Amazon.AIF-C01.v2026-02-28.q145

Exam Code:	AIF-C01
Exam Name:	AWS Certified AI Practitioner
Certification Provider:	Amazon
Free Question Number:	145
Version:	v2026-02-28
# of views:	136
# of Questions views:	1450
https://www.freepdfdumps.com/Amazon.AIF-C01.v2026-02-28.q145.html	

NEW QUESTION: 1
A company wants to classify images of different objects based on custom features extracted from a dataset.
Which solution will meet this requirement with the LEAST development effort?
A. Use traditional ML algorithms with custom features extracted from the dataset.
B. Use a pre-trained deep learning model and fine-tune the model on the dataset.
C. Use a generative adversarial network (GAN) model to classify the images.
D. Use a support vector machine (SVM) with manually engineered features for classification.
Answer: B (LEAVE A REPLY)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
Fine-tuning a pre-trained deep learning model (transfer learning) provides the lowest development effort because:
The model already understands general visual features
Only minimal training is required on the custom dataset
Feature extraction is handled automatically
AWS guidance strongly recommends transfer learning for image classification tasks when development speed and efficiency are priorities.
Why the other options are incorrect:
A and D require extensive feature engineering and tuning.
C (GANs) are designed for data generation, not classification.
AWS AI document references:
Transfer Learning for Computer Vision on AWS
Image Classification Best Practices with SageMaker
Deep Learning Model Reuse and Fine-Tuning

NEW QUESTION: 2

A company wants to label training datasets by using human feedback to fine-tune a foundation model (FM).
The company does not want to develop labeling applications or manage a labeling workforce.
Which AWS service or feature meets these requirements?
A. Amazon SageMaker Data Wrangler
B. Amazon SageMaker Ground Truth Plus
C. Amazon Transcribe
D. Amazon Macie
Answer: B (LEAVE A REPLY)
Amazon SageMaker Ground Truth Plus provides a fully managed data labeling service where AWS manages the workforce, tools, and processes.
Data Wrangler is for data preparation and transformation.
Transcribe is for speech-to-text.
Macie is for sensitive data discovery, not labeling.
Reference:
AWS Documentation - SageMaker Ground Truth Plus

NEW QUESTION: 3
A company wants to improve multiple ML models.
Select the correct technique from the following list of use cases. Each technique should be selected one time or not at all. (Select THREE.) Few-shot learning Fine-tuning Retrieval Augmented Generation (RAG) Zero-shot learning
Answer:
AWS References:
Amazon Bedrock - Retrieval Augmented Generation (RAG)
AWS Generative AI Guide - Zero-shot & Few-shot learning

NEW QUESTION: 4
A company is using a pre-trained large language model (LLM). The LLM must perform multiple tasks that require specific domain knowledge. The LLM does not have information about several technical topics in the domain. The company has unlabeled data that the company can use to fine-tune the model.
Which fine-tuning method will meet these requirements?
A. Full training
B. Supervised fine-tuning
C. Continued pre-training
D. Retrieval Augmented Generation (RAG)
Answer: (SHOW ANSWER)
The correct answer is C because Continued Pre-training (also known as domain-adaptive pre-training) involves training a pre-trained model further on unlabeled domain-specific data. This method helps adapt the LLM to a specific domain without needing labeled datasets, making it

ideal for cases where the goal is to enhance the model's understanding of technical language or terminology.
From AWS documentation:
"Continued pre-training allows an LLM to ingest large volumes of domain-specific text without labels to improve contextual understanding in a particular area. This is effective when adapting a foundation model to new knowledge without altering the model architecture." Explanation of other options:
A). Full training refers to building a model from scratch, which is extremely resource-intensive and unnecessary if a strong base model already exists.
B). Supervised fine-tuning requires labeled data, which the scenario explicitly lacks.
D). RAG is a method to retrieve external information at inference time, not a training technique using unlabeled data.
Referenced AWS AI/ML Documents and Study Guides:
* AWS Bedrock Model Customization Documentation - Continued Pre-training
* Amazon SageMaker JumpStart - Domain Adaptation Techniques
* AWS Machine Learning Specialty Study Guide - Foundation Model Customization Section

NEW QUESTION: 5
An ecommerce company is deploying a chatbot. The chatbot will give users the ability to ask questions about the company's products and receive details on users' orders. The company must implement safeguards for the chatbot to filter harmful content from the input prompts and chatbot responses.
Which AWS feature or resource meets these requirements?
A. Amazon Bedrock Guardrails
B. Amazon Bedrock Agents
C. Amazon Bedrock inference APIs
D. Amazon Bedrock custom models
Answer: A (LEAVE A REPLY)
The ecommerce company is deploying a chatbot that needs safeguards to filter harmful content from input prompts and responses. Amazon Bedrock Guardrails provide mechanisms to ensure responsible AI usage by filtering harmful content, such as hate speech, violence, or misinformation, making it the appropriate feature for this requirement.
Exact Extract from AWS AI Documents:
From the AWS Bedrock User Guide:
"Amazon Bedrock Guardrails enable developers to implement safeguards for generative AI applications, such as chatbots, by filtering harmful content in input prompts and model responses. Guardrails include content filters, word filters, and denied topics to ensure safe and responsible outputs." (Source: AWS Bedrock User Guide, Guardrails for Responsible AI) Detailed Explanation:

* Option A: Amazon Bedrock GuardrailsThis is the correct answer. Amazon Bedrock Guardrails are specifically designed to filter harmful content from chatbot inputs and responses, ensuring safe interactions for users.
* Option B: Amazon Bedrock AgentsAmazon Bedrock Agents are used to automate tasks and integrate with external tools, not to filter harmful content. This option does not meet the requirement.
* Option C: Amazon Bedrock inference APIsAmazon Bedrock inference APIs allow users to invoke foundation models for generating responses, but they do not provide built-in content filtering mechanisms.
* Option D: Amazon Bedrock custom modelsCustom models on Amazon Bedrock allow users to fine- tune models, but they do not inherently include safeguards for filtering harmful content unless explicitly implemented.
References:
AWS Bedrock User Guide: Guardrails for Responsible AI (https://docs.aws.amazon.com/bedrock/latest /userguide/guardrails.html)
AWS AI Practitioner Learning Path: Module on Responsible AI and Model Safety Amazon Bedrock Developer Guide: Building Safe AI Applications (https://aws.amazon.com/bedrock/)

NEW QUESTION: 6
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 (LEAVE A REPLY)
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: 7
A company is building a generative AI application to help customers make travel reservations. The application will process customer requests and invoke the appropriate API calls to complete reservation transactions.
Which Amazon Bedrock resource will meet these requirements?
A. Agents
B. Intelligent prompt routing
C. Knowledge Bases
D. Guardrails
Answer: (SHOW ANSWER)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
Amazon Bedrock Agents are designed to:
Interpret user intent
Plan multi-step actions
Invoke APIs and backend services
Complete transactional workflows
This makes Agents ideal for task-oriented applications such as travel reservations, where the system must:
Understand customer requests
Call reservation APIs
Orchestrate end-to-end transactions
Why the other options are incorrect:
Prompt routing (B) selects prompts or models but does not invoke APIs.
Knowledge Bases (C) retrieve information but do not execute actions.
Guardrails (D) enforce safety and compliance, not task execution.
AWS AI document references:
Amazon Bedrock Agents Documentation
Building Task-Oriented Generative AI Applications
Foundation Model Orchestration on AWS

NEW QUESTION: 8
A hospital is developing an AI system to assist doctors in diagnosing diseases based on patient records and medical images. To comply with regulations, the sensitive patient data must not leave the country the data is located in.
A. Data residency
B. Data quality

C. Data discoverability
D. Data enrichment
Answer: A (LEAVE A REPLY)
The correct answer is A - Data residency. AWS defines data residency as ensuring that regulated or sensitive data-such as patient medical records under healthcare laws-remains physically stored and processed within specific national or regional boundaries. This aligns with regulatory frameworks such as HIPAA, GDPR, and country-specific health data protection acts. AWS allows customers to deploy all ML workloads, including Amazon SageMaker, Amazon Bedrock, and Amazon S3, within a chosen AWS Region, ensuring no cross- border data movement unless explicitly configured. Data quality (B) refers to accuracy and consistency, discoverability (C) relates to cataloging, and enrichment (D) refers to enhancing datasets. None of these address the compliance requirement to prevent data from leaving the country. Data residency is a core component of AWS's Shared Responsibility Model and foundational for healthcare AI compliance.
Referenced AWS Documentation:
AWS Data Privacy Whitepaper - Data Residency Controls
AWS Compliance Programs - Regional Data Handling Requirements

NEW QUESTION: 9
A manufacturing company uses AI to inspect products and find any damages or defects.
Which type of AI application is the company using?
A. Recommendation system
B. Natural language processing (NLP)
C. Computer vision
D. Image processing
Answer: (SHOW ANSWER)
The manufacturing company uses AI to inspect products for damages or defects, which involves analyzing visual data (e.g., images or videos of products). This task falls under computer vision, a type of AI application that enables machines to interpret and understand visual information, such as identifying defects in manufacturing.
Exact Extract from AWS AI Documents:
From the AWS AI Practitioner Learning Path:
"Computer vision enables machines to interpret and understand visual data from the world, such as images or videos. Common applications include defect detection in manufacturing, where AI models analyze product images to identify damages or anomalies." (Source: AWS AI Practitioner Learning Path, Module on AI Concepts) Detailed Explanation:
Option A: Recommendation systemRecommendation systems suggest items or actions based on user preferences (e.g., product recommendations). They are not relevant for inspecting products for defects.
Option B: Natural language processing (NLP)NLP focuses on processing and understanding text or speech, not visual data like product images. This option is incorrect.

Option C: Computer visionThis is the correct answer. Computer vision is used for tasks like defect detection in manufacturing by analyzing visual data to identify damages or defects.
Option D: Image processingWhile image processing involves manipulating images (e.g., filtering, resizing), it is a lower-level technique, not an AI application. Computer vision, which often uses image processing as a component, is the broader AI application here.
References:
AWS AI Practitioner Learning Path: Module on AI Concepts
Amazon Rekognition Developer Guide: Image Analysis
(https://docs.aws.amazon.com/rekognition/latest/dg
/what-is.html)
AWS Documentation: Introduction to Computer Vision (https://aws.amazon.com/computer-vision/)

NEW QUESTION: 10
Which AWS service or feature stores embeddings In a vector database for use with foundation models (FMs) and Retrieval Augmented Generation (RAG)?
A. Amazon SageMaker Ground Truth
B. Amazon Textract
C. Amazon OpenSearch Service
D. Amazon Transcribe
Answer: C (LEAVE A REPLY)

NEW QUESTION: 11
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 (LEAVE A REPLY)
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: 12
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: (SHOW ANSWER)
Amazon SageMaker Feature Store is the correct solution for sharing and managing variables (features) across multiple teams during model development.
Amazon SageMaker Feature Store:
A fully managed repository for storing, sharing, and managing machine learning features across different teams and models.
It enables collaboration and reuse of features, ensuring consistent data usage and reducing redundancy.
Why Option A is Correct:
Centralized Feature Management: Provides a central repository for managing features, making it easier to share them across teams.
Collaboration and Reusability: Improves efficiency by allowing teams to reuse existing features instead of creating them from scratch.
Why Other Options are Incorrect:
B . SageMaker Data Wrangler: Helps with data preparation and analysis but does not provide a centralized feature store.
C . SageMaker Clarify: Used for bias detection and explainability, not for managing variables across teams.
D . SageMaker Model Cards: Provide model documentation, not feature management.

NEW QUESTION: 13

A company is using large language models (LLMs) to develop online tutoring applications. The company needs to apply configurable safeguards to the LLMs. These safeguards must ensure that the LLMs follow standard safety rules when creating applications.
Which solution will meet these requirements with the LEAST effort?
A. Amazon Bedrock playgrounds
B. Amazon SageMaker Clarify
C. Amazon Bedrock Guardrails
D. Amazon SageMaker JumpStart
Answer: C (LEAVE A REPLY)
The correct answer is C because Amazon Bedrock Guardrails provides out-of-the-box configurable safety mechanisms to control the behavior of LLMs in generative AI applications. Guardrails can be configured with denylists, content filters, sensitive topics, and tone enforcement, all without retraining the model.
From AWS documentation:
"Amazon Bedrock Guardrails allows developers to define safety and responsible AI policies directly in the model inference layer, making it easy to prevent harmful, biased, or unsafe outputs with minimal configuration." Explanation of other options:
A). Bedrock playgrounds are interactive environments for testing prompts and models but do not provide production-grade safety enforcement.
B). SageMaker Clarify focuses on bias detection and explainability for supervised ML models - it does not directly apply guardrails to LLM outputs.
D). SageMaker JumpStart is for model fine-tuning and deployment, not for enforcing safety policies on LLM responses.
Referenced AWS AI/ML Documents and Study Guides:
Amazon Bedrock Documentation - Guardrails Overview
AWS Responsible AI Whitepaper
AWS Certified ML Specialty Study Guide - Safety in Generative AI

NEW QUESTION: 14
A company wants to develop an AI assistant for employees to query internal data.
Which AWS service will meet this requirement?
A. Amazon Rekognition
B. Amazon Textract
C. Amazon Lex
D. Amazon Q Business
Answer: (SHOW ANSWER)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
Amazon Q Business is a managed AI assistant designed to:
Allow employees to query internal enterprise data
Provide conversational answers
Respect enterprise security and access controls

AWS guidance positions Amazon Q Business as the solution for internal knowledge discovery and enterprise AI assistance.
Why the other options are incorrect:
Rekognition (A) analyzes images.
Textract (B) extracts text from documents.
Lex (C) builds conversational interfaces but does not provide enterprise data integration out of the box.
AWS AI document references:
Amazon Q Business Overview
Enterprise AI Assistants on AWS
Secure Access to Internal Data with AI

NEW QUESTION: 15
A company is building a job recommendation system based on job posting data and job seeker user profiles.
The system shows bias in job recommendations based on gender for user profiles that are otherwise equivalent.
Which principle should the company follow to address this issue, according to AWS best practices for responsible AI?
A. Governance
B. Explainability
C. Controllability
D. Fairness
Answer: (SHOW ANSWER)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
Fairness is a core principle of AWS Responsible AI. It requires that AI systems:
Do not produce biased or discriminatory outcomes
Treat individuals and groups equitably
Are evaluated and mitigated for bias across protected attributes
AWS Responsible AI guidance specifically highlights fairness as the principle used to identify, measure, and mitigate bias in AI systems such as recommendation engines.
Why the other options are incorrect:
Governance (A) focuses on oversight and policies.
Explainability (B) explains decisions but does not directly mitigate bias.
Controllability (C) focuses on human oversight and intervention.
AWS AI document references:
AWS Responsible AI Principles
Bias and Fairness in ML Systems
Responsible AI Best Practices on AWS

NEW QUESTION: 16

A company is implementing the Amazon Titan foundation model (FM) by using Amazon Bedrock. The company needs to supplement the model by using relevant data from the company's private data sources.
Which solution will meet this requirement?
A. Use a different FM
B. Choose a lower temperature value
C. Create an Amazon Bedrock knowledge base
D. Enable model invocation logging
Answer: C (LEAVE A REPLY)
Creating an Amazon Bedrock knowledge base allows the integration of external or private data sources with a foundation model (FM) like Amazon Titan. This integration helps supplement the model with relevant data from the company's private data sources to enhance its responses.
* Option C (Correct): "Create an Amazon Bedrock knowledge base": This is the correct answer as it enables the company to incorporate private data into the FM to improve its effectiveness.
* Option A: "Use a different FM" is incorrect because it does not address the need to supplement the current model with private data.
* Option B: "Choose a lower temperature value" is incorrect as it affects output randomness, not the integration of private data.
* Option D: "Enable model invocation logging" is incorrect because logging does not help in supplementing the model with additional data.
AWS AI Practitioner References:
* Amazon Bedrock and Knowledge Integration: AWS explains how creating a knowledge base allows Amazon Bedrock to use external data sources to improve the FM's relevance and accuracy.

NEW QUESTION: 17
A medical company is customizing a foundation model (FM) for diagnostic purposes. The company needs the model to be transparent and explainable to meet regulatory requirements.
Which solution will meet these requirements?
A. Configure the security and compliance by using Amazon Inspector.
B. Generate simple metrics, reports, and examples by using Amazon SageMaker Clarify.
C. Encrypt and secure training data by using Amazon Macie.

D. Gather more data. Use Amazon Rekognition to add custom labels to the data.

Answer: B (LEAVE A REPLY)

Amazon SageMaker Clarify provides transparency and explainability for machine learning models by generating metrics, reports, and examples that help to understand model predictions. For a medical company that needs a foundation model to be transparent and explainable to meet regulatory requirements, SageMaker Clarify is the most suitable solution.

* Amazon SageMaker Clarify:
* It helps in identifying potential bias in the data and model, and also explains model behavior by generating feature attributions, providing insights into which features are most influential in the model's predictions.
* These capabilities are critical in medical applications where regulatory compliance often mandates transparency and explainability to ensure that decisions made by the model can be trusted and audited.
* Why Option B is Correct:
* Transparency and Explainability: SageMaker Clarify is explicitly designed to provide insights into machine learning models' decision-making processes, helping meet regulatory requirements by explaining why a model made a particular prediction.
* Compliance with Regulations: The tool is suitable for use in sensitive domains, such as healthcare, where there is a need for explainable AI.
* Why Other Options are Incorrect:
* A. Amazon Inspector: Focuses on security assessments, not on explainability or model transparency.
* C. Amazon Macie: Provides data security by identifying and protecting sensitive data, but does not help in making models explainable.
* D. Amazon Rekognition: Used for image and video analysis, not relevant to making models explainable.

Thus, B is the correct answer for meeting transparency and explainability requirements for the foundation model

NEW QUESTION: 18

A company runs a website for users to make travel reservations. The company wants an AI solution to help create consistent branding for hotels on the website. The AI solution needs to generate hotel descriptions for the website in a consistent writing style. Which AWS service will meet these requirements?

A. Amazon Comprehend

B. Amazon Personalize

C. Amazon Rekognition

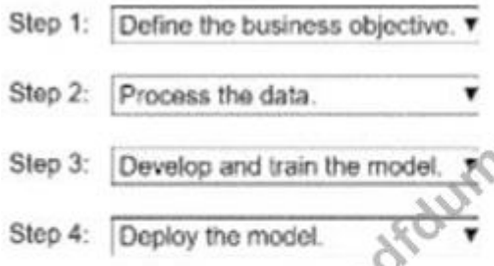
D. Amazon Bedrock

Answer: D (LEAVE A REPLY)

The correct answer is D because Amazon Bedrock provides access to foundation models (FMs) from various providers for generative AI use cases, including text generation. It supports

generating content in a consistent tone, voice, or writing style using prompts or few-shot examples.
From AWS documentation:
"Amazon Bedrock allows you to build and scale generative AI applications using foundation models from AI21 Labs, Anthropic, Cohere, Meta, Mistral, Stability AI, and Amazon. These models can generate text with controlled tone and style for applications like branding, content creation, and copywriting." Explanation of other options:
A . Amazon Comprehend is for natural language understanding, such as sentiment analysis and entity recognition, not generation.
B . Amazon Personalize is for building recommendation systems, not content generation.
C . Amazon Rekognition is for image and video analysis, not text generation.
Referenced AWS AI/ML Documents and Study Guides:
Amazon Bedrock Developer Guide - Generative AI Use Cases
AWS Certified Machine Learning Specialty Guide - Content Generation with FMs

NEW QUESTION: 19
Sated and order the steps from the following bat to correctly describe the ML Lifecycle for a new custom modal Select each step one time. (Select and order FOUR.)
* Define the business objective.
* Deploy the modal.
* Develop and tram the model.
* Process the data.
Answer:



Step 1: Define the business objective.
Step 2: Process the data.
Step 3: Develop and train the model.
Step 4: Deploy the model.
The correct order represents the machine learning lifecycle as defined by AWS in the Amazon SageMaker documentation and AWS Certified Machine Learning Specialty Study Guide. The lifecycle describes the sequence of tasks required to build, train, and deploy a custom ML model effectively.
From AWS documentation:

"The machine learning process begins with defining the business problem, followed by collecting and processing data, developing and training models, and finally deploying them into production for inference." Step 1 - Define the business objective:
This step involves clearly identifying the business problem to be solved and determining the measurable outcomes expected from the ML model. This ensures alignment between business goals and ML outputs.
Step 2 - Process the data:
Data is collected, cleaned, transformed, and prepared for training. This includes handling missing values, normalizing data, and performing feature engineering - a crucial phase that influences model performance.
Step 3 - Develop and train the model:
The model is built and trained on the processed data using algorithms appropriate to the problem (e.g., regression, classification, clustering). Hyperparameters are tuned to optimize model accuracy.
Step 4 - Deploy the model:
Once validated, the model is deployed to a production environment (e.g., Amazon SageMaker endpoint) to make predictions on new data. Continuous monitoring and retraining ensure the model remains effective.
Referenced AWS AI/ML Documents and Study Guides:
Amazon SageMaker Developer Guide - Machine Learning Lifecycle
AWS Certified Machine Learning Specialty Study Guide - Model Development Lifecycle AWS ML Best Practices Whitepaper - End-to-End ML Workflow

NEW QUESTION: 20
A company is developing an ML model to make loan approvals. The company must implement a solution to detect bias in the model. The company must also be able to explain the model's predictions.
Which solution will meet these requirements?
A. Amazon SageMaker Clarify
B. Amazon SageMaker Data Wrangler
C. Amazon SageMaker Model Cards
D. AWS AI Service Cards
Answer: A (LEAVE A REPLY)
Amazon SageMaker Clarify provides built-in tools to detect bias in data and models, and to generate detailed explainability reports for model predictions, including SHAP values and feature importance.
A is correct:
"Amazon SageMaker Clarify provides bias detection, explainability for ML models, and comprehensive reports to satisfy regulatory and ethical requirements." (Reference: Amazon SageMaker Clarify Overview) B (Data Wrangler) is for data preparation, not bias/explainability.
C (Model Cards) document models, but don't detect bias or explain predictions.

D (AI Service Cards) provide transparency for AWS AI services, not custom model explainability.

NEW QUESTION: 21
A company needs to use Amazon SageMaker AI for model training and inference. The company must comply with regulatory requirements to run SageMaker jobs in an isolated environment without internet access.
Which solution will meet these requirements?
A. Run SageMaker training and inference by using SageMaker Experiments.
B. Run SageMaker training and inference by using network isolation.
C. Encrypt the data at rest by using encryption for SageMaker geospatial capabilities.
D. Associate appropriate AWS Identity and Access Management (IAM) roles with the SageMaker jobs.
Answer: B (LEAVE A REPLY)
Network isolation is a key security feature for SageMaker. It ensures that training and inference jobs run in a VPC and are not accessible from the internet. Per the official SageMaker documentation:
"When you enable network isolation, your model can't make any outbound network calls. This is useful for security and regulatory compliance when working with sensitive data."

NEW QUESTION: 22
An online media streaming company wants to give its customers the ability to perform natural language-based image search and filtering. The company needs a vector database that can help with similarity searches and nearest neighbor queries.
Which AWS service meets these requirements?
A. Amazon Comprehend
B. Amazon Personalize
C. Amazon Polly
D. Amazon OpenSearch Service
Answer: D (LEAVE A REPLY)
The correct answer is D because Amazon OpenSearch Service supports k-Nearest Neighbor (k-NN) search and vector similarity search, which are required for semantic search tasks, such as matching natural language queries to image embeddings.
From AWS documentation:
"Amazon OpenSearch Service supports k-NN search, which allows you to run efficient similarity searches on large-scale datasets using vector embeddings generated by models. This enables applications like natural language-based image search and personalized recommendations." In this use case, image data can be encoded into vectors using foundation models (e.g., via Amazon Bedrock or SageMaker), and OpenSearch Service can index and retrieve results based on vector similarity.
Explanation of other options:

A). Amazon Comprehend is for text-based NLP tasks and does not provide vector similarity or search functionality.
B). Amazon Personalize is for user-item recommendations and personalization, not vector-based semantic search.
C). Amazon Polly is a text-to-speech service and not related to image search or vector databases.
Referenced AWS AI/ML Documents and Study Guides:
Amazon OpenSearch Service Documentation - k-NN and Vector Search
AWS ML Specialty Study Guide - Semantic Search and Vector Indexing
AWS Generative AI Best Practices - Embeddings and Vector Databases

NEW QUESTION: 23
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 (LEAVE A REPLY)
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: 24
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: (SHOW ANSWER)

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

* Amazon Comprehend:

* A natural language processing (NLP) service that uses machine learning to uncover insights and relationships in text.

* Capable of extracting key phrases, detecting entities, analyzing sentiment, and identifying topics from text data, making it ideal for analyzing customer support interactions.

* Why Option B is Correct:

* Text Analysis Capabilities: Can process large volumes of text to identify common topics, phrases, and sentiment, providing valuable insights.

* Suitable for Customer Support Analysis: Specifically designed to understand the content and meaning of text, which is key for identifying frequently asked questions.

* Why Other Options are Incorrect:

* A. Amazon Lex: Used for building conversational interfaces, not for text analysis.

* C. Amazon Transcribe: Converts speech to text but does not perform text analysis.

* D. Amazon Translate: Used for translating text between languages, not for analyzing content.

NEW QUESTION: 25

A publishing company built a Retrieval Augmented Generation (RAG) based solution to give its users the ability to interact with published content. New content is published daily. The company wants to provide a near real-time experience to users.

Which steps in the RAG pipeline should the company implement by using offline batch processing to meet these requirements? (Select TWO.)

A. Generation of content embeddings

B. Generation of embeddings for user queries

C. Creation of the search index

D. Retrieval of relevant content

E. Response generation for the user

Answer: A,C (LEAVE A REPLY)

Comprehensive and Detailed Explanation From Exact Extract:

In a RAG (Retrieval Augmented Generation) architecture, there are steps that can be optimized using offline batch processing, particularly for operations that do not require real-time updates:

A). Generation of content embeddings:When new content is published, it can be processed in batches to generate embeddings (vector representations) offline. These embeddings are then

used at query time for similarity search. As new documents come in daily, batch processing is ideal for generating embeddings for all new content together.
"Content/document embeddings are typically generated offline, as this operation can be computationally expensive and does not need to happen in real-time." (Reference: AWS GenAI RAG Blog, Amazon Bedrock RAG Pattern) C). Creation of the search index:After generating the content embeddings, these are indexed in a vector database or search service. This indexing is also typically performed in batch as part of the offline pipeline.
"Building or updating the vector index is often performed as a batch operation, reflecting the latest state of the content repository." (Reference: AWS RAG Pattern Whitepaper) B, D, and E are real-time steps. Embeddings for user queries (B), retrieval of relevant content (D), and response generation (E) must be processed in real-time to provide an interactive experience.
References:
Retrieval Augmented Generation (RAG) on AWS
Amazon Bedrock RAG Documentation

NEW QUESTION: 26
An ecommerce company is developing an AI application that categorizes product images and extracts specifications. The application will use a high-quality labeled dataset to customize a foundation model (FM) to generate accurate responses.
Which ML technique will meet these requirements by using Amazon Bedrock?
A. Apply continued pre-training
B. Create an agent
C. Perform fine-tuning
D. Develop prompt engineering
Answer: (SHOW ANSWER)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
The correct technique is fine-tuning, which is explicitly supported by Amazon Bedrock for customizing foundation models using high-quality labeled datasets.
Fine-tuning involves:
Starting with a pre-trained foundation model
Training it further using domain-specific, labeled data
Improving accuracy for specialized tasks, such as product classification, image-based understanding, and specification extraction In this use case:
The company has labeled data
They want to customize model behavior
They require high accuracy and domain adaptation
These conditions match the definition of fine-tuning, not prompt-only methods.
Why the other options are incorrect:
A). Continued pre-training typically requires massive unlabeled datasets and is not the standard customization method exposed in Amazon Bedrock.

B). Creating an agent orchestrates model interactions and tools but does not customize the model's learned parameters.
D). Prompt engineering improves responses through prompt design but does not modify the underlying model weights, making it insufficient for deep domain adaptation.
AWS AI document references (for exact extracts):
Amazon Bedrock Documentation - section on Model customization and fine-tuning AWS Generative AI Study Guide - comparison of prompt engineering vs fine-tuning Foundation Models on AWS - explanation of fine-tuning with labeled datasets

NEW QUESTION: 27
A company wants to generate synthetic data responses for multiple prompts from a large volume of data. The company wants to use an API method to generate the responses. The company does not need to generate the responses immediately.
A. Input the prompts into the model. Generate responses by using real-time inference.
B. Use Amazon Bedrock batch inference. Generate responses asynchronously.
C. Use Amazon Bedrock agents. Build an agent system to process the prompts recursively.
D. Use AWS Lambda functions to automate the task. Submit one prompt after another and store each response.
Answer: (SHOW ANSWER)
The correct answer is B - Use Amazon Bedrock batch inference, which allows asynchronous generation of large-scale model outputs through APIs without requiring low-latency performance. According to AWS Bedrock documentation, batch inference is ideal for high-volume workloads that can tolerate delay, such as bulk content generation or summarization jobs. Unlike real-time inference, it processes requests in bulk, reducing cost and operational load. AWS handles the queuing, processing, and scaling automatically. Bedrock Agents (option C) are for workflow orchestration, not large-scale generation. AWS Lambda (option D) can automate tasks but is not optimized for high-volume LLM calls. Batch inference provides cost efficiency, scalability, and simplicity for delayed, asynchronous generation needs.
Referenced AWS AI/ML Documents and Study Guides:
Amazon Bedrock Developer Guide - Batch Inference
AWS ML Specialty Study Guide - Scalable Inference Options

NEW QUESTION: 28
An AI practitioner wants to generate more diverse and more creative outputs from a large language model (LLM).
How should the AI practitioner adjust the inference parameter?
A. Increase the temperature value.
B. Decrease the Top K value.
C. Increase the response length.
D. Decrease the prompt length.
Answer: A (LEAVE A REPLY)

Comprehensive and Detailed Explanation From Exact AWS AI documents:
The temperature parameter controls randomness in model outputs.
AWS generative AI guidance explains:
Higher temperature # more randomness and creativity
Lower temperature # more deterministic and predictable outputs
To increase diversity and creativity, the temperature should be increased.
Why the other options are incorrect:
Lower Top K (B) reduces output diversity.
Response length (C) affects size, not creativity.
Prompt length (D) does not directly control randomness.
AWS AI document references:
Inference Parameters for Foundation Models
Controlling Creativity in LLMs
Text Generation Configuration on AWS

NEW QUESTION: 29
A company's large language model (LLM) is experiencing hallucinations.
How can the company decrease hallucinations?
A. Set up Agents for Amazon Bedrock to supervise the model training.
B. Use data pre-processing and remove any data that causes hallucinations.
C. Decrease the temperature inference parameter for the model.
D. Use a foundation model (FM) that is trained to not hallucinate.
Answer: (SHOW ANSWER)
Hallucinations in large language models (LLMs) occur when the model generates outputs that are factually incorrect, irrelevant, or not grounded in the input data. To mitigate hallucinations, adjusting the model's inference parameters, particularly the temperature, is a well-documented approach in AWS AI Practitioner resources. The temperature parameter controls the randomness of the model's output. A lower temperature makes the model more deterministic, reducing the likelihood of generating creative but incorrect responses, which are often the cause of hallucinations.
Exact Extract from AWS AI Documents:
From the AWS documentation on Amazon Bedrock and LLMs:
"The temperature parameter controls the randomness of the generated text. Higher values (e.g., 0.8 or above) increase creativity but may lead to less coherent or factually incorrect outputs, while lower values (e.g., 0.2 or
0.3) make the output more focused and deterministic, reducing the likelihood of hallucinations." (Source: AWS Bedrock User Guide, Inference Parameters for Text Generation)
Detailed Explanation:
Option A: Set up Agents for Amazon Bedrock to supervise the model training.Agents for Amazon Bedrock are used to automate tasks and integrate LLMs with external tools, not to supervise

model training or directly address hallucinations. This option is incorrect as it does not align with the purpose of Agents in Bedrock.
Option B: Use data pre-processing and remove any data that causes hallucinations.While data pre-processing can improve model performance, identifying and removing specific data that causes hallucinations is impractical because hallucinations are often a result of the model's generative process rather than specific problematic data points. This approach is not directly supported by AWS documentation for addressing hallucinations.
Option C: Decrease the temperature inference parameter for the model.This is the correct approach. Lowering the temperature reduces the randomness in the model's output, making it more likely to stick to factual and contextually relevant responses. AWS documentation explicitly mentions adjusting inference parameters like temperature to control output quality and mitigate issues like hallucinations.
Option D: Use a foundation model (FM) that is trained to not hallucinate.No foundation model is explicitly trained to "not hallucinate," as hallucinations are an inherent challenge in LLMs. While some models may be fine-tuned for specific tasks to reduce hallucinations, this is not a standard feature of foundation models available on Amazon Bedrock.
References:
AWS Bedrock User Guide: Inference Parameters for Text Generation
(https://docs.aws.amazon.com/bedrock
/latest/userguide/model-parameters.html)
AWS AI Practitioner Learning Path: Module on Large Language Models and Inference Configuration Amazon Bedrock Developer Guide: Managing Model Outputs
(https://docs.aws.amazon.com/bedrock/latest
/devguide/)

NEW QUESTION: 30
Which technique can a company use to lower bias and toxicity in generative AI applications during the post- processing ML lifecycle?
A. Human-in-the-loop
B. Data augmentation
C. Feature engineering
D. Adversarial training
Answer: A (LEAVE A REPLY)
The correct answer is A because Human-in-the-loop (HITL) is a post-processing strategy used to monitor, review, and filter outputs from generative AI models for toxicity, bias, or inappropriate content. It allows human reviewers to approve or reject model responses before they are delivered to end-users, ensuring alignment with ethical guidelines and company policies.
From the AWS documentation:
"Human-in-the-loop (HITL) workflows in generative AI are used to validate and approve outputs of models, especially in applications where content quality, compliance, or harm reduction is critical.

HITL is a key step in responsible AI implementations to mitigate hallucinations, bias, and unsafe content." Explanation of other options:
B). Data augmentation is a pre-processing technique to increase data diversity, not typically used in post- processing stages.
C). Feature engineering is relevant in traditional ML, especially structured data tasks, not typically used in generative AI post-processing.
D). Adversarial training is a model training strategy, not a post-processing mitigation approach.
Referenced AWS AI/ML Documents and Study Guides:
AWS Responsible AI Practices Whitepaper
AWS Generative AI Developer Guide - Human-in-the-loop and Post-processing Amazon A2I Documentation - Integrating Human Review in ML Workflows

NEW QUESTION: 31
Which prompting attack directly exposes the configured behavior of a large language model (LLM)?
A. Prompted persona switches
B. Exploiting friendliness and trust
C. Ignoring the prompt template
D. Extracting the prompt template
Answer: D (LEAVE A REPLY)
A prompt template defines how the model is structured and guided (system prompts, roles, guardrails).
An attack that reveals or leaks this prompt template is known as a prompt extraction attack.
The other options (persona switching, exploiting friendliness, ignoring prompts) describe adversarial techniques but do not directly expose the internal configured behavior.
Reference:
AWS Responsible AI - Prompt Injection & Extraction Attacks

NEW QUESTION: 32
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?

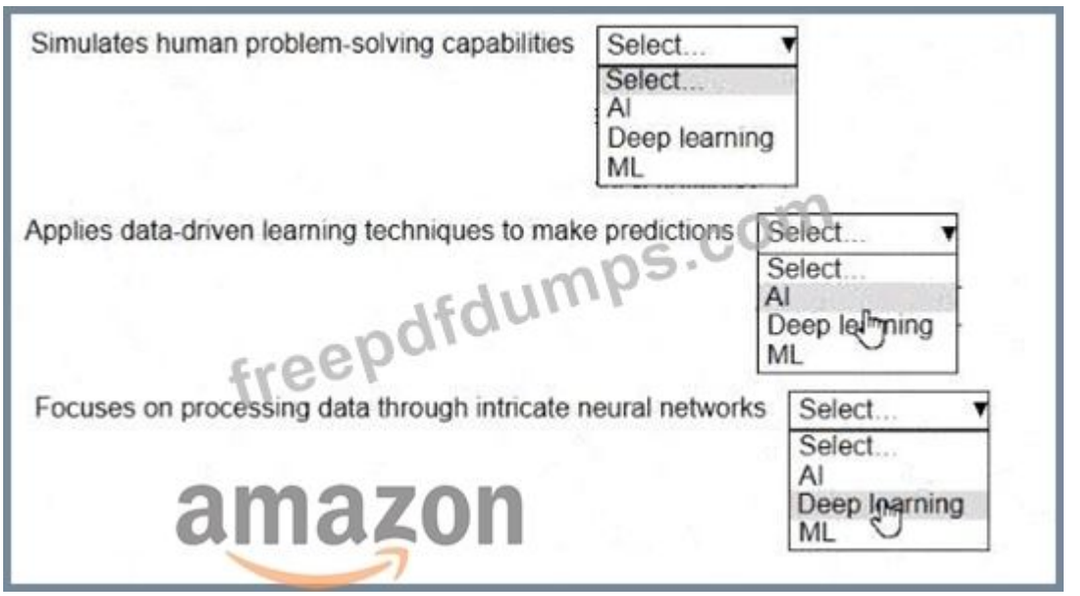
A. Build an automatic named entity recognition system.
B. Create a recommendation engine.
C. Develop a summarization chatbot.
D. Develop a multi-language translation system.
Answer: C (LEAVE A REPLY)
A summarization chatbot is ideal for extracting key points from legal documents. Large language models (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: 33
What is an example of structured data?
A. A file of text comments from an online forum
B. A compilation of video files that contains news broadcasts
C. A CSV file that consists of measurement data
D. Transcribed conversations between call center agents and customers
Answer: C (LEAVE A REPLY)
The correct answer is C - A CSV file that consists of measurement data, which is a classic example of structured data. According to AWS documentation, structured data is data that is organized in a predefined schema, typically stored in tabular formats with fixed rows, columns, and data types. CSV files, relational databases, and spreadsheets fall into this category because their structure allows deterministic querying and processing. Measurement data contained in a CSV file is easy to analyze with SQL, Amazon Athena, Amazon Redshift, or SageMaker Data Wrangler. In contrast, options A and D (text comments and transcriptions) are examples of unstructured data, requiring NLP techniques. Option B (video files) represents unstructured multimedia data, requiring computer vision. Structured data is essential for supervised ML models, business analytics, and statistical modeling. AWS emphasizes that structured data provides the simplest path to feature engineering and model training due to its consistent format.
Referenced AWS Documentation:
* AWS Data Analytics Whitepaper - Structured vs. Unstructured Data

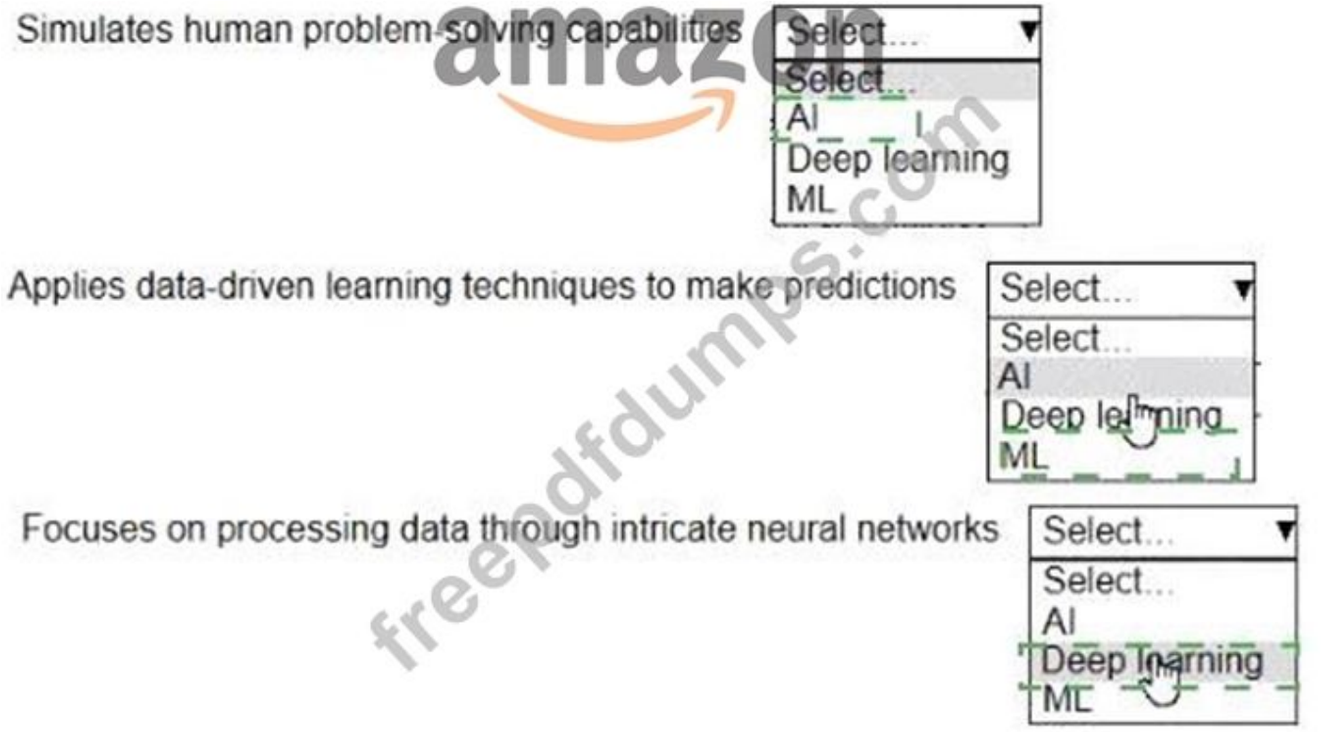
* Amazon SageMaker Data Wrangler Documentation

NEW QUESTION: 34
A company has implemented a generative AI solution to create personalized exercise routines for premium subscription users. The company offers free basic subscriptions and paid premium subscriptions. The company wants to evaluate the AI solution's return on investment over time.
A. The average revenue per user (ARPU) over the past month
B. The number of daily interactions by basic subscription users
C. The conversion rate and the customer retention rate
D. The decrease in the number of premium customer queries and issue volume
Answer: C (LEAVE A REPLY)
The correct answer is C, as conversion rate (how many users upgrade to premium) and retention rate (how many stay subscribed) are primary indicators of a generative AI system's business impact and ROI.
According to AWS documentation, evaluating AI performance should go beyond technical accuracy to include business metrics that show real-world value. For a premium service, higher conversion implies successful personalization that attracts free users, while strong retention reflects sustained engagement and user satisfaction. Metrics like ARPU or query reduction are secondary indicators. AWS emphasizes outcome- based measurement to ensure AI initiatives contribute measurable ROI tied to customer adoption, loyalty, and profitability. Therefore, tracking conversion and retention gives the clearest long-term measure of financial success from generative AI.
Referenced AWS AI/ML Documents and Study Guides:
AWS Machine Learning Specialty Guide - ML Business Metrics
AWS AI Adoption Framework - Value Realization and ROI Tracking

NEW QUESTION: 35
HOTSPOT
Select the correct AI term from the following list for each statement. Each AI term should be selected one time. (Select THREE.)
* AI
* Deep learning
* ML



Answer:



Explanation:

Artificial Intelligence (AI) is the broad field focused on simulating human problem-solving and cognitive abilities, including reasoning, perception, and decision-making.

(Reference: AWS Certified AI Practitioner Official Study Guide)

Machine Learning (ML) is a subset of AI that uses data-driven algorithms to identify patterns and make predictions without explicit programming for each specific task.

(Reference: AWS Machine Learning Overview)

Deep learning is a subset of ML that uses neural networks with many layers (deep neural networks) to process complex data and extract high-level features.

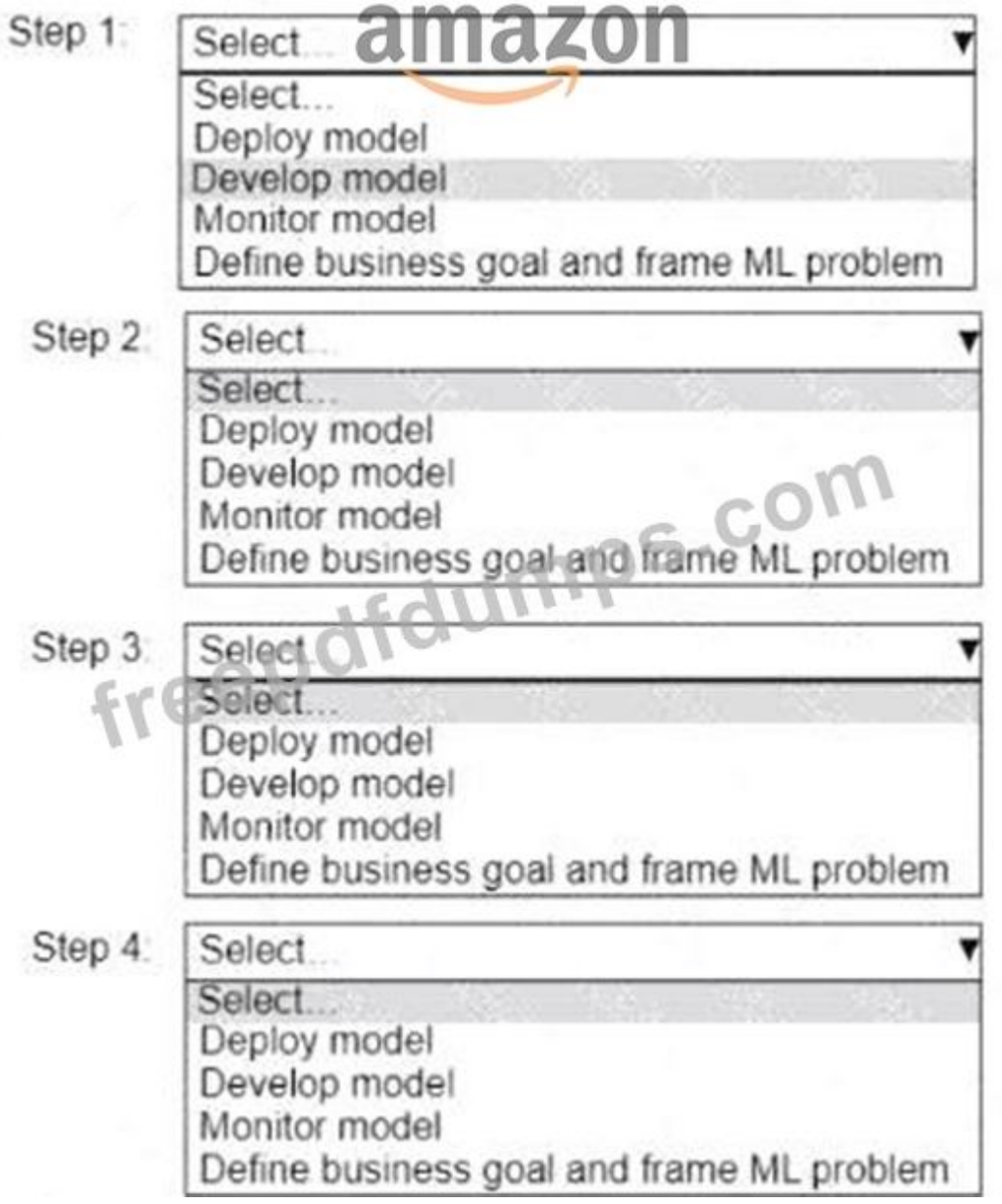
(Reference: AWS Deep Learning on AWS)

NEW QUESTION: 36

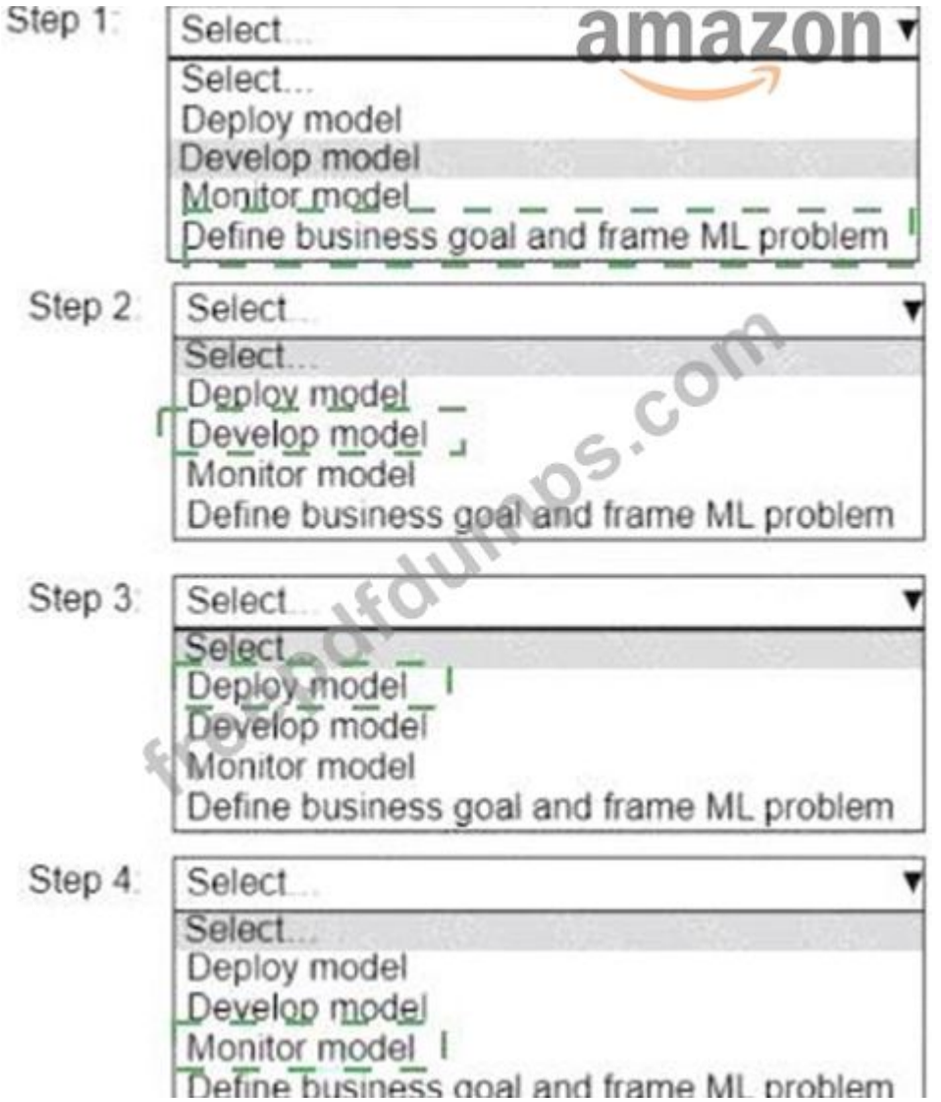
A company wants to build an ML application.

Select and order the correct steps from the following list to develop a well-architected ML workload. Each step should be selected one time. (Select and order FOUR.)

* Deploy model
* Develop model
* Monitor model
* Define business goal and frame ML problem



Answer:



Explanation:

Building a well-architected ML workload follows a structured lifecycle as outlined in AWS best practices.

The process begins with defining the business goal and framing the ML problem to ensure the project aligns with organizational objectives. Next, the model is developed, which includes data preparation, training, and evaluation. Once the model is ready, it is deployed tomake predictions in a production environment. Finally, the model is monitored to ensure it performs as expected and to address any issues like drift or degradation over time. This order ensures a systematic approach to ML development.

Exact Extract from AWS AI Documents:

From the AWS AI Practitioner Learning Path:

"The machine learning lifecycle typically follows these stages: 1) Define the business goal and frame the ML problem, 2) Develop the model (including data preparation, training, and evaluation), 3) Deploy the model to production, and 4) Monitor the model for performance and drift to ensure it continues to meet business needs." (Source: AWS AI Practitioner Learning Path, Module on Machine Learning Lifecycle) Detailed Explanation:

Step 1: Define business goal and frame ML problemThis is the first step in any ML project. It involves understanding the business objective (e.g., reducing churn) and framing the ML problem

(e.g., classification or regression). Without this step, the project lacks direction. The hotspot lists this option as "Define business goal and frame ML problem," which matches this stage.
Step 2: Develop modelAfter defining the problem, the next step is to develop the model. This includes collecting and preparing data, selecting an algorithm, training the model, and evaluating its performance. The hotspot lists "Develop model" as an option, aligning with this stage.
Step 3: Deploy modelOnce the model is developed and meets performance requirements, it is deployed to a production environment to make predictions or automate decisions. The hotspot includes "Deploy model" as an option, which fits this stage.
Step 4: Monitor modelAfter deployment, the model must be monitored to ensure it performs well over time, addressing issues like data drift or performance degradation. The hotspot lists "Monitor model" as an option, completing the lifecycle.
Hotspot Selection Analysis:
The hotspot provides four steps, each with the same dropdown options: "Select...," "Deploy model," "Develop model," "Monitor model," and "Define business goal and frame ML problem."
The correct selections are:
Step 1: Define business goal and frame ML problem
Step 2: Develop model
Step 3: Deploy model
Step 4: Monitor model
Each option is used exactly once, as required, and follows the logical order of the ML lifecycle.
References:
AWS AI Practitioner Learning Path: Module on Machine Learning Lifecycle Amazon SageMaker Developer Guide: Machine Learning Workflow (https://docs.aws.amazon.com/sagemaker/latest/dg/how-it-works-mlconcepts.html)
AWS Well-Architected Framework: Machine Learning Lens (https://docs.aws.amazon.com/wellarchitected/latest/machine-learning-lens/)

NEW QUESTION: 37
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: (SHOW ANSWER)

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: 38
A customer service team is developing an application to analyze customer feedback and automatically classify the feedback into different categories. The categories include product quality, customer service, and delivery experience.
Which AI concept does this scenario present?
A. Computer vision
B. Natural language processing (NLP)
C. Recommendation systems
D. Fraud detection
Answer: B (LEAVE A REPLY)
The scenario involves analyzing customer feedback and automatically classifying it into categories such as product quality, customer service, and delivery experience. This task requires processing and understanding textual data, which is a core application of natural language processing (NLP). NLP encompasses techniques for analyzing, interpreting, and generating human language, including tasks like text classification, sentiment analysis, and topic modeling, all of which are relevant to this use case.
Exact Extract from AWS AI Documents:
From the AWS AI Practitioner Learning Path:
"Natural Language Processing (NLP) enables machines to understand and process human language. Common NLP tasks include text classification, sentiment analysis, named entity recognition, and topic modeling.

Services like Amazon Comprehend can be used to classify text into predefined categories based on content." (Source: AWS AI Practitioner Learning Path, Module on AI and ML Concepts)
Detailed Explanation:
Option A: Computer visionComputer vision involves processing and analyzing visual data, such as images or videos. Since the scenario deals with textual customer feedback, computer vision is not applicable.
Option B: Natural language processing (NLP)This is the correct answer. The task of classifying customer feedback into categories requires understanding and processing text, which is an NLP task. AWS services like Amazon Comprehend are specifically designed for such text classification tasks.
Option C: Recommendation systemsRecommendation systems suggest items or content based on user preferences or behavior. The scenario does not involve recommending products or services but rather classifying feedback, so this option is incorrect.
Option D: Fraud detectionFraud detection involves identifying anomalous or fraudulent activities, typically in financial or transactional data. The scenario focuses on text classification, not anomaly detection, making this option irrelevant.
References:
AWS AI Practitioner Learning Path: Module on AI and ML Concepts
Amazon Comprehend Developer Guide: Text Classification
(https://docs.aws.amazon.com/comprehend/latest
/dg/how-classification.html)
AWS Documentation: Introduction to NLP (https://aws.amazon.com/what-is/natural-language-processing/)

NEW QUESTION: 39
An AI practitioner is writing software code. The AI practitioner wants to quickly develop a test case and create documentation for the code.
A. Upload the code to an online coding assistant.
B. Develop an application to use foundation models (FMs).
C. Use Amazon Q Developer in an integrated development environment (IDE).
D. Research and write test cases. Then, create test cases and add documentation.
Answer: C (LEAVE A REPLY)
* Amazon Q Developer is an AI-powered coding assistant integrated into IDEs (e.g., VS Code, JetBrains). It can:
* Generate unit tests.
* Create documentation.
* Suggest code completions.
* This is the fastest and most effective solution for this scenario.
# Reference:
Amazon Q Developer - AWS Documentation

NEW QUESTION: 40
A company wants to control employee access to publicly available foundation models (FMs).
Which solution meets these requirements?
A. Analyze cost and usage reports in AWS Cost Explorer.
B. Download AWS security and compliance documents from AWS Artifact.
C. Configure Amazon SageMaker JumpStart to restrict discoverable FMs.
D. Build a hybrid search solution by using Amazon OpenSearch Service.
Answer: C (LEAVE A REPLY)
The correct answer is C because Amazon SageMaker JumpStart provides administrative controls that allow organizations to manage and restrict access to foundation models within the AWS environment.
According to the official AWS documentation:
"Amazon SageMaker JumpStart provides model access management capabilities that enable administrators to control which foundation models are visible and usable by end users. Using AWS Identity and Access Management (IAM) policies, you can restrict access to specific models or completely disable model discovery in JumpStart." This allows companies to enforce governance over which FMs their users can see and interact with, satisfying the requirement to control employee access to publicly available foundation models.
Explanation of other options:
A . AWS Cost Explorer is used to analyze billing and usage data but does not control access to services or models. It is helpful for budgeting and visibility, not access control.
B . AWS Artifact provides access to compliance reports and certifications, not tools for controlling user access to ML models.
D . Amazon OpenSearch Service is used for search and analytics on structured and unstructured data. It does not provide access control mechanisms for foundation models.
Referenced AWS AI/ML Documents and Study Guides:
Amazon SageMaker JumpStart Documentation - Model Access Management
AWS IAM Documentation - Restricting Access to SageMaker Resources
AWS Machine Learning Specialty Certification Guide - Security and Governance Section

NEW QUESTION: 41
A company has a large amount of unlabeled data. The company wants to group the data based on feature similarities.
Which algorithm will meet this requirement?
A. XGBoost
B. K-means
C. DeepAR forecasting
D. Linear learner
Answer: B (LEAVE A REPLY)
Comprehensive and Detailed Explanation (AWS AI documents):

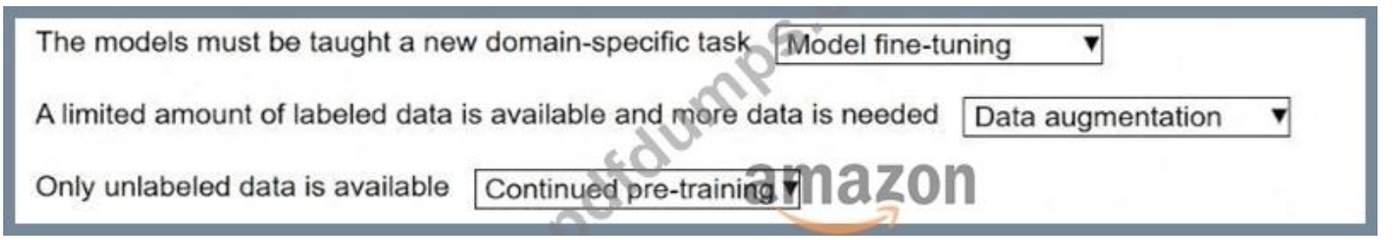
AWS machine learning fundamentals classify K-means as an unsupervised learning algorithm used to group unlabeled data into clusters based on feature similarity and distance metrics. Because the data is unlabeled and the goal is grouping rather than prediction, K-means is the most appropriate choice.
Why the other options are incorrect:
* XGBoost is a supervised learning algorithm that requires labeled data.
* DeepAR forecasting is designed for time series forecasting.
* Linear learner is typically used for supervised regression or classification tasks.
AWS AI Study Guide References:
* AWS unsupervised learning concepts
* AWS clustering algorithms overview

NEW QUESTION: 42
A company is developing a new image classification model by using a dataset of photos. The dataset must follow the AWS principles of responsible AI.
Which characteristics should the dataset have to meet this requirement?
A. The dataset should be diverse, sourced from reputable sources, and have balanced categories.
B. The dataset should contain over 5 million photos, and 1% of photos should be labeled.
C. The dataset should include photos from a limited source.
D. The dataset should be curated entirely by the company's own engineers and researchers.
Answer: (SHOW ANSWER)
Comprehensive and Detailed Explanation (AWS AI documents):
AWS Responsible AI principles stress the importance of fairness, robustness, and accountability, which begin with the quality of the data used to train models. A responsible dataset should:
* Be diverse, representing different populations, environments, and conditions to reduce bias
* Be sourced from reputable and appropriate sources, ensuring data quality and ethical use
* Have balanced categories, preventing the model from favoring one class over others and reducing the risk of discriminatory outcomes These characteristics help ensure that the resulting image classification model behaves fairly, performs reliably across groups, and aligns with AWS Responsible AI best practices.
Why the other options are incorrect:
* B focuses on dataset size rather than quality, balance, or representativeness.
* C increases the risk of bias and poor generalization.
* D limits diversity and does not inherently ensure fairness or accountability.
AWS AI Study Guide References:
* AWS Responsible AI principles: fairness and data quality
* AWS guidance on dataset selection and preparation for responsible ML

NEW QUESTION: 43

A company is using an Amazon Bedrock base model to summarize documents for an internal use case. The company trained a custom model to improve the summarization quality.
Which action must the company take to use the custom model through Amazon Bedrock?
A. Purchase Provisioned Throughput for the custom model.
B. Deploy the custom model in an Amazon SageMaker endpoint for real-time inference.
C. Register the model with the Amazon SageMaker Model Registry.
D. Grant access to the custom model in Amazon Bedrock.
Answer: (SHOW ANSWER)
To use a custom model that has been trained to improve summarization quality, the company must deploy the model on an Amazon SageMaker endpoint. This allows the model to be used for real-time inference through Amazon Bedrock or other AWS services. By deploying the model in SageMaker, the custom model can be accessed programmatically via API calls, enabling integration with Amazon Bedrock.
Option B (Correct): "Deploy the custom model in an Amazon SageMaker endpoint for real-time inference":
This is the correct answer because deploying the model on SageMaker enables it to serve real-time predictions and be integrated with Amazon Bedrock.
Option A: "Purchase Provisioned Throughput for the custom model" is incorrect because provisioned throughput is related to database or storage services, not model deployment.
Option C: "Register the model with the Amazon SageMaker Model Registry" is incorrect because while the model registry helps with model management, it does not make the model accessible for real-time inference.
Option D: "Grant access to the custom model in Amazon Bedrock" is incorrect because Bedrock does not directly manage custom model access; it relies on deployed endpoints like those in SageMaker.
AWS AI Practitioner References:
Amazon SageMaker Endpoints: AWS recommends deploying models to SageMaker endpoints to use them for real-time inference in various applications.

NEW QUESTION: 44
A company wants more customized responses to its generative AI models' prompts.
Select the correct customization methodology from the following list for each use case. Each use case should be selected one time. (Select THREE.)
* Continued pre-training
* Data augmentation
* Model fine-tuning
Answer:



Model fine-tuning adapts a pre-trained model to a specific domain or task using labeled data. This is the preferred approach when you want highly customized behavior for a particular application or subject area.
(Reference: Amazon Bedrock Fine-Tuning)
Data augmentation is used to artificially increase the size of a labeled dataset, usually by transforming or generating variations of the original data. This helps improve model generalization when labeled data is limited.
(Reference: AWS Data Preparation Techniques, AWS AI Practitioner Guide) Continued pre-training (also called domain-adaptive pre-training) further trains a foundation model on large amounts of unlabeled data from a specific domain, improving the model's language understanding or generation for that domain.
(Reference: Amazon Bedrock Customization Options)

NEW QUESTION: 45
What is tokenization used for in natural language processing (NLP)?
A. To encrypt text data
B. To compress text files
C. To break text into smaller units for processing
D. To translate text between languages
Answer: C (LEAVE A REPLY)
The correct answer is C because tokenization is the NLP process of breaking down text into smaller units, such as words, subwords, or characters, that can be processed by ML models.
From AWS documentation:
"Tokenization is the process of splitting text into meaningful units, such as words or subwords, that are used as input tokens in NLP tasks. Tokenization is an essential step in preparing text data for models." This is a foundational concept used in language models, including those on Amazon Bedrock and SageMaker.
Explanation of other options:
A). Encryption is not related to NLP and is used for data security.
B). Compression reduces file size but is unrelated to language processing.
D). Translation is a separate NLP task that uses tokenization as a preprocessing step but is not the definition of tokenization itself.
Referenced AWS AI/ML Documents and Study Guides:
AWS NLP on Amazon SageMaker Documentation
AWS Machine Learning Specialty Guide - NLP Fundamentals
Amazon Bedrock Foundation Model Documentation - Tokenization and Input Formatting

NEW QUESTION: 46
A hospital is developing an AI system to assist doctors in diagnosing diseases based on patient records and medical images. To comply with regulations, the sensitive patient data must not leave the country the data is located in.
Which data governance strategy will ensure compliance and protect patient privacy?
A. Data residency
B. Data quality
C. Data discoverability
D. Data enrichment
Answer: (SHOW ANSWER)
The correct answer is Data residency, which ensures that data remains stored and processed within specific geographical or jurisdictional boundaries. AWS defines data residency as the practice of keeping sensitive or regulated data, such as healthcare records, inside designated regions to meet local privacy laws like HIPAA or GDPR. Amazon SageMaker, Bedrock, and other AWS services allow region-specific resource deployment, guaranteeing data never leaves the country. Data quality refers to accuracy and consistency, while discoverability and enrichment concern accessibility and augmentation, not compliance. Data residency is central to AWS's Shared Responsibility Model, ensuring organizations maintain sovereignty over healthcare data.
Referenced AWS AI/ML Documents and Study Guides:
AWS Data Privacy Whitepaper - Data Residency and Compliance
AWS ML Specialty Guide - Data Governance and Security

NEW QUESTION: 47
A company wants to use AI for budgeting. The company made one budget manually and one budget by using an AI model. The company compared the budgets to evaluate the performance of the AI model. The AI model budget produced incorrect numbers.
Which option represents the AI model's problem?
A. Hallucinations
B. Safety
C. Interpretability
D. Cost

Answer: A (LEAVE A REPLY)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
Hallucinations occur when an AI model generates incorrect, fabricated, or misleading outputs that appear plausible but are factually wrong.
AWS generative AI guidance identifies hallucinations as:
A common limitation of generative models
A risk when models generate numerical or factual data
A key reason for validation and human review in critical use cases
Why the other options are incorrect:
Safety (B) relates to harmful or restricted content.
Interpretability (C) refers to understanding how a model makes decisions.
Cost (D) concerns operational expenses.
AWS AI document references:
Generative AI Risks and Limitations
Responsible Use of Foundation Models
Model Validation Best Practices

NEW QUESTION: 48
A company wants to learn about generative AI applications in an experimental environment.
Which solution will meet this requirement MOST cost-effectively?
A. Amazon Q Developer
B. Amazon SageMaker JumpStart
C. Amazon Bedrock PartyRock
D. Amazon Q Business
Answer: (SHOW ANSWER)
The correct answer is Amazon Bedrock PartyRock, a playground for building and experimenting with generative AI apps in a low-cost, no-code environment. PartyRock is designed for innovation and learning. It enables users to try out prompts, LLM apps, and templates using Amazon Bedrock under a free-tier friendly setup. According to AWS, PartyRock abstracts infrastructure and allows rapid prototyping using models from Bedrock providers. This makes it ideal for early experimentation, especially for non-developers or those not ready to invest in full production pipelines. In contrast, Amazon Q Developer is for software engineering tasks, SageMaker JumpStart focuses on deploying ML models, and Q Business targets enterprise knowledge workers. None of those are as cost-effective and experimental-focused as PartyRock.
Referenced AWS AI/ML Documents and Study Guides:
Amazon Bedrock Documentation - PartyRock Overview
AWS Generative AI Learning Path - Getting Started Tools

NEW QUESTION: 49

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: (SHOW ANSWER)
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:
Modifying the prompt can guide the LLM to produce outputs that are shorter and tailored to the desired language.
A well-crafted prompt can provide specific instructions to the model, such as "Answer in a short sentence in Spanish." Why Option A is Correct:
Control Over Output: Adjusting the prompt allows for direct control over the style, length, and language of the LLM outputs.
Flexibility: Prompt engineering is a flexible approach to refining the model's behavior without modifying the model itself.
Why Other Options are Incorrect:
B). Choose an LLM of a different size: The model size does not directly impact the response length or language.
C). Increase the temperature: Increases randomness in responses but does not ensure brevity or specific language.
D). Increase the Top K value: Affects diversity in model output but does not align directly with response length or language specificity.

NEW QUESTION: 50
A company is working on a large language model (LLM) and noticed that the LLM's outputs are not as diverse as expected. Which parameter should the company adjust?
A. Temperature
B. Batch size
C. Learning rate
D. Optimizer type
Answer: A (LEAVE A REPLY)
The correct answer is A because temperature controls the randomness of a language model's output. A higher temperature increases diversity by making the model more likely to explore less probable tokens, while a lower temperature results in more deterministic and repetitive outputs.
From AWS documentation:

"The temperature parameter in LLMs adjusts the randomness of generated responses. Higher values (e.g., 0.8-
1.0) produce more creative and diverse output, while lower values (e.g., 0.1-0.3) make output more focused and repetitive." Explanation of other options:
B). Batch size is related to training efficiency, not output diversity.
C). Learning rate affects the training convergence rate, not inference-time output variety.
D). Optimizer type is a training configuration that influences how the model learns during training, not diversity during inference.
Referenced AWS AI/ML Documents and Study Guides:
Amazon Bedrock - Parameter Tuning Guide
AWS Machine Learning Specialty Guide - LLM Inference Parameters

NEW QUESTION: 51
A company is training ML models on datasets. The datasets contain some classes that have more examples than other classes. The company wants to measure how well the model balances detecting and labeling the classes.
Which metric should the company use?
A. Accuracy
B. Recall
C. Precision
D. F1 score
Answer: D (LEAVE A REPLY)
The correct answer is D - F1 score, which is the harmonic mean of precision and recall. AWS documentation states that the F1 score is the preferred metric when evaluating ML models trained on imbalanced datasets, where one class appears significantly more often than others. Accuracy becomes misleading in these situations because a model can achieve high accuracy simply by predicting the majority class. Precision alone measures correctness of positive predictions, while recall measures completeness-neither fully captures class balancing performance. The F1 score provides a combined measurement that reflects how well the model identifies minority classes while avoiding excessive false positives. AWS's ML Specialty Guide recommends using F1 for classification tasks where "class distribution is uneven or where both false negatives and false positives matter." This metric provides a fairer assessment of overall model performance across all class categories.
Referenced AWS Documentation:
AWS Certified Machine Learning Specialty Guide - Metrics for Imbalanced Classes Amazon SageMaker Developer Guide - Evaluating Classification Models

NEW QUESTION: 52
A company is using a generative AI model to develop a digital assistant. The model's responses occasionally include undesirable and potentially harmful content. Select the correct Amazon

Bedrock filter policy from the following list for each mitigation action. Each filter policy should be selected one time. (Select FOUR.)
* Content filters
* Contextual grounding check
* Denied topics
* Word filters

Answer:

Block input prompts or model responses that contain harmful content such as hate, insults, violence, or misconduct:Content filters Avoid subjects related to illegal investment advice or legal advice:Denied topics Detect and block specific offensive terms:Word filters Detect and filter out information in the model's responses that is not grounded in the provided source information:Contextual grounding check The company is using a generative AI model on Amazon Bedrock and needs to mitigate undesirable and potentially harmful content in the model's responses. Amazon Bedrock provides several guardrail mechanisms, including content filters, denied topics, word filters, and contextual grounding checks, to ensure safe and accurate outputs. Each mitigation action in the hotspot aligns with a specific Bedrock filter policy, and each policy must be used exactly once.
Exact Extract from AWS AI Documents:
From the AWS Bedrock User Guide:
*"Amazon Bedrock guardrails provide mechanisms to control model outputs, including:
Content filters: Block harmful content such as hate speech, violence, or misconduct.
Denied topics: Prevent the model from generating responses on specific subjects, such as illegal activities or advice.
Word filters: Detect and block specific offensive or inappropriate terms.
Contextual grounding check: Ensure responses are grounded in the provided source information, filtering out ungrounded or hallucinated content."*(Source: AWS Bedrock User Guide, Guardrails for Responsible AI) Detailed Explanation:
Block input prompts or model responses that contain harmful content such as hate, insults, violence, or misconduct: Content filtersContent filters in Amazon Bedrock are designed to detect and block harmful content, such as hate speech, insults, violence, or misconduct, ensuring the model's outputs are safe and appropriate. This matches the first mitigation action.
Avoid subjects related to illegal investment advice or legal advice: Denied topicsDenied topics allow users to specify subjects the model should avoid, such as illegal investment advice or legal advice, which could have regulatory implications. This policy aligns with the second mitigation action.
Detect and block specific offensive terms: Word filtersWord filters enable the detection and blocking of specific offensive or inappropriate terms defined by the user, making them ideal for this mitigation action focused on specific terms.
Detect and filter out information in the model's responses that is not grounded in the provided source information: Contextual grounding checkThe contextual grounding check ensures that the

model's responses are based on the provided source information, filtering out ungrounded or hallucinated content. This matches the fourth mitigation action.
Hotspot Selection Analysis:
The hotspot lists four mitigation actions, each with the same dropdown options: "Select...," "Content filters,"
"Contextual grounding check," "Denied topics," and "Word filters." The correct selections are:
First action: Content filters
Second action: Denied topics
Third action: Word filters
Fourth action: Contextual grounding check
Each filter policy is used exactly once, as required, and aligns with Amazon Bedrock's guardrail capabilities.
References:
AWS Bedrock User Guide: Guardrails for Responsible AI
(https://docs.aws.amazon.com/bedrock/latest
/userguide/guardrails.html)
AWS AI Practitioner Learning Path: Module on Responsible AI and Model Safety Amazon Bedrock Developer Guide: Configuring Guardrails (https://aws.amazon.com/bedrock/)

NEW QUESTION: 53
A software company wants to use a large language model (LLM) for workflow automation. The application will transform user messages into JSON files. The company will use the JSON files as inputs for data pipelines.
The company has a labeled dataset that contains user messages and output JSON files.
Which solution will train the LLM for workflow automation?
A. Unsupervised learning
B. Continued pre-training
C. Fine-tuning
D. Reinforcement learning from human feedback (RLHF)
Answer: C (LEAVE A REPLY)
Fine-tuning is the process of training a pre-trained LLM with a labeled dataset specific to a desired task-in this case, mapping user messages to JSON outputs. Fine-tuning leverages supervised learning to specialize the model's outputs.
C is correct:
"Fine-tuning is a supervised learning approach in which a model is further trained on a custom, labeled dataset to adapt to a specific use case." (Reference: Amazon Bedrock Fine-Tuning, AWS Certified AI Practitioner Study Guide) A is incorrect-unsupervised learning does not use labeled data.
B (continued pre-training) uses unlabeled data.
D (RLHF) uses reward signals and human feedback, not direct labeled input/output pairs.

NEW QUESTION: 54
Which prompting technique can protect against prompt injection attacks?
A. Adversarial prompting
B. Zero-shot prompting
C. Least-to-most prompting
D. Chain-of-thought prompting
Answer: (SHOW ANSWER)
The correct answer is A because adversarial prompting is a defensive technique used to identify and protect against prompt injection attacks in large language models (LLMs). In adversarial prompting, developers intentionally test the model with manipulated or malicious prompts to evaluate how it behaves under attack and to harden the system by refining prompts, filters, and validation logic.
From AWS documentation:
"Adversarial prompting is used to evaluate and defend generative AI models against harmful or manipulative inputs (prompt injections). By testing with adversarial examples, developers can identify vulnerabilities and apply safeguards such as Guardrails or context filtering to prevent model misuse." Prompt injection occurs when an attacker tries to override system or developer instructions within a prompt, leading the model to disclose restricted information or behave undesirably. Adversarial prompting helps uncover and mitigate these risks before deployment.
Explanation of other options:
B). Zero-shot prompting provides no examples and does not protect against injection attacks.
C). Least-to-most prompting is a reasoning technique used to break down complex problems step-by-step, not a security measure.
D). Chain-of-thought prompting encourages detailed reasoning by the model but can actually increase exposure to prompt injection if not properly constrained.
Referenced AWS AI/ML Documents and Study Guides:
AWS Responsible AI Practices - Prompt Injection and Safety Testing
Amazon Bedrock Developer Guide - Secure Prompt Design and Evaluation
AWS Generative AI Security Whitepaper - Adversarial Testing and Guardrails

NEW QUESTION: 55
Which AWS service makes foundation models (FMs) available to help users build and scale generative AI applications?
A. Amazon Q Developer
B. Amazon Bedrock
C. Amazon Kendra
D. Amazon Comprehend
Answer: B (LEAVE A REPLY)
The correct answer is Amazon Bedrock, AWS's fully managed service for building and scaling generative AI applications using foundation models (FMs). Bedrock gives developers access to models from leading providers such as Anthropic (Claude), Meta (Llama), Mistral, Cohere, and

Amazon Titan. Users can invoke these models via API without managing infrastructure or model training. According to AWS documentation, Bedrock supports tasks such as text generation, summarization, question answering, image generation, and RAG workflows with minimal setup. It supports both on-demand and provisioned throughput modes and integrates with features like Guardrails, Knowledge Bases, and Agents for secure, enterprise-grade applications. Amazon Q Developer is a generative AI tool for developers, but it doesn't host or scale models.
Amazon Kendra is an intelligent search engine, and Amazon Comprehend is used for NLP tasks like entity extraction-not foundation model access.
Referenced AWS AI/ML Documents and Study Guides:
Amazon Bedrock Developer Guide - Foundation Models and Use Cases
AWS Certified ML Specialty Guide - Generative AI on AWS

NEW QUESTION: 56
A company uses Amazon Comprehend to analyze customer feedback. A customer has several unique trained models. The company uses Comprehend to assign each model an endpoint. The company wants to automate a report on each endpoint that is not used for more than 15 days.
A. AWS Trusted Advisor
B. Amazon CloudWatch
C. AWS CloudTrail
D. AWS Config
Answer: B (LEAVE A REPLY)
The correct answer is B because Amazon CloudWatch provides monitoring capabilities that include tracking usage metrics for Amazon Comprehend endpoints, such as invocation counts. You can configure CloudWatch to collect these metrics and create custom dashboards or alarms to report when an endpoint has zero usage over a period (e.g., 15 days).
From AWS documentation:
"Amazon CloudWatch enables you to collect and track metrics for Comprehend endpoints, create alarms, and automatically react to changes in your AWS resources." This allows automated reporting and alerting for underused or idle endpoints.
Explanation of other options:
A). AWS Trusted Advisor focuses on general AWS resource optimization, security, and limits but does not track endpoint usage.
C). AWS CloudTrail tracks API calls but does not provide time-based monitoring or usage analysis over time.
D). AWS Config is used to track configuration changes and compliance, not endpoint usage metrics.
Referenced AWS AI/ML Documents and Study Guides:
Amazon CloudWatch Metrics for Amazon Comprehend
AWS Certified Machine Learning Specialty Guide - Monitoring and Logging Section AWS Cloud Operations Guide - Resource Utilization Monitoring

NEW QUESTION: 57
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 (LEAVE A REPLY)
AWS Glue is the correct service for transforming petabytes of unstructured data into a structured format suitable for machine learning tasks.
AWS Glue:
A fully managed extract, transform, and load (ETL) service that makes it easy to prepare and transform unstructured data into a structured format.
Provides a range of tools for cleaning, enriching, and cataloging data, making it ready for data scientists to use in ML models.
Why Option D is Correct:
Data Transformation: AWS Glue can handle large volumes of data and transform unstructured data into structured formats efficiently.
Integrated ML Support: Glue integrates with other AWS services to support ML workflows.
Why Other Options are Incorrect:
A . Amazon Lex: Used for building chatbots, not for data transformation.
B . Amazon Rekognition: Used for image and video analysis, not for data transformation.
C . Amazon Kinesis Data Streams: Handles real-time data streaming, not suitable for batch transformation of large volumes of unstructured data.

NEW QUESTION: 58
A company wants to build and deploy ML models on AWS without writing any code.
Which AWS service or feature meets these requirements?
A. Amazon SageMaker Canvas
B. Amazon Rekognition
C. AWS DeepRacer
D. Amazon Comprehend
Answer: A (LEAVE A REPLY)
Amazon SageMaker Canvas is a visual, no-code tool for building and deploying ML models.
According to the official SageMaker Canvas documentation:
"SageMaker Canvas provides a visual point-and-click interface that allows business analysts to generate accurate ML predictions without writing any code."

NEW QUESTION: 59

A company is developing an editorial assistant application that uses generative AI. During the pilot phase, usage is low and application performance is not a concern. The company cannot predict application usage after the application is fully deployed and wants to minimize application costs.
Which solution will meet these requirements?
A. Use GPU-powered Amazon EC2 instances.
B. Use Amazon Bedrock with Provisioned Throughput.
C. Use Amazon Bedrock with On-Demand Throughput.
D. Use Amazon SageMaker JumpStart.
Answer: C (LEAVE A REPLY)
The correct answer is C because Amazon Bedrock with On-Demand Throughput is designed for variable or unpredictable workloads. It allows you to pay only for what you use, which is ideal for early-stage or pilot-phase generative AI applications.
From AWS documentation:
"On-Demand Throughput in Amazon Bedrock provides flexibility and cost-efficiency for unpredictable or low-volume generative AI applications. You are charged based on the number of input/output tokens processed, with no need to provision dedicated capacity." Explanation of other options:
A . Using GPU-powered EC2 instances requires managing infrastructure and is generally more expensive, especially for inconsistent usage.
B . Provisioned Throughput is best suited for predictable, high-volume production workloads where guaranteed throughput is needed.
D . SageMaker JumpStart is for training and deploying ML models, not for efficient inference of foundation models.
Referenced AWS AI/ML Documents and Study Guides:
Amazon Bedrock Pricing and Throughput Models
AWS Generative AI Best Practices - Cost Optimization
AWS ML Specialty Guide - Bedrock Deployment Models

NEW QUESTION: 60
A healthcare company wants to create a model to improve disease diagnostics by analyzing patient voices.
The company has recorded hundreds of patient voices for this project. The company is currently filtering voice recordings according to duration and language.
A. Data collection
B. Data preprocessing
C. Feature engineering
D. Model training
Answer: B (LEAVE A REPLY)
The correct answer is B - Data preprocessing, which involves cleaning, filtering, and transforming raw data before model training. AWS defines preprocessing as the step where irrelevant or poor-

quality data (e.g., excessively short or misrecorded audio) is removed or normalized. The described task-filtering recordings by duration and language-is a textbook example of preprocessing since it prepares the dataset for downstream feature extraction. Feature engineering would occur afterward when converting audio signals into numerical representations (like Mel-frequency coefficients). Model training and data collection come after and before preprocessing respectively. Proper preprocessing ensures data consistency, reduces noise, and improves model accuracy, especially in audio-based healthcare models where clarity and consistency are critical.
Referenced AWS AI/ML Documents and Study Guides:
AWS Machine Learning Specialty Guide - Data Preprocessing and Cleaning
Amazon SageMaker Data Wrangler Documentation - Data Preparation Best Practices

NEW QUESTION: 61
A pharmaceutical company wants to analyze user reviews of new medications and provide a concise overview for each medication. Which solution meets these requirements?
A. Create a time-series forecasting model to analyze the medication reviews by using Amazon Personalize.
B. Create medication review summaries by using Amazon Bedrock large language models (LLMs).
C. Create a classification model that categorizes medications into different groups by using Amazon SageMaker.
D. Create medication review summaries by using Amazon Rekognition.
Answer: (SHOW ANSWER)
Amazon Bedrock provides large language models (LLMs) that are optimized for natural language understanding and text summarization tasks, making it the best choice for creating concise summaries of user reviews. Time-series forecasting, classification, and image analysis (Rekognition) are not suitable for summarizing textual data. References: AWS Bedrock Documentation.

NEW QUESTION: 62
A company wants to assess internet quality in remote areas of the world. The company needs to collect internet speed data and store the data in Amazon RDS. The company will analyze internet

speed variation throughout each day. The company wants to create an AI model to predict potential internet disruptions.
Which type of data should the company collect for this task?
A. Time series data
B. Audio data
C. Text data
D. Tabular data
Answer: (SHOW ANSWER)

NEW QUESTION: 63
A company has an ML model. The company wants to know how the model makes predictions.
Which term refers to understanding model predictions?
A. Model interpretability
B. Model training
C. Model interoperability
D. Model performance
Answer: (SHOW ANSWER)
The correct answer is A because model interpretability refers to the ability to understand and explain how an ML model arrives at a particular prediction or decision.
From AWS documentation:
"Model interpretability is the degree to which a human can understand the cause of a decision made by a machine learning model. Interpretability techniques help explain which features influenced a prediction and how much they contributed." This is essential in areas like financial services, healthcare, or compliance-heavy industries, where decision transparency is critical.
Explanation of other options:
B). Model training refers to the process of teaching a model from data and doesn't explain how predictions are made.
C). Model interoperability refers to the ability of systems or models to work across different platforms or environments.
D). Model performance refers to how accurate or effective the model is but doesn't relate to the explanation of its decisions.
Referenced AWS AI/ML Documents and Study Guides:
Amazon SageMaker Clarify Documentation - Explainability and Bias Detection AWS Machine Learning Specialty Study Guide - Responsible AI and Model Explainability

NEW QUESTION: 64
Which AI technique combines large language models (LLMs) with external knowledge bases to improve response accuracy?
A. Reinforcement learning (RL)
B. Natural language processing (NLP)
C. Retrieval Augmented Generation (RAG)

D. Transfer learning

Answer: C (LEAVE A REPLY)

Comprehensive and Detailed Explanation From Exact AWS AI documents:

Retrieval Augmented Generation (RAG) enhances LLM responses by:

Retrieving relevant information from external knowledge sources

Injecting retrieved content into the prompt context

Reducing hallucinations and improving factual accuracy

AWS generative AI guidance describes RAG as a best practice when models must use up-to-date or domain- specific knowledge that is not embedded in the model weights.

Why the other options are incorrect:

RL (A) focuses on reward-based learning.

NLP (B) is a broad field, not a specific technique.

Transfer learning (D) adapts model weights but does not retrieve external data at inference time.

AWS AI document references:

Retrieval Augmented Generation on AWS

Improving LLM Accuracy with External Knowledge

Generative AI Architectures

NEW QUESTION: 65

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 Showcone

Answer: A (LEAVE A REPLY)

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: 66
How can companies use large language models (LLMs) securely on Amazon Bedrock?
A. Design clear and specific prompts. Configure AWS Identity and Access Management (IAM) roles and policies by using least privilege access.
B. Enable AWS Audit Manager for automatic model evaluation jobs.
C. Enable Amazon Bedrock automatic model evaluation jobs.
D. Use Amazon CloudWatch Logs to make models explainable and to monitor for bias.
Answer: A (LEAVE A REPLY)
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: 67
A company wants to use large language models (LLMs) to create a chatbot. The chatbot will assist customers with product inquiries, order tracking, and returns. The chatbot must be able to process text inputs and image inputs to generate responses.
Which AWS service meets these requirements?
A. Amazon Bedrock
B. Amazon Comprehend
C. Amazon Q
D. Amazon Rekognition
Answer: A (LEAVE A REPLY)

Comprehensive and Detailed Explanation From Exact AWS AI documents:
Amazon Bedrock provides access to multimodal foundation models that can process both text and image inputs and generate intelligent responses.
Bedrock is designed for:
Building chatbots and conversational AI
Handling multimodal inputs
Integrating LLMs with enterprise applications
Why the other options are incorrect:
Amazon Comprehend (B) performs text analysis only.
Amazon Q (C) is a managed assistant, not a general chatbot platform.
Amazon Rekognition (D) analyzes images but does not generate conversational responses.
AWS AI document references:
Amazon Bedrock Overview
Multimodal Foundation Models on AWS
Building Chatbots with Amazon Bedrock

NEW QUESTION: 68
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 (LEAVE A REPLY)
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: 69
A company acquires International Organization for Standardization (ISO) accreditation to manage AI risks and to use AI responsibly. What does this accreditation certify?
A. All members of the company are ISO certified.
B. All AI systems that the company uses are ISO certified.
C. All AI application team members are ISO certified.
D. The company's development framework is ISO certified.
Answer: D (LEAVE A REPLY)
ISO certifications apply to processes, frameworks, and systems - not individuals or every piece of software.
When a company is ISO-certified, its development framework and governance processes comply with ISO standards for security, risk, or AI responsibility.
Reference:
AWS Compliance Programs - ISO

NEW QUESTION: 70
A documentary filmmaker wants to reach more viewers. The filmmaker wants to automatically add subtitles and voice-overs in multiple languages to their films.
Which combination of steps will meet these requirements? (Select TWO.)
A. Use Amazon Transcribe and Amazon Translate to generate subtitles in other languages
B. Use Amazon Textract and Amazon Translate to generate subtitles in other languages
C. Use Amazon Polly to generate voice-overs in other languages
D. Use Amazon Translate to generate voice-overs in other languages
E. Use Amazon Textract to generate voice-overs in other languages
Answer: A,C (LEAVE A REPLY)
The correct answers are A and C because:
Amazon Transcribe converts spoken dialogue from video into text (captions or subtitles).
Amazon Translate translates that transcribed text into other languages.
Amazon Polly can then convert the translated text into lifelike speech for voice-overs in different languages.
From AWS documentation:
"Amazon Transcribe is used to create accurate speech-to-text transcripts. You can feed this text into Amazon Translate to support multilingual subtitles. To generate audio output, Amazon Polly provides neural text-to- speech in multiple languages." Explanation of other options:
B). Amazon Textract is used for extracting text from documents/images, not audio or video, and is not applicable here.
D). Amazon Translate does not generate speech - it only translates text.
E). Amazon Textract does not generate audio.
Referenced AWS AI/ML Documents and Study Guides:
Amazon Transcribe Developer Guide - Media Transcription
Amazon Translate Developer Guide - Multilingual Support

Amazon Polly Documentation - Text-to-Speech in Multiple Languages
AWS ML Specialty Guide - Multimedia AI Workflows

NEW QUESTION: 71
A company wants to build an ML model to detect abnormal patterns in sensor data. The company does not have labeled data for training. Which ML method will meet these requirements?
A. Linear regression
B. Classification
C. Decision tree
D. Autoencoders
Answer: (SHOW ANSWER)
The correct answer is D because autoencoders are an unsupervised machine learning method commonly used for anomaly detection when labeled data is not available.
From AWS documentation:
"Autoencoders learn to compress and reconstruct input data. During anomaly detection, they learn normal patterns in data. Data points that the model cannot accurately reconstruct are flagged as anomalies." This approach is ideal when there is no labeled data and when patterns must be learned based on normal behavior alone - a common situation in IoT sensor data environments.
Explanation of other options:
A). Linear regression requires labeled data and is used for predicting continuous values.
B). Classification requires labeled data to assign inputs into categories.
C). Decision trees are supervised learning models and also require labeled datasets.
Referenced AWS AI/ML Documents and Study Guides:
AWS Machine Learning Specialty Guide - Unsupervised Learning Techniques Amazon SageMaker Examples - Anomaly Detection Using Autoencoders

NEW QUESTION: 72
An e-commerce company wants to build a solution to determine customer sentiments based on written customer reviews of products.
Which AWS services meet these requirements? (Select TWO.)
A. Amazon Lex
B. Amazon Comprehend
C. Amazon Polly
D. Amazon Bedrock
E. Amazon Rekognition
Answer: B,D (LEAVE A REPLY)
To determine customer sentiments based on written customer reviews, the company can use Amazon Comprehend and Amazon Bedrock.
* Amazon Comprehend:

* A natural language processing (NLP) service that uses machine learning to uncover insights and relationships in text.
* Can analyze customer reviews to detect sentiments (positive, negative, neutral, or mixed) automatically.
* Amazon Bedrock:
* Provides access to foundational models (FMs) from multiple AI companies for tasks such as text generation, summarization, and sentiment analysis.
* The company can use a pre-trained sentiment analysis model available on Amazon Bedrock for processing customer reviews.
* Why Other Options are Incorrect:
* A. Amazon Lex: Used for building conversational interfaces like chatbots, not for sentiment analysis.
* C. Amazon Polly: Converts text to speech; it doesn't analyze sentiment.
* E. Amazon Rekognition: Analyzes images and videos, not text.

NEW QUESTION: 73
Which type of AI model makes numeric predictions?
A. Diffusion
B. Regression
C. Transformer
D. Multi-modal
Answer: B (LEAVE A REPLY)
The correct answer is regression. In machine learning, regression models are designed to predict continuous numerical values based on input features. Common use cases include predicting house prices, sales forecasting, temperature trends, or medical risk scores. According to AWS SageMaker documentation, regression tasks fall under supervised learning where the output is a real-valued number rather than a class label. For instance, linear regression is one of the most commonly used models for predicting a single continuous output. By contrast, diffusion models are typically used in generative image tasks, transformers are architectures (not specific to numeric output), and multi-modal models process various data types like text, images, and audio. Only regression models are purpose-built for making precise numeric predictions, which aligns with AWS best practices when the output is a quantity, not a category.
Referenced AWS AI/ML Documents and Study Guides:
AWS Machine Learning Specialty Guide - Regression Models
Amazon SageMaker Built-in Algorithms - Linear Learner (Regression and Classification)

NEW QUESTION: 74
An airline company wants to use a generative AI model to convert a flight booking system from one coding language into another coding language. The company must select a model for this task.
Which criteria should the company use to select the correct generative AI model for this task?

A. Syntax, semantic understanding, and code optimization capabilities
B. Code generation speed and error handling capabilities
C. Ability to generate creative content
D. Model size and resource requirements
Answer: A (LEAVE A REPLY)
The correct answer is A because code translation tasks require a model with deep understanding of syntax, semantic meaning, and the ability to maintain functional equivalence in different programming languages.
Some foundation models offered via Amazon Bedrock (like CodeWhisperer or Meta's Code LLMs) are designed with these capabilities in mind.
From AWS documentation:
"Generative AI models used for code generation and translation must understand programming semantics and syntax to generate accurate and secure code. These models are trained to recognize language-specific patterns and preserve logic when converting between languages."
Explanation of other options:
B). Speed and error handling are secondary to correctness and comprehension in code translation tasks.
C). Creative content generation is not relevant for deterministic tasks like source code translation.
D). Model size may affect performance or latency but is not a primary selection criterion for translation accuracy.
Referenced AWS AI/ML Documents and Study Guides:
* Amazon Bedrock Model Selection Guide - Code Generation and Translation
* AWS Developer Tools - CodeWhisperer Capabilities
* AWS ML Specialty Study Guide - LLM Use Cases in Software Engineering

NEW QUESTION: 75
A company stores customer personally identifiable information (PII) data. The company must store the PII data within the company's AWS Region.
Which aspect of governance does this describe?
A. Data mining
B. Data residency
C. Pre-training bias
D. Geolocation routing
Answer: B (LEAVE A REPLY)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
Data residency refers to requirements that dictate where data is physically stored and processed, often to meet regulatory, legal, or organizational policies.
AWS governance guidance emphasizes that:
PII data may be legally required to remain within a specific geographic region Customers can control data location by selecting AWS Regions Data residency is a core component of data governance and compliance Why the other options are incorrect:

Data mining (A) refers to extracting insights from data.
Pre-training bias (C) relates to model training data, not storage location.
Geolocation routing (D) concerns request routing, not data governance.
AWS AI document references:
AWS Data Governance and Residency
Responsible AI and Data Management on AWS
Security and Compliance in AWS Regions

NEW QUESTION: 76
A company deploys a custom ML model on Amazon SageMaker AI. The company uses the model to build a generative AI application for a healthcare recommendation system.
The company tests the application and finds a potential bias issue. The application consistently recommends different treatment approaches for patients who have identical medical conditions based on patient demographic information.
The company needs a solution to ensure that the application does not generate biased recommendations.
Which solution will meet this requirement?
A. Use SageMaker Clarify to detect bias patterns. Collect and use additional balanced training data. Use the data to retrain the model.
B. Implement prompt engineering techniques to explicitly instruct the model to provide fair recommendations regardless of demographics.
C. Apply content filtering by using Amazon Comprehend to remove potentially biased recommendations before they reach users.
D. Create separate foundation model (FM) endpoints for each demographic group to provide specialized care recommendations.
Answer: A (LEAVE A REPLY)
Comprehensive and Detailed Explanation (AWS AI documents):
AWS Responsible AI best practices emphasize that bias should be detected, measured, mitigated, and monitored throughout the ML lifecycle, especially for sensitive domains such as healthcare. When biased outcomes are observed, AWS guidance recommends addressing bias at the data and model level, not only at the output level.
Using Amazon SageMaker Clarify aligns directly with AWS Responsible AI principles because it is designed to:
* Detect and quantify bias in datasets and model predictions across sensitive attributes such as demographic groups
* Provide pre-training and post-training bias metrics, allowing practitioners to identify where bias originates
* Support data-centric mitigation, including improving dataset balance and representativeness
After identifying bias with SageMaker Clarify, collecting additional balanced training data and retraining the model helps ensure that:
* The model learns from a more representative dataset

* Disparate treatment recommendations based on demographics are reduced
* Fairness is improved while maintaining clinical accuracy
Why the other options are not sufficient or aligned with AWS best practices:
* B. Prompt engineering can influence outputs but does not address underlying data or model bias and is not sufficient for regulated, high-risk domains like healthcare.
* C. Content filtering removes outputs after generation but does not prevent biased decision-making by the model itself.
* D. Separate FM endpoints by demographic group increases the risk of reinforcing bias and violates fairness principles rather than mitigating them.
AWS AI Study Guide References:
* AWS Responsible AI principles: Fairness and Governance
* Amazon SageMaker Clarify: bias detection and mitigation
* AWS best practices for ML in high-risk domains such as healthcare

NEW QUESTION: 77
A company wants to develop an educational game where users answer questions such as the following: "A jar contains six red, four green, and three yellow marbles. What is the probability of choosing a green marble from the jar?" Which solution meets these requirements with the LEAST operational overhead?
A. Use supervised learning to create a regression model that will predict probability.
B. Use reinforcement learning to train a model to return the probability.
C. Use code that will calculate probability by using simple rules and computations.
D. Use unsupervised learning to create a model that will estimate probability density.
Answer: C (LEAVE A REPLY)
The problem involves a simple probability calculation that can be handled efficiently by straightforward mathematical rules and computations. Using machine learning techniques would introduce unnecessary complexity and operational overhead.
Option C (Correct): "Use code that will calculate probability by using simple rules and computations": This is the correct answer because it directly solves the problem with minimal overhead, using basic probability rules.
Option A: "Use supervised learning to create a regression model" is incorrect as it overcomplicates the solution for a simple probability problem.

Option B: "Use reinforcement learning to train a model" is incorrect because reinforcement learning is not needed for a simple probability calculation.
Option D: "Use unsupervised learning to create a model" is incorrect as unsupervised learning is not applicable to this task.
AWS AI Practitioner References:
Choosing the Right Solution for AI Tasks: AWS recommends using the simplest and most efficient approach to solve a given problem, avoiding unnecessary machine learning techniques for straightforward tasks.

NEW QUESTION: 78
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 (LEAVE A REPLY)
The average response time is the correct metric for measuring the runtime efficiency of operating AI models.
* Average Response Time:
* Refers to the time taken by the model to generate an output after receiving an input. It is a key metric for evaluating the performance and efficiency of AI models in production.
* A lower average response time indicates a more efficient model that can handle queries quickly.
* Why Option C is Correct:
* Measures Runtime Efficiency: Directly indicates how fast the model processes inputs and delivers outputs, which is critical for real-time applications.
* Performance Indicator: Helps identify potential bottlenecks and optimize model performance.
* Why Other Options are Incorrect:
* A. Customer satisfaction score (CSAT): Measures customer satisfaction, not model runtime efficiency.
* B. Training time for each epoch: Measures training efficiency, not runtime efficiency during model operation.
* D. Number of training instances: Refers to data used during training, not operational efficiency.

NEW QUESTION: 79
A financial company has offices in different countries worldwide. The company requires that all API calls between generative AI applications and foundation models (FM) must not travel across the public internet.
Which AWS service should the company use?
A. AWS PrivateLink
B. Amazon
C. Amazon CloudFront

D. AWS CloudTrail

Answer: (SHOW ANSWER)

AWS PrivateLink provides private connectivity between VPCs, AWS services, and on-premises networks, ensuring traffic does not traverse the public internet.

A is correct:

"AWS PrivateLink provides private connectivity to services across VPCs, keeping API traffic off the public internet." (Reference: AWS PrivateLink Overview) B (Amazon Q) is a generative AI assistant, not a network security/control tool.

C (CloudFront) is a CDN, not for private API calls.

D (CloudTrail) is for logging and monitoring, not secure connectivity.

NEW QUESTION: 80

A company is building a mobile app for users who have a visual impairment. The app must be able to hear what users say and provide voice responses.

Which solution will meet these requirements?

A. Use a deep learning neural network to perform speech recognition.

B. Build ML models to search for patterns in numeric data.

C. Use generative AI summarization to generate human-like text.

D. Build custom models for image classification and recognition.

Answer: A (LEAVE A REPLY)

The mobile app for users with visual impairment needs to hear user speech and provide voice responses, requiring speech-to-text (speech recognition) and text-to-speech capabilities. Deep learning neural networks are widely used for speech recognition tasks, as they can effectively process and transcribe spoken language.

AWS services like Amazon Transcribe, which uses deep learning for speech recognition, can fulfill this requirement by converting user speech to text, and Amazon Polly can generate voice responses.

Exact Extract from AWS AI Documents:

From the AWS Documentation on Amazon Transcribe:

"Amazon Transcribe uses deep learning neural networks to perform automatic speech recognition (ASR), converting spoken language into text with high accuracy. This is ideal for applications requiring voice input, such as accessibility features for visually impaired users." (Source: Amazon Transcribe Developer Guide, Introduction to Amazon Transcribe) Detailed Explanation:

Option A: Use a deep learning neural network to perform speech recognition.This is the correct answer. Deep learning neural networks are the foundation of modern speech recognition systems, as used in AWS services like Amazon Transcribe. They enable the app to hear and transcribe user speech, and a service like Amazon Polly can handle voice responses, meeting the requirements.

Option B: Build ML models to search for patterns in numeric data.This option is irrelevant, as the task involves processing speech (audio data) and generating voice responses, not analyzing numeric data patterns.

Option C: Use generative AI summarization to generate human-like text.Generative AI summarization focuses on summarizing text, not processing speech orgenerating voice responses. This option does not address the core requirement of speech recognition.
Option D: Build custom models for image classification and recognition.Image classification and recognition are unrelated to processing speech or generating voice responses, making this option incorrect for an app focused on audio interaction.
References:
Amazon Transcribe Developer Guide: Introduction to Amazon Transcribe
(https://docs.aws.amazon.com
/transcribe/latest/dg/what-is.html)
Amazon Polly Developer Guide: Text-to-Speech Overview
(https://docs.aws.amazon.com/polly/latest/dg
/what-is.html)
AWS AI Practitioner Learning Path: Module on Speech Recognition and Synthesis

NEW QUESTION: 81
Which option is a disadvantage of using generative AI models in production systems?
A. Possible high accuracy and reliability
B. Deterministic and consistent behavior
C. Negligible computational resource requirements
D. Hallucinations and inaccuracies
Answer: D (LEAVE A REPLY)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
A known limitation of generative AI models is their tendency to hallucinate, meaning they may generate plausible but incorrect or fabricated information.
AWS generative AI guidance highlights:
Non-deterministic behavior
Risk of incorrect outputs
Need for validation, monitoring, and guardrails in production systems
Why the other options are incorrect:
High accuracy (A) is an advantage, not a disadvantage.
Deterministic behavior (B) is generally not true for generative models.
Negligible resource usage (C) is incorrect; generative models are resource-intensive.
AWS AI document references:
Generative AI Risks and Mitigations
Responsible Deployment of Foundation Models
Amazon Bedrock Guardrails Guidance

NEW QUESTION: 82
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 (LEAVE A REPLY)
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:
* This strategy, known as few-shot learning, involves giving the model examples of input-output pairs (e.g., text passages with their sentiment labels) to help it understand the task context.
* It allows the model to learn from these examples and apply the learned pattern to classify new text passages correctly.
* Why Option A is Correct:
* Guides the Model: Providing labeled examples teaches the model how to perform sentiment analysis effectively, increasing accuracy.
* Contextual Relevance: Aligns the model's responses to the specific task of classifying sentiment.
* Why Other Options are Incorrect:
* B. Detailed explanation of sentiment analysis: Unnecessary for the model's operation; it requires examples, not explanations.
* C. New text passage without context: Provides no guidance or learning context for the model.
* D. Unrelated task examples: Would confuse the model and lead to inaccurate results.

NEW QUESTION: 83
An online learning company with large volumes of educational materials wants to use enterprise search.
Which AWS service meets these requirements?
A. Amazon Comprehend
B. Amazon Textract
C. Amazon Kendra
D. Amazon Personalize
Answer: C (LEAVE A REPLY)
Amazon Kendra is an enterprise search service that uses machine learning to index and search unstructured data such as documents, manuals, and course materials.
Amazon Comprehend performs NLP tasks like sentiment analysis and entity recognition, not enterprise search.
Amazon Textract extracts structured text from scanned docs and PDFs.

Amazon Personalize builds recommendation systems, not search engines.
Reference:
AWS Documentation - Amazon Kendra

NEW QUESTION: 84
A medical company wants to develop an AI application that can access structured patient records, extract relevant information, and generate concise summaries.
Which solution will meet these requirements?
A. Use Amazon Comprehend Medical to extract relevant medical entities and relationships. Apply rule- based logic to structure and format summaries.
B. Use Amazon Personalize to analyze patient engagement patterns. Integrate the output with a general purpose text summarization tool.
C. Use Amazon Textract to convert scanned documents into digital text. Design a keyword extraction system to generate summaries.
D. Implement Amazon Kendra to provide a searchable index for medical records. Use a template-based system to format summaries.
Answer: A (LEAVE A REPLY)
Amazon Comprehend Medical is designed for processing medical records and extracting key clinical entities, useful for summaries. Per the AWS Comprehend Medical documentation: "Amazon Comprehend Medical enables extraction of relevant medical information from unstructured clinical text such as medications, conditions, and relationships, making it ideal for summarization tasks."

NEW QUESTION: 85
A company wants to implement a generative AI solution to improve its marketing operations. The company wants to increase its revenue in the next 6 months.
Which approach will meet these requirements?
A. Immediately start training a custom FM by using the company's existing data.
B. Conduct stakeholder interviews to refine use cases and set measurable goals.
C. Implement a prebuilt AI assistant solution and measure its impact on customer satisfaction.
D. Analyze industry AI implementations and replicate the most successful features.
Answer: B (LEAVE A REPLY)
The correct answer is B, because AWS recommends that any generative AI initiative begin with use-case clarification, stakeholder alignment, and measurable business objectives. According to the AWS Generative AI Adoption Framework and the AWS Machine Learning Journey guidelines, successful AI projects start with defining specific business outcomes-such as revenue growth, reduced time-to-market, or improved campaign efficiency. Conducting stakeholder interviews ensures the organization identifies the highest-value marketing use cases, such as personalized campaigns or content generation, and sets KPIs to measure revenue impact within the target timeframe. AWS documentation emphasizes avoiding premature model training (option A), which is costly and unnecessary without validated requirements. Implementing a prebuilt assistant

(option C) may help operations but does not guarantee alignment with the core revenue objective. Replicating competitor features (option D) may lack strategic alignment. Therefore, refining use cases and measurable goals is the foundational AWS-recommended approach for a targeted revenue-driven generative AI initiative.
Referenced AWS Documentation:
AWS Generative AI Adoption Framework
AWS ML Business Value and Use-Case Identification Guidelines

NEW QUESTION: 86
A company uses an open-source pre-trained model to analyze user sentiment for a newly released product.
Which action must the company perform, according to MLOps best practices?
A. Use deep learning to perform hyperparameter tuning.
B. Collect user reviews and label each review as positive or negative.
C. Continuously monitor outputs in production.
D. Perform feature engineering on the input dataset.
Answer: (SHOW ANSWER)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
According to MLOps best practices, once an ML model is deployed to production-especially an open-source pre-trained model-the organization must continuously monitor model outputs to ensure:
Model predictions remain accurate over time
No performance degradation due to data drift or concept drift
Outputs remain aligned with business and ethical expectations
AWS MLOps guidance emphasizes monitoring in production as a mandatory step for maintaining model reliability and governance.
Why the other options are incorrect:
A (Hyperparameter tuning) is optional and model-dependent.
B (Labeling data) is required only when training or fine-tuning, not when using a pre-trained model directly.
D (Feature engineering) is less relevant for modern pre-trained NLP models.
AWS AI document references:
MLOps Best Practices on AWS
Amazon SageMaker Model Monitoring
Operationalizing Machine Learning on AWS

NEW QUESTION: 87
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 (LEAVE A REPLY)

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: 88

A bank is fine-tuning a large language model (LLM) on Amazon Bedrock to assist customers with questions about their loans. The bank wants to ensure that the model does not reveal any private customer data.

Which solution meets these requirements?

A. Use Amazon Bedrock Guardrails.

B. Remove personally identifiable information (PII) from the customer data before fine-tuning the LLM.

C. Increase the Top-K parameter of the LLM.

D. Store customer data in Amazon S3. Encrypt the data before fine-tuning the LLM.

Answer: (SHOW ANSWER)

The goal is to prevent a fine-tuned large language model (LLM) on Amazon Bedrock from revealing private customer data. Let's analyze the options:

A). Amazon Bedrock Guardrails: Guardrails in Amazon Bedrock allow users to define policies to filter harmful or sensitive content in model inputs and outputs. While useful for real-time content moderation, they do not address the risk of private data being embedded in the model during fine-tuning, as the model could still memorize sensitive information.

B). Remove personally identifiable information (PII) from the customer data before fine-tuning the LLM:

Removing PII (e.g., names, addresses, account numbers) from the training dataset ensures that the model does not learn or memorize sensitive customer data, reducing the risk of data leakage. This is a proactive and effective approach to data privacy during model training.

C). Increase the Top-K parameter of the LLM: The Top-K parameter controls the randomness of the model's output by limiting the number of tokens considered during generation. Adjusting this parameter affects output diversity but does not address the privacy of customer data embedded in the model.

D). Store customer data in Amazon S3. Encrypt the data before fine-tuning the LLM: Encrypting data in Amazon S3 protects data at rest and in transit, but during fine-tuning, the data is

decrypted and used to train the model. If PII is present, the model could still learn and potentially expose it, so encryption alone does not solve the problem.
Exact Extract Reference: AWS emphasizes data privacy in AI/ML workflows, stating, "To protect sensitive data, you can preprocess datasets to remove personally identifiable information (PII) before using them for model training. This reduces the risk of models inadvertently learning or exposing sensitive information." (Source: AWS Best Practices for Responsible AI, https://aws.amazon.com/machine-learning/responsible-ai/).
Additionally, the Amazon Bedrock documentation notes that users are responsible for ensuring compliance with data privacy regulations during fine-tuning (https://docs.aws.amazon.com/bedrock/latest/userguide/model-customization.html).
Removing PII before fine-tuning is the most direct and effective way to prevent the model from revealing private customer data, making B the correct answer.
References:
AWS Bedrock Documentation: Model Customization (https://docs.aws.amazon.com/bedrock/latest/userguide/model-customization.html)
AWS Responsible AI Best Practices (https://aws.amazon.com/machine-learning/responsible-ai/)
AWS AI Practitioner Study Guide (emphasis on data privacy in LLM fine-tuning)

NEW QUESTION: 89
A financial company uses AWS to host its generative AI models. The company must generate reports to show adherence to international regulations for handling sensitive customer data.
A. Amazon Macie
B. AWS Artifact
C. AWS Secrets Manager
D. AWS Config
Answer: B (LEAVE A REPLY)
* AWS Artifact provides compliance reports and certifications (ISO, SOC, GDPR-related documentation) to prove regulatory adherence.

NEW QUESTION: 90
A global financial company has developed an ML application to analyze stock market data and provide stock market trends. The company wants to continuously monitor the application development phases and ensure that company policies and industry regulations are followed.
Which AWS services will help the company assess compliance with these requirements? (Select TWO.)
A. AWS Audit Manager
B. AWS Config
C. Amazon Inspector
D. Amazon CloudWatch

E. AWS CloudTrail

Answer: A,B,E (LEAVE A REPLY)

Comprehensive and Detailed Explanation From Exact AWS AI documents:

To assess compliance, governance, and regulatory adherence throughout the ML application lifecycle, AWS recommends using services that provide audit evidence, configuration tracking, and activity logging.

AWS Audit Manager continuously evaluates AWS resource usage against predefined regulatory frameworks and internal controls, helping organizations demonstrate compliance.

AWS Config enables continuous monitoring and recording of AWS resource configurations, allowing organizations to assess whether resources comply with company policies.

AWS CloudTrail provides a complete audit trail of API calls and user actions, supporting traceability and regulatory investigations.

Together, these services provide end-to-end compliance assessment and monitoring across development and operational phases.

Why the other options are incorrect:

Amazon Inspector (C) focuses on vulnerability scanning, not governance or regulatory compliance.

Amazon CloudWatch (D) is primarily used for performance monitoring and logging, not compliance assessment.

AWS AI document references (for exact extracts):

AWS Governance, Risk, and Compliance (GRC) Overview

Auditing AI and ML Workloads on AWS

Security, Compliance, and Monitoring in AWS AI Systems

NEW QUESTION: 91

A food service company wants to collect a dataset to predict customer food preferences. The company wants to ensure that the food preferences of all demographics are included in the data.

A. Accuracy

B. Diversity

C. Recency bias

D. Reliability

Answer: B (LEAVE A REPLY)

Diversity in datasets ensures representation of all demographics, reducing bias and improving fairness.

Accuracy is model performance.

Recency bias skews towards recent data.

Reliability measures consistency of results, not representation.

Reference:

AWS Responsible AI Guidelines - Data Diversity

NEW QUESTION: 92
An AI practitioner trained a custom model on Amazon Bedrock by using a training dataset that contains confidential data. The AI practitioner wants to ensure that the custom model does not generate inference responses based on confidential data.
How should the AI practitioner prevent responses based on confidential data?
A. Delete the custom model. Remove the confidential data from the training dataset. Retrain the custom model.
B. Mask the confidential data in the inference responses by using dynamic data masking.
C. Encrypt the confidential data in the inference responses by using Amazon SageMaker.
D. Encrypt the confidential data in the custom model by using AWS Key Management Service (AWS KMS).
Answer: A (LEAVE A REPLY)
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 beingtrained 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: 93
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. Set up Amazon CloudWatch alarms for notification of policy violations.
D. Implement Amazon SageMaker Model Monitor to detect data drift and receive alerts when model quality degrades.
Answer: (SHOW ANSWER)
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: 94
A company is building a custom AI solution in Amazon SageMaker Studio to analyze financial transactions for fraudulent activity in real time. The company needs to ensure that the connectivity from SageMaker Studio to Amazon Bedrock traverses the company's VPC.
Which solution meets these requirements?
A. Configure AWS Identity and Access Management (IAM) roles and policies for SageMaker Studio to access Amazon Bedrock.
B. Configure Amazon Macie to proxy requests from SageMaker Studio to Amazon Bedrock.
C. Configure AWS PrivateLink endpoints for the Amazon Bedrock API endpoints in the VPC that SageMaker Studio is connected to.
D. Configure a new VPC for the Amazon Bedrock usage. Register the VPCs as peers.
Answer: C (LEAVE A REPLY)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
AWS PrivateLink enables private connectivity between AWS services through VPC endpoints, ensuring traffic does not traverse the public internet.
AWS guidance recommends PrivateLink for:
Secure, private service access
Regulatory and compliance requirements
VPC-based architecture
Why the other options are incorrect:
IAM (A) controls access, not network routing.
Macie (B) is a data security service.
VPC peering (D) is not required for Bedrock access.
AWS AI document references:
Amazon Bedrock Networking and Security
Private Connectivity with AWS PrivateLink
Secure AI Architectures on AWS

NEW QUESTION: 95

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 (LEAVE A REPLY)
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: 96
A company is using Amazon Bedrock Agents to build an application to automate business workflows.
A. To invoke foundation models (FMs) to process visual, audio, and text inputs
B. To enhance foundation models (FMs) with a prompting strategy
C. To provide users with full control of querying external data sources and APIs
D. To evaluate user inputs and orchestrate actions for multiple tasks
Answer: D (LEAVE A REPLY)
The correct answer is D. Amazon Bedrock Agents are used to orchestrate and execute complex workflows by connecting foundation models with APIs, databases, and tools. According to AWS documentation, agents interpret user inputs, plan the necessary steps, call external APIs or systems, and return structured results. This allows the model to go beyond text generation into full automation workflows-such as booking tasks, querying internal systems, or summarizing reports. Option A describes multi-modal models, B refers to prompt tuning, and C misstates

control delegation; agents act autonomously based on model reasoning. Thus, Bedrock Agents function as intelligent orchestrators, handling multi-step task execution through integrated tool use.
Referenced AWS AI/ML Documents and Study Guides:
Amazon Bedrock Developer Guide - Agents Overview
AWS Generative AI Best Practices - Workflow Orchestration

NEW QUESTION: 97
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. Use the fine-tuned model to process user prompts.
D. Upload PDF documents to an Amazon Bedrock knowledge base. Use the knowledge base to provide context when users submit prompts to Amazon Bedrock.
Answer: A (LEAVE A REPLY)
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: 98
A financial company is developing a generative AI application for loan approval decisions. The company needs the application output to be responsible and fair.
Which solution meets these requirements?
A. Review the training data to check for biases. Include data from all demographics in the training data.
B. Use a deep learning model with many hidden layers.
C. Keep the model's decision-making process a secret to protect proprietary algorithms.
D. Continuously monitor the model's performance on a static test dataset.
Answer: A (LEAVE A REPLY)
The correct answer is A because ensuring responsibility and fairness in ML begins with bias detection in the training data. Including a balanced representation of all demographics ensures the model learns fairly across different groups, which is critical in regulated industries like finance.
From AWS documentation:
"A key principle of responsible AI is building models that do not propagate or amplify bias. Fairness begins with training data. Reviewing and augmenting data for representation is essential." Explanation of other options:
B). The number of hidden layers doesn't inherently improve fairness or responsibility.
C). Keeping decisions opaque violates explainability principles in responsible AI.
D). A static dataset can become outdated and may not reflect real-world shifts, which limits fairness assessment over time.
Referenced AWS AI/ML Documents and Study Guides:
Amazon SageMaker Clarify Documentation - Bias Detection and Explainability AWS Responsible AI Guidelines AWS ML Specialty Study Guide - Fairness and Governance

NEW QUESTION: 99
A company that streams media is selecting an Amazon Nova foundation model (FM) to process documents and images. The company is comparing Nova Micro and Nova Lite. The company wants to minimize costs.
A. Nova Micro uses transformer-based architectures. Nova Lite does not use transformer-based architectures.
B. Nova Micro supports only text data. Nova Lite is optimized for numerical data.
C. Nova Micro supports only text. Nova Lite supports images, videos, and text.
D. Nova Micro runs only on CPUs. Nova Lite runs only on GPUs.
Answer: (SHOW ANSWER)
The correct answer is C, because Amazon Nova Micro is a smaller, lower-cost foundation model that is text- only, while Nova Lite is a more capable multimodal model that supports images, videos, and text. According to AWS Bedrock documentation, the Nova model family includes variants that differ in capability and cost.

Nova Micro is optimized for lightweight text-based tasks, including summarization, question answering, and basic reasoning. This makes it cheaper to operate and well-suited for cost-sensitive workloads. Nova Lite, on the other hand, is a multimodal FM that can analyze documents, screenshots, photographs, charts, and videos, making it ideal for media companies requiring cross-format understanding. AWS clarifies that both Micro and Lite use transformer-based architectures, and run on managed infrastructure that abstracts hardware considerations. Therefore, the main differentiator is capability-and Nova Micro being text-only is the more cost-effective option. Nova Lite is appropriate only when image or video analysis is required.
Referenced AWS Documentation:
* Amazon Bedrock - Nova Model Family Overview
* AWS Generative AI Model Selection Guide

NEW QUESTION: 100
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 (LEAVE A REPLY)
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:
* Can process and integrate different types of data (e.g., text and images) into a common representation space, enabling a unified search capability.
* Suitable for applications where queries or content involve multiple data modalities.
* Why Option A is Correct:
* Handles Multiple Modalities: Supports both text and image data, aligning with the application's requirement.
* Improves Search Relevance: Allows for more accurate and relevant search results across different types of input data.
* Why Other Options are Incorrect:
* B. Text embedding model: Only handles text data, not images.
* C. Multi-modal generation model: Focuses on generating outputs rather than embedding for search tasks.
* D. Image generation model: Only handles image data, not suitable for text queries.

NEW QUESTION: 101

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: (SHOW ANSWER)
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:
* Few-shot learning aims to enable the model to learn from a small number of examples. For intent detection, the model needs to understand the relationship between user messages and the intended action or meaning.
* Providing examples of user messages and the correct user intents allows the model to learn patterns in the phrasing or language that corresponds to each intent.
* Why Option C is Correct:
* User Messages and Intents: These examples directly teach the model how to map a user's input to the appropriate intent, which is the goal of intent detection in chatbots.
* Improves Accuracy: By using few-shot learning with these examples, the model can generalize better from limited data, improving intent detection.
* Why Other Options are Incorrect:
* A. Pairs of chatbot responses and correct user intents: Incorrect because it does not focus on user input but rather on outputs.
* B. Pairs of user messages and correct chatbot responses: This would be useful for response generation, not intent detection.
* D. Pairs of user intents and correct chatbot responses: Again, this is not aligned with detecting intents but with generating responses.

NEW QUESTION: 102
A company is developing an AI solution to help make hiring decisions.
Which strategy complies with AWS guidance for responsible AI?
A. Use the AI solution to make final hiring decisions without human review.
B. Train the AI solution exclusively on data from previous successful hires.
C. Test the AI solution to ensure that it does not discriminate against any protected groups.
D. Keep the AI decision-making process confidential to maintain a competitive advantage.
Answer: C (LEAVE A REPLY)
The correct answer is C - Test the AI solution to ensure that it does not discriminate against any protected groups. According to AWS Responsible AI principles, fairness and bias mitigation are

essential when AI is used for high-impact decisions such as hiring. AWS documentation emphasizes evaluating datasets, model outputs, and demographic performance to ensure that AI systems do not reinforce or reproduce discriminatory patterns. Services such as Amazon SageMaker Clarify support automated bias detection and explainability, helping teams identify and mitigate unwanted correlations in training data or model predictions. Option A violates AWS guidance, as human-in-the-loop review is required for sensitive decisions. Option B risks amplifying historical bias because training on only "successful" hires can create feedback loops. Option D contradicts transparency principles, which AWS states are crucial for accountability in regulated or ethical decision-making domains. Therefore, rigorous fairness testing aligns with AWS's recommended practices for responsible AI in hiring workflows.
Referenced AWS Documentation:
* AWS Responsible AI Whitepaper - Fairness and Bias Mitigation
* Amazon SageMaker Clarify Documentation

NEW QUESTION: 103
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 (LEAVE A REPLY)
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: 104
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: (SHOW ANSWER)
The scenario where a student copies content from generative AI to write essays represents the challenge of plagiarism in responsible AI use.
* Plagiarism:
* Occurs when someone uses content generated by AI (or any source) without proper attribution, claiming it as their own.
* This is a key challenge with generative AI models, which can produce human-like text that might be misused for academic or other purposes.
* Why Option C is Correct:
* Represents Unauthorized Use: Copying content directly from AI without attribution is a clear case of plagiarism.
* Ethical Concern: Highlights the ethical considerations around using AI-generated content responsibly.
* Why Other Options are Incorrect:
* A. Toxicity: Refers to harmful or offensive content generation, not content copying.
* B. Hallucinations: When AI generates incorrect or nonsensical information, not plagiarism.
* D. Privacy: Involves the misuse or exposure of personal information, not copying content.

NEW QUESTION: 105
Which option is an example of unsupervised learning?
A. Clustering data points into groups based on their similarity
B. Training a model to recognize images of animals
C. Predicting the price of a house based on the house's features
D. Generating human-like text based on a given prompt
Answer: (SHOW ANSWER)
Unsupervised learning involves discovering hidden patterns without labeled data. Example: clustering.
Image recognition (B) is supervised learning.
House price prediction (C) is regression (supervised).

NEW QUESTION: 106
A company wants to set up private access to Amazon Bedrock APIs from the company's AWS account. The company also wants to protect its data from internet exposure.
A. Use Amazon CloudFront to restrict access to the company's private content
B. Use AWS Glue to set up data encryption across the company's data catalog

C. Use AWS Lake Formation to manage centralized data governance and cross-account data sharing
D. Use AWS PrivateLink to configure a private connection between the company's VPC and Amazon Bedrock
Answer: D (LEAVE A REPLY)
AWS PrivateLink enables private connectivity between your VPC and supported AWS services (like Amazon Bedrock) without sending traffic over the public internet.
CloudFront (A) is for CDN and content delivery, not private service connections.
AWS Glue (B) is for ETL/data catalog, not networking.
Lake Formation (C) provides governance for data lakes, not API network isolation.
Reference:
AWS Documentation - Access Amazon Bedrock with PrivateLink

NEW QUESTION: 107
A company is building an AI application to automate business processes. The company uses a foundation model (FM) to support the application.
The company needs to select datasets to assess the quality of the AI model's behavior.
Which type of datasets will meet these requirements?
A. Curated datasets that have had all outliers and correlations removed
B. Synthetic datasets that have been generated by the newest FM
C. Diverse datasets that cover various use cases and usage scenarios
D. Randomized datasets that have arbitrary features and skewed distributions
Answer: C (LEAVE A REPLY)
Comprehensive and Detailed Explanation (AWS AI documents):
AWS Responsible AI and generative AI evaluation guidance emphasizes that assessing the quality and behavior of a foundation model requires representative and diverse evaluation datasets. These datasets should reflect real-world usage patterns, edge cases, and multiple business scenarios to properly evaluate how the model behaves in production.
Using diverse datasets that cover various use cases and usage scenarios allows organizations to:
* Evaluate robustness and generalization of the FM
* Identify failure modes, bias, and unsafe behavior across different inputs

* Validate that the model performs consistently across business workflows Why the other options are incorrect:
* A may remove meaningful real-world patterns and does not reflect realistic usage.
* B risks reinforcing model biases and does not provide independent evaluation.
* D does not reflect real or meaningful business scenarios and can distort evaluation results.
AWS AI Study Guide References:
* AWS Responsible AI evaluation practices
* AWS guidance on foundation model testing and validation

NEW QUESTION: 108
An education company wants to build a private tutor application. The application will give users the ability to enter text or provide a picture of a question. The application will respond with a written answer and an explanation of the written answer.
Which model type meets these requirements?
A. Computer vision model
B. Multimodal LLM
C. Diffusion model
D. Text-to-speech model
Answer: (SHOW ANSWER)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
A multimodal large language model (LLM) can:
Accept both text and image inputs
Understand visual and textual context
Generate coherent written explanations
AWS generative AI guidance positions multimodal LLMs as the best choice for applications requiring cross- modal understanding and text generation.
Why the other options are incorrect:
Computer vision (A) does not generate text explanations.
Diffusion models (C) generate images.
Text-to-speech (D) converts text to audio.
AWS AI document references:
Multimodal Foundation Models on AWS
Building AI Tutors with Generative Models
Text and Image Understanding with LLMs

NEW QUESTION: 109
A financial company is training a generative AI model to predict outcomes of loan applications. The training dataset is small. The dataset categorizes loan applicants as "younger-aged," "middle-aged," or "older-aged." Most individuals in the dataset are characterized as "middle-aged." The company removes the age range feature from the training dataset.
Which model behavior will likely happen as a result of this change to the dataset?

A. The model will inaccurately predict outcomes for younger and older age groups.
B. The model will require less training data.
C. The model will predict accurate outcomes for only younger age groups.
D. The model will accurately predict outcomes for all ages.
Answer: A (LEAVE A REPLY)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
Removing a relevant feature from a small and imbalanced dataset can reduce the model's ability to:
Learn meaningful distinctions between groups
Generalize well for underrepresented categories
AWS Responsible AI guidance explains that insufficient or removed features can lead to reduced predictive performance, especially for minority groups.
Why the other options are incorrect:
B is unrelated to model behavior.
C is unsupported by the scenario.
D is unlikely given data imbalance and feature removal.
AWS AI document references:
Bias and Data Representation in ML
Responsible Feature Selection
Dataset Imbalance and Model Performance

NEW QUESTION: 110
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 (LEAVE A REPLY)
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:
* Adjusting and refining the prompt allows for better control over the FM's outputs, ensuring they align with the desired tone and style.
* This iterative process involves testing different prompts and modifying them based on the model's responses to achieve the desired outcome.
* Why Option C is Correct:
* Directly Influences Output Quality: Allows for fine-tuning of the model's responses to match the company's tone.

* Cost-Effective: Does not require modifying the model itself, only the inputs to it.
* Why Other Options are Incorrect:
* A. Low limit on tokens: Limits response length but not the adherence to company tone.
* B. Batch inferencing: Relates to processing multiple inputs, not controlling response tone.
* D. Higher temperature: Increases randomness in responses, which could deviate from the desired tone.

NEW QUESTION: 111
A company wants to use AWS services to build an AI assistant for internal company use. The AI assistant's responses must reference internal documentation. The company stores internal documentation as PDF, CSV, and image files.
Which solution will meet these requirements with the LEAST operational overhead?
A. Use Amazon SageMaker AI to fine-tune a model.
B. Use Amazon Bedrock Knowledge Bases to create a knowledge base.
C. Configure a guardrail in Amazon Bedrock Guardrails.
D. Select a pre-trained model from Amazon SageMaker JumpStart.
Answer: B (LEAVE A REPLY)
The best solution is Amazon Bedrock Knowledge Bases, which allows for the seamless integration of structured and unstructured internal documents-such as PDFs, CSVs, and extracted image text-into a retrieval-augmented generation (RAG) pipeline. According to AWS documentation, Bedrock Knowledge Bases offer a no-code or low-code setup to link your enterprise data with foundation models for context-aware responses, without needing to fine-tune or retrain models. The system indexes documents in an Amazon S3 bucket, creates embeddings, and stores them in a vector store. At inference time, the model retrieves relevant context and incorporates it into its response. This approach provides dynamic and up-to-date responses while maintaining data privacy, with minimal operational overhead. Unlike fine-tuning or building a model from scratch in SageMaker, which requires considerable compute resources and model management, Bedrock Knowledge Bases are serverless and easy to configure. It is designed exactly for internal knowledge AI assistants.
Referenced AWS AI/ML Documents and Study Guides:
Amazon Bedrock Developer Guide - Knowledge Bases
AWS Generative AI Best Practices - RAG Patterns for Enterprise Search

NEW QUESTION: 112
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 (LEAVE A REPLY)

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:

This metric measures the time taken to handle a customer call or query. A successful LLM chatbot should reduce the call duration by efficiently providing answers, minimizing the need for human intervention.

By decreasing the average call duration, the company can improve call center efficiency, reduce costs, and enhance the user experience.

Why Option B is Correct:

Direct Impact: The objective aligns directly with the goal of reducing the number of actions call center employees must take.

Operational Efficiency: Reducing call duration is a clear indicator of the chatbot's effectiveness in assisting customers without human help.

Why Other Options are Incorrect:

A . Website engagement rate: Is unrelated to call center operations.

C . Corporate social responsibility: Does not relate to call center efficiency.

D . Regulatory compliance: Is important but does not measure the effectiveness of a chatbot in reducing employee actions.

NEW QUESTION: 113

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

Which technique will solve the problem?

A. Data augmentation for imbalanced classes

B. Model monitoring for class distribution

C. Retrieval Augmented Generation (RAG)

D. Watermark detection for images

Answer: A (LEAVE A REPLY)

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

* Data Augmentation for Imbalanced Classes:

* Involves generating new data samples by modifying existing ones, such as flipping, rotating, or cropping images, to balance the representation of different classes.

* Helps mitigate bias by ensuring that the training data is more representative of diverse characteristics and scenarios.

* Why Option A is Correct:

* Balances Data Distribution: Addresses class imbalance by augmenting underrepresented classes, which reduces bias in the model.

* Improves Model Fairness: Ensures that the model is exposed to a more diverse set of training examples, promoting fairness in image generation.
* Why Other Options are Incorrect:
* B. Model monitoring for class distribution: Helps identify bias but does not actively correct it.
* C. Retrieval Augmented Generation (RAG): Involves combining retrieval and generation but is unrelated to mitigating bias in image generation.
* D. Watermark detection for images: Detects watermarks in images, not a technique for addressing bias.

NEW QUESTION: 114
A company needs an automated solution to group its customers into multiple categories. The company does not want to manually define the categories. Which ML technique should the company use?
A. Classification
B. Linear regression
C. Logistic regression
D. Clustering
Answer: D (LEAVE A REPLY)
Comprehensive and Detailed Explanation from AWS AI Documents:
* Classification requires predefined labels (categories). The company explicitly does not want to define categories.
* Regression (linear or logistic) predicts numerical values or probabilities, not groups.
* Clustering is an unsupervised learning technique that groups similar data points together based on their features without needing labeled categories.
AWS defines clustering as:
"Clustering is an unsupervised machine learning algorithm that automatically groups data points into clusters based on their similarities." This makes Clustering the correct choice for segmenting customers into groups without predefined labels.
Reference:
AWS ML Glossary - Clustering

NEW QUESTION: 115
A company stores millions of PDF documents in an Amazon S3 bucket. The company needs to extract the text from the PDFs, generate summaries of the text, and index the summaries for fast searching.
Which combination of AWS services will meet these requirements? (Select TWO.)
A. Amazon Translate
B. Amazon Bedrock
C. Amazon Transcribe
D. Amazon Polly
E. Amazon Textract

Answer: (SHOW ANSWER)
Amazon Textract (E) automatically extracts text and structured data from scanned documents, such as PDFs.
Amazon Bedrock (B) offers access to LLMs (such as Amazon Titan or Anthropic Claude) for tasks like summarization and generating embeddings for search.
Workflow:
Amazon Textract extracts text from PDFs in S3.
Amazon Bedrock LLMs summarize the extracted text.
(Optional: Summaries can be indexed using Amazon OpenSearch or another search solution.) A (Translate) is for language translation, not extraction or summarization.
C (Transcribe) is for audio to text, not PDFs.
D (Polly) is for text-to-speech.
"Amazon Textract extracts text, forms, and tables from scanned documents... Bedrock provides generative AI models to perform summarization and other text generation tasks." (Reference: Amazon Textract, Amazon Bedrock, AWS GenAI RAG Reference)

NEW QUESTION: 116
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: (SHOW ANSWER)
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:
* Extracts text, forms, and tables from scanned documents accurately.
* Ideal for automating the process of converting PDF resumes into plain text format.
* Why Option A is Correct:
* Automation of Text Extraction: Textract is designed to handle large volumes of documents and convert them into machine-readable text, perfect for the company's need.
* Scalability and Efficiency: Supports scalability to handle a growing volume of resumes as the company expands.
* Why Other Options are Incorrect:
* B. Amazon Personalize: Used for creating personalized recommendations, not for text extraction.

* C. Amazon Lex: A service for building conversational interfaces, not for processing documents.
* D. Amazon Transcribe: Used for converting speech to text, not for extracting text from documents.

NEW QUESTION: 117
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 (LEAVE A REPLY)
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: 118
Which AW5 service makes foundation models (FMs) available to help users build and scale generative AI applications?
A. Amazon Q Developer
B. Amazon Bedrock
C. Amazon Kendra

D. Amazon Comprehend

Answer: B (LEAVE A REPLY)

Amazon Bedrock is a fully managed service that provides access to foundation models (FMs) from various providers, enabling users to build and scale generative AI applications. It simplifies the process of integrating FMs into applications for tasks like text generation, chatbots, and more.

Exact Extract from AWS AI Documents:

From the AWS Bedrock User Guide:

"Amazon Bedrock is a fully managed service that makes foundation models (FMs) from leading AI providers available through a single API, enabling developers to build and scale generative AI applications with ease." (Source: AWS Bedrock User Guide, Introduction to Amazon Bedrock)

Detailed Explanation:

* Option A: Amazon Q DeveloperAmazon Q Developer is an AI-powered assistant for coding and AWS service guidance, not a service for hosting or providing foundation models.
* Option B: Amazon BedrockThis is the correct answer. Amazon Bedrock provides access to foundation models, making it the primary service for building and scaling generative AI applications.
* Option C: Amazon KendraAmazon Kendra is an intelligent search service powered by machine learning, not a service for providing foundation models or building generative AI applications.
* Option D: Amazon ComprehendAmazon Comprehend is an NLP service for text analysis tasks like sentiment analysis, not for providing foundation models or supporting generative AI.

References:

AWS Bedrock User Guide: Introduction to Amazon Bedrock (https://docs.aws.amazon.com/bedrock/latest/userguide/what-is-bedrock.html)

AWS AI Practitioner Learning Path: Module on Generative AI Services

AWS Documentation: Generative AI on AWS (https://aws.amazon.com/generative-ai/)

NEW QUESTION: 119

A company has documents that are missing some words because of a database error. The company wants to build an ML model that can suggest potential words to fill in the missing text. Which type of model meets this requirement?

A. Topic modeling

B. Clustering models

C. Prescriptive ML models

D. BERT-based models

Answer: D (LEAVE A REPLY)

BERT-based models (Bidirectional Encoder Representations from Transformers) are suitable for tasks that involve understanding the context of words in a sentence and suggesting missing words. These models use bidirectional training, which considers the context from both directions (left and right of the missing word) to predict the appropriate word to fill in the gaps.

* BERT-based Models:

* BERT is a pre-trained transformer model designed for natural language understanding tasks, including text completion, where certain words are missing.
* It excels at understanding context and relationships between words in a sentence, making it ideal for suggesting potential words to fill in missing text.
* Why Option D is Correct:
* Contextual Understanding: BERT uses its bidirectional training to understand the context around missing words, making it highly accurate in suggesting suitable replacements.
* Text Completion Capability: BERT's architecture is explicitly designed for tasks like masked language modeling, where certain words in a text are masked (or missing), and the model predicts the missing words.
* Why Other Options are Incorrect:
* A. Topic modeling: Focuses on identifying topics in a text corpus, not on predicting missing words.
* B. Clustering models: Group similar data points together, which is not suitable for predicting missing text.
* C. Prescriptive ML models: Focus on providing recommendations based on data analysis, not on natural language processing tasks like filling in missing text.

NEW QUESTION: 120
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 (LEAVE A REPLY)
Transfer learning is the correct strategy for adapting pre-trained models for new, related tasks without creating models from scratch.
* Transfer Learning:
* Involves taking a pre-trained model and fine-tuning it on a new dataset for a related task.
* This approach is efficient because it leverages existing knowledge from a model trained on a large dataset, requiring less data and computational resources than training a new model from scratch.
* Why Option B is Correct:
* Adaptation of Pre-trained Models: Allows for adapting existing models to new tasks, which aligns with the company's goal of not starting from scratch.
* Efficiency and Speed: Speeds up the model development process by building on the knowledge of pre-trained models.
* Why Other Options are Incorrect:

* A. Increase the number of epochs: Does not address the strategy of reusing pre-trained models.
* C. Decrease the number of epochs: Similarly, does not apply to adapting pre-trained models.
* D. Use unsupervised learning: Does not involve using pre-trained models for new tasks.

NEW QUESTION: 121
An ecommerce company wants to improve search engine recommendations by customizing the results for each user of the company's ecommerce platform. Which AWS service meets these requirements?
A. Amazon Personalize
B. Amazon Kendra
C. Amazon Rekognition
D. Amazon Transcribe
Answer: A (LEAVE A REPLY)
The ecommerce company wants to improve search engine recommendations by customizing results for each user. Amazon Personalize is a machine learning service that enables personalized recommendations, tailoring search results or product suggestions based on individual user behavior and preferences, making it the best fit for this requirement.
Exact Extract from AWS AI Documents:
From the Amazon Personalize Developer Guide:
"Amazon Personalize enables developers to build applications with personalized recommendations, such as customized search results or product suggestions, by analyzing user behavior and preferences to deliver tailored experiences." (Source: Amazon Personalize Developer Guide, Introduction to Amazon Personalize) Detailed Explanation:
* Option A: Amazon PersonalizeThis is the correct answer. Amazon Personalize specializes in creating personalized recommendations, ideal for customizing search results for each user on an ecommerce platform.
* Option B: Amazon KendraAmazon Kendra is an intelligent search service for enterprise data, focusing on retrieving relevant documents or answers, not on personalizing search results for individual users.
* Option C: Amazon RekognitionAmazon Rekognition is for image and video analysis, such as object detection or facial recognition, and is unrelated to search engine recommendations.
* Option D: Amazon TranscribeAmazon Transcribe converts speech to text, which is not relevant for improving search engine recommendations.
References:
Amazon Personalize Developer Guide: Introduction to Amazon Personalize
(https://docs.aws.amazon.com
/personalize/latest/dg/what-is-personalize.html)
AWS AI Practitioner Learning Path: Module on Recommendation Systems
AWS Documentation: Personalization with Amazon Personalize
(https://aws.amazon.com/personalize/)

NEW QUESTION: 122
A company is building a new generative AI chatbot. The chatbot uses an Amazon Bedrock foundation model (FM) to generate responses. During testing, the company notices that the chatbot is prone to prompt injection attacks.
What can the company do to secure the chatbot with the LEAST implementation effort?
A. Fine-tune the FM to avoid harmful responses.
B. Use Amazon Bedrock Guardrails content filters and denied topics.
C. Change the FM to a more secure FM.
D. Use chain-of-thought prompting to produce secure responses.
Answer: B (LEAVE A REPLY)
Amazon Bedrock Guardrails allow developers to create safeguards that filter harmful content and prevent sensitive topics from being discussed. This functionality helps mitigate prompt injection attacks with minimal implementation effort. According to the official Amazon Bedrock documentation:
"You can configure Guardrails for Amazon Bedrock to define denied topics, use content filters, and apply sensitive information filters, offering protection against prompt injection attacks with minimal development effort."

NEW QUESTION: 123
Which scenario describes a potential risk and limitation of prompt engineering In the context of a generative AI model?
A. Prompt engineering does not ensure that the model will consistently generate highly reliable outputs when working with real-world data.
B. Prompt engineering does not ensure that the model always produces consistent and deterministic outputs, eliminating the need for validation.
C. Prompt engineering could expose the model to vulnerabilities such as prompt injection attacks.
D. Properly designed prompts reduce but do not eliminate the risk of data poisoning or model hijacking.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 124

A financial company wants to build workflows for human review of ML predictions. The company wants to define confidence thresholds for its use case and adjust the threshold over time.
Which AWS service meets these requirements?
A. Amazon Personalize
B. Amazon Augmented AI (Amazon A2I)
C. Amazon Inspector
D. AWS Audit Manager
Answer: B (LEAVE A REPLY)
The correct answer is B because Amazon Augmented AI (Amazon A2I) allows developers to integrate human review workflows into ML systems. It supports defining confidence thresholds, such that only low-confidence predictions are sent to human reviewers.
From AWS documentation:
"Amazon A2I provides built-in human review workflows for ML predictions. You can configure confidence thresholds to determine when human review is triggered, enabling continual adjustment based on accuracy needs." This supports use cases where business decisions (like financial approvals) require manual oversight for edge cases.
Explanation of other options:
A). Amazon Personalize is used for recommendation systems, not human review.
C). Amazon Inspector is for security assessment and vulnerability scanning.
D). AWS Audit Manager is used for compliance audits, not ML prediction reviews.
Referenced AWS AI/ML Documents and Study Guides:
Amazon A2I Developer Guide - Confidence Threshold Configuration
AWS ML Specialty Guide - Human-in-the-Loop Systems
AWS Responsible AI Documentation - Review and Escalation Workflows

NEW QUESTION: 125
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 (LEAVE A REPLY)
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: 126
A financial company uses a generative AI model to assign credit limits to new customers. The company wants to make the decision-making process of the model more transparent to its customers.
A. Use a rule-based system instead of an ML model.
B. Apply explainable AI techniques to show customers which factors influenced the model's decision.
C. Develop an interactive UI for customers and provide clear technical explanations about the system.
D. Increase the accuracy of the model to reduce the need for transparency.
Answer: B (LEAVE A REPLY)
The correct answer is B because explainable AI (XAI) provides transparency into how models reach specific decisions. According to AWS documentation, techniques such as SHAP values (SHapley Additive exPlanations) or LIME can identify which input features (e.g., income, debt ratio, or credit history) most influenced a model's prediction. This helps financial institutions comply with fairness and transparency requirements under regulatory frameworks like the Equal Credit Opportunity Act. AWS SageMaker Clarify is a built-in service that offers explainability reports and bias detection to enhance trust. Rule-based systems and UIs alone do not satisfy transparency standards, and accuracy improvements do not replace explainability. By implementing explainable AI, customers can understand and trust credit limit decisions, reducing bias concerns and ensuring compliance.
Referenced AWS AI/ML Documents and Study Guides:
Amazon SageMaker Clarify Documentation - Explainability and Feature Attribution AWS Responsible AI Practices - Transparency and Accountability

NEW QUESTION: 127
Which option is a benefit of ongoing pre-training when fine-tuning a foundation model (FM)?
A. Helps decrease the model's complexity
B. Improves model performance over time
C. Decreases the training time requirement
D. Optimizes model inference time

Answer: B (LEAVE A REPLY)
Ongoing pre-training when fine-tuning a foundation model (FM) improves model performance over time by continuously learning from new data.
* Ongoing Pre-Training:
* Involves continuously training a model with new data to adapt to changing patterns, enhance generalization, and improve performance on specific tasks.
* Helps the model stay updated with the latest data trends and minimize drift over time.
* Why Option B is Correct:
* Performance Enhancement: Continuously updating the model with new data improves its accuracy and relevance.
* Adaptability: Ensures the model adapts to new data distributions or domain-specific nuances.
* Why Other Options are Incorrect:
* A. Decrease model complexity: Ongoing pre-training typically enhances complexity by learning new patterns, not reducing it.
* C. Decreases training time requirement: Ongoing pre-training may increase the time needed for training.
* D. Optimizes inference time: Does not directly affect inference time; rather, it affects model performance.

NEW QUESTION: 128
What does inference refer to in the context of AI?
A. The process of creating new AI algorithms
B. The use of a trained model to make predictions or decisions on unseen data
C. The process of combining multiple AI models into one model
D. The method of collecting training data for AI systems
Answer: B (LEAVE A REPLY)
* Inference = applying a trained ML model to new, unseen data to make predictions, classifications, or generate outputs.
* A is algorithm research, C refers to ensemble learning, D is data collection.
# Reference:
AWS ML Glossary - Inference

NEW QUESTION: 129
A company wants to use generative AI to increase developer productivity and software development. The company wants to use Amazon Q Developer.
What can Amazon Q Developer do to help the company meet these requirements?
A. Create software snippets, reference tracking, and open-source license tracking.
B. Run an application without provisioning or managing servers.
C. Enable voice commands for coding and providing natural language search.
D. Convert audio files to text documents by using ML models.
Answer: C (LEAVE A REPLY)

Amazon Q Developer is a tool designed to assist developers in increasing productivity by generating code snippets, managing reference tracking, and handling open-source license tracking. These features help developers by automating parts of the software development process.
Option A (Correct): "Create software snippets, reference tracking, and open-source license tracking": This is the correct answer because these are key features that help developers streamline and automate tasks, thus improving productivity.
Option B: "Run an application without provisioning or managing servers" is incorrect as it refers to AWS Lambda or AWS Fargate, not Amazon Q Developer.
Option C: "Enable voice commands for coding and providing natural language search" is incorrect because this is not a function of Amazon Q Developer.
Option D: "Convert audio files to text documents by using ML models" is incorrect as this refers to Amazon Transcribe, not Amazon Q Developer.
AWS AI Practitioner References:
Amazon Q Developer Features: AWS documentation outlines how Amazon Q Developer supports developers by offering features that reduce manual effort and improve efficiency.

NEW QUESTION: 130
A company is developing an ML model to predict heart disease risk. The model uses patient data, such as age, cholesterol, blood pressure, smoking status, and exercise habits. The dataset includes a target value that indicates whether a patient has heart disease.
Which ML technique will meet these requirements?
A. Unsupervised learning
B. Supervised learning
C. Reinforcement learning
D. Semi-supervised learning
Answer: (SHOW ANSWER)
This scenario describes a classic supervised learning problem. Supervised learning involves training a model on a labeled dataset where the correct output (also called the target variable or ground truth) is already known.
In this case, the company has a dataset containing features such as age, cholesterol, and exercise habits, and a label indicating whether a patient has heart disease or not. This matches the definition of a binary classification task, which is a type of supervised learning. According to the AWS Machine Learning Specialty Study Guide, supervised learning is most effective when labels are available for all training examples, and the goal is to map input variables to a known output. This method enables the model to learn patterns and predict the presence or absence of heart disease. By contrast, unsupervised learning is used when labels are not available, and reinforcement learning involves agents learning via rewards and punishments, which is not applicable here. Semi-supervised learning is used when only some data is labeled.
Referenced AWS AI/ML Documents and Study Guides:

AWS Certified Machine Learning Specialty Guide - Supervised vs. Unsupervised Learning
Amazon SageMaker Documentation - Classification Models

NEW QUESTION: 131
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 (LEAVE A REPLY)
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:
* This type of machine learning is used when the data has no labels or pre-defined categories. The goal is to identify patterns, clusters, or associations within the data.
* In this case, the company has petabytes of unlabeled customer data and needs to classify customers into different tiers. Unsupervised learning techniques like clustering (e.g., K-Means, Hierarchical Clustering) can group similar customers based on various attributes without any prior knowledge or labels.
* Why Option B is Correct:
* Handling Unlabeled Data: Unsupervised learning is specifically designed to work with unlabeled data, making it ideal for the company's need to classify customer data.
* Customer Segmentation: Techniques in unsupervised learning can be used to find natural groupings within customer data, such as identifying high-value vs. low-value customers or segmenting based on purchasing behavior.
* Why Other Options are Incorrect:
* A. Supervised learning: Requires labeled data with input-output pairs to train the model, which is not suitable since the company's data is unlabeled.
* C. Reinforcement learning: Focuses on training an agent to make decisions by maximizing some notion of cumulative reward, which does not align with the company's need for customer classification.
* D. Reinforcement learning from human feedback (RLHF): Similar to reinforcement learning but involves human feedback to refine the model's behavior; it is also not appropriate for classifying unlabeled customer data.

NEW QUESTION: 132

A company stores its AI datasets in Amazon S3 buckets. The company wants to share the S3 buckets with its business partners. The company needs to avoid accidentally sharing sensitive data.
Which AWS service should the company use to discover sensitive data in the dataset?
A. Amazon Kendra
B. Amazon Macie
C. Amazon Textract
D. AWS Data Exchange
Answer: (SHOW ANSWER)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
Amazon Macie uses machine learning to:
Discover sensitive data such as PII
Classify data stored in Amazon S3
Help prevent unintended data exposure
AWS security guidance recommends Macie before data sharing to ensure compliance and privacy protection.
Why the other options are incorrect:
Kendra (A) is a search service.
Textract (C) extracts text from documents.
Data Exchange (D) shares datasets, not analyzes sensitivity.
AWS AI document references:
Amazon Macie Overview
Protecting Sensitive Data in S3
Data Privacy and Governance on AWS

NEW QUESTION: 133
An ecommerce company is using a chatbot to automate the customer order submission process. The chatbot is powered by AI and Is available to customers directly from the company's website 24 hours a day, 7 days a week.
Which option is an AI system input vulnerability that the company needs to resolve before the chatbot is made available?
A. Data leakage
B. Prompt injection
C. Large language model (LLM) hallucinations
D. Concept drift
Answer: A (LEAVE A REPLY)
The ecommerce company's chatbot, powered by AI, automates customer order submissions and is accessible
24/7 via the website. Prompt injection is an AI system input vulnerability where malicious users craft inputs to manipulate the chatbot's behavior, such as bypassing safeguards or accessing unauthorized information.

This vulnerability must be resolved before the chatbot is made available to ensure security.
Exact Extract from AWS AI Documents:
From the AWS Bedrock User Guide:
"Prompt injection is a vulnerability in AI systems, particularly chatbots, where malicious inputs can manipulate the model's behavior, potentially leading to unauthorized actions or harmful outputs. Implementing guardrails and input validation can mitigate this risk."
(Source: AWS Bedrock User Guide, Security Best Practices)
Detailed Explanation:
Option A: Data leakageData leakage refers to the unintended exposure of sensitive data during model training or inference, not an input vulnerability affecting a chatbot's operation.
Option B: Prompt injectionThis is the correct answer. Prompt injection is a critical input vulnerability for chatbots, where malicious prompts can exploit the AI to produce harmful or unauthorized responses, a risk that must be addressed before launch.
Option C: Large language model (LLM) hallucinationsLLM hallucinations refer to the model generating incorrect or ungrounded responses, which is an output issue, not an input vulnerability.
Option D: Concept driftConcept drift occurs when the data distribution changes over time, affecting model performance. It is not an input vulnerability but a long-term performance issue.
References:
AWS Bedrock User Guide: Security Best Practices
(https://docs.aws.amazon.com/bedrock/latest/userguide
/security.html)
AWS AI Practitioner Learning Path: Module on AI Security and Vulnerabilities AWS Documentation: Securing AI Systems (https://aws.amazon.com/security/)

NEW QUESTION: 134
Which statement presents an advantage of using Retrieval Augmented Generation (RAG) for natural language processing (NLP) tasks?
A. RAG can use external knowledge sources to generate more accurate and informative responses
B. RAG is designed to improve the speed of language model training
C. RAG is primarily used for speech recognition tasks
D. RAG is a technique for data augmentation in computer vision tasks
Answer: (SHOW ANSWER)
Retrieval-Augmented Generation (RAG) integrates external knowledge sources (databases, vector stores, document repositories) with LLMs, enabling them to generate contextually accurate and up-to-date responses without retraining.
B is incorrect: RAG does not speed up training; it improves inference results.
C is incorrect: speech recognition is not an RAG use case.
D is incorrect: computer vision augmentation is unrelated to RAG.
Reference:

AWS Documentation - Knowledge Bases for RAG in Amazon Bedrock

NEW QUESTION: 135
A company is using AI to build a toy recommendation website that suggests toys based on a customer's interests and age. The company notices that the AI tends to suggest stereotypically gendered toys.
Which AWS service or feature should the company use to investigate the bias?
A. Amazon Rekognition
B. Amazon Q Developer
C. Amazon Comprehend
D. Amazon SageMaker Clarify
Answer: D (LEAVE A REPLY)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
Amazon SageMaker Clarify is designed to detect and explain bias in ML models and datasets.
AWS Responsible AI guidance recommends Clarify to:
Identify bias in predictions
Analyze feature attribution
Support fairness and ethical AI practices
Why the other options are incorrect:
Rekognition (A) analyzes images, not recommendation bias.
Amazon Q Developer (B) assists developers with code.
Comprehend (C) performs NLP tasks, not bias analysis.
AWS AI document references:
Amazon SageMaker Clarify Documentation
Detecting Bias in AI Systems
Responsible AI on AWS

NEW QUESTION: 136
A research group wants to test different generative AI models to create research papers. The research group has defined a prompt and needs a method to assess the models' output. The research group wants to use a team of scientists to perform the output assessments.
Which solution will meet these requirements?
A. Use automatic evaluation on Amazon Personalize.
B. Use content moderation on Amazon Rekognition.
C. Use model evaluation on Amazon Bedrock.
D. Use sentiment analysis on Amazon Comprehend.
Answer: C (LEAVE A REPLY)
The correct answer is C because Amazon Bedrock's model evaluation feature allows users to compare outputs from different foundation models using human evaluation or automatic metrics. It enables the creation of structured evaluations where human reviewers (in this case, scientists) can assess model responses based on custom criteria like relevance, coherence, or accuracy.

From AWS documentation:
"Amazon Bedrock provides model evaluation capabilities that support both automatic and human evaluation.
You can define custom evaluation prompts and collect assessments from reviewers to compare foundation model outputs for tasks such as summarization, text generation, and more." This solution is ideal for research workflows requiring domain experts to provide feedback on LLM-generated content.
Explanation of other options:
A). Amazon Personalize is used for building recommendation systems, not for evaluating model output.
B). Amazon Rekognition is used for analyzing images and videos (e.g., moderation, facial recognition), not textual output.
D). Amazon Comprehend provides NLP services like sentiment analysis, but sentiment is not sufficient for full quality evaluation of research paper generation.
Referenced AWS AI/ML Documents and Study Guides:
Amazon Bedrock Developer Guide - Model Evaluation Overview
AWS Generative AI Best Practices
AWS ML Specialty Study Guide - Evaluation and Feedback Loops in LLMs

NEW QUESTION: 137
A media company wants to analyze viewer behavior and demographics to recommend personalized content.
The company wants to deploy a customized ML model in its production environment. The company also wants to observe if the model quality drifts over time.
Which AWS service or feature meets these requirements?
A. Amazon Rekognition
B. Amazon SageMaker Clarify
C. Amazon Comprehend
D. Amazon SageMaker Model Monitor
Answer: (SHOW ANSWER)
The requirement is to deploy a customized machine learning (ML) model and monitor its quality for potential drift over time in a production environment. Let's evaluate each option:

A). Amazon Rekognition: This service is designed for image and video analysis, such as object detection, facial recognition, and text extraction. It is not suited for deploying custom ML models or monitoring model quality drift.
B). Amazon SageMaker Clarify: This feature helps detect bias in ML models and explains model predictions.
While it addresses fairness and interpretability, it does not specifically focus on monitoring model quality drift over time in production.
C). Amazon Comprehend: This is a natural language processing (NLP) service for extracting insights from text, such as sentiment analysis or entity recognition. It does not support deploying custom ML models or monitoring model performance drift.
D). Amazon SageMaker Model Monitor: This feature is part of Amazon SageMaker and is specifically designed to monitor ML models in production. It tracks metrics such as data drift, model drift, and performance degradation over time, alerting users when issues are detected.
Exact Extract Reference: According to the AWS documentation on Amazon SageMaker, "Amazon SageMaker Model Monitor allows you to detect and remediate data and model quality issues in production. It continuously monitors the performance of deployed models, capturing data and model predictions to detect deviations from expected behavior, such as data drift or model performance degradation." (Source: AWS SageMaker Documentation - Model Monitoring, https://docs.aws.amazon.com/sagemaker/latest/dg/model- monitor.html).
This directly aligns with the requirement to observe model quality drift, making Amazon SageMaker Model Monitor the correct choice.
References:
AWS SageMaker Documentation: Model Monitoring
(https://docs.aws.amazon.com/sagemaker/latest/dg
/model-monitor.html)
AWS AI Practitioner Study Guide (conceptual alignment with monitoring deployed ML models)

NEW QUESTION: 138
A hospital wants to use a generative AI solution with speech-to-text functionality to help improve employee skills in dictating clinical notes.
A. Amazon Q Developer
B. Amazon Polly
C. Amazon Rekognition
D. AWS HealthScribe
Answer: D (LEAVE A REPLY)
AWS HealthScribe provides speech-to-text and medical documentation generation, specifically designed for healthcare applications.
Amazon Polly is text-to-speech, not speech-to-text.
Amazon Rekognition is computer vision.
Amazon Q Developer is a generative AI assistant for developers, not healthcare.
Reference:

AWS Documentation - AWS HealthScribe

NEW QUESTION: 139
An AI practitioner has trained a model on a training dataset. The model performs well on the training data.
However, the model does not perform well on evaluation data. What is the MOST likely cause of this issue?
A. The model is underfit.
B. The model requires prompt engineering.
C. The model is biased.
D. The model is overfit.
Answer: (SHOW ANSWER)
When a model performs well on training data but poorly on evaluation/test data, it indicates overfitting.
Overfitting: The model memorizes the training data patterns instead of generalizing.
Underfitting (A) means the model performs poorly on both training and test data.
Bias (C) refers to systemic errors in predictions, not this training/test mismatch.
Prompt engineering (B) applies to generative AI, not general ML training models.
Reference:
AWS ML Glossary - Overfitting and Underfitting

NEW QUESTION: 140
A company has developed an ML model to predict real estate sale prices. The company wants to deploy the model to make predictions without managing servers or infrastructure.
Which solution meets these requirements?
A. Deploy the model on an Amazon EC2 instance.
B. Deploy the model on an Amazon Elastic Kubernetes Service (Amazon EKS) cluster.
C. Deploy the model by using Amazon CloudFront with an Amazon S3 integration.
D. Deploy the model by using an Amazon SageMaker AI endpoint.
Answer: (SHOW ANSWER)
Amazon SageMaker endpoints provide fully managed, serverless model deployment for real-time and batch predictions, allowing companies to deploy ML models without handling any servers or infrastructure management.
D is correct: SageMaker endpoints let you deploy, scale, and monitor ML models with no infrastructure overhead.
A and B require infrastructure management.
C (CloudFront/S3) is not for model deployment, but for static content delivery.
"Amazon SageMaker endpoints allow you to deploy machine learning models for inference without the need to manage underlying infrastructure." (Reference: AWS SageMaker Model Deployment, AWS Certified AI Practitioner Study Guide)

NEW QUESTION: 141
A company is developing a mobile ML app that uses a phone's camera to diagnose and treat insect bites. The company wants to train an image classification model by using a diverse dataset of insect bite photos from different genders, ethnicities, and geographic locations around the world.
Which principle of responsible AI does the company demonstrate in this scenario?
A. Fairness
B. Explainability
C. Governance
D. Transparency
Answer: (SHOW ANSWER)
The company is training an image classification model for diagnosing insect bites using a diverse dataset that includes photos from different genders, ethnicities, and geographic locations. This approach demonstrates the principle of fairness in responsible AI, as it aims to reduce bias and ensure the model performs equitably across diverse populations.
Exact Extract from AWS AI Documents:
From the AWS AI Practitioner Learning Path:
"Fairness in AI involves ensuring that models do not exhibit bias against certain groups and perform equitably across diverse populations. This can be achieved by training models on diverse datasets that represent various demographics, such as gender, ethnicity, and geographic location." (Source: AWS AI Practitioner Learning Path, Module on Responsible AI) Detailed Explanation:
* Option A: FairnessThis is the correct answer. By using a diverse dataset, the company ensures the model is less likely to be biased against specific groups, promoting fairness in its predictions and treatments for insect bites.
* Option B: ExplainabilityExplainability refers to making the model's decisions understandable to users, such as byproviding insights into how predictions are made. The scenario focuses on dataset diversity, not explainability.
* Option C: GovernanceGovernance involves establishing policies and processes to manage AI systems, such as compliance and oversight. The scenario does not describe governance mechanisms.
* Option D: TransparencyTransparency involves disclosing how a model works, its limitations, and its data sources. While transparency is important, the scenario specifically highlights the diversity of the dataset, which aligns more directly with fairness.
References:
AWS AI Practitioner Learning Path: Module on Responsible AI
AWS Documentation: Responsible AI Principles (https://aws.amazon.com/machine-learning/responsible-ai/) Amazon SageMaker Developer Guide: Bias and Fairness in ML (https://docs.aws.amazon.com/sagemaker
/latest/dg/clarify-bias.html)

NEW QUESTION: 142
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 with the 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 (LEAVE A REPLY)
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: 143
Which AWS service or feature can help an AI development team quickly deploy and consume a foundation model (FM) within the team's VPC?
A. Amazon Personalize
B. Amazon SageMaker JumpStart
C. PartyRock, an Amazon Bedrock Playground
D. Amazon SageMaker endpoints
Answer: B (LEAVE A REPLY)
Amazon SageMaker JumpStart is the correct service for quickly deploying and consuming a foundation model (FM) within a team's VPC.

* Amazon SageMaker JumpStart:
* Provides access to a wide range of pre-trained models and solutions that can be easily deployed and consumed within a VPC.
* Designed to simplify and accelerate the deployment of machine learning models, including foundation models.
* Why Option B is Correct:
* Rapid Deployment: JumpStart allows for quick deployment of models with minimal configuration, directly within a secure VPC environment.
* Ease of Use: Provides a user-friendly interface to select and deploy models, reducing the time to value.
* Why Other Options are Incorrect:
* A. Amazon Personalize: Focuses on creating personalized recommendations, not deploying foundation models.
* C. PartyRock: Not a recognized AWS service.
* D. Amazon SageMaker endpoints: Endpoints are for deploying specific models, not a feature for quickly starting with pre-trained foundation models.

NEW QUESTION: 144
A company is deploying AI/ML models by using AWS services. The company wants to offer transparency into the models' decision-making processes and provide explanations for the model outputs.
A. Amazon SageMaker Model Cards
B. Amazon Rekognition
C. Amazon Comprehend
D. Amazon Lex
Answer: A (LEAVE A REPLY)
Amazon SageMaker Model Cards document model details, performance, intended use cases, and risk considerations. They support responsible AI by improving transparency and governance.
Rekognition is computer vision.
Comprehend is NLP for entity/sentiment.
Lex is conversational AI.
Reference:
AWS Documentation - SageMaker Model Cards

NEW QUESTION: 145
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. Refine the prompt until the FM produces the desired responses.
D. Define a higher number for the temperature parameter.
Answer: C (LEAVE A REPLY)
Comprehensive and Detailed Explanation From Exact AWS AI documents:
Prompt engineering is the primary method for controlling tone, style, and behavior of foundation model responses.
AWS generative AI guidance explains that:
Prompts can define tone, voice, and response structure
Iterative refinement ensures consistent outputs
Prompt refinement requires no model retraining
Why the other options are incorrect:
Token limits (A) affect length, not tone.
Batch inferencing (B) affects processing mode, not response style.
Higher temperature (D) increases randomness, reducing consistency.
AWS AI document references:
Prompt Engineering Best Practices
Controlling Model Output Tone
Using Prompts with Foundation Models on AWS