated using pre-trained contextual embeddings from BERT models. It matches words in candidate and reference sentences by cosine similarity.
Incorrect options:

Automatic model evaluation is valuable for assessing qualitative aspects of the model, whereas, human model valuation is valuable for assessing quantitative aspects of the model - This option contradicts the explanation provided above. Human evaluators can give qualitative feedback on the overall quality of the model's outputs.

Human model evaluation provides model scores that are calculated using various statistical methods such as BERT Score and F1 - As mentioned earlier, it is only for automated model evaluation job report cards, you get details on model scores that are calculated using various statistical methods such as BERT Score and F1.

For human model evaluation, you can use either built-in prompt datasets or your own prompt datasets -

To evaluate a model's performance for automatic model evaluation jobs, you can use either built-in prompt datasets or your own prompt datasets. For model evaluation jobs that use workers, you must use your own dataset. So, this option is incorrect.

References:
https://docs.aws.amazon.com/bedrock/latest/userguide/model-evaluation.html
https://docs.aws.amazon.com/bedrock/latest/userguide/model-evaluation-report-programmatic.html

Question: 19

Which of the following use cases is NOT the right fit for Amazon Rekognition?

A. Searchable media libraries
B. Celebrity recognition
C. Face-based user identity verification
D. Enable multilingual user experiences in your applications
Answer: D
Explanation:
Correct option:

Enable multilingual user experiences in your applications

Amazon Translate is a text translation service that uses advanced machine learning technologies to provide high-quality translation on demand. You can enable multilingual user experiences in your applications by integrating Amazon Translate.

Amazon Rekognition cannot be used to create multilingual user experiences.

Incorrect options:

Face-based user identity verification
Searchable media libraries
Celebrity recognition

Face-based user identity verification, searchable media libraries, and celebrity recognition are classic use cases of Amazon Rekognition.

Reference:
https://docs.aws.amazon.com/translate/latest/dg/what-is.html

Question: 20

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 model. Remove the confidential data from the training dataset. Retrain the custom model.

B. Mask the confidential data in the inference responses by using dynamic data masking.
C. Encrypt the confidential data in the inference responses by using Amazon SageMaker.
D. 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.

Question: 21

A loan company is building a generative AI-based solution to offer new applicants discounts based on specific business criteria. The company wants to build and use an AI model responsibly to minimize bias that could negatively affect some customers.
Which actions should the company take to meet these requirements? (Select TWO.)

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

Answer: A, C

Explanation:

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

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

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

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

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

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

AWS AI Practitioner References:

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

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

Question: 22

An AI practitioner is using an Amazon Bedrock base model to summarize session chats from the customer service department. The AI practitioner wants to store invocation logs to monitor model input and output data.
Which strategy should the AI practitioner use?

A. Configure AWS CloudTrail as the logs destination for the model.
B. Enable invocation logging in Amazon Bedrock.
C. Configure AWS Audit Manager as the logs destination for the model.
D. Configure model invocation logging in Amazon EventBridge.
Answer: B
Explanation:

Amazon Bedrock provides an option to enable invocation logging to capture and store the input and output data of the models used. This is essential for monitoring and auditing purposes, particularly when handling customer data.

Option B (Correct): "Enable invocation logging in Amazon Bedrock": This is the correct answer as it directly enables the logging of all model invocations, ensuring transparency and traceability.

Option A: "Configure AWS CloudTrail" is incorrect because CloudTrail logs API calls but does not provide specific logging for model inputs and outputs.

Option C: "Configure AWS Audit Manager" is incorrect as Audit Manager is used for compliance reporting, not specific invocation logging for AI models.

Option D: "Configure model invocation logging in Amazon EventBridge" is incorrect as EventBridge is for event-driven architectures, not specifically designed for logging AI model inputs and outputs.

AWS AI Practitioner References:

Amazon Bedrock Logging Capabilities: AWS emphasizes using built-in logging features in Bedrock to maintain data integrity and transparency in model operations.

Question: 23

A company is implementing the Amazon Titan foundation model (FM) by using Amazon Bedrock. The company needs to supplement the model by using relevant data from the company's private data sources.
Which solution will meet this requirement?

A. Use a different FM
B. Choose a lower temperature value
C. Create an Amazon Bedrock knowledge base
D. Enable model invocation logging
Answer: C
Explanation:

Creating an Amazon Bedrock knowledge base allows the integration of external or private data sources

with a foundation model (FM) like Amazon Titan. This integration helps supplement the model with relevant data from the company's private data sources to enhance its responses.

Option C (Correct): "Create an Amazon Bedrock knowledge base": This is the correct answer as it enables the company to incorporate private data into the FM to improve its effectiveness.

Option A: "Use a different FM" is incorrect because it does not address the need to supplement the current model with private data.

Option B: "Choose a lower temperature value" is incorrect as it affects output randomness, not the integration of private data.

Option D: "Enable model invocation logging" is incorrect because logging does not help in supplementing the model with additional data.

AWS AI Practitioner References:

Amazon Bedrock and Knowledge Integration: AWS explains how creating a knowledge base allows Amazon Bedrock to use external data sources to improve the FMâ€™s relevance and accuracy.

Question: 24

A company needs to choose a model from Amazon Bedrock to use internally. The company must identify a model that generates responses in a style that the company's employees prefer.
What should the company do to meet these requirements?

A. Evaluate the models by using built-in prompt datasets.
B. Evaluate the models by using a human workforce and custom prompt datasets.
C. Use public model leaderboards to identify the model.
D. Use the model InvocationLatency runtime metrics in Amazon CloudWatch when trying models.

Answer: B

Explanation:

To determine which model generates responses in a style that the company's employees prefer, the best approach is to use a human workforce to evaluate the models with custom prompt datasets. This method allows for subjective evaluation based on the specific stylistic preferences of the company's employees, which cannot be effectively assessed through automated methods or pre-built datasets.

Option B (Correct): "Evaluate the models by using a human workforce and custom prompt datasets": This is the correct answer as it directly involves human judgment to evaluate the style and quality of the responses, aligning with employee preferences.

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

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

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

AWS AI Practitioner References:

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

Question: 25

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: B, C

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.

Thank You for trying AIF-C01 PDF Demo

https://www.certifiedumps.com/amazon/aif-c01-dumps.html

Start Your AIF-C01 Preparation

[Limited Time Offer] Use Coupon "cert20" for extra 20% discount on the purchase of PDF file. Test your AIF-C01 preparation with actual exam questions