DUMPSARENA

AWS Certified AI Practitioner Exam(AI1-C01)

Amazon AWS AIF-C01

Version Demo

Total Demo Questions: 35

Total Premium Questions: 406

Buy Premium PDF

https://dumpsarena.co

sales@dumpsarena.co

sales@dumpsarena.co
dumpsarena.co

QUESTION NO: 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

Explanation:

The least-effort approach is to start with a pre-trained computer vision model and fine-tune it on your own images. Pre-trained models already “know” a lot of general visual patterns (edges, shapes, textures), so you don’t need to design feature extractors by hand. You mainly focus on preparing your labeled dataset and running a fine-tuning job, which is much faster than building a model and features from scratch.

The other choices require more work. Traditional ML or SVM approaches usually mean you must manually engineer features (like HOG, SIFT, color histograms) and then tune the classifier, which is time-consuming and often less accurate than modern deep learning for images. GANs are mainly for generating new data, not for straightforward image classification.

References: https://docs.aws.amazon.com/sagemaker/latest/dg/image-classification.html and https://docs.aws.amazon.com/sagemaker/latest/dg/transfer-learning.html

QUESTION NO: 2

A company uses Amazon SageMaker AI to generate article summaries in multiple languages. The company needs a metric to evaluate the quality of the summary translations in multiple languages. Which evaluation metric will meet these requirements?

A. Recall-Oriented Understudy for Gisting Evaluation (ROUGE)

B. Bilingual evaluation understudy (BLEU)

C. Area Under the ROC Curve (AUC)

D. Precision

ANSWER: B

Explanation:

To judge how good the translations are across different languages, you want a metric that’s designed for machine translation specifically. BLEU is the classic choice here—it compares the model’s translated text against one or more human “reference”

translations and scores how much they overlap (usually using n-grams). That makes it a solid, widely used way to evaluate translation quality in a consistent, language-agnostic way.

ROUGE is more about summarization quality (how well a summary matches a reference summary), not whether a translation is accurate. AUC is for classification problems, and precision is also mainly used for classification/retrieval tasks, so neither fits translation evaluation.

References: https://en.wikipedia.org/wiki/BLEU and https://huggingface.co/spaces/evaluate-metric/bleu

QUESTION NO: 3

A financial company is using ML to help with some of the company's tasks.

Which option is a use of generative AI models?

A. Summarizing customer complaints

B. Classifying customers based on product usage

C. Segmenting customers based on type of investments D. Forecasting revenue for certain products

D. Forecasting revenue for certain products

ANSWER: A

Explanation:

Generative AI models are built to create new content—like writing text, summarizing long passages, or generating images—based on what you give them as input. So when you take a bunch of long customer complaints and ask the model to produce a shorter, readable summary, that's a classic generative AI task because it's generating new text that didn't exist before.

On the other hand, things like classifying customers by usage or segmenting them by investment type are usually traditional ML tasks (classification or clustering). They're about assigning labels or grouping data, not creating new content. Forecasting revenue is also standard predictive modeling (regression/time series), not generative AI.

For more background, see the AWS overview on generative AI here: https://aws.amazon.com/what-is/generative-ai/

QUESTION NO: 4

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

Explanation:

The best fit here is **AWS Audit Manager** and **AWS Config**. Audit Manager helps you pull together evidence and map it to common compliance frameworks (and your own controls), which is exactly what a regulated financial company needs when auditors ask “prove it.” It’s built for ongoing compliance checks and reporting, not just one-time reviews.

AWS Config is the service that continuously tracks how your AWS resources are configured over time. That makes it really useful for enforcing internal policies (like “S3 buckets must not be public” or “encryption must be enabled”) and for spotting drift during development and deployment phases.

AWS CloudTrail is also very important for auditing API activity, but since the question says “Select TWO,” Audit Manager + Config most directly cover compliance assessment plus continuous policy monitoring.

References: https://docs.aws.amazon.com/audit-manager/latest/userguide/what-is.html and https://docs.aws.amazon.com/config/latest/developerguide/WhatIsConfig.html

QUESTION NO: 5

A company wants to use Amazon Q Business for its data. The company needs to ensure the security and privacy of the data.

Which combination of steps will meet these requirements? (Select TWO.)

A. Enable AWS Key Management Service (AWS KMS) keys for the Amazon Q Business enterprise index.

B. Set up cross-account access to the Amazon Q index.

C. Configure Amazon Inspector for authentication.

D. Allow public access to the Amazon Q index.

E. Configure AWS Identity and Access Management (IAM) for authentication.

ANSWER: A E

Explanation:

To keep Amazon Q Business data secure and private, you typically want two things: strong encryption and tight access control. Using AWS KMS for the Amazon Q Business enterprise index helps protect data at rest by encrypting it with customer-managed keys, so you control key policies, rotation, and who can use the keys.

On top of that, IAM is the right tool for authentication and authorization. With IAM (and often IAM Identity Center in real deployments), you can limit who can access Q Business resources and what they’re allowed to do, following least-privilege practices.

The other choices don’t really solve the core security/privacy need. Cross-account access is a sharing pattern and can increase complexity/risk if not carefully controlled. Amazon Inspector is for vulnerability management, not user authentication. And public access would obviously be the opposite of privacy.

References: https://docs.aws.amazon.com/kms/latest/developerguide/overview.html and https://docs.aws.amazon.com/IAM/latest/UserGuide/introduction.html and https://docs.aws.amazon.com/amazonq/latest/qbusiness-ug/security.html

QUESTION NO: 6

A software company has deployed an AI model to translate paragraphs of text into a user's chosen language. The model can produce a confidence score for the translations. The company wants to incorporate its employees into a review process to validate and improve the model's translations. Which AWS solution will meet these requirements?

A. Amazon SageMaker Clarify

B. Amazon Augmented AI (Amazon A2I)

C. Amazon SageMaker Model Monitor

D. Amazon Bedrock Agents

ANSWER: B

Explanation:

Amazon Augmented AI (A2I) is built for exactly this kind of “human-in-the-loop” setup. When your translation model returns a confidence score, you can use that score to automatically send low-confidence translations (or random samples) to employees for review. That way, people only step in when the model is unsure or when you want extra quality checks.

A2I gives you the managed workflow pieces: a reviewer UI, task routing, and tracking/audit history. Reviewers can approve or correct the translation, and those human-labeled results can then be used to improve the model over time (for example, by retraining or tuning your evaluation process). It’s a practical way to keep translation quality high without making humans review everything.

The other options don’t really fit the goal. SageMaker Clarify is about bias and explainability, not human review queues. Model Monitor watches for drift and production issues, not employee validation of outputs. Bedrock Agents help orchestrate tasks with foundation models, but they’re not a managed human review service.

Reference: https://docs.aws.amazon.com/sagemaker/latest/dg/a2i.html

QUESTION NO: 7

A loan company is building a generative AI-based solution to offer new applicants discounts based on specific business criteria. The company wants to build and use an AI model responsibly to minimize bias that could

negatively affect some customers.

Which actions should the company take to meet these requirements? (Choose two.)

A. Detect imbalances or disparities in the data.

B. Ensure that the model runs frequently.

C. Evaluate the model's behavior so that the company can provide transparency to stakeholders.

D. Use the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) technique to ensure that the model is 100% accurate.

E. Ensure that the model's inference time is within the accepted limits.

ANSWER: A C

Explanation:

The best two actions are **A** and **C**. If you want to reduce bias, you start with the data. Checking for imbalances or disparities helps you spot cases where certain groups are underrepresented or where historical patterns in the data could lead the model to treat some applicants unfairly. Once you know where the data is skewed, you can fix it (for example, by collecting better data, rebalancing, or removing problematic proxy features).

After that, you also need to evaluate how the model behaves in the real world and be able to explain it. Testing outcomes across different groups and documenting what you find is how you create transparency for business owners, risk teams, and compliance. On AWS, a common approach is using Amazon SageMaker Clarify to detect bias and support explainability, which makes it easier to communicate what the model is doing and monitor it over time.

Useful references: https://docs.aws.amazon.com/sagemaker/latest/dg/clarify-detect-data-bias.html and https://docs.aws.amazon.com/sagemaker/latest/dg/clarify-model-explainability.html

QUESTION NO: 8

An airline company wants to build a conversational AI assistant to answer customer questions about flight schedules, booking, and payments. The company wants to use large language models (LLMs) and a knowledge base to create a text-based chatbot interface.

Which solution will meet these requirements with the LEAST development effort?

A. Train models on Amazon SageMaker Autopilot.

B. Develop a Retrieval Augmented Generation (RAG) agent by using Amazon Bedrock.

C. Create a Python application by using Amazon Q Developer.

D. Fine-tune models on Amazon SageMaker Jumpstart.

ANSWER: B

Explanation:

The easiest path here is Amazon Bedrock with a RAG setup (often done with Knowledge Bases for Amazon Bedrock). Bedrock lets you pick a ready-to-use foundation model for chat and connect it to your company content so the bot can “look up” the right info (schedules, policies, payment rules) before it answers. That means you don’t have to build your own retrieval pipeline, prompt stitching, or model hosting from scratch.

SageMaker Autopilot is mainly for automated training on tabular/structured ML problems, not for building an LLM chatbot with grounded answers. JumpStart can help you deploy or fine-tune models, but fine-tuning still doesn’t automatically give you a knowledge-base-backed chatbot—and it’s usually more work than you need when your data changes often. Amazon Q Developer is aimed at helping developers write code, not running a customer-facing airline support bot.

References: https://docs.aws.amazon.com/bedrock/latest/userguide/what-is-bedrock.html and https://docs.aws.amazon.com/bedrock/latest/userguide/knowledge-base.html.

QUESTION NO: 9

A company wants to deploy a conversational chatbot to answer customer questions. The chatbot is based on a fine-tuned Amazon SageMaker JumpStart model. The application must comply with multiple regulatory frameworks.

Which capabilities can the company show compliance for? (Choose two.)

A. Auto scaling inference endpoints

B. Threat detection

C. Data protection

D. Cost optimization

E. Loosely coupled microservices

ANSWER: B C

Explanation:

For a regulated, customer-facing chatbot, the easiest compliance story to tell is usually around protecting data and spotting suspicious activity. That's why **data protection** and **threat detection** are the best fits here.

Data protection maps directly to common requirements in many frameworks (privacy, confidentiality, auditability). On AWS you can show encryption at rest and in transit, strong access controls, and auditable key usage—often using services like AWS KMS, IAM, and CloudTrail. Those are the kinds of controls auditors expect when customer prompts or chat logs might include sensitive info.

Threat detection is also a common requirement because regulators want proof you can monitor and respond to suspicious behavior. Services like Amazon GuardDuty can continuously analyze logs and network activity for indicators of compromise, and AWS Security Hub can centralize findings to help with reporting and evidence.

Auto scaling, cost optimization, and microservices are useful design choices, but they don't directly prove regulatory compliance the way security monitoring and data safeguards do.

References: https://aws.amazon.com/guardduty/ https://aws.amazon.com/security-hub/ https://aws.amazon.com/kms/ https://aws.amazon.com/compliance/

QUESTION NO: 10

A publishing company built a Retrieval Augmented Generation (RAG) solution that lets users interact with published content. New content is published daily, and the company wants a near real-time user experience. Which steps in the RAG pipeline should the company implement using offline batch processing to meet these requirements? (Choose 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

Explanation:

In a RAG setup, anything that depends only on the content library (and not on a specific user question) is a great fit for offline batch work. That way, the “live” user request stays fast and only does what it absolutely must do at query time.

First, the company should generate embeddings for the published content offline. When new articles are added each day, you can chunk them, create embeddings, and store them in a vector database ahead of time. This avoids doing heavy embedding work while a user is waiting for an answer. See https://docs.aws.amazon.com/bedrock/latest/userguide/model-parameters-embeddings.html

Second, the search index should be created/updated offline as well. Building and maintaining a vector index (and its supporting structures) can be compute-intensive, so it’s better handled as a background batch or incremental indexing process. Then, when a user asks something, retrieval is quick because the index is already ready to search. See https://docs.aws.amazon.com/opensearch-service/latest/developerguide/knn.html

By contrast, embedding the user’s query, retrieving the top matches, and generating the final response all have to happen online because they depend on the user’s prompt and the retrieved passages.

QUESTION NO: 11

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

Explanation:

For code translation, the biggest risk is changing what the program actually does. So the model needs strong syntax and semantic understanding—meaning it can read code correctly, understand the intent behind it, and reproduce the same behavior in a different language. If it can also handle code optimization (without breaking logic), that’s a bonus because translated code often needs cleanup to match best practices in the new language.

Speed and general “error handling” matter, but they’re secondary. A fast model that produces subtly wrong code is more expensive in the long run because engineers have to debug and retest everything. Creativity isn’t the goal here either—translation is mostly a correctness and equivalence problem, not an open-ended writing task. Model size and resources can affect cost/latency, but they’re not the main criteria for choosing a model that translates code accurately.

References: https://docs.aws.amazon.com/bedrock/latest/userguide/models-supported.html and https://docs.aws.amazon.com/amazonq/latest/qdeveloper-ug/what-is.html

QUESTION NO: 12

A company wants to use foundation models (FMs) to develop and deploy an AI model.

Which AWS service or resource will meet these requirements with the LEAST development effort?

A. Amazon Bedrock

B. Amazon SageMaker AI

C. Amazon Bedrock PartyRock

D. Amazon Q Developer

ANSWER: A

Explanation:

Amazon Bedrock is the best fit when you want to use foundation models quickly and with the least hands-on work. Bedrock gives you managed access to popular FMs through simple APIs, so you can start building right away without setting up GPUs, managing model hosting, or stitching together a bunch of services just to get to “hello world.” It’s basically the fastest path from idea to a working generative AI app on AWS.

Amazon SageMaker AI can absolutely be used for building and deploying models, but it typically involves more setup and ML workflow management (training jobs, endpoints, pipelines, tuning, etc.). That’s great for custom ML, but it’s not the lowest-effort option for using existing FMs.

Bedrock PartyRock is more of a playground for experimenting and prototyping apps. It’s not the main service you’d pick for a production deployment strategy.

Amazon Q Developer is a coding assistant and productivity tool, not a service for deploying foundation models as part of an AI product.

References: https://aws.amazon.com/bedrock/ and https://docs.aws.amazon.com/bedrock/latest/userguide/what-is-bedrock.html

QUESTION NO: 13

A company has a team of AI practitioners that builds and maintains AI applications in an AWS account. The company must keep records of the actions that each AI practitioner takes in the AWS account for audit purposes.

Which AWS service will meet these requirements?

A. AWS CloudTrail

B. AWS Config

C. AWS Audit Manager

D. AWS Trusted Advisor

ANSWER: A

Explanation:

AWS CloudTrail is the right service here because it keeps an audit trail of activity in your AWS account. It records API calls and console actions and captures details like who did the action (the IAM user/role), what they did, when they did it, and where they did it from. That's exactly what you need when an auditor asks, "Show me what each practitioner changed and when."

The other choices don't really fit. AWS Config is great for tracking how resources change over time (like security group rules or S3 bucket settings), but it's not meant to be a full record of every user action. AWS Audit Manager helps you gather evidence and map it to compliance frameworks, but it depends on underlying logs like CloudTrail—it doesn't replace them. Trusted Advisor is more of a health check and recommendations tool, not an audit log.

References: https://docs.aws.amazon.com/awscloudtrail/latest/userguide/cloudtrail-user-guide.html and https://aws.amazon.com/cloudtrail/

QUESTION NO: 14

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

Explanation:

When your classes are imbalanced, plain accuracy can look great even if the model is basically ignoring the smaller (minority) classes. For example, if 95% of your data is one class, a model that always predicts that class will get ~95% accuracy while doing a terrible job overall.

That's why the **F1 score** is a better fit here. It combines **precision** (how often the model's predicted labels are correct) and **recall** (how many of the real items in a class the model actually finds). Since it's the harmonic mean of precision and recall, it only scores high when *both* are reasonably strong—so it's a solid way to check whether the model is balancing "detecting" and "labeling" well, especially for minority classes.

References: https://en.wikipedia.org/wiki/F-score and https://docs.aws.amazon.com/sagemaker/latest/dg/autopilot-metrics-validation.html

QUESTION NO: 15

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

Explanation:

Ongoing pre-training (sometimes called continued or domain-adaptive pre-training) helps the model get better at the kind of language and content it will see in the real world. By feeding it a lot of in-domain text (often unlabeled), the model picks up domain terms, writing style, and common patterns before you do the task-specific fine-tuning. That usually translates into better accuracy and more reliable outputs once you fine-tune for your exact use case.

It's not really about making the model "smaller" or less complex—its parameter count and architecture don't change just because you keep training it. And it doesn't automatically make training faster either; continued pre-training is extra training, so total compute often goes up even if fine-tuning becomes more data-efficient.

Inference speed also doesn't improve just because of ongoing pre-training. If you want faster inference, you typically look at things like quantization, distillation, or better serving setups—not more pre-training.

References: https://huggingface.co/docs/transformers/en/training and
https://docs.aws.amazon.com/bedrock/latest/userguide/what-is-bedrock.html

QUESTION NO: 16

A company uses Amazon Bedrock for its generative AI application. The company wants to use Amazon Bedrock Guardrails to detect and filter harmful user inputs and model-generated outputs.

Which content categories can the guardrails filter? (Choose two.)

A. Hate

B. Politics

C. Violence

D. Gambling

E. Religion

ANSWER: A C

Explanation:

Amazon Bedrock Guardrails includes built-in *content filters* that can scan both what users type (inputs) and what the model tries to return (outputs). These filters are designed around common safety categories, and two of the core ones are **Hate** and **Violence**. So if someone tries to prompt the model to generate hateful language or violent content, guardrails can detect it and block it (or handle it based on the policy you set).

Hate filtering helps prevent content that targets people based on protected characteristics, uses slurs, or promotes hateful ideologies. This is a big deal for customer-facing apps because it reduces the chance of harmful or discriminatory content showing up in chat responses.

Violence filtering focuses on violent threats, graphic violence, encouragement of harm, or instructions that could enable wrongdoing. It's a common safety requirement, especially if the app is used by the public or in workplaces.

Options like politics, gambling, and religion aren't listed as the standard built-in content filter categories in the same way. If a company wants to control those areas, it would usually be done through other guardrail features (like denied topics) or additional app logic.

References: https://docs.aws.amazon.com/bedrock/latest/userguide/guardrails.html and https://aws.amazon.com/bedrock/guardrails/

QUESTION NO: 17

A company has a generative AI application that uses a pre-trained foundation model (FM) on Amazon Bedrock. The company wants the FM to include more context by using company information. Which solution meets these requirements MOST cost-effectively?

A. Use Amazon Bedrock Knowledge Bases.

B. Choose a different FM on Amazon Bedrock.

C. Use Amazon Bedrock Agents.

D. Deploy a custom model on Amazon Bedrock.

ANSWER: A

Explanation:

The most cost-effective way to add company-specific context to a pre-trained foundation model in Amazon Bedrock is to use Knowledge Bases for Amazon Bedrock. This is a classic Retrieval Augmented Generation (RAG) setup: your app retrieves relevant snippets from your company documents (like PDFs, wikis, or S3 data) and feeds them into the model at prompt time.

This approach is cheaper and faster than fine-tuning or deploying a custom model because you're not training anything—you're just grounding the model's responses in your own data when needed. It also tends to be easier to keep up to date: when your internal docs change, you update the data source rather than retrain a model.

Picking a different FM doesn't magically add your private company knowledge, Agents are more about tool use and workflow orchestration (not primarily about injecting document context), and deploying a custom model is usually overkill and more expensive for this requirement.

References: https://docs.aws.amazon.com/bedrock/latest/userguide/knowledge-base.html and https://docs.aws.amazon.com/bedrock/latest/userguide/knowledge-base-rag.html

QUESTION NO: 18

Which option is a characteristic of transformer-based language models?

A. Transformer-based language models use convolutional layers to apply filters across an input to capture local patterns through filtered views.

B. Transformer-based language models can process only text data.

C. Transformer-based language models use self-attention mechanisms to capture contextual relationships.

D. Transformer-based language models process data sequences one element at a time in cyclic iterations.

ANSWER: C

Explanation:

Transformer-based language models are best known for using **self-attention**. Instead of reading a sentence strictly left-to-right and carrying state step by step (like an RNN), self-attention lets the model look at all tokens and decide which other tokens matter most for understanding each one. That's how transformers get strong "context awareness," like linking pronouns to the right nouns or handling long-range meaning in a paragraph.

This is why option C is the key characteristic. The other options describe different ideas: convolution is more of a CNN thing (local filters), "only text" is wrong because transformers can be used for images/audio too, and "one element at a time in cycles" describes recurrent models, not transformers.

References: https://arxiv.org/abs/1706.03762, https://www.tensorflow.org/text/tutorials/transformer

QUESTION NO: 19

A company is building a chatbot to improve user experience. The company is using a large language model (LLM) from Amazon Bedrock for intent detection. The company wants to use few-shot learning to improve intent detection accuracy.

Which additional data does the company need to meet these requirements?

A. Pairs of chatbot responses and correct user intents

B. Pairs of user messages and correct chatbot responses

C. Pairs of user messages and correct user intents

D. Pairs of user intents and correct chatbot responses

ANSWER: C

Explanation:

Few-shot learning for intent detection is basically "show the model a few good examples of what you want" right inside the prompt. For intent detection, the input is what the user says, and the output you want back is the intent label (like *reset_password* or *track_order*). So the extra data you need is a small set of example user messages, each paired with the correct intent.

The other choices don't quite fit because they focus on chatbot responses. Responses are downstream of intent detection and can vary a lot even when the intent is the same, so they're not the clean signal you want for classification. And pairs like "intent → response" help with response generation, not figuring out the intent from a user's text.

In Amazon Bedrock, this is typically done through prompt engineering: you include a handful of labeled examples (user message → intent) and then ask the model to classify a new message the same way.

References: https://docs.aws.amazon.com/bedrock/latest/userguide/prompt-engineering.html and https://docs.aws.amazon.com/bedrock/latest/userguide/what-is-bedrock.html

QUESTION NO: 20

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

Explanation:

To create subtitles automatically, you first need to turn the spoken audio in the film into text. That's exactly what Amazon Transcribe is for—it listens to the dialogue and produces a time-stamped transcript. Once you have the text, Amazon Translate can convert those subtitles into multiple languages so you can reach viewers who don't speak the original language. See https://docs.aws.amazon.com/transcribe/ and https://docs.aws.amazon.com/translate/.

For voice-overs, you need text-to-speech. After you translate the script (or the transcript) into the target language, Amazon Polly can read that text out loud using natural-sounding voices in many languages. That gives you audio tracks you can line up with the video as dubbed narration. Reference: https://docs.aws.amazon.com/polly/.

The other options don't fit because Amazon Textract is meant for pulling text out of images and documents (like scanned PDFs), not for transcribing movie audio or generating speech. Also, Amazon Translate only translates text—it doesn't produce audio by itself.

QUESTION NO: 21

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

Explanation:

In a RAG setup, the parts that make sense to run “offline” are the ones tied to your documents, not the user’s live question. Since new content arrives daily, you can batch-process that new content to create document (content) embeddings ahead of time. That way, when a user asks something, you’re not doing heavy embedding work on the entire library in the moment.

After you generate those embeddings, you also want to batch-update (or rebuild) the vector/search index. Indexing is another compute-heavy step that’s perfect for scheduled jobs, and it keeps retrieval fast at query time. Together, these two offline steps help the system feel near real-time for users because the “slow prep work” is already done.

The remaining steps—embedding the user query, retrieving the most relevant chunks, and generating the final answer—must happen online in real time, because they depend on what the user just asked.

References: https://docs.aws.amazon.com/prescriptive-guidance/latest/retrieval-augmented-generation-options/rag-architecture.html and https://aws.amazon.com/what-is/retrieval-augmented-generation/

QUESTION NO: 22 - (HOTSPOT)

HOTSPOT

-

A company is training its employees on how to structure prompts for foundation models.

Select the correct prompt engineering technique from the following list for each prompt template. Each prompt engineering technique should be selected one time.

ANSWER:

Explanation:

1. "Classify the following text as either sports, politics, or entertainment: [input text]." : Zero-shot Learning .

Zero-shot learning refers to a model making predictions without having seen labeled examples beforehand.

Reasoning: In this case, the model is asked to classify text without any prior examples. It relies solely on its pre-trained knowledge to determine the correct category.

2. "A [image 1], [image 2], and [image 3] are examples of [target class]. Classify the following image as [target

class]." :Few-shot Learning .

Few-shot learning involves providing the model with a few labeled examples before asking it to make a prediction.

Reasoning: The prompt provides three examples of a given class before asking the model to classify a new instance. This aligns with few-shot learning, where the model generalizes from a small set of examples.

3. "[Question.] [Instructions to follow.] Think step by step and walk me through your thinking." :Chain-of- Thought Reasoning .

Chain-of-thought reasoning is a prompting technique that encourages the model to break down complex problems into step-by-step explanations.

Reasoning: The phrase "Think step by step and walk me through your thinking" explicitly instructs the model to generate a structured, logical explanation rather than providing a direct answer. This is a characteristic of chain-of-thought reasoning.

QUESTION NO: 23

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: B E

Explanation:

To pull text out of PDFs at scale, the best fit is **Amazon Textract**. Textract is built for document text extraction (including scanned PDFs) and can read not just plain text but also forms and tables, which is often exactly what you want when you're dealing with real-world PDFs in S3.

Once you have the extracted text, you need a service that can create summaries. **Amazon Bedrock** is the easiest way to do that on AWS because it gives you access to foundation models that can summarize large chunks of text without you managing any model infrastructure.

For the "fast searching" part, you'd typically index those summaries in something like Amazon OpenSearch Service. That said, since the question asks you to pick only TWO services, the core combo is Textract (extract) + Bedrock (summarize). The indexing piece is commonly added as the next step in a real architecture.

References: https://docs.aws.amazon.com/textract/latest/dg/what-is.html and https://docs.aws.amazon.com/bedrock/latest/userguide/what-is-bedrock.html

QUESTION NO: 24 - (HOTSPOT)

HOTSPOT

A company wants to improve multiple ML models.

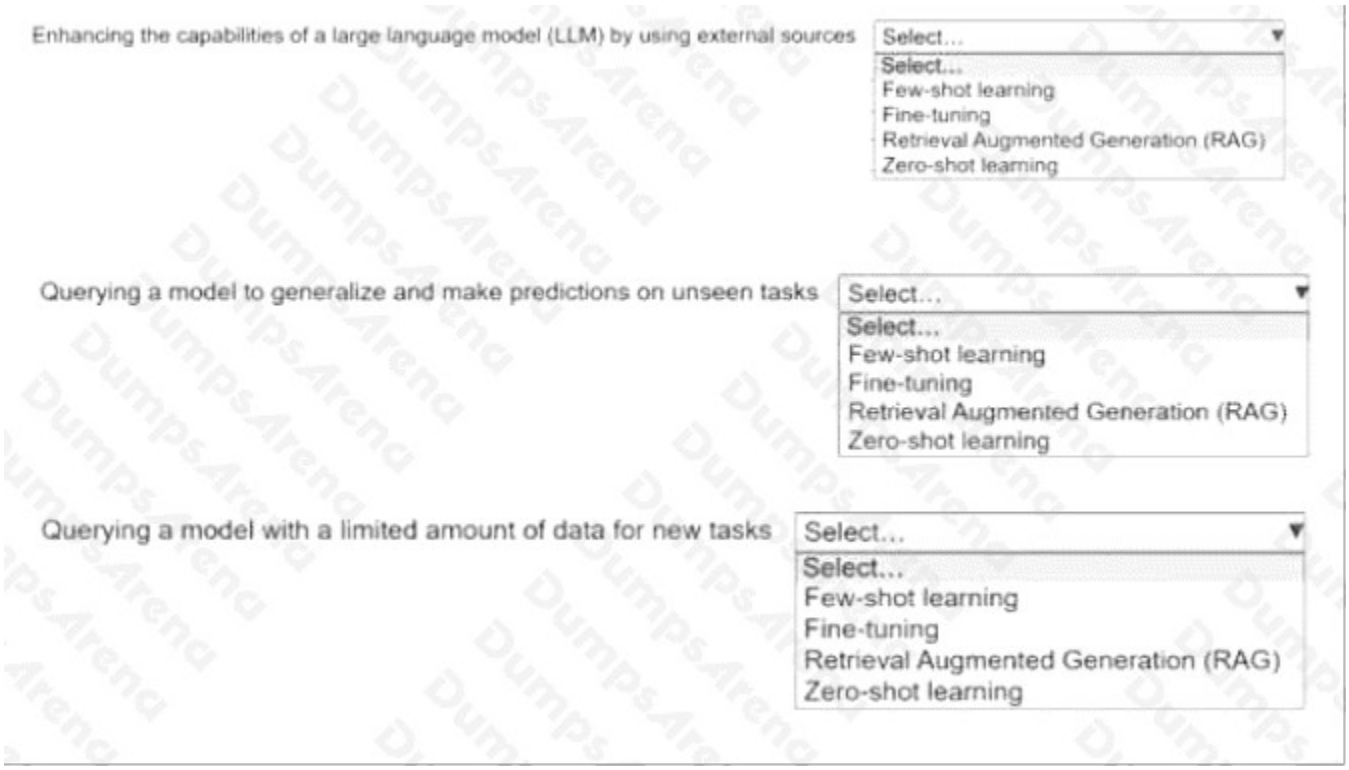
Select the correct technique from the following list of use cases. Each technique should be selected one time or not at all. (Select THREE.)

Few-shot learning

Fine-tuning

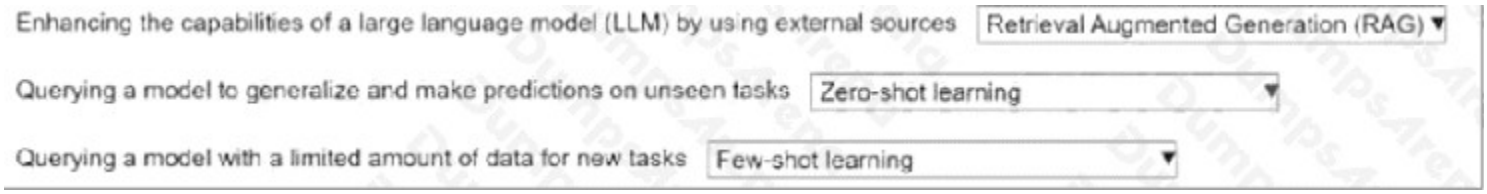
Retrieval Augmented Generation (RAG)

Zero-shot learning



ANSWER:

Explanation:



AWS Reference:

Amazon Bedrock " Retrieval Augmented Generation (RAG) AWS Generative AI Guide " Zero-shot & Few-shot learning

---

QUESTION NO: 25

An ecommerce company wants to build a solution to determine customer sentiments based on written customer reviews of products.

Which AWS services meet these requirements? (Choose two.)

A. Amazon Lex

B. Amazon Comprehend

C. Amazon Polly

D. Amazon Bedrock

E. Amazon Rekognition

ANSWER: B D

Explanation:

For sentiment analysis on written product reviews, **Amazon Comprehend** is the most straightforward fit. It's built for natural language processing tasks like detecting sentiment (positive, negative, neutral, mixed) directly from text, without you having to train a model. You can run it in real time for new reviews or in batch for a big backlog, which is perfect for ecommerce review pipelines. See https://aws.amazon.com/comprehend/ and https://docs.aws.amazon.com/comprehend/latest/dg/how-sentiment.html.

Amazon Bedrock also works well here because foundation models can classify sentiment from text via prompting (and can go further, like giving aspect-based sentiment such as "delivery: negative, quality: positive"). It's a good choice when you want more flexible or richer outputs than a standard sentiment API. More details at https://aws.amazon.com/bedrock/ and https://docs.aws.amazon.com/bedrock/latest/userguide/what-is-bedrock.html.

The other options don't really match: **Lex** is for chatbots, **Polly** turns text into speech, and **Rekognition** analyzes images/video—not written reviews.

QUESTION NO: 26

A company needs to log all requests made to its Amazon Bedrock API. The company must retain the logs securely for 5 years at the lowest possible cost.

Which combination of AWS service and storage class meets these requirements? (Select TWO.)

A. AWS CloudTrail

B. Amazon CloudWatch

C. AWS Audit Manager

D. Amazon S3 Intelligent-Tiering

E. Amazon S3 Standard

ANSWER: A D

Explanation:

To capture every request made to the Amazon Bedrock API, the right tool is AWS CloudTrail. CloudTrail is designed specifically to record AWS API activity (who called what, when, from where, and with which parameters). That's exactly what you want when the requirement says "log all requests made to its API."

For keeping those logs for 5 years as cheaply as possible while still staying secure, storing the CloudTrail logs in Amazon S3 Intelligent-Tiering is a solid fit. Intelligent-Tiering automatically moves objects between access tiers based on usage, so you don't have to guess how often you'll need the logs, and you avoid paying "hot storage" prices for data you rarely touch.

You can also lock things down with S3 features like encryption and bucket policies, and use S3 Object Lock if you need extra protection against deletion or tampering during the retention period.

References: https://docs.aws.amazon.com/awscloudtrail/latest/userguide/cloudtrail-user-guide.html and https://docs.aws.amazon.com/AmazonS3/latest/userguide/intelligent-tiering.html

QUESTION NO: 27

An AI company periodically evaluates its systems and processes with the help of independent software vendors (ISVs). The company needs to receive email message notifications when an ISV's compliance reports become available.

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

A. AWS Audit Manager

B. AWS Artifact

C. AWS Trusted Advisor

D. AWS Data Exchange

ANSWER: B

Explanation:

AWS Artifact is the right fit here because it's AWS's central place to get compliance documents and audit reports. When your process depends on checking whether new compliance reports are available, Artifact is designed for that exact "grab the latest compliance paperwork" workflow, instead of making you hunt through emails or portals.

The question also calls out needing notifications when reports become available. While services like Audit Manager help you collect evidence from your own AWS environment, Artifact is the service meant for accessing published compliance reports and agreements, which is what you'd typically use when working with auditors or third-party assessors.

The other options don't really match the need. **AWS Audit Manager** focuses on automating audit evidence collection in your AWS accounts. **Trusted Advisor** gives best-practice recommendations, not compliance report publishing. **AWS Data Exchange** is for subscribing to third-party datasets and data products, not compliance artifacts.

References: https://aws.amazon.com/artifact/ and https://docs.aws.amazon.com/artifact/latest/ug/what-is-aws-artifact.html

QUESTION NO: 28

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

Explanation:

The safest way to use LLMs on Amazon Bedrock is to control *who* can call the model and *how* the app talks to it. IAM is the main security gate here: you give your app an IAM role and only allow the exact Bedrock actions it needs (least privilege). That way, if something goes wrong, the "blast radius" stays small because the role can't suddenly access extra models or do admin tasks. Bedrock security still follows normal AWS patterns like IAM + auditing, so this is the right foundation. See https://docs.aws.amazon.com/bedrock/latest/userguide/security.html and https://docs.aws.amazon.com/IAM/latest/UserGuide/best-practices.html.

On top of access control, clear prompts and guardrails help reduce real-world risks like prompt injection and accidental data leakage. Good prompts won't replace IAM, but they do make the model's behavior more predictable and easier to keep within policy. Bedrock Guardrails can also filter or block unwanted content at runtime. See https://docs.aws.amazon.com/bedrock/latest/userguide/guardrails.html.

The other choices don't really solve "secure use." Audit Manager is for compliance evidence, not preventing misuse (https://docs.aws.amazon.com/audit-manager/latest/userguide/what-is.html). Model evaluation helps with quality testing, not runtime security (https://docs.aws.amazon.com/bedrock/latest/userguide/model-evaluation.html). CloudWatch Logs are useful for monitoring, but logging alone doesn't make the system secure.

QUESTION NO: 29

A company makes forecasts each quarter to decide how to optimize operations to meet expected demand. The company uses ML models to make these forecasts.

An AI practitioner is writing a report about the trained ML models to provide transparency and explainability to company stakeholders.

What should the AI practitioner include in the report to meet the transparency and explainability requirements?

A. Code for model training

B. Partial dependence plots (PDPs)

C. Sample data for training

D. Model convergence tables

ANSWER: B

Explanation:

To make an ML forecast feel transparent to stakeholders, you usually need something that shows how the model's inputs influence its outputs. Partial dependence plots (PDPs) do exactly that: they visualize how the predicted demand changes as one feature (like price, promotion level, or lead time) changes, while averaging out the effects of the other features. That gives people an easy, "if this goes up, the forecast tends to do that" kind of explanation.

The other items don't really meet the explainability goal. Training code and sample training data can help with reproducibility and auditing, but they don't clearly explain the model's behavior in a way most business readers can interpret. Convergence tables mainly show the training process stabilized (loss/metric over epochs), which is useful for engineers, but it doesn't explain what drives the forecast or how sensitive it is to key inputs.

For more background, see https://scikit-learn.org/stable/modules/partial_dependence.html and AWS's overview of explainability approaches in SageMaker Clarify at https://docs.aws.amazon.com/sagemaker/latest/dg/clarify-model-explainability.html.

QUESTION NO: 30

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 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

Explanation:

For regulated use cases like medical diagnostics, you usually need to show why a model made a prediction, not just what it predicted. Amazon SageMaker Clarify is built for this: it helps explain model behavior (for example, which features most influenced an outcome) and can also help detect potential bias. Those kinds of explanation reports and metrics are exactly what auditors and regulators typically look for when they ask for transparency.

The other services don't really solve the "explainable model" problem. Amazon Inspector is about finding security issues in workloads, not interpreting ML predictions. Amazon Macie focuses on discovering and protecting sensitive data (like PII) in S3, which is important for privacy but doesn't explain model decisions. Amazon Rekognition is for analyzing images and adding labels, which might help with data preparation, but it still won't provide the required explainability for the diagnostic model itself.

References: https://docs.aws.amazon.com/sagemaker/latest/dg/clarify.html and https://aws.amazon.com/sagemaker/clarify/

QUESTION NO: 31

A company needs to log all requests made to its Amazon Bedrock API. The company must retain the logs securely for 5 years at the lowest possible cost.

Which combination of AWS service and storage class meets these requirements? (Choose two.)

A. AWS CloudTrail

B. Amazon CloudWatch

C. AWS Audit Manager

D. Amazon S3 Intelligent-Tiering

E. Amazon S3 Standard

ANSWER: A D

Explanation:

To capture a record of every call made to the Amazon Bedrock API, the right tool is **AWS CloudTrail**. CloudTrail is built specifically for tracking AWS API activity (who called what, when, from where, and with which parameters). That's exactly what "log all requests made to its API" usually means in an audit/compliance sense. See https://aws.amazon.com/cloudtrail/ and https://docs.aws.amazon.com/awscloudtrail/latest/userguide/cloudtrail-user-guide.html.

For keeping those logs for 5 years at the lowest cost, you typically deliver CloudTrail logs to **Amazon S3** and use a cost-optimized storage class. **S3 Intelligent-Tiering** is a solid fit because it automatically moves objects into cheaper tiers as they become rarely accessed, without you having to guess access patterns up front. You still get S3's strong security controls (encryption, bucket policies, IAM) for "retain securely." References: https://aws.amazon.com/s3/storage-classes/intelligent-tiering/ and https://docs.aws.amazon.com/AmazonS3/latest/userguide/storage-class-intro.html.

QUESTION NO: 32

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 to ensure that company policies and industry regulations are followed.

Which AWS services will help the company assess compliance requirements? (Choose two.) A. AWS Audit Manager

B. AWS Config

C. Amazon Inspector

D. Amazon CloudWatch

E. AWS CloudTrail

Answer: AB

Explanation:

The correct answer is AB. Here's why:

A. AWS Audit Manage

B. AWS Config

C. Amazon Inspector

D. Amazon CloudWatch

E. AWS CloudTrail

ANSWER: A B

Explanation:

AWS Audit Manager is built specifically for compliance. It helps you map your controls to common frameworks (or your own internal policies) and then automatically gathers evidence from AWS services so you can show auditors what happened and when. For a regulated financial company, that “audit-ready” evidence trail is a big deal because it reduces manual screenshots, spreadsheets, and back-and-forth during audits. See: https://aws.amazon.com/audit-manager/ and https://docs.aws.amazon.com/audit-manager/latest/userguide/what-is.html

AWS Config focuses on continuous compliance at the resource configuration level. It records configuration changes over time and can evaluate resources against rules (for example: “S3 buckets must not be public,” “encryption must be enabled,” “security groups can’t allow 0.0.0.0/0 on certain ports”). That’s exactly what you want during ongoing development, where small changes can accidentally drift out of compliance. See: https://aws.amazon.com/config/ and https://docs.aws.amazon.com/config/latest/developerguide/WhatIsConfig.html

The other services are useful, but not as direct for “assessing compliance requirements.” Inspector is mainly vulnerability scanning, CloudWatch is operational monitoring, and CloudTrail is activity logging—great for investigations, but it doesn’t evaluate compliance on its own.

QUESTION NO: 33

An AI company periodically evaluates its systems and processes with the help of independent software vendors (ISVs). The company needs to receive email message notifications when an ISV's compliance reports become available.

Which AWS service meets this requirement?

A. AWS Audit Manager

B. AWS Artifact

C. AWS Trusted Advisor

D. AWS Data Exchange

ANSWER: B

Explanation:

AWS Artifact is the right fit here because it's basically AWS's "library" for compliance and audit documents. It gives you access to reports like SOC, ISO, PCI, and other assurance documentation, and it's designed for situations where you need to pull official compliance reports when they're published or updated.

What matters in this question is the email notification part. Artifact supports subscribing to notifications so you can be alerted when new compliance reports or updates become available, which matches the need to know when an ISV's compliance report is ready.

The other options don't really line up. AWS Audit Manager helps you collect evidence and map it to audit frameworks, but it's more about managing your own audits than notifying you about third-party compliance reports. Trusted Advisor is for best-practice checks (cost, security, performance), not compliance report distribution. AWS Data Exchange is for subscribing to datasets, not compliance documentation alerts.

References: https://docs.aws.amazon.com/artifact/latest/ug/what-is-aws-artifact.html and https://aws.amazon.com/artifact/

QUESTION NO: 34

A company wants to use Amazon Q Business for its data. The company needs to ensure the security and privacy of the data. Which combination of steps will meet these requirements? (Select TWO.)

A. Enable AWS Key Management Service (AWS KMS) keys for the Amazon Q Business Enterprise index.

B. Set up cross-account access to the Amazon Q index.

C. Configure Amazon Inspector for authentication.

D. Allow public access to the Amazon Q index.

E. Configure AWS Identity and Access Management (IAM) for authentication.

ANSWER: A E

Explanation:

To keep Amazon Q Business data secure and private, you generally want two things: strong access control and strong encryption. Configuring IAM (and typically IAM Identity Center behind the scenes) makes sure only the right users and roles can sign in and query the Q Business app, following least-privilege access.

On top of that, enabling AWS KMS keys for the Amazon Q Business index helps protect data at rest. Using customer-managed KMS keys is a common requirement for compliance and gives you more control over key policies, rotation, and auditing.

The other options don't really match the goal. Cross-account access is only needed if you're intentionally sharing resources across AWS accounts, not as a core security requirement. Amazon Inspector is for finding vulnerabilities on compute workloads, not for authenticating users to Q Business. And public access would obviously work against privacy.

References: https://docs.aws.amazon.com/amazonq/latest/qbusiness-ug/security.html and https://docs.aws.amazon.com/kms/latest/developerguide/overview.html

QUESTION NO: 35

A company wants to display the total sales for its top-selling products across various retail locations in the past 12 months.

Which AWS solution should the company use to automate the generation of graphs?

A. Amazon Q in Amazon EC2

B. Amazon Q Developer

C. Amazon Q in Amazon QuickSight

D. Amazon Q in AWS Chatbot

ANSWER: C

Explanation:

The best fit here is **Amazon Q in Amazon QuickSight**. QuickSight is AWS's managed BI and visualization service, built specifically for turning sales data into charts, dashboards, and reports. It can connect to common data sources (like Amazon S3, Amazon Redshift, and databases), and it supports scheduled refreshes—handy when you're always looking at a rolling "last 12 months" view.

What makes it really "automated" in this question is the **Amazon Q** part. With Amazon Q in QuickSight, users can ask questions in plain language (for example, "Show top-selling products by location over the last year"), and QuickSight can generate the visuals without someone manually building every graph and filter.

The other choices don't match the goal. EC2 is just compute—you'd still need to build and maintain your own dashboarding tool. Amazon Q Developer is for coding help, not BI charts. AWS Chatbot is mainly for ops notifications and commands in chat apps, not business reporting dashboards.

References: https://aws.amazon.com/quicksight/ and https://aws.amazon.com/amazon-q/