



Amazon-Web-Services

Exam Questions AIF-C01

AWS Certified AI Practitioner

About ExamBible

Your Partner of IT Exam

Found in 1998

ExamBible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, ExamBible has its unique advantages that other companies could not achieve.

Our Advances

* 99.9% Uptime

All examinations will be up to date.

* 24/7 Quality Support

We will provide service round the clock.

* 100% Pass Rate

Our guarantee that you will pass the exam.

* Unique Gurantee

If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

NEW QUESTION 1

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.

NEW QUESTION 2

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

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

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

Answer: B

Explanation:

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

? Average Call Duration:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 3

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

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

Which solution will meet these requirements?

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

Answer: A

Explanation:

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

? Amazon SageMaker Serverless Inference provides a fully managed environment

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

? Why Option A is Correct:

? Why Other Options are Incorrect:

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

NEW QUESTION 4

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

Which solution meets these requirements?

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

Answer: C

Explanation:

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

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

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

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

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

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

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

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

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

AWS AI Practitioner References:

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

NEW QUESTION 5

An AI practitioner wants to use a foundation model (FM) to design a search application. The search application must handle queries that have text and images. Which type of FM should the AI practitioner use to power the search application?

- A. Multi-modal embedding model
- B. Text embedding model
- C. Multi-modal generation model
- D. Image generation model

Answer: A

Explanation:

A multi-modal embedding model is the correct type of foundation model (FM) for powering a search application that handles queries containing both text and images.

? Multi-Modal Embedding Model:

? Why Option A is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 6

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

What should the company do to meet these requirements?

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

Answer: B

Explanation:

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

? Option B (Correct): "Evaluate the models by using a human workforce and custom

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

? Option A: "Evaluate the models by using built-in prompt datasets" is incorrect

because built-in datasets may not capture the company's specific stylistic requirements.

? Option C: "Use public model leaderboards to identify the model" is incorrect as

leaderboards typically measure model performance on standard benchmarks, not on stylistic preferences.

? Option D: "Use the model InvocationLatency runtime metrics in Amazon

CloudWatch" is incorrect because latency metrics do not provide any information about the style of the model's responses.

AWS AI Practitioner References:

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

NEW QUESTION 7

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

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

Answer: A

Explanation:

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

resources and minimizes the potential impact of security incidents.

? Option A (Correct): "Design clear and specific prompts. Configure AWS Identity and Access Management (IAM) roles and policies by using least privilege access": This is the correct answer as it directly addresses both security practices in prompt design and access management.

? Option B: "Enable AWS Audit Manager for automatic model evaluation jobs" is incorrect because Audit Manager is for compliance and auditing, not directly related to secure LLM usage.

? Option C: "Enable Amazon Bedrock automatic model evaluation jobs" is incorrect because Bedrock does not provide automatic model evaluation jobs specifically for security purposes.

? Option D: "Use Amazon CloudWatch Logs to make models explainable and to monitor for bias" is incorrect because CloudWatch Logs are used for monitoring and not directly for making models explainable or secure.

AWS AI Practitioner References:

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

NEW QUESTION 8

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

Which SageMaker feature meets these requirements?

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

Answer: A

Explanation:

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

? Amazon SageMaker Feature Store:

? Why Option A is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 9

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

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

Answer: A

Explanation:

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

NEW QUESTION 10

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

Which ML algorithm meets these requirements?

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

Answer: A

Explanation:

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

? Option A (Correct): "Decision trees": This is the correct answer because decision trees provide a clear and interpretable representation of how input features influence the model's output, making it ideal for understanding the inner mechanisms affecting predictions.

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

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

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

AWS AI Practitioner References:

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

NEW QUESTION 10

A company wants to assess the costs that are associated with using a large language model (LLM) to generate inferences. The company wants to use Amazon Bedrock to build generative AI applications.

Which factor will drive the inference costs?

- A. Number of tokens consumed
- B. Temperature value
- C. Amount of data used to train the LLM

D. Total training time

Answer: A

Explanation:

In generative AI models, such as those built on Amazon Bedrock, inference costs are driven by the number of tokens processed. A token can be as short as one character or as long as one word, and the more tokens consumed during the inference process, the higher the cost.

? Option A (Correct): "Number of tokens consumed": This is the correct answer because the inference cost is directly related to the number of tokens processed by the model.

? Option B: "Temperature value" is incorrect as it affects the randomness of the model's output but not the cost directly.

? Option C: "Amount of data used to train the LLM" is incorrect because training data size affects training costs, not inference costs.

? Option D: "Total training time" is incorrect because it relates to the cost of training the model, not the cost of inference.

AWS AI Practitioner References:

? Understanding Inference Costs on AWS: AWS documentation highlights that inference costs for generative models are largely based on the number of tokens processed.

NEW QUESTION 14

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 19

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 21

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

Which solution meets these requirements MOST cost-effectively?

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

Answer: A

Explanation:

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

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

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

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

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

AWS AI Practitioner References:

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

NEW QUESTION 22

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

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

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

Answer: B

Explanation:

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

? Amazon Comprehend:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 27

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 31

Which metric measures the runtime efficiency of operating AI models?

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

Answer: C

Explanation:

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

? Average Response Time:

? Why Option C is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 36

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 39

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 41

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 45

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
- B. Remove the confidential data from the training dataset
- 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 48

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 50

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 53

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 58

A company has petabytes of unlabeled customer data to use for an advertisement campaign. The company wants to classify its customers into tiers to advertise and promote the company's products.

Which methodology should the company use to meet these requirements?

- A. Supervised learning
- B. Unsupervised learning
- C. Reinforcement learning
- D. Reinforcement learning from human feedback (RLHF)

Answer: B

Explanation:

Unsupervised learning is the correct methodology for classifying customers into tiers when the data is unlabeled, as it does not require predefined labels or outputs.

? Unsupervised Learning:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 61

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 66

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 68

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

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

Answer: C

Explanation:

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

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

Here is why option C is the correct Answer:

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

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

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

? AWS AI Practitioner References:

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

- ? A. Integration with Amazon S3 for object storage is about storing data objects, not vector-based searching or indexing.
- ? B. Support for geospatial indexing and queries is related to location-based data, not vectors used in machine learning.
- ? D. Ability to perform real-time analysis on streaming data relates to analyzing incoming data streams, which is different from the vector search capabilities.

NEW QUESTION 70

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

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

- A. Data pre-processing
- B. Feature engineering
- C. Exploratory data analysis
- D. Hyperparameter tuning

Answer: C

Explanation:

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

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

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

? Option B: "Feature engineering" is incorrect because it involves creating new features from raw data, not analyzing the data's existing structure.

? Option D: "Hyperparameter tuning" is incorrect because it refers to optimizing model parameters, not analyzing the data.

AWS AI Practitioner References:

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

NEW QUESTION 73

A company wants to build an interactive application for children that generates new stories based on classic stories. The company wants to use Amazon Bedrock and needs to ensure that the results and topics are appropriate for children.

Which AWS service or feature will meet these requirements?

- A. Amazon Rekognition
- B. Amazon Bedrock playgrounds
- C. Guardrails for Amazon Bedrock
- D. Agents for Amazon Bedrock

Answer: C

Explanation:

Amazon Bedrock is a service that provides foundational models for building generative AI applications. When creating an application for children, it is crucial to ensure that the generated content is appropriate for the target audience. "Guardrails" in Amazon Bedrock provide mechanisms to control the outputs and topics of generated content to align with desired safety standards and appropriateness levels.

? Option C (Correct): "Guardrails for Amazon Bedrock": This is the correct answer because guardrails are specifically designed to help users enforce content moderation, filtering, and safety checks on the outputs generated by models in Amazon Bedrock. For a children's application, guardrails ensure that all content generated is suitable and appropriate for the intended audience.

? Option A: "Amazon Rekognition" is incorrect. Amazon Rekognition is an image and video analysis service that can detect inappropriate content in images or videos, but it does not handle text or story generation.

? Option B: "Amazon Bedrock playgrounds" is incorrect because playgrounds are environments for experimenting and testing model outputs, but they do not inherently provide safeguards to ensure content appropriateness for specific audiences, such as children.

? Option D: "Agents for Amazon Bedrock" is incorrect. Agents in Amazon Bedrock facilitate building AI applications with more interactive capabilities, but they do not provide specific guardrails for ensuring content appropriateness for children.

AWS AI Practitioner References:

? Guardrails in Amazon Bedrock: Designed to help implement controls that ensure generated content is safe and suitable for specific use cases or audiences, such as children, by moderating and filtering inappropriate or undesired content.

? Building Safe AI Applications: AWS provides guidance on implementing ethical AI practices, including using guardrails to protect against generating inappropriate or biased content.

NEW QUESTION 76

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

Which SageMaker inference option meets these requirements?

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

Answer: A

Explanation:

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

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

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

? Option C: "Asynchronous inference" is incorrect because it is used for workloads that do not require immediate responses.
? Option D: "Batch transform" is incorrect as it is intended for offline, large-batch processing where immediate response is not necessary.
AWS AI Practitioner References:
? Amazon SageMaker Inference Options: AWS documentation describes real-time inference as the best solution for applications that require immediate prediction results with low latency.

NEW QUESTION 80

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

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

Answer: C

Explanation:

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

? Anomaly Detection:

? Why Option C is Correct:

? Why Other Options are Incorrect:

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

NEW QUESTION 83

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 85

A company is using domain-specific models. The company wants to avoid creating new models from the beginning. The company instead wants to adapt pre-trained models to create models for new, related tasks. Which ML strategy meets these requirements?

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

Answer: B

Explanation:

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

? Transfer Learning:

? Why Option B is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 87

A company is building a chatbot to improve user experience. The company is using a large language model (LLM) from Amazon Bedrock for intent detection. The company wants to use few-shot learning to improve intent detection accuracy. Which additional data does the company need to meet these requirements?

- A. Pairs of chatbot responses and correct user intents
- B. Pairs of user messages and correct chatbot responses
- C. Pairs of user messages and correct user intents
- D. Pairs of user intents and correct chatbot responses

Answer: C

Explanation:

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

? Few-shot Learning for Intent Detection:

? Why Option C is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 88

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

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

Answer: C

Explanation:

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

? Plagiarism:

? Why Option C is Correct:

? Why Other Options are Incorrect:

NEW QUESTION 89

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 92

A company is developing a new model to predict the prices of specific items. The model performed well on the training dataset. When the company deployed the model to production, the model's performance decreased significantly.

What should the company do to mitigate this problem?

- A. Reduce the volume of data that is used in training.
- B. Add hyperparameters to the model.
- C. Increase the volume of data that is used in training.
- D. Increase the model training time.

Answer: C

Explanation:

When a model performs well on the training data but poorly in production, it is often due to overfitting. Overfitting occurs when a model learns patterns and noise specific to the training data, which does not generalize well to new, unseen data in production. Increasing the volume of data used in training can help mitigate this problem by providing a more diverse and representative dataset, which helps the model generalize better.

? Option C (Correct): "Increase the volume of data that is used in training":

Increasing the data volume can help the model learn more generalized patterns rather than specific features of the training dataset, reducing overfitting and improving performance in production.

? Option A: "Reduce the volume of data that is used in training" is incorrect, as reducing data volume would likely worsen the overfitting problem.

? Option B: "Add hyperparameters to the model" is incorrect because adding hyperparameters alone does not address the issue of data diversity or model generalization.

? Option D: "Increase the model training time" is incorrect because simply increasing training time does not prevent overfitting; the model needs more diverse data.

AWS AI Practitioner References:

? Best Practices for Model Training on AWS: AWS recommends using a larger and more diverse training dataset to improve a model's generalization capability and reduce the risk of overfitting.

NEW QUESTION 96

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 98

.....

Relate Links

100% Pass Your AIF-C01 Exam with ExamBible Prep Materials

<https://www.exambible.com/AIF-C01-exam/>

Contact us

We are proud of our high-quality customer service, which serves you around the clock 24/7.

Viste - <https://www.exambible.com/>