

PRACTICE QUESTIONS

AWS

AIF-C01

AWS Certified AI Practitioner

SHARED BY [u/throwaway_c01](#) on aiexamcerts

EDITION Free Set 1

QUESTIONS 10

SHARED October 15, 2026

EXAM AWS AIF-C01, AWS Certified AI Practitioner

Shared by u/throwaway_c01 on aiexamcerts

10 practice questions for the AWS AIF-C01 exam, AWS Certified AI Practitioner, each followed by its answer and a short explanation. More free sets and the full practice exam: www.aiexamcerts.com/aws/aif-c01. Spotted a mistake? Email support@aiexamcerts.com with the exam code AIF-C01 and the question number.

-
1. A retailer wants to group customers by purchasing behavior without predefined categories. Which technique fits best?
- A Binary classification
 - B Clustering
 - C Linear regression
 - D Sentiment analysis

Answer: B

Clustering is an unsupervised method that groups records by similarity when no labels exist. Binary classification is tempting because customers could be split into two segments, but it requires labeled examples of each class. Sentiment analysis applies to opinion text, so it does not describe purchase histories.

-
2. In supervised learning, the value the model is trained to predict is called the _____.

Answer: target (also accepted: target variable, label)

The target, also called the label, is the column the model learns to predict from the input features. Features are the input columns used to make that prediction, which makes them easy to confuse with the target. A dataset with no target column can only be used for unsupervised methods.

-
3. What does the self-attention mechanism in a transformer do?
- A It merges rare tokens into a shared vocabulary entry
 - B It weighs how strongly each token relates to the other tokens in the sequence
 - C It stores every training example in a lookup table
 - D It sorts the model output with a fixed ranking step

Answer: B

Self-attention computes relevance scores between tokens so each position can draw context from the whole sequence. Compressing tokens into bits misrepresents how attention works. A lookup table describes memorization rather than attention, and a fixed sorting step is unrelated.

4. A higher temperature setting makes a model's output more random.
- A True
 - B False

Answer: A (True)

Temperature scales the probability distribution before sampling, so higher values flatten it and make less likely tokens more probable. The statement is true. Lower temperature moves generation closer to deterministic selection of the most likely token.

5. Which two metrics are commonly used to evaluate the retrieval step of a RAG application? (Choose two.)
- A Silhouette score of the training clusters
 - B Recall at k for the retrieved passages
 - C Mean squared error of the fine-tuning loss
 - D Context precision of the retrieved passages
 - E Token count of the system prompt

Answer: B, D

Recall at k shows whether the passages needed for the answer were retrieved, and context precision shows how many retrieved passages were relevant. The silhouette score is tempting because it also uses distances, but it measures cluster quality rather than retrieval. Mean squared error of the loss and token counts do not score retrieval.

6. Which Amazon Bedrock feature manages ingestion, chunking, embedding, and retrieval for a RAG application?
- A Amazon Bedrock Agents action groups
 - B Amazon Bedrock Guardrails
 - C Amazon Bedrock model evaluation jobs
 - D Knowledge Bases for Amazon Bedrock

Answer: D

Knowledge Bases for Amazon Bedrock connects to a data source, handles chunking and embeddings, and returns relevant passages to the model. Agent action groups are tempting because agents also use retrieved information, but action groups call APIs or functions to take actions. Guardrails filter content, and model evaluation jobs score outputs.

7. A lending model approves applicants from one demographic group less often, even though the group attribute is not an input. Which concern does this describe?
- A Proxy bias in the model's outputs
 - B Model drift in the inference endpoint
 - C Prompt injection through user input
 - D Overfitting to the noise in the training set

Answer: A

Proxy bias arises when neutral-looking features correlate with a protected attribute, so the model reproduces a disparity without the attribute being present. Model drift is tempting because outcomes changed, but drift describes shifting input data over time rather than a group disparity. Prompt injection concerns malicious input, and overfitting concerns generalization.

8. The NIST AI Risk Management Framework organizes its core functions as Govern, Map, Measure, and _____.

Answer: Manage

The four core functions of the NIST AI RMF are Govern, Map, Measure, and Manage, and they apply across the AI lifecycle. Monitor is a tempting near-match, but it is not one of the four functions. Assess is not a core function name either.

9. A company must show auditors that its AI workloads meet a compliance framework. Which AWS service automates collecting evidence mapped to that framework?
- A AWS Trusted Advisor
 - B Amazon Inspector
 - C AWS Audit Manager
 - D AWS Systems Manager

Answer: C

AWS Audit Manager automates evidence collection and maps it to audit frameworks, which makes audits faster. AWS Trusted Advisor is tempting because it checks accounts against best practices, but it does not organize evidence for an audit framework. Inspector scans workloads for vulnerabilities and Systems Manager handles operational tasks.

10. Which practice best protects an AI application from prompt injection attempts that try to override its system instructions?
- A Raising the max tokens limit so instructions can be longer
 - B Caching every response for longer to avoid recomputation
 - C Trusting any user who has a valid account login
 - D Treating user input as untrusted and validating it before it reaches any tools

Answer: D

Prompt injection is reduced by treating user content as untrusted, limiting what tools the model can call, and validating inputs and outputs. A higher max tokens limit is tempting if you think more space means more rules, but it does not stop malicious instructions. Caching and login status do not address injected instructions.