AWS AI Practitioner Exam Questions

Q1: 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

Q2: 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

Q3: A company wants to classify human genes into 20 categories based on gene characteristics. The company needs an ML algorithm to document how the inner mechanism of the model affects the output. Which ML algorithm meets these requirements?

- A. Decision trees

- B. Linear regression
- C. Logistic regression
- D. Neural networks

  Answer: A

Q4: A company has built an image classification model to predict plant diseases from photos of plant leaves. The company wants to evaluate how many images the model classified correctly. Which evaluation metric should the company use to measure the model's performance?

- A. R-squared score
- B. Accuracy
- C. Root mean squared error (RMSE)
- D. Learning rate

  Answer: B

Q5: A company is using a pre-trained large language model (LLM) to build a chatbot for product recommendations. The company needs the LLM outputs to be short and written in a specific language. Which solution will align the LLM response quality with the company's expectations?

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

  Answer: A

Q6: A company uses Amazon SageMaker for its ML pipeline in a production environment. The company has large input data sizes up to 1 GB and processing times up to 1 hour. The company needs near real-time latency. Which SageMaker inference option meets these requirements?

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

Answer: C

Q7: 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

Q8: A company is building a solution to generate images for protective eyewear. The solution must have high accuracy and must minimize the risk of incorrect annotations. Which solution will meet these requirements?

- A. Human-in-the-loop validation by using Amazon SageMaker Ground Truth Plus
- B. Data augmentation by using an Amazon Bedrock knowledge base
- C. Image recognition by using Amazon Rekognition
- D. Data summarization by using Amazon QuickSight Q

Answer: A

Q9: 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

Q10: 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

Q11: A company wants to build an ML model by using Amazon SageMaker. The company needs to share and manage variables for model development across multiple teams. Which SageMaker feature meets these requirements?

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

Answer: A

Q12: 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: A

Q13: 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

Q14: 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

Q15: 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

Q16: A company is building a contact center application and wants to gain insights from customer conversations. The company wants to analyze and

extract key information from the audio of the customer calls. Which solution meets these requirements?

- A. Build a conversational chatbot by using Amazon Lex
- B. Transcribe call recordings by using Amazon Transcribe
- C. Extract information from call recordings by using Amazon SageMaker Model Monitor
- D. Create classification labels by using Amazon Comprehend

  Answer: B

Q17: 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

Q18: 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

Q19: A company uses a foundation model (FM) from Amazon Bedrock for an AI search tool. The company wants to fine-tune the model to be more accurate by using the company's data. Which strategy will successfully fine-tune the model?

- A. Provide labeled data with the prompt field and the completion field

- B. Prepare the training dataset by creating a .txt file that contains multiple lines in .csv format
- C. Purchase Provisioned Throughput for Amazon Bedrock
- D. Train the model on journals and textbooks

  Answer: A

Q20: A company wants to use AI to protect its application from threats. The AI solution needs to check if an IP address is from a suspicious source. Which solution meets these requirements?

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

  Answer: C

Q21: Which feature of Amazon OpenSearch Service gives companies the ability to build vector database applications?

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

  Answer: C

Q22: Which option is a use case for generative AI models?

- A. Improving network security by using intrusion detection systems
- B. Creating photorealistic images from text descriptions for digital marketing
- C. Enhancing database performance by using optimized indexing
- D. Analyzing financial data to forecast stock market trends

  Answer: B

Q23: A company wants to build a generative AI application by using Amazon Bedrock and needs to choose a foundation model (FM). The company wants to

know how much information can fit into one prompt. Which consideration will inform the company's decision?

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

  Answer: B

Q24: 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

Q25: 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

Q26: A security company is using Amazon Bedrock to run foundation models (FMs). The company wants to ensure that only authorized users invoke the models. The company needs to identify any unauthorized access attempts to set appropriate AWS Identity and Access Management (IAM) policies and roles for future iterations of the FMs. Which AWS service should the company use to identify unauthorized users that are trying to access Amazon Bedrock?

- A. AWS Audit Manager
- B. AWS CloudTrail
- C. Amazon Fraud Detector
- D. AWS Trusted Advisor

  Answer: B

Q27: A company has developed an ML model for image classification. The company wants to deploy the model to production so that a web application can use the model. The company needs to implement a solution to host the model and serve predictions without managing any of the underlying infrastructure. Which solution will meet these requirements?

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

  Answer: A

Q28: 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

Q29: A company wants to use a large language model (LLM) to develop a conversational agent. The company needs to prevent the LLM from being manipulated with common prompt engineering techniques to perform undesirable actions or expose sensitive information. Which action will reduce these risks?

- A. Create a prompt template that teaches the LLM to detect attack patterns
- B. Increase the temperature parameter on invocation requests to the LLM
- C. Avoid using LLMs that are not listed in Amazon SageMaker
- D. Decrease the number of input tokens on invocations of the LLM

  Answer: A

Q30: A company is using the Generative AI Security Scoping Matrix to assess security responsibilities for its solutions. The company has identified four different solution scopes based on the matrix. Which solution scope gives the company the MOST ownership of security responsibilities?

- A. Using a third-party enterprise application that has embedded generative AI features
- B. Building an application by using an existing third-party generative AI foundation model (FM)
- C. Refining an existing third-party generative AI foundation model (FM) by fine-tuning the model by using data specific to the business
- D. Building and training a generative AI model from scratch by using specific data that a customer owns

  Answer: D

Q31: An AI practitioner has a database of animal photos. The AI practitioner wants to automatically identify and categorize the animals in the photos without manual human effort. Which strategy meets these requirements?

- A. Object detection
- B. Anomaly detection
- C. Named entity recognition
- D. Inpainting

  Answer: A

Q32: 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

Q33: 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

Q34: 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

Q35: A company has terabytes of data in a database that the company can use for business analysis. The company wants to build an AI-based application that can build a SQL query from input text that employees provide. The employees have minimal experience with technology. Which solution meets these requirements?

- A. Generative pre-trained transformers (GPT)

- B. Residual neural network
- C. Support vector machine
- D. WaveNet

  Answer: A

Q36: A company built a deep learning model for object detection and deployed the model to production. Which AI process occurs when the model analyzes a new image to identify objects?

- A. Training
- B. Inference
- C. Model deployment
- D. Bias correction

  Answer: B

Q37: 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

Q38: 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

Q39: 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

Q40: 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

Q41: A company is training a foundation model (FM). The company wants to increase the accuracy of the model up to a specific acceptance level. Which solution will meet these requirements?

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

  Answer: B

Q42: 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

Q43: Which functionality does Amazon SageMaker Clarify provide?

- A. Integrates a Retrieval Augmented Generation (RAG) workflow
- B. Monitors the quality of ML models in production
- C. Documents critical details about ML models
- D. Identifies potential bias during data preparation

  Answer: D

Q44: A company is developing a new model to predict the prices of specific items. The model performed well on the training dataset. When the company deployed the model to production, the model's performance decreased significantly. What should the company do to mitigate this problem?

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

  Answer: C

Q45: 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

Q46: 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: D

Q47: A social media company wants to use a large language model (LLM) for content moderation. The company wants to evaluate the LLM outputs for bias and potential discrimination against specific groups or individuals. Which data source should the company use to evaluate the LLM outputs with the LEAST administrative effort?

- A. User-generated content
- B. Moderation logs
- C. Content moderation guidelines
- D. Benchmark datasets

  Answer: D

Q48: A company wants to use a pre-trained generative AI model to generate content for its marketing campaigns. The company needs to ensure that the generated content aligns with the company's brand voice and messaging requirements. Which solution meets these requirements?

- A. Optimize the model's architecture and hyperparameters to improve the model's overall performance

- B. Increase the model's complexity by adding more layers to the model's architecture
- C. Create effective prompts that provide clear instructions and context to guide the model's generation
- D. Select a large, diverse dataset to pre-train a new generative model

  Answer: C

Q49: 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 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

Q50: 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: A

Q51: A company needs to choose a model from Amazon Bedrock to use internally. The company must identify a model that generates responses in a style that the company's employees prefer. What should the company do to meet these requirements?

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

  Answer: B

Q52: A student at a university is copying content from generative AI to write essays. Which challenge of responsible generative AI does this scenario represent?

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

  Answer: C

Q53: A company needs to build its own large language model (LLM) based on only the company's private data. The company is concerned about the environmental effect of the training process. Which Amazon EC2 instance type has the LEAST environmental effect when training LLMs?

- A. Amazon EC2 C series
- B. Amazon EC2 G series
- C. Amazon EC2 P series
- D. Amazon EC2 Trn series

  Answer: D

Q54: A company wants to build an interactive application for children that generates new stories based on classic stories. The company wants to use Amazon Bedrock and needs to ensure that the results and topics are appropriate for children. Which AWS service or feature will meet these requirements?

- A. Amazon Rekognition
- B. Amazon Bedrock playgrounds

- C. Guardrails for Amazon Bedrock
- D. Agents for Amazon Bedrock

  Answer: C

Q55: A company is building an application that needs to generate synthetic data that is based on existing data. Which type of model can the company use to meet this requirement?

- A. Generative adversarial network (GAN)
- B. XGBoost
- C. Residual neural network
- D. WaveNet

  Answer: A

Q56: A digital devices company wants to predict customer demand for memory hardware. The company does not have coding experience or knowledge of ML algorithms and needs to develop a data-driven predictive model. The company needs to perform analysis on internal data and external data. Which solution will meet these requirements?

- A. Store data in Amazon S3 and use SageMaker built-in algorithms
- B. Use SageMaker Data Wrangler and built-in algorithms
- C. Use Data Wrangler with Amazon Personalize Trending-Now
- D. Use Amazon SageMaker Canvas to build models visually

  Answer: D

Q57: A company has installed a security camera. The company uses an ML model to evaluate the security camera footage for potential thefts. The company has discovered that the model disproportionately flags people who are members of a specific ethnic group. Which type of bias is affecting the model output?

- A. Measurement bias
- B. Sampling bias
- C. Observer bias
- D. Confirmation bias

  Answer: B

Q58: A company is building a customer service chatbot. The company wants the chatbot to improve its responses by learning from past interactions and online resources. Which AI learning strategy provides this self-improvement capability?

- A. Supervised learning with curated good/bad responses
- B. Reinforcement learning with rewards for positive customer feedback
- C. Unsupervised clustering of similar inquiries
- D. Supervised learning with a continuously updated FAQ database

  Answer: B

Q59: An AI practitioner has built a deep learning model to classify the types of materials in images. The AI practitioner now wants to measure the model performance. Which metric will help the AI practitioner evaluate the performance of the model?

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

  Answer: A

Q60: A company has built a chatbot that can respond to natural language questions with images. The company wants to ensure that the chatbot does not return inappropriate or unwanted images. Which solution will meet these requirements?

- A. Implement moderation APIs
- B. Retrain the model with a general public dataset
- C. Perform model validation
- D. Automate user feedback integration

  Answer: A

Q61: An AI practitioner is using an Amazon Bedrock base model to summarize session chats from the customer service department. The AI practitioner wants to store invocation logs to monitor model input and output data. Which strategy should the AI practitioner use?

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

  Answer: B

Q62: 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

Q63: Which term describes the numerical representations of real-world objects and concepts that AI and NLP models use to improve understanding of textual information?

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

  Answer: A

Q64: A research company implemented a chatbot by using a foundation model (FM) from Amazon Bedrock. The chatbot searches for answers from a large database of research papers. After multiple prompt engineering attempts, the company notices poor performance due to complex scientific terms. How can the company improve performance?

- A. Use few-shot prompting for answer style
- B. Use domain adaptation fine-tuning for complex scientific terms
- C. Change FM inference parameters

- D. Clean data to remove complex terms

  Answer: B

Q65: A company wants to use a large language model (LLM) on Amazon Bedrock for sentiment analysis. The company needs the LLM to produce more consistent responses to the same input prompt. Which adjustment should be made?

- A. Decrease the temperature value
- B. Increase the temperature value
- C. Decrease the length of output tokens
- D. Increase the maximum generation length

  Answer: A

Q66: A company wants to develop an LLM application by using Amazon Bedrock and customer data in Amazon S3. Security policy: each team can access only its own customers' data. Which solution meets these requirements?

- A. Create an Amazon Bedrock custom service role per team with least-privilege to the team's S3 data
- B. One role with S3 access; teams pass customer name per request
- C. Redact personal data in S3 and open access
- D. One Bedrock role with full S3 and per-team IAM roles for folders

  Answer: A

Q67: 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 responses and receive notifications on violations. Which solution meets these requirements?

- A. Amazon Macie to scan outputs with alerts
- B. AWS CloudTrail to monitor responses
- C. Guardrails for Amazon Bedrock with CloudWatch alarms
- D. SageMaker Model Monitor for data drift

  Answer: C

Q68: A company manually reviews resumes in PDF format. Volume will exceed capacity. The company needs an automated system to convert PDF resumes into plain text for processing. Which AWS service meets this requirement?

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

  Answer: A

Q69: An education provider is building a Q&A application with a generative AI model. The response style should change based on the age range provided. Which solution meets these requirements with the least effort?

- A. Fine-tune with age-range training data
- B. Add a role description in the prompt instructing target age range
- C. Use chain-of-thought reasoning to deduce style
- D. Summarize to be shorter for younger users

  Answer: B

Q70: Which strategy evaluates the accuracy of a foundation model (FM) used in image classification tasks?

- A. Calculate total cost of resources
- B. Measure accuracy against a predefined benchmark dataset
- C. Count the number of layers
- D. Assess color accuracy of processed images

  Answer: B

Q71: An accounting firm wants to implement an LLM to automate document processing responsibly. What should the firm do when developing and deploying the LLM? (Choose two.)

- A. Include fairness metrics for evaluation
- B. Adjust the temperature parameter
- C. Modify training data to mitigate bias
- D. Avoid overfitting on the training data

- E. Apply prompt engineering techniques

  Answer: A, C

Q72: A company is building an ML model. The company collected new data and analyzed it by creating a correlation matrix, calculating statistics, and visualizing the data. Which stage is this?

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

  Answer: C

Q73: A company has documents that are missing some words because of a database error. The company wants an ML model that can suggest potential words to fill in the missing text. Which model meets this requirement?

- A. Topic modeling
- B. Clustering models
- C. Prescriptive ML models
- D. BERT-based models

  Answer: D

Q74: A company wants to display total sales for top-selling products across various retail locations in the past 12 months. Which AWS solution automates graph generation?

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

  Answer: C

Q75: A company is building a chatbot and using an LLM from Amazon Bedrock for intent detection. The company wants to use few-shot learning to improve accuracy. Which additional data is needed?

- 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

Q76: A company is using few-shot prompting on a base model on Amazon Bedrock with 10 examples in the prompt. Invoked once daily and performing well; wants lower monthly cost. Which solution meets these requirements?

- A. Customize with fine-tuning
- B. Decrease the number of tokens in the prompt
- C. Increase the number of tokens in the prompt
- D. Use Provisioned Throughput

  Answer: B

Q77: An AI practitioner is using an LLM to create marketing content. The content sounds plausible and factual but is incorrect. Which problem is this?

- A. Data leakage
- B. Hallucination
- C. Overfitting
- D. Underfitting

  Answer: B

Q78: An AI practitioner trained a custom model on Amazon Bedrock by using a training dataset that contains confidential data. The practitioner wants to ensure the model does not generate responses based on that confidential data. How should this be prevented?

- A. Delete the custom model, remove confidential data from training, retrain
- B. Mask confidential data in inference responses
- C. Encrypt confidential data in inference responses using SageMaker
- D. Encrypt confidential data in the model using AWS KMS

  Answer: A

Q79: A company's generative AI solution translates training manuals. The company wants to evaluate accuracy by examining generated text. Which

evaluation strategy meets these requirements?

- A. BLEU
- B. RMSE
- C. ROUGE
- D. F1 score

  Answer: A

Q80: A large retailer receives thousands of customer support inquiries daily. The company wants to implement Agents for Amazon Bedrock. What are the key benefits?

- A. Generation of custom FMs to predict customer needs
- B. Automation of repetitive tasks and orchestration of complex workflows
- C. Automatically calling multiple FMs and consolidating results
- D. Selecting the FM based on predefined criteria and metrics

  Answer: B

Q81: 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

Q82: What are tokens in the context of generative AI models?

- A. Basic units of input/output (words, subwords, or other linguistic units)
- B. Mathematical representations (embeddings)
- C. Pre-trained weights fine-tuned for tasks
- D. Specific prompts or instructions

  Answer: A

Q83: A company wants to assess the costs associated with using an LLM to generate inferences on Amazon Bedrock. Which factor drives 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

Q84: A company is using SageMaker Studio notebooks and stores data in Amazon S3. The company needs to manage the flow of data from S3 to SageMaker Studio notebooks. Which solution meets this requirement?

- A. Use Amazon Inspector to monitor Studio
- B. Use Amazon Macie to monitor Studio
- C. Configure SageMaker to use a VPC with an S3 endpoint
- D. Configure SageMaker to use S3 Glacier Deep Archive

  Answer: C

Q85: A company has an FM customized via Amazon Bedrock to answer customer queries about products. The company wants to validate responses to new query types and upload a new dataset for validation. Which AWS service meets these requirements?

- A. Amazon S3
- B. Amazon EBS
- C. Amazon EFS
- D. AWS Snowcone

  Answer: A

Q86: 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

Q87: A company wants to use Amazon Bedrock and needs to review which security aspects the company is responsible for. Which aspect is the company responsible for?

- A. Patching and updating Amazon Bedrock
- B. Protecting the infrastructure that hosts Bedrock
- C. Securing the company's data in transit and at rest
- D. Provisioning Amazon Bedrock within the company network

  Answer: C

Q88: A social media company wants to compare the generated output toxicity of LLMs available on SageMaker JumpStart. Which strategy gives the ability to evaluate with the least operational overhead?

- A. Crowd-sourced evaluation
- B. Automatic model evaluation
- C. Model evaluation with human workers
- D. RLHF

  Answer: B

Q89: A company is testing the security of an FM and wants to get around safety features to make harmful content. Which security technique is this?

- A. Fuzzing training data
- B. Denial of service (DoS)
- C. Penetration testing with authorization
- D. Jailbreak

  Answer: D

Q90: A company must run SageMaker training and inference in an isolated environment without internet access to meet regulatory requirements. Which solution meets these requirements?

- A. Use SageMaker Experiments
- B. Use network isolation for training and inference
- C. Encrypt data at rest (geospatial capabilities)
- D. Associate appropriate IAM roles

Answer: B

Q91: An ML research team develops custom ML models and shares artifacts with other teams. They want an auditable mechanism when publishing models. Which solution should be used?

- A. Create documents and store in S3
- B. Use AWS AI Service Cards
- C. Create Amazon SageMaker Model Cards with intended uses and training/inference details
- D. Commit training scripts to Git

Answer: C

Q92: A software company wants to use AI to increase software development productivity. Which solution meets these requirements?

- A. Use a binary classification model for code reviews
- B. Install code recommendation software in developer tools
- C. Install a code forecasting tool
- D. Use an NLP tool to generate code

Answer: B

Q93: A retail store wants to predict product demand using the SageMaker DeepAR forecasting algorithm. Which type of data is required?

- A. Text data
- B. Image data
- C. Time series data
- D. Binary data

Answer: C

Q94: A large retail bank wants to develop an ML system to help decide loan allocations across demographics. What must the bank do to develop an unbiased ML model?

- A. Reduce the size of the training dataset
- B. Ensure predictions are consistent with historical results
- C. Create a different model for each demographic

- D. Measure class imbalance and adapt training accordingly

  Answer: D

Q95: 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: A

Q96: A company fine-tuned an LLM for a help desk and wants to determine if fine-tuning enhanced accuracy. Which metric should be used?

- A. Precision
- B. Time to first token
- C. F1 score
- D. Word error rate

  Answer: C

Q97: A company uses RAG with Amazon Bedrock and Stable Diffusion to generate product images from text. Results are often random and lack specific details. How to increase specificity?

- A. Increase number of generation steps
- B. Use MASK_IMAGE_BLACK
- C. Increase classifier-free guidance (CFG) scale
- D. Increase prompt strength

  Answer: C

Q98: A company wants an LLM-based chatbot to provide real-time contextual responses to customer inquiries using company policies as the knowledge base. Which solution is most cost-effective?

- A. Retrain the LLM on company policy data
- B. Fine-tune the LLM on company policy data

- C. Implement Retrieval Augmented Generation (RAG) for in-context responses
- D. Use pre-training and data augmentation

  Answer: C

Q99: A company wants to create a new solution using AWS Glue but has minimal programming experience with Glue. Which AWS service can help?

- A. Amazon Q Developer
- B. AWS Config
- C. Amazon Personalize
- D. Amazon Comprehend

  Answer: A

Q100: 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 this demonstrate?

- A. Fairness
- B. Explainability
- C. Governance
- D. Transparency

  Answer: A

Q101: 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

Q102: A company has developed a generative text summarization model by using Amazon Bedrock. The company will use Amazon Bedrock automatic model evaluation capabilities. Which metric should the company use to evaluate the accuracy of the model?

- A. Area Under the ROC Curve (AUC) score
- B. F1 score
- C. BERTScore
- D. Real world knowledge (RWK) score

  Answer: C

Q103: An AI practitioner wants to predict the classification of flowers based on petal length, petal width, sepal length, and sepal width. Which algorithm meets these requirements?

- A. K-nearest neighbors (k-NN)
- B. K-mean
- C. Autoregressive Integrated Moving Average (ARIMA)
- D. Linear regression

  Answer: A

Q104: A company is using custom models in Amazon Bedrock for a generative AI application. The company wants to use a company managed encryption key to encrypt the model artifacts that the model customization jobs create. Which AWS service meets these requirements?

- A. AWS Key Management Service (AWS KMS)
- B. Amazon Inspector
- C. Amazon Macie
- D. AWS Secrets Manager

  Answer: A

Q105: A company wants to use large language models (LLMs) to produce code from natural language code comments. Which LLM feature meets these requirements?

- A. Text summarization

- B. Text generation
- C. Text completion
- D. Text classification

  Answer: B

Q106: A company is introducing a mobile app that helps users learn foreign languages. The app makes text more coherent by calling a large language model (LLM). The company collected a diverse dataset of text and supplemented the dataset with examples of more readable versions. The company wants the LLM output to resemble the provided examples. Which metric should the company use?

- A. Value of the loss function
- B. Semantic robustness
- C. ROUGE score
- D. Latency of the text generation

  Answer: C

Q107: A company notices that its foundation model (FM) generates images that are unrelated to the prompts. The company wants to modify the prompt techniques to decrease unrelated images. Which solution meets these requirements?

- A. Use zero-shot prompts
- B. Use negative prompts
- C. Use positive prompts
- D. Use ambiguous prompts

  Answer: B

Q108: A company wants to use an LLM to generate concise, feature-specific descriptions for products. Which prompt engineering technique meets these requirements?

- A. One generic prompt; manually edit outputs
- B. Prompts per product category with key features and desired format/length
- C. Include a diverse range of features in each prompt for creativity

- D. Provide detailed, product-specific prompts only

  Answer: B

Q109: A company is developing an ML model to predict customer churn. The model performs well on training data but not on new data. Which solution resolves this issue?

- A. Decrease the regularization parameter to increase complexity
- B. Increase the regularization parameter to decrease complexity
- C. Add more features
- D. Train for more epochs

  Answer: B

Q110: A company is implementing intelligent agents to provide conversational search experiences and needs a database to store and query embeddings as vectors. Which AWS service meets these requirements?

- A. Amazon Athena
- B. Amazon Aurora PostgreSQL
- C. Amazon Redshift
- D. Amazon EMR

  Answer: B

Q111: A financial institution is building an AI solution to make loan approval decisions by using an FM. For security and audit purposes, the company needs decisions to be explainable. Which factor relates to explainability?

- A. Model complexity
- B. Training time
- C. Number of hyperparameters
- D. Deployment time

  Answer: A

Q112: 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. Time-series forecasting with Amazon Personalize

- B. Medication review summaries by using Amazon Bedrock LLMs
- C. Classification model to categorize medications
- D. Summaries by using Amazon Rekognition

  Answer: B

Q113: A company wants to build a lead prioritization application where employees can view and adjust the weights assigned to different variables based on domain knowledge. Which model type meets these requirements?

- A. Logistic regression model
- B. Deep learning model on principal components
- C. k-NN model
- D. Neural network

  Answer: A

Q114: HOTSPOT — A company wants to build an ML application. Select and order the correct steps to develop a well-architected ML workload. (No explicit options provided in sheet.)

Answer: Not specified in sheet content

Q115: Which strategy will determine if a foundation model (FM) effectively meets business objectives?

- A. Evaluate performance on benchmark datasets
- B. Analyze architecture and hyperparameters
- C. Assess alignment with specific use cases
- D. Measure computational resources for deployment

  Answer: C

Q116: A company needs to train an ML model to classify images of animals. The company has a large dataset of labeled images and will not label more data. Which type of learning should be used?

- A. Supervised learning
- B. Unsupervised learning
- C. Reinforcement learning
- D. Active learning

Answer: A

Q117: Which phase of the ML lifecycle determines compliance and regulatory requirements?

- A. Feature engineering
- B. Model training
- C. Data collection
- D. Business goal identification

Answer: D

Q118: A food service company wants to develop an ML model to help decrease daily food waste and increase sales revenue. The company needs to continuously improve accuracy. Which solution meets these requirements?

- A. Use Amazon SageMaker and iterate with newer data
- B. Use Amazon Personalize and iterate with historical data
- C. Use Amazon CloudWatch to analyze orders
- D. Use Amazon Rekognition to optimize the model

Answer: A

Q119: A company has developed an ML model to predict real estate sale prices and wants to deploy it for predictions without managing servers. Which solution meets these requirements?

- A. Deploy on Amazon EC2
- B. Deploy on Amazon EKS
- C. Use CloudFront with S3
- D. Use an Amazon SageMaker endpoint

Answer: D

Q120: A company wants to develop an AI application to help employees check open customer claims, identify details, and access documents. Which solution meets these requirements?

- A. Agents for Amazon Bedrock with Amazon Fraud Detector
- B. Agents for Amazon Bedrock with Amazon Bedrock knowledge bases
- C. Amazon Personalize with Bedrock knowledge bases

- D. Train a new ML model on SageMaker

  Answer: B

Q121: A manufacturing company uses AI to inspect products and find damages or defects. Which type of AI application is this?

- A. Recommendation system
- B. Natural language processing (NLP)
- C. Computer vision
- D. Image processing

  Answer: C

Q122: A company wants to create an ML model to predict customer satisfaction and needs fully automated model tuning. Which AWS service meets these requirements?

- A. Amazon Personalize
- B. Amazon SageMaker
- C. Amazon Athena
- D. Amazon Comprehend

  Answer: B

Q123: Which technique can a company use to lower bias and toxicity in generative AI applications during the post-processing lifecycle?

- A. Human-in-the-loop
- B. Data augmentation
- C. Feature engineering
- D. Adversarial training

  Answer: A

Q124: A bank fine-tuned an LLM to expedite loan approvals. An external audit found faster approvals for a specific demographic. How should the bank fix this issue most cost-effectively?

- A. Include more diverse training data and fine-tune again
- B. Use RAG with the fine-tuned model

- C. Use AWS Trusted Advisor to eliminate bias
- D. Pre-train a new LLM with more diverse data

  Answer: A

Q125: HOTSPOT — A company wants to make the LLM available to internal teams; select the appropriate inference mode per use case. (Details not present in sheet row.)

Answer: Not specified in sheet content

Q126: A company needs to log all requests to its Amazon Bedrock API and retain the logs securely for 5 years at the lowest cost. Which combination 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

Q127: An ecommerce company wants to improve search engine recommendations by customizing results for each user. Which AWS service meets these requirements?

- A. Amazon Personalize
- B. Amazon Kendra
- C. Amazon Rekognition
- D. Amazon Transcribe

  Answer: A

Q128: A hospital is developing an AI system to assist diagnoses based on patient records and images. Sensitive data must not leave the country. Which data governance strategy ensures compliance and privacy?

- A. Data residency
- B. Data quality
- C. Data discoverability

- D. Data enrichment

  Answer: A

Q129: A company needs to monitor the performance of its ML systems using a highly scalable AWS service. Which service meets these requirements?

- A. Amazon CloudWatch
- B. AWS CloudTrail
- C. AWS Trusted Advisor
- D. AWS Config

  Answer: A

Q130: An AI practitioner is developing a prompt for an Amazon Titan model on Bedrock to solve numerical reasoning and adds “show work step by step.” Which technique is this?

- A. Chain-of-thought prompting
- B. Prompt injection
- C. Few-shot prompting
- D. Prompt templating

  Answer: A

Q131: Which AWS service makes foundation models available to help build and scale generative AI applications?

- A. Amazon Q Developer
- B. Amazon Bedrock
- C. Amazon Kendra
- D. Amazon Comprehend

  Answer: B

Q132: A company is building a mobile app for visually impaired users. The app must hear users and provide voice responses. Which solution meets these requirements?

- A. Use a deep learning neural network to perform speech recognition
- B. Build models to search numeric patterns

- C. Use generative AI summarization to generate text
- D. Build image classification models

  Answer: A

Q133: A company wants to enhance response quality for complex problem-solving tasks requiring step-by-step explanations. Which prompt technique meets these requirements?

- A. Few-shot prompting
- B. Zero-shot prompting
- C. Directional stimulus prompting
- D. Chain-of-thought prompting

  Answer: D

Q134: A company wants to keep its FM relevant by using recent data and implement a training strategy with regular updates. Which solution meets these requirements?

- A. Batch learning
- B. Continuous pre-training
- C. Static training
- D. Latent training

  Answer: B

Q135: HOTSPOT — Select the correct ML paradigm for each use case. (Details not present in sheet row.)

Answer: Not specified in sheet content

Q136: Which option is a characteristic of AI governance frameworks for building trust and deploying human-centered AI?

- A. Expanding initiatives across business units
- B. Ensuring alignment with business standards and revenue goals
- C. Overcoming challenges to drive transformation and growth
- D. Developing policies and guidelines for data, transparency, responsible AI, and compliance

  Answer: D

Q137: An ecommerce company is using a generative AI chatbot for customer inquiries and wants to measure financial effect. Which metric should be used?

- A. Number of inquiries handled
- B. Cost of training AI models
- C. Cost for each customer conversation
- D. Average handled time (AHT)

  Answer: C

Q138: A company wants to find groups for customers based on demographics and buying patterns. Which algorithm should be used?

- A. k-NN
- B. K-means
- C. Decision tree
- D. SVM

  Answer: B

Q139: A company's LLM is experiencing hallucinations. How can the company decrease hallucinations?

- A. Set up Agents for Bedrock to supervise training
- B. Pre-process data to remove hallucinatory samples
- C. Decrease the temperature parameter
- D. Use an FM trained not to hallucinate

  Answer: C

Q140: A Bedrock chatbot processes support requests requiring multiple back-and-forth messages. Which solution gives the LLM the ability to use content from previous messages?

- A. Turn on invocation logging
- B. Add previous messages to the model prompt
- C. Use Amazon Personalize to save conversation history
- D. Use Provisioned Throughput

  Answer: B

Q141: Employees provide product descriptions and recommendations based on location; the company wants to automate using FMs. Which AWS service meets these requirements?

- A. Amazon Macie
- B. Amazon Transcribe
- C. Amazon Bedrock
- D. Amazon Textract

  Answer: C

Q142: A company uploads customer service emails to S3 and wants alerts whenever sensitive data is found. Which solution fully automates detection with least effort?

- A. Configure Amazon Macie for S3 documents
- B. Deploy an LLM on SageMaker to redact
- C. Develop regex patterns on a SageMaker notebook
- D. Ask customers to avoid sharing sensitive info

  Answer: A

Q143: HOTSPOT — Train employees on prompt structuring; select technique for each template. (Details not present.)

Answer: Not specified in sheet content

Q144: HOTSPOT — Model responses sometimes include harmful content; select the correct Bedrock filter policy for each mitigation. (Details not present.)

Answer: Not specified in sheet content

Q145: Which option is a benefit of using SageMaker Model Cards to document AI models?

- A. Visually appealing summary of capabilities
- B. Standardizing information about purpose, performance, and limitations
- C. Reducing computational requirements
- D. Physically storing models for archival

  Answer: B

Q146: What does an F1 score measure in the context of FM performance?

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

  Answer: A

Q147: A company deployed an AI/ML FAQ assistant; questions change over time. Agents need to ask and receive automatically generated answers to common questions. Which strategy is most cost-effective?

- A. Fine-tune regularly
- B. Train with context data
- C. Pre-train and benchmark with context data
- D. Use RAG with prompt engineering

  Answer: D

Q148: A resume screening system was trained on a dataset not representative of all demographics. Which responsible AI dimension does this present?

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

  Answer: A

Q149: A global financial company needs to continuously monitor development phases and ensure policies/regulations are followed. Which AWS services help assess compliance? (Choose two.)

- A. AWS Audit Manager
- B. AWS Config
- C. Amazon Inspector
- D. Amazon CloudWatch
- E. AWS CloudTrail

  Answer: A, B

Q150: A company wants to improve the accuracy of responses from a generative AI application that uses an FM on Amazon Bedrock. Which solution is most cost-effective?

- A. Fine-tune the FM
- B. Retrain the FM
- C. Train a new FM
- D. Use prompt engineering

  Answer: D

Q151: A company wants to identify harmful language in the comments section of social media posts by using an ML model. The company will not use labeled data to train the model. Which strategy should the company use to identify harmful language?

- A. Use Amazon Rekognition moderation
- B. Use Amazon Comprehend toxicity detection
- C. Use Amazon SageMaker built-in algorithms to train the model
- D. Use Amazon Polly to monitor comments

  Answer: B

Q152: A media company wants to analyze viewer behavior and demographics to recommend personalized content. The company wants to deploy a customized ML model in production and observe if 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: D

Q153: A company is using Amazon EC2 Auto Scaling to scale its Amazon EC2 instances. Which benefit of the AWS Cloud does this example illustrate?

- A. High availability
- B. Elasticity

- C. Reliability
- D. Global reach

  Answer: B

Q154: A manufacturing company wants to create product descriptions in multiple languages. Which AWS service will automate this task?

- A. Amazon Translate
- B. Amazon Transcribe
- C. Amazon Kendra
- D. Amazon Polly

  Answer: A

Q155: Which of the following is the least relevant consideration in assessing whether users should be given the right to opt out from an AI system?

- A. Feasibility
- B. Risk to users
- C. Industry practice
- D. Cost of alternative mechanisms

  Answer: C

Q156: Which AWS feature records details about ML instance data for governance and reporting?

- A. Amazon SageMaker Model Cards
- B. Amazon SageMaker Debugger
- C. Amazon SageMaker Model Monitor
- D. Amazon SageMaker JumpStart

  Answer: A

Q157: A financial company is using ML to help with some 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 investments

- D. Forecasting revenue for certain products

  Answer: A

Q158: A medical company wants an AI application to access structured patient records, extract relevant information, and generate concise summaries. Which solution meets these requirements?

- A. Use Amazon Comprehend Medical to extract medical entities/relationships and apply rule-based summary formatting
- B. Use Amazon Personalize with a general-purpose summarizer
- C. Use Amazon Textract plus keyword extraction to generate summaries
- D. Use Amazon Kendra indexing plus a template-based summary system

  Answer: A

Q159: Which option describes embeddings in the context of AI?

- A. A method for compressing large datasets
- B. An encryption method for securing sensitive data
- C. A method for visualizing high-dimensional data
- D. A numerical method for data representation in a reduced dimensionality space

  Answer: D

Q160: A company is building an AI application to summarize books of varying lengths. During testing, the application fails to summarize some books. Why?

- A. Temperature is set too high
- B. Selected model does not support fine-tuning
- C. Top P value is too high
- D. Input tokens exceed the model’s context size

  Answer: D

Q161: An airline company wants a conversational AI assistant for flight schedules, booking, and payments with LLMs and a knowledge base. Which solution meets requirements with least development effort?

- A. Train models on SageMaker Autopilot
- B. Develop a RAG agent by using Amazon Bedrock

- C. Create a Python app with Amazon Q Developer
- D. Fine-tune models on SageMaker JumpStart

  Answer: B

Q162: What is tokenization used for in 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

Q163: Which option is a characteristic of transformer-based language models?

- A. Use convolutional layers to capture local patterns
- B. Can process only text data
- C. Use self-attention mechanisms to capture contextual relationships
- D. Process sequences one element at a time in cyclic iterations

  Answer: C

Q164: A financial company uses AI systems to obtain customer credit scores. The company is expanding to a new geographic area and must ensure it can operate there. Which compliance laws should be reviewed?

- A. Local health data protection laws
- B. Local payment card data protection laws
- C. Local education privacy laws
- D. Local algorithm accountability laws

  Answer: D

Q165: A company wants to use Bedrock Guardrails to detect and filter harmful user inputs and outputs. Which categories can be filtered? (Choose two.)

- A. Hate
- B. Politics
- C. Violence
- D. Gambling

- E. Religion

  Answer: A, C

Q166: Which scenario describes a potential risk/limitation of prompt engineering?

- A. Ensures consistent deterministic outputs, removing need for validation
- B. Could expose the model to prompt injection attacks
- C. Properly designed prompts eliminate data poisoning/model hijacking
- D. Guarantees highly reliable outputs with real-world data

  Answer: B

Q167: A publisher built a RAG solution with new content daily and wants near real-time experience. Which steps should be implemented with offline batch processing? (Choose two.)

- A. Generation of content embeddings
- B. Generation of query embeddings
- C. Creation of the search index
- D. Retrieval of relevant content
- E. Response generation for the user

  Answer: A, C

Q168: Which technique breaks a complex task into smaller subtasks sent sequentially to an LLM?

- A. One-shot prompting
- B. Prompt chaining
- C. Tree of thoughts
- D. RAG

  Answer: B

Q169: An AI practitioner needs to improve the accuracy of a natural language generation model using rapidly changing inventory data. Which technique will help?

- A. Transfer learning

- B. Federated learning
- C. Retrieval Augmented Generation (RAG)
- D. One-shot prompting

  Answer: C

Q170: A company needs standardized documentation of model version tracking and a record of model development for collaboration with research institutes. Which solution meets these requirements?

- A. Track changes with Git
- B. Use Amazon Fraud Detector
- C. Amazon SageMaker Model Cards
- D. Amazon Comprehend

  Answer: C

Q171: A company using multiple ML models wants to identify changes in model quality. Which service/feature meets this requirement?

- A. Amazon SageMaker JumpStart
- B. Amazon SageMaker HyperPod
- C. Amazon SageMaker Data Wrangler
- D. Amazon SageMaker Model Monitor

  Answer: D

Q172: What is the purpose of chunking in RAG?

- A. Avoid database storage limits for large documents
- B. Avoid converting large text into vector embeddings
- C. Improve contextual relevancy of results from the vector index
- D. Decrease storage cost by storing chunks

  Answer: C

Q173: A company's editorial assistant app usage is unpredictable post-deployment and performance is not a concern now. They want to minimize cost. Which solution meets these requirements?

- A. GPU-powered EC2

- B. Bedrock with Provisioned Throughput
- C. Bedrock with On-Demand Throughput
- D. SageMaker JumpStart

  Answer: C

Q174: A RAG app on Bedrock gathers financial news for newsletters. Users reported politically influenced ideas. Which guardrail can identify and filter this content?

- A. Word filters
- B. Denied topics
- C. Sensitive information filters
- D. Content filters

  Answer: B

Q175: A fraud detection system flags potential fraud; employees review flagged cases. The company wants to minimize time spent on non-fraudulent flags. Which metric meets this requirement?

- A. Recall
- B. Accuracy
- C. Precision
- D. Lift chart

  Answer: C

Q176: A product-manuals QA agent should improve customer confidence. Which strategy helps?

- A. Include confidence level in the response
- B. Include referenced manual links in the response
- C. Use an avatar that looks like a computer
- D. Match company language style

  Answer: B

Q177: A hospital AI system provides personalized treatment recommendations and must provide rationale accessible to doctors and patients. Which human-centered design principle is this?

- A. Explainability
- B. Privacy and security
- C. Fairness
- D. Data governance

  Answer: A

Q178: Which statement presents an advantage of using RAG for NLP tasks?

- A. Uses external knowledge to generate more accurate, informative responses
- B. Designed to improve training speed
- C. Primarily used for speech recognition
- D. Technique for data augmentation in computer vision

  Answer: A

Q179: A company fine-tuned an LLM from Bedrock and wants to handle a steady rate of requests each minute cost-effectively. Which approach fits best?

- A. Deploy on EC2 compute optimized instance
- B. Use On-Demand throughput on Bedrock
- C. Store model in S3 and host with Lambda
- D. Purchase Provisioned Throughput on Bedrock

  Answer: D

Q180: Which technique involves training AI models on labeled datasets to adapt models to specific industry terminology and requirements?

- A. Data augmentation
- B. Fine-tuning
- C. Model quantization
- D. Continuous pre-training

  Answer: B

Q181: A company is creating an agent with Amazon Bedrock Agents and wants to improve accuracy by providing specific examples. Which solution meets these requirements?

- A. Modify the advanced prompts for the agent to include the examples
- B. Create a guardrail including the examples
- C. Use SageMaker Ground Truth to label the examples
- D. Run a Lambda script to add examples to the training dataset

  Answer: A

Q182: Which option is a benefit of using infrastructure as code (IaC) in MLOps?

- A. Eliminates need for hyperparameter tuning
- B. Always provisions powerful compute
- C. Streamlines deployment of scalable and consistent ML workloads in cloud environments
- D. Minimizes expenses by deploying only low-cost instances

  Answer: C

Q183: A company wants to fine-tune an FM to answer questions for a specific domain using instruction-based fine-tuning. How should training data be prepared?

- A. Merge internal documents and industry materials into a single file
- B. Collect external company reviews and label positive/negative
- C. Create pairs of questions and answers specific to the domain
- D. Create few-shot prompts to restrict to domain knowledge

  Answer: C

Q184: Which ML technique ensures data compliance and privacy when training AI models on AWS?

- A. Reinforcement learning
- B. Transfer learning
- C. Federated learning
- D. Unsupervised learning

  Answer: C

Q186: A company ingests consumer complaints using complex hard-coded logic and wants to scale across markets and product lines. Which advantage do

generative AI models offer here?

- A. Predictability of outputs
- B. Adaptability
- C. Less sensitivity to input changes
- D. Explainability

  Answer: B

Q187: A financial company wants to flag all credit card activity as possibly fraudulent or non-fraudulent. Which ML model type meets these requirements?

- A. Regression
- B. Diffusion
- C. Binary classification
- D. Multi-class classification

  Answer: C

Q188: HOTSPOT — Designing a customer service chatbot ensuring responsible AI characteristics. (Details not present.)

Answer: Not specified in sheet content

Q189: A hospital wants a generative AI solution with speech-to-text to improve dictation of clinical notes. Which AWS service meets these requirements?

- A. Amazon Q Developer
- B. Amazon Polly
- C. Amazon Rekognition
- D. AWS HealthScribe

  Answer: D

Q190: Which type of AI model makes numeric predictions?

- A. Diffusion
- B. Regression
- C. Transformer
- D. Multi-modal

  Answer: B

Q191: HOTSPOT — Use Amazon SageMaker features for various use cases. (Details not present.)

Answer: Not specified in sheet content

Q192: What is the purpose of vector embeddings in an LLM?

- A. Splitting text into manageable pieces
- B. Grouping characters as a single unit
- C. Providing the ability to mathematically compare texts
- D. Providing the count of every word in input

  Answer: C

Q193: A company wants to fine-tune an FM on AWS, ensuring data stays private, safe, and in-region, cost-effectively. Which steps meet these requirements? (Choose two.)

- A. Host on-premises with AWS Outposts
- B. Use the Amazon Bedrock API
- C. Use AWS PrivateLink and a VPC
- D. Host the Bedrock API on premises
- E. Use Amazon CloudWatch logs and metrics

  Answer: B, C

Q194: A financial company must generate reports to show adherence to international regulations for handling sensitive customer data. Which AWS service meets these requirements?

- A. Amazon Macie
- B. AWS Artifact
- C. AWS Secrets Manager
- D. AWS Config

  Answer: B

Q195: A medical company wants to modernize an onsite information processing app and use generative AI to respond to medical questions responsibly. Which AWS service should be used?

- A. Guardrails for Amazon Bedrock

- B. Amazon Inspector
- C. Amazon Rekognition
- D. AWS Trusted Advisor

  Answer: A

Q196: Which metric is used to evaluate FM performance for text summarization tasks?

- A. F1 score
- B. BLEU score
- C. Accuracy
- D. Mean squared error (MSE)

  Answer: B

Q197: What is the benefit of fine-tuning a foundation model (FM)?

- A. Reduces size/complexity and enables slower inference
- B. Retrains the FM from scratch with specific data
- C. Keeps knowledge up to date by pre-training on recent data
- D. Improves performance on a specific task by further training on new labeled data

  Answer: D

Q198: A company wants to improve a chatbot's tone using 100 examples of high-quality agent conversations. Which solution meets these requirements?

- A. Use Amazon Personalize
- B. SageMaker HyperPod pre-training job
- C. Host model on SageMaker and use TensorRT for LLM deployment
- D. Create an Amazon Bedrock fine-tuning job

  Answer: D

Q199: An ecommerce company is using a chatbot to automate order submission. Which input vulnerability must be resolved before launch?

- A. Data leakage
- B. Prompt injection

- C. LLM hallucinations
- D. Concept drift

  Answer: B

Q200: A social media company wants to prevent users from posting discriminatory content and use Amazon Bedrock as part of the solution. How can Bedrock be used to meet these requirements?

- A. Give users the ability to interact based on preferences
- B. Block interactions related to predefined topics
- C. Restrict conversations to predefined topics
- D. Provide multiple responses to select from

  Answer: B

Q201: An education company wants a conversational AI application where users can enter text or provide a picture of a question, and the application responds with a written answer and an explanation. Which model type meets these requirements?

- A. Computer vision model
- B. Large multi-modal language model
- C. Diffusion model
- D. Text-to-speech model

  Answer: B

Q202: In which stage of the generative AI model lifecycle are tests performed to examine the model's accuracy?

- A. Deployment
- B. Data selection
- C. Fine-tuning
- D. Evaluation

  Answer: D

Q203: Which statement correctly describes embeddings in generative AI?

- A. Embeddings represent data as high-dimensional vectors that capture semantic relationships

- B. Embeddings search data to find the most helpful information to answer questions
- C. Embeddings reduce hardware requirements by using less precise data types
- D. Embeddings store and retrieve data for generative AI applications

  Answer: A

Q204: A company wants LLM responses to be as deterministic and stable as possible. Which solution meets these requirements?

- A. Set the temperature parameter to 0 on prompt submission
- B. Add “make your response deterministic” at the end of the prompt
- C. Add “make your response deterministic” at the beginning of the prompt
- D. Set the temperature parameter to 1 on prompt submission

  Answer: A

Q205: A company needs a generative AI model for an application that must provide responses in real time. Which characteristic should be considered?

- A. Model complexity
- B. Innovation speed
- C. Inference speed
- D. Training time

  Answer: C

Q207: A retail company wants to build a recommendation model responsibly and reduce bias. Which practice should be applied when collecting data?

- A. Use data only from demographics matching the overall base
- B. Collect data only from customers with a past purchase history
- C. Ensure data is balanced and collected from a diverse group
- D. Use a public dataset only

  Answer: C

Q208: A company is developing an ML model to predict customer churn. Which evaluation metric assesses the performance of this binary classification task?

- A. F1 score
- B. Mean squared error (MSE)
- C. R-squared
- D. Time used to train the model

  Answer: A

Q209: A practitioner needs a metric that shows the ratio of correctly classified items to the total of correct and incorrect items. Which metric meets this requirement?

- A. Accuracy
- B. Precision
- C. F1 score
- D. Recall

  Answer: A

Q210: An ecommerce company receives gigabytes of data daily and performs inferences once per day to forecast demand. Which inference type meets these requirements?

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

  Answer: A

Q211: A company’s generative AI model has been in production for a long time; recently responses became inconsistent. The company wants to evaluate bias and drift. Which service/feature meets these requirements?

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

  Answer: A

Q212: A company wants to restrict employee access to specific models available on Amazon Bedrock. Which solution meets these requirements?

- A. Use IAM policies to restrict model access
- B. Use temporary credentials service to gate models
- C. Use service roles to restrict subscriptions
- D. Use a security scanner to monitor access

  Answer: A

Q213: Which ML technique uses training data labeled with correct outputs?

- A. Supervised learning
- B. Unsupervised learning
- C. Reinforcement learning
- D. Transfer learning

  Answer: A

Q214: Which LLM parameter controls the number of possible next tokens considered at each generation step?

- A. Maximum tokens
- B. Top K
- C. Temperature
- D. Batch size

  Answer: B

Q215: A company uses Amazon Lex and Amazon OpenSearch Service and needs to convert private text data into vectors before storing in a database. Which FM meets this requirement?

- A. Text completion model
- B. Instruction following model
- C. Text embeddings model
- D. Image generation model

  Answer: C

Q216: A company wants LLM-generated product descriptions to match a format shown in examples. Which prompt technique will achieve this?

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

  Answer: D

Q217: A bank is fine-tuning an LLM on Amazon Bedrock to assist with customer loan questions and must ensure that private customer data is not revealed. Which solution meets these requirements?

- A. Use Bedrock Guardrails
- B. Remove PII from data before fine-tuning
- C. Increase the Top-K parameter
- D. Store data in S3 and encrypt before fine-tuning

  Answer: B

Q218: A grocery store wants a chatbot to find products in-store, check inventory in real time, and provide location. Which prompt technique should be used?

- A. Zero-shot prompting
- B. Few-shot prompting
- C. Least-to-most prompting
- D. ReAct (reasoning and acting) prompting

  Answer: D

Q219: A company uses a third-party model on Amazon Bedrock to analyze confidential documents and is concerned about data privacy. How does Bedrock protect data privacy?

- A. Inputs/outputs anonymized and shared with providers
- B. Inputs/outputs are not shared with any third-party model providers
- C. Inputs confidential but outputs shared with providers
- D. Inputs/outputs redacted before sharing with providers

Answer: B

Q220: An animation company wants to provide subtitles for its content. Which AWS service meets this requirement?

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

Answer: C

Q221: An ecommerce company wants to group customers based on purchase history and preferences to personalize the app experience. Which ML technique should be used?

- A. Classification
- B. Clustering
- C. Regression
- D. Content generation

Answer: B

Q222: A company wants to control employee access to publicly available foundation models. Which solution meets these requirements?

- A. Analyze cost and usage reports
- B. Download security/compliance docs
- C. Configure SageMaker JumpStart to restrict discoverable FMs
- D. Build a hybrid search with OpenSearch Service

Answer: C

Q223: A company compares a translation tool's responses to human responses on the same documents via a parallel data process. Which strategy should be used to evaluate the tool?

- A. Use BLEU to estimate absolute translation quality of both methods
- B. Use BLEU to estimate relative translation quality between methods
- C. Use BERTScore to estimate absolute quality

- D. Use BERTScore to estimate relative quality

  Answer: B

Q224: An AI practitioner wants more diverse and creative outputs from an LLM. How should the inference parameter be adjusted?

- A. Increase the temperature value
- B. Decrease Top K
- C. Increase response length
- D. Decrease prompt length

  Answer: A

Q225: A company built custom computer vision models and needs a user-friendly interface for data labeling to minimize mistakes on new real-world data. Which AWS service/tool meets these requirements?

- A. Amazon SageMaker Ground Truth
- B. Amazon SageMaker Canvas
- C. Amazon Bedrock playground
- D. Amazon Bedrock Agents

  Answer: A

Q226: A company is integrating AI into recruitment and hiring and wants to mitigate bias risks and ensure responsible AI while prioritizing equitable hiring decisions. Which core dimensions should be considered? (Choose two.)

- A. Fairness
- B. Tolerance
- C. Flexibility
- D. Open source
- E. Transparency

  Answer: A, E

Q227: A company deployed a churn prediction model and wants to evaluate how accurately it predicts churn compared to actual behavior after 1 week. Which metric meets this requirement?

- A. RMSE

- B. ROI
- C. F1 score
- D. BLEU score

  Answer: C

Q228: A company is using a pre-trained FM on Amazon Bedrock and wants the FM to include more context using company information. Which is MOST cost-effective?

- A. Use Amazon Bedrock Knowledge Bases
- B. Choose a different FM on Bedrock
- C. Use Amazon Bedrock Agents
- D. Deploy a custom model on Bedrock

  Answer: A

Q229: HOTSPOT — Using SageMaker across the AI model lifecycle steps. (Details not present.)

Answer: Not specified in sheet content

Q230: A food service company wants a dataset to predict food preferences and ensure all demographics' preferences are included. Which dataset characteristic is this?

- A. Accuracy
- B. Diversity
- C. Recency bias
- D. Reliability

  Answer: B

Q231: A company wants a chatbot that answers HR policy questions using an LLM and a large documentation base. Which technique optimizes responses?

- A. Retrieval Augmented Generation (RAG)
- B. Few-shot prompting
- C. Set temperature to 1
- D. Decrease token size

Answer: A

Q232: An education company is building a chatbot for teenagers and wants it to use creative spelling/shortened words. Which metric will assess performance?

- A. F1 score
- B. BERTScore
- C. ROUGE
- D. BLEU

Answer: B

Q233: A customer service team wants to analyze customer feedback and automatically classify into categories (product quality, customer service, delivery experience). Which AI concept is this?

- A. Computer vision
- B. Natural language processing (NLP)
- C. Recommendation systems
- D. Fraud detection

Answer: B

Q234: Planning a conversational AI surfaced in Microsoft Teams, Cortana, and Amazon Alexa. Which service should be used?

- A. Azure Bot Service
- B. Azure Cognitive Search
- C. Speech
- D. Language service

Answer: A

Q235: HOTSPOT — True/False selections. (Details not present.)

Answer: Not specified in sheet content

Q236: HOTSPOT — Complete the sentence. (Details not present.)

Answer: Not specified in sheet content

Q237: A smart device that answers “What is the stock price of Contoso, Ltd.?” is an example of which AI workload?

- A. Knowledge mining
- B. Natural language processing
- C. Computer vision
- D. Anomaly detection

  Answer: B

Q238: DRAG DROP — Match ML models to descriptions. (Details not present.)

Answer: Not specified in sheet content