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
A company wants to use a large language model (LLM) on Amazon Bedrock for sentiment analysis. The company wants to know how much information can fit into one prompt.
Which consideration will inform the company's decision?

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

Answer: B

Explanation:
The context window determines how much information can fit into a single prompt when using a large language model (LLM) like those on Amazon Bedrock.
? Context Window:
? Why Option B is Correct:
? Why Other Options are Incorrect:

NEW QUESTION 2
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 3
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 4
A law firm wants to build an AI application by using large language models (LLMs). The application will read legal documents and extract key points from the documents. Which solution meets these requirements?

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

Answer: C

Explanation:
A summarization chatbot is ideal for extracting key points from legal documents. Large language models (LLMs) can be used to summarize complex texts, such as legal documents, making them more accessible and understandable.
? Option C (Correct): "Develop a summarization chatbot": This is the correct answer
because a summarization chatbot uses LLMs to condense and extract key information from text, which is precisely the requirement for reading and summarizing

legal documents.
? Option A: "Build an automatic named entity recognition system" is incorrect
because it focuses on identifying specific entities, not summarizing documents.
? Option B: "Create a recommendation engine" is incorrect as it is used to suggest products or content, not summarize text.
? Option D: "Develop a multi-language translation system" is incorrect because translation is unrelated to summarizing text.
AWS AI Practitioner References:
? Using LLMs for Text Summarization on AWS: AWS supports developing summarization tools using its AI services, including Amazon Bedrock.

NEW QUESTION 5
A company wants to display the total sales for its top-selling products across various retail locations in the past 12 months.
Which AWS solution should the company use to automate the generation of graphs?

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

Answer: C

Explanation:
Amazon QuickSight is a fully managed business intelligence (BI) service that allows users to create and publish interactive dashboards that include visualizations like graphs, charts, and tables. "Amazon Q" is the natural language query feature within Amazon QuickSight. It enables users to ask questions about their data in natural language and receive visual responses such as graphs.
? Option C (Correct): "Amazon Q in Amazon QuickSight": This is the correct answer
because Amazon QuickSight Q is specifically designed to allow users to explore their data through natural language queries, and it can automatically generate graphs to display sales data and other metrics. This makes it an ideal choice for the company to automate the generation of graphs showing total sales for its top-selling products across various retail locations.
? Option A, B, and D: These options are incorrect:
AWS AI Practitioner References:
? Amazon QuickSight Q is designed to provide insights from data by using natural language queries, making it a powerful tool for generating automated graphs and visualizations directly from queried data.
? Business Intelligence (BI) on AWS: AWS services such as Amazon QuickSight
provide business intelligence capabilities, including automated reporting and visualization features, which are ideal for companies seeking to visualize data like sales trends over time.

NEW QUESTION 6
An AI practitioner has built a deep learning model to classify the types of materials in images. The AI practitioner now wants to measure the model performance.
Which metric will help the AI practitioner evaluate the performance of the model?

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

Answer: A

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

NEW QUESTION 7
A company wants to make a chatbot to help customers. The chatbot will help solve technical problems without human intervention. The company chose a foundation model (FM) for the chatbot. The chatbot needs to produce responses that adhere to company tone.
Which solution meets these requirements?

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

Answer: C

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

NEW QUESTION 8
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 9
A company has a foundation model (FM) that was customized by using Amazon Bedrock to answer customer queries about products. The company wants to validate the model's responses to new types of queries. The company needs to upload a new dataset that Amazon Bedrock can use for validation.
Which AWS service meets these requirements?

A. Amazon S3
B. Amazon Elastic Block Store (Amazon EBS)
C. Amazon Elastic File System (Amazon EFS)
D. AWS Snowcone

Answer: A

Explanation:
Amazon S3 is the optimal choice for storing and uploading datasets used for machine learning model validation and training. It offers scalable, durable, and secure storage, making it ideal for holding datasets required by Amazon Bedrock for validation purposes.
? Option A (Correct): "Amazon S3": This is the correct answer because Amazon S3
is widely used for storing large datasets that are accessed by machine learning models, including those in Amazon Bedrock.
? Option B: "Amazon Elastic Block Store (Amazon EBS)" is incorrect because EBS
is a block storage service for use with Amazon EC2, not for directly storing datasets for Amazon Bedrock.
? Option C: "Amazon Elastic File System (Amazon EFS)" is incorrect as it is
primarily used for file storage with shared access by multiple instances.
? Option D: "AWS Snowcone" is incorrect because it is a physical device for offline data transfer, not suitable for directly providing data to Amazon Bedrock.
AWS AI Practitioner References:
? Storing and Managing Datasets on AWS for Machine Learning: AWS recommends using S3 for storing and managing datasets required for ML model training and validation.

NEW QUESTION 10
A company built an AI-powered resume screening system. The company used a large dataset to train the model. The dataset contained resumes that were not representative of all demographics. Which core dimension of responsible AI does this scenario present?

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

Answer: A

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

NEW QUESTION 10
A company wants to classify human genes into 20 categories based on gene characteristics. The company needs an ML algorithm to document how the inner mechanism of the model affects the output.
Which ML algorithm meets these requirements?

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

Answer: A

Explanation:
Decision trees are an interpretable machine learning algorithm that clearly documents the decision-making process by showing how each input feature affects the output. This transparency is particularly useful when explaining how the model arrives at a certain decision, making it suitable for classifying genes into categories.
? Option A (Correct): "Decision trees": This is the correct answer because decision
trees provide a clear and interpretable representation of how input features influence the model's output, making it ideal for understanding the inner mechanisms affecting predictions.
? Option B: "Linear regression" is incorrect because it is used for regression tasks,
not classification.
? Option C: "Logistic regression" is incorrect as it does not provide the same level of interpretability in documenting decision-making processes.
? Option D: "Neural networks" is incorrect because they are often considered "black boxes" and do not easily explain how they arrive at their outputs.
AWS AI Practitioner References:
? Interpretable Machine Learning Models on AWS: AWS supports using interpretable models, such as decision trees, for tasks that require clear documentation of how input data affects output decisions.

NEW QUESTION 13

Which term describes the numerical representations of real-world objects and concepts that AI and natural language processing (NLP) models use to improve understanding of textual information?

A. Embeddings
B. Tokens
C. Models
D. Binaries

Answer: A

Explanation:
Embeddings are numerical representations of objects (such as words, sentences, or documents) that capture the objects' semantic meanings in a form that AI and NLP models can easily understand. These representations help models improve their understanding of textual information by representing concepts in a continuous vector space.
? Option A (Correct): "Embeddings": This is the correct term, as embeddings provide
a way for models to learn relationships between different objects in their input space, improving their understanding and processing capabilities.
? Option B: "Tokens" are pieces of text used in processing, but they do not capture
semantic meanings like embeddings do.
? Option C: "Models" are the algorithms that use embeddings and other inputs, not the representations themselves.
? Option D: "Binaries" refer to data represented in binary form, which is unrelated to the concept of embeddings.
AWS AI Practitioner References:
? Understanding Embeddings in AI and NLP: AWS provides resources and tools, like Amazon SageMaker, that utilize embeddings to represent data in formats suitable for machine learning models.

NEW QUESTION 16
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 18
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 23
A research company implemented a chatbot by using a foundation model (FM) from Amazon Bedrock. The chatbot searches for answers to questions from a large database of research papers.
After multiple prompt engineering attempts, the company notices that the FM is performing poorly because of the complex scientific terms in the research papers.
How can the company improve the performance of the chatbot?

A. Use few-shot prompting to define how the FM can answer the questions.
B. Use domain adaptation fine-tuning to adapt the FM to complex scientific terms.
C. Change the FM inference parameters.
D. Clean the research paper data to remove complex scientific terms.

Answer: B

Explanation:
Domain adaptation fine-tuning involves training a foundation model (FM) further using a specific dataset that includes domain-specific terminology and content, such as scientific terms in research papers. This process allows the model to better understand and handle complex terminology, improving its performance on specialized tasks.
? Option B (Correct): "Use domain adaptation fine-tuning to adapt the FM to complex
scientific terms": This is the correct answer because fine-tuning the model on domain-specific data helps it learn and adapt to the specific language and terms used in the research papers, resulting in better performance.
? Option A: "Use few-shot prompting to define how the FM can answer the
questions" is incorrect because while few-shot prompting can help in certain
scenarios, it is less effective than fine-tuning for handling complex domain-specific terms.
? Option C: "Change the FM inference parameters" is incorrect because adjusting
inference parameters will not resolve the issue of the model's lack of understanding of complex scientific terminology.
? Option D: "Clean the research paper data to remove complex scientific terms" is
incorrect because removing the complex terms would result in the loss of important information and context, which is not a viable solution.
AWS AI Practitioner References:
? Domain Adaptation in Amazon Bedrock: AWS recommends fine-tuning models with domain-specific data to improve their performance on specialized tasks involving unique terminology.

NEW QUESTION 28
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 Bedroc
D. Use the fine- tuned model to process user prompts.
E. Upload PDF documents to an Amazon Bedrock knowledge bas
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 32
A company wants to develop a large language model (LLM) application by using Amazon Bedrock and customer data that is uploaded to Amazon S3. The company's security policy states that each team can access data for only the team's own customers.
Which solution will meet these requirements?

A. Create an Amazon Bedrock custom service role for each team that has access to only the team's customer data.
B. Create a custom service role that has Amazon S3 acces
C. Ask teams to specify the customer name on each Amazon Bedrock request.
D. Redact personal data in Amazon S3. Update the S3 bucket policy to allow team access to customer data.
E. Create one Amazon Bedrock role that has full Amazon S3 acces
F. Create IAM roles for each team that have access to only each team's customer folders.

Answer: A

Explanation:
To comply with the company's security policy, which restricts each team to access data for only their own customers, creating an Amazon Bedrock custom service role for each team is the correct solution.
? Custom Service Role Per Team:
? Why Option A is Correct:
? Why Other Options are Incorrect:
Thus, A is the correct answer to meet the company's security requirements.

NEW QUESTION 36
A company deployed an AI/ML solution to help customer service agents respond to frequently asked questions. The questions can change over time. The company wants to give customer service agents the ability to ask questions and receive automatically generated answers to common customer questions. Which strategy will meet these requirements MOST cost-effectively?

A. Fine-tune the model regularly.
B. Train the model by using context data.
C. Pre-train and benchmark the model by using context data.

D. Use Retrieval Augmented Generation (RAG) with prompt engineering techniques.

Answer: D

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

NEW QUESTION 41
A company wants to use language models to create an application for inference on edge devices. The inference must have the lowest latency possible.
Which solution will meet these requirements?

A. Deploy optimized small language models (SLMs) on edge devices.
B. Deploy optimized large language models (LLMs) on edge devices.
C. Incorporate a centralized small language model (SLM) API for asynchronous communication with edge devices.
D. Incorporate a centralized large language model (LLM) API for asynchronous communication with edge devices.

Answer: A

Explanation:
To achieve the lowest latency possible for inference on edge devices, deploying optimized small language models (SLMs) is the most effective solution. SLMs require fewer
resources and have faster inference times, making them ideal for deployment on edge devices where processing power and memory are limited.
? Option A (Correct): "Deploy optimized small language models (SLMs) on edge
devices": This is the correct answer because SLMs provide fast inference with low latency, which is crucial for edge deployments.
? Option B: "Deploy optimized large language models (LLMs) on edge devices" is
incorrect because LLMs are resource-intensive and may not perform well on edge devices due to their size and computational demands.
? Option C: "Incorporate a centralized small language model (SLM) API for
asynchronous communication with edge devices" is incorrect because it introduces network latency due to the need for communication with a centralized server.
? Option D: "Incorporate a centralized large language model (LLM) API for
asynchronous communication with edge devices" is incorrect for the same reason, with even greater latency due to the larger model size.
AWS AI Practitioner References:
? Optimizing AI Models for Edge Devices on AWS: AWS recommends using small, optimized models for edge deployments to ensure minimal latency and efficient performance.

NEW QUESTION 45
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 46
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 49
A company wants to use a pre-trained generative AI model to generate content for its marketing campaigns. The company needs to ensure that the generated content aligns with the company's brand voice and messaging requirements.
Which solution meets these requirements?

A. Optimize the model's architecture and hyperparameters to improve the model's overall performance.
B. Increase the model's complexity by adding more layers to the model's architecture.

C. Create effective prompts that provide clear instructions and context to guide the model's generation.
D. Select a large, diverse dataset to pre-train a new generative model.

Answer: C

Explanation:
Creating effective prompts is the best solution to ensure that the content generated by a pre-trained generative AI model aligns with the company's brand voice and messaging requirements.
? Effective Prompt Engineering:
? Why Option C is Correct:
? Why Other Options are Incorrect:

NEW QUESTION 53
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 56
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 58
A company wants to deploy a conversational chatbot to answer customer questions. The chatbot is based on a fine-tuned Amazon SageMaker JumpStart model. The application must comply with multiple regulatory frameworks.
Which capabilities can the company show compliance for? (Select TWO.)

A. Auto scaling inference endpoints
B. Threat detection
C. Data protection
D. Cost optimization
E. Loosely coupled microservices

Answer: BC

Explanation:
To comply with multiple regulatory frameworks, the company must ensure data protection and threat detection. Data protection involves safeguarding sensitive customer information, while threat detection identifies and mitigates security threats to the application.
? Option C (Correct): "Data protection": This is correct because data protection is
critical for compliance with privacy and security regulations.
? Option B (Correct): "Threat detection": This is correct because detecting and mitigating threats is essential to maintaining the security posture required for regulatory compliance.
? Option A: "Auto scaling inference endpoints" is incorrect because auto-scaling does not directly relate to regulatory compliance.
? Option D: "Cost optimization" is incorrect because it is focused on managing expenses, not compliance.
? Option E: "Loosely coupled microservices" is incorrect because this architectural approach does not directly address compliance requirements.
AWS AI Practitioner References:
? AWS Compliance Capabilities: AWS offers services and tools, such as data protection and threat detection, to help companies meet regulatory requirements for security and privacy.

NEW QUESTION 62
A company manually reviews all submitted resumes in PDF format. As the company grows, the company expects the volume of resumes to exceed the company's review capacity. The company needs an automated system to convert the PDF resumes into plain text format for additional processing.
Which AWS service meets this requirement?

A. Amazon Textract
B. Amazon Personalize
C. Amazon Lex
D. Amazon Transcribe

Answer: A

Explanation:
Amazon Textract is a service that automatically extracts text and data from scanned documents, including PDFs. It is the best choice for converting resumes from PDF format to plain text for further processing.
? Amazon Textract:
? Why Option A 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 70
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 71
A company is using a pre-trained large language model (LLM) to build a chatbot for product recommendations. The company needs the LLM outputs to be short and written in a specific language.

Which solution will align the LLM response quality with the company's expectations?

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

Answer: A

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

NEW QUESTION 72
A financial institution is using Amazon Bedrock to develop an AI application. The application is hosted in a VPC. To meet regulatory compliance standards, the VPC is not allowed access to any internet traffic.
Which AWS service or feature will meet these requirements?

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

Answer: A

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

NEW QUESTION 75
A company has terabytes of data in a database that the company can use for business analysis. The company wants to build an AI-based application that can build a SQL query from input text that employees provide. The employees have minimal experience with technology.
Which solution meets these requirements?

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

Answer: A

Explanation:
Generative Pre-trained Transformers (GPT) are suitable for building an AI-based application that can generate SQL queries from natural language input provided by employees.
? GPT for Natural Language Processing:
? Why Option A is Correct:
? Why Other Options are Incorrect:

NEW QUESTION 80
A company is building an ML model to analyze archived data. The company must perform inference on large datasets that are multiple GBs in size. The company does not need to access the model predictions immediately.
Which Amazon SageMaker inference option will meet these requirements?

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

Answer: A

Explanation:
Batch transform in Amazon SageMaker is designed for offline processing of large datasets. It is ideal for scenarios where immediate predictions are not required, and the inference can be done on large datasets that are multiple gigabytes in size. This method processes data in batches, making it suitable for analyzing archived data without the need for real- time access to predictions.
? Option A (Correct): "Batch transform": This is the correct answer because batch
transform is optimized for handling large datasets and is suitable when immediate access to predictions is not required.
? Option B: "Real-time inference" is incorrect because it is used for low-latency, real-
time prediction needs, which is not required in this case.
? Option C: "Serverless inference" is incorrect because it is designed for small-scale, intermittent inference requests, not for large batch processing.

? Option D: "Asynchronous inference" is incorrect because it is used when immediate predictions are required, but with high throughput, whereas batch transform is more suitable for very large datasets.
AWS AI Practitioner References:
? Batch Transform on AWS SageMaker: AWS recommends using batch transform for large datasets when real-time processing is not needed, ensuring cost-effectiveness and scalability.

NEW QUESTION 82
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 83
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 86
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 87
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 89
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 92
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 93
A company makes forecasts each quarter to decide how to optimize operations to meet expected demand. The company uses ML models to make these forecasts.
An AI practitioner is writing a report about the trained ML models to provide transparency and explainability to company stakeholders.
What should the AI practitioner include in the report to meet the transparency and explainability requirements?

A. Code for model training
B. Partial dependence plots (PDPs)
C. Sample data for training
D. Model convergence tables

Answer: B

Explanation:
Partial dependence plots (PDPs) are visual tools used to show the relationship between a feature (or a set of features) in the data and the predicted outcome of a machine learning model. They are highly effective for providing transparency and explainability of the model's behavior to stakeholders by illustrating how different input variables impact the model's predictions.
? Option B (Correct): "Partial dependence plots (PDPs)": This is the correct answer because PDPs help to interpret how the model's predictions change with varying values of input features, providing stakeholders with a clearer understanding of the model's decision-making process.
? Option A: "Code for model training" is incorrect because providing the raw code for model training may not offer transparency or explainability to non-technical stakeholders.
? Option C: "Sample data for training" is incorrect as sample data alone does not explain how the model works or its decision-making process.
? Option D: "Model convergence tables" is incorrect. While convergence tables can show the training process, they do not provide insights into how input features affect the model's predictions.

AWS AI Practitioner References:
? Explainability in AWS Machine Learning: AWS provides various tools for model explainability, such as Amazon SageMaker Clarify, which includes PDPs to help explain the impact of different features on the model??s predictions.

NEW QUESTION 97
A company is training a foundation model (FM). The company wants to increase the accuracy of the model up to a specific acceptance level.
Which solution will meet these requirements?

A. Decrease the batch size.
B. Increase the epochs.
C. Decrease the epochs.
D. Increase the temperature parameter.

Answer: B

Explanation:
Increasing the number of epochs during model training allows the model to learn from the data over more iterations, potentially improving its accuracy up to a certain point. This is a common practice when attempting to reach a specific level of accuracy.
? Option B (Correct): "Increase the epochs": This is the correct answer because
increasing epochs allows the model to learn more from the data, which can lead to higher accuracy.
? Option A: "Decrease the batch size" is incorrect as it mainly affects training speed
and may lead to overfitting but does not directly relate to achieving a specific accuracy level.
? Option C: "Decrease the epochs" is incorrect as it would reduce the training time,
possibly preventing the model from reaching the desired accuracy.
? Option D: "Increase the temperature parameter" is incorrect because temperature affects the randomness of predictions, not model accuracy.
AWS AI Practitioner References:
? Model Training Best Practices on AWS: AWS suggests adjusting training parameters, like the number of epochs, to improve model performance.

NEW QUESTION 100
A company wants to create a chatbot by using a foundation model (FM) on Amazon Bedrock. The FM needs to access encrypted data that is stored in an Amazon S3 bucket.
The data is encrypted with Amazon S3 managed keys (SSE-S3).
The FM encounters a failure when attempting to access the S3 bucket data. Which solution will meet these requirements?

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

Answer: A

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

NEW QUESTION 102
.....

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/